Poverty, Inequality and Development

Essays in Honor of Erik Thorbecke

Alain de Janvry
Ravi Kanbur

POVERTY, INEQUALITY AND DEVELOPMENT

Essays in Honor of Erik Thorbecke

ECONOMIC STUDIES IN INEQUALITY, SOCIAL EXCLUSION AND WELL-BEING

This series will publish volumes that go beyond the traditional concepts of consumption, income or wealth and will offer a broad, inclusive view of inequality and well-being. Specific areas of interest will include Capabilities and Inequalities, Discrimination and Segregation in the Labor Market, Equality of Opportunities, Globalization and Inequality, Human Development and the Quality of Life, Income and Social Mobility, Inequality and Development, Inequality and Happiness, Inequality and Malnutrition, Income and Social Mobility, Inequality in Consumption and Time Use, Inequalities in Health and Education, Multidimensional Inequality and Poverty Measurement, Polarization among Children and Elderly People, Social Policy and the Welfare State, and Wealth Distribution.

Volume I
de Janvry, Alain and Kanbur, Ravi
Poverty, Inequality and Development: Essays in Honor of Erik Thorbecke

POVERTY, INEQUALITY AND DEVELOPMENT
Essays in Honor of Erik Thorbecke

Edited by

Alain de Janvry[1] and Ravi Kanbur[2]
[1]University of California, Berkeley and [2]Cornell University

Springer

Library of Congress Control Number: 2005934374

ISBN:10: 1-4020-7850-1
ISBN-13: 978-1-4020-7850-7

e-ISBN-10: 0-387-29748-0
e-ISBN-13: 978-0387-29748-4

Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springeronline.com

Contents

Contributing Authors

Christopher B. Barrett, Cornell University
Kaushik Basu, Cornell University
Gary S. Fields, Cornell University
James E. Foster, Vanderbilt University
Alain de Janvry, University of California, Berkeley
Tjeerd Jellema, European Central Bank
Ravi Kanbur, Cornell University
Steven Keuning, European Central Bank
Jacob Kol, Eramus University
Michael Lokshin, World Bank
Peter McAdam, European Central Bank
John G. McPeak, Syracuse University
Reimund Mink, European Central Bank
Christian Morrisson, Université de Paris I and Delta
Jeffrey B. Nugent, University of Southern California
Per Pinstrup-Andersen, Cornell University
Graham Pyatt, University of Warwick
Gustav Ranis, Yale University
Martin Ravallion, World Bank
Sherman Robinson, University of Sussex
David Roland-Holst, Mills College & University of California, Berkeley
Jeffery I. Round, University of Warwick
Elisabeth Sadoulet, University of California, Berkley
David E. Sahn, Cornell University
Shailender Swaminathan, University of Alabama, Birmingham

Finn Tarp, University of Copenhagen
Matti Tuomala, University of Tampere

Preface

This collection of essays honors a remarkable man and his work. Erik Thorbecke has made significant contributions to the microeconomic and the macroeconomic analysis of poverty, inequality and development, ranging from theory to empirics and policy. The essays in this volume display the same range. As a collection they make the fundamental point that deep understanding of these phenomena requires both the micro and the macro perspectives together, utilizing the strengths of each but also the special insights that come when the two are linked together. After an overview section which contains the introductory chapter and a chapter examining the historical roots of Erik Thorbecke's motivations, the essays in this volume are grouped into four parts, each part identifying a major strand of Erik's work—Measurement of Poverty and Inequality, Micro Behavior and Market Failure, SAMs and CGEs, and Institutions and Development. The range of topics covered in the essays, written by leading authorities in their own areas, highlight the extraordinary depth and breadth of Erik Thorbecke's influence in research and policy on poverty, inequality and development.

Acknowledgements

These papers were presented at a conference in honor of Erik Thorbecke held at Cornell University on October 10-11, 2003. The conference was supported by the funds of the H.E. Babcock Chair in Food, Nutrition and Public Policy, and the T.H. Lee Chair in World Affairs at Cornell University. We are grateful to these funds for their financial support, to Joyce Knuutila for managing the conference and to Susan Snyder for preparing the volume for publication.

Apart from the paper presenters, the conference benefited from the presence of a number of colleagues and friends of Erik Thorbecke who chaired sessions and acted as discussants and referees. We would like to acknowledge the contribution of the following to the conference and the volume: Irma Adelman, Iwan Azis, Alok Bhargava, Peter Cornelisse, Alfred Field, Augustin Fosu, Patrick Guillaumont, Jan Willem Gunning, Jere Haas, Walter Isard, Stefan Klonner, Arie Kuyvenhoven, Mukul Majumadar, Germano Mwabu, Uri Possen, Vernon Ruttan, Richard Schuler, Karl Shell, John Strauss, Jan Svejnar and Henry Wan.

Chapter 1

POVERTY, INEQUALITY AND DEVELOPMENT: MICRO-MACRO PERSPECTIVES AND LINKAGES

Alain de Janvry[1] and Ravi Kanbur[2]
[1]University of California, Berkeley, alain@are.berkeley.edu and [2]Cornell University, sk145@cornell.edu

The essays in this volume honor Erik Thorbecke, a remarkable man who has devoted his life's work to analyzing issues of poverty, inequality and development. The roots of his motivation are examined by Jacob Kol in Chapter 2 of this volume. The output of this motivation is all around us. The most commonly used family of poverty measures in applied work today carries his name—the T in the "FGT" family of poverty indices stands for Thorbecke. It was in 1984 that the famous paper proposing this index, by Erik and two of his then graduate students, was published. But in the twenty years before and since, Erik has produced a mountain of work that establishes his place in the literature on poverty, inequality and development. The papers in this volume pay tribute to Erik's accomplishments by themselves making significant contributions to that literature.

After the beginning overview section, comprising this introduction to the chapters by the editors and Jacob Kol's introduction to Erik Thorbecke, the volume is divided into four parts, each reflecting a dimension of Erik's contribution to the literature. We begin with the measurement of poverty and inequality in Part I. Part II moves to individual behavior and market failure. Part III makes the transition to economy wide issues and Part IV addresses the role of institutions.

The four papers in Part I take up the issues connected with the measurement of poverty and inequality, to which Erik has made fundamental contributions. The paper by Martin Ravallion and Michael Lokshin (Chapter 3) addresses a central building block of poverty measurement—the poverty line. The two most commonly used procedures for determining poverty lines, the "Food-Energy Intake" (FEI) method and the "Cost-of-

Basic Needs" (CBN) method, are assessed with respect to "utility consistency". Both methods do badly in the assessment. The FEI method is shown to be theoretically problematic, and unlikely to be utility consistent. The utility consistency of the CBN method is an empirical matter, but it is shown to be violated in the application to Russian regional poverty lines. Interestingly, it is observed that this may be because of the decentralized administrative process followed in Russia in setting the poverty lines, leading to a tension between local autonomy and consistency across the nation.

Given the poverty line, how is the information in the distribution of wellbeing to be aggregated to produce a poverty index? This is where Erik Thorbecke has made perhaps his most famous contribution, and his co-author in that effort, James Foster, gives an overview of the axiomatic approach to poverty measurement in Chapter 4. He provides an illuminating categorization of axioms sets—Invariance axioms, Dominance axioms, Continuity axioms and Subgroup axioms. This categorization will prove useful to those taking up deeper investigation of different aspects of poverty measurement. For example, the subgroup axioms are closely related to the decomposability of poverty measures—a well known and most useful property of the FGT index. At the end of his paper, Foster takes up the intriguing question of the role of generalized means in poverty measurement, playing off their role in inequality measurement. He argues that just as many inequality measures are a function of a ratio of two generalized means, many poverty measures are functions of a ratio of a generalized mean to the poverty line. This observation must surely open up a rich seam of investigation in the theory of poverty measurement.

The contribution by Gary Fields, Chapter 5, addresses a long standing concern about what information is conveyed by poverty and inequality measures separately, and whether they can and should be combined. Fields asks two question on this combining—can it be done and should it be done? His answers are—yes, and (by and large) no. Using simple stylized growth paths, all of which exhibit first order dominance, he argues that in most cases he would not want to combine poverty and inequality indices into a single "blend" index since it can give intuitively misleading results. Great caution is thus called for in taking standard measures of poverty and standard measures of inequality and constructing (for example) a weighted combination of the two as a welfare measure.

While the papers by Ravallion and Lokshin and by Foster are entirely about poverty, and the paper by Fields is about combining measures of poverty and inequality, the focus of the paper by David Sahn, Chapter 6, is on inequality. However, the domain of measurement is different from the standard one of income. The focus is on inequality in health, with an application to data on Indian states. Health is measured by standardized height for children, and is used analogously to income in computation of inequality. This inequality is first of all compared across Indian states. Then intertemporal changes in "nutrition poverty" (or "health poverty"), defined as the percentage of children below two standard deviations of the standardized height distribution, are decomposed into components reflecting change in the mean and change in inequality. This procedure is familiar from income poverty analysis, where the FGT poverty measure has been decomposed in this manner, but its application to non-income dimensions of wellbeing is novel.

Part II of the volume turns to micro and market behavior in understanding poverty and inequality in developing countries. Chapter 7, by Jeffrey Nugent and Shailender Swaminathan, reflects not only Erik Thorbecke's work on inequality and the microeconomics of poor households, but also his long standing interest in Indonesia. Using data from Indonesia, the authors address the important issue of how inequality within a community affects investments in education by households in that community. There are several channels, suggested by theory, through which a group characteristic like inequality could affect individual level decisions—and the effects do not necessarily all go in the same direction. The issue is thus primarily an empirical one. They show that greater inequality can indeed be associated with lower household level investments in education. Moreover, they identify some very interesting channels for this effect—for example, through the influence of community level inequality on the supply of educational infrastructure in that community. There is now a growing literature on inequality and the supply of local public goods, and this paper is an important empirical contribution to that literature.

Chapter 8, by Christopher Barrett and John McPeak, addresses a topic of growing importance in the literature on poverty, inequality and development—risk and vulnerability. Erik Thorbecke has himself contributed to the growth of this part of the literature. Barrett and McPeak explore three questions: "What is the etiology of chronic poverty and vulnerability? How does nutrition-related health risk affect patterns of chronic poverty and vulnerability? What are the implications for the design of development policy, especially safety net implications?" Based on their own work and on the growing literature, they begin to develop answers to these questions. As important as the specific answers, however, is their general conclusion that this is a hugely under researched area relative to its importance in the actual lives of poor people in poor countries. An important implication is that we will have to focus on risk and vulnerability relative to critical thresholds that emerge from the context in which households live, and not (just) relative to standard poverty lines, as is the dominant practice in the literature to date. The design of safety nets relative to these thresholds is an important policy question waiting to be addressed.

A constant theme in Erik's work on development has been the importance of various types of market failure in explaining household behavior and household outcomes. The importance of this perspective is now well recognized. In Chapter 9, Alain de Janvry and Elisabeth Sadoulet present an overview of the literature on farm households that has developed in response to this challenge. In demonstrating that these households in poor countries are "systematically embedded in complex configurations characterized by incomplete and failing markets," they draw three important conclusions. The first is that a whole range of behaviors, such as the interrelatioships between health and education decisions and production choices, can be understood only in the context of behavioral responses to market failures. The second conclusion, however, is that we do not know "how close to a first best situation do these countervailing strategies allow households to get." This is a call for detailed and targeted empirical research. But in order to get at these estimates we will need in the first place improved structural models. This is the third conclusion and it brings us back full circle to the authors' central point—to construct these models, we need a much better understanding than we currently have of household behavior in the presence of market failure.

Kaushik Basu continues the theme of the impact of markets on the well being of the poor in Chapter 10, which starts with the statement: "Markets often work in peculiar ways." His central thesis is that "India's myriad labor laws, meant to protect laborers, may actually hurt them." Thus market failures may call for intervention to help the poor, but if it is not the right intervention then the interaction between individual behavior and market response may end up actually making them worse off. This chapter highlights a method that involves an intricate interplay between deep institutional knowledge of a specific context, in this case Indian labor law, combined with parsimonious modeling of the key features of that situation to address the question at hand. The conclusions show that extreme positions are not supported in terms of policy implications. Basu argues not that there should be complete freedom of contracting between employers and workers, but somewhat more freedom than there is now in the Indian case. Similarly, he argues that minimum wages can benefit workers provided they are not too high. The issue is not therefore an ideological one of minimum wages or not, but a case and context specific one of the level at which it should be set.

Chapters 7-10, each in its own way, began the process of linking individual or household behaviors and outcomes to the broader context of communities, markets and the nation. The chapters in Parts III and IV of the volume continue this process and take it to the economy wide and even global level in systematic fashion, addressing macroeconomic and institutional issues. The papers in Part III contribute to a literature with which Erik Thorbecke's name is linked very closely as founder. This is the literature on Social Accounting Matrices (SAMs) and Computable General Equilibrium Models (CGEs). Erik has been one of the pioneers not only in conceptualizing and modeling, but also in application of these models and frameworks to many countries and to a wide range of specific policy problems.

In the first paper in Part III of the volume, Chapter 11, Sherman Robinson takes an overview of a range of models in the SAM-CGE tradition. He places the models on a spectrum between SAM fix-price multiplier models on the one hand and a full blown Walrasian CGE model on the other. The tradition of structuralist macro economic models has much in common with the first, while the neoclassical general equilibrium tradition is clearly linked to the second. Models along the spectrum are identified by particular macro closure rules. Robinson explores the tension between the two ends of the spectrum, and asks if this can be reduced in some way. He argues that while the tension can never be eliminated, with judicious implementation, "many features of Keynesian demand-driven multiplier models can be accommodated within the flow-equilibrium structure of a CGE model..." But Robinson concludes with the view that it is considerations of dynamics, including forward looking expectations, that may well provide fertile ground for integration.

Chapter 12, by Graham Pyatt and Jeffery Round, stays within the fix-price SAM multiplier framework but explicitly links the macro economy to poverty impacts. Once again, Erik Thorbecke was one of the early contributors to this literature, as made clear by the authors. The paper first of all derives the multipliers—which measure the impact of injections into any one account on all of the other accounts. It then goes on to apply these to an analysis of the poverty impacts of such injections, and to present empirical results using a SAM for Indonesia originally compiled by Keuning (a co-author of Chapter 13 in this volume) and Thorbecke. The paper concludes by taking up an issue raised recently by Erik—the need to move away from fixed-price formulations to a general equilibrium setting where prices are determined endogenously. This of course relates back to the tensions emphasized by Robinson in Chapter 11. Pyatt and Thorbecke argue that to the extent that changing prices are important for capacity constraints, this can be incorporated in fixed-price models under various assumptions. However, so far as the implications of changing prices for simulation of changes in poverty are concerned, they argue that micro simulation techniques can improve poverty impact estimates.

All work on SAMs recognizes that SAM-based analysis, however sophisticated in theory, is only as good as the actual SAM to which it is applied. Indeed, Pyatt and Round end Chapter 12 with the observation that compared to issues of "model closure", for example, "improving primary data and greater sophistication in the construction of social accounting matrices are probably more important avenues towards a better understanding of basic issues and mechanisms." In their contribution to this volume in Chapter 13, Tjeerd Jellema, Steven Keuning, Peter McAdam and Reimund Mink take us through the development of an actual accounting matrix, this time for the Euro area. They show how existing data sets can be used in an overall framework of linking national accounting matrices together. The specific issue they consider is the European Central Bank's monetary policy analysis, and the Euro Area Accounting Matrix (EAAM) is constructed with this emphasis. But what the paper illustrates is the careful work needed in providing a solid empirical basis for SAMs.

Chapter 14 of the volume uses a SAM for Vietnam to address a crucial issue for that country—the economy-wide impact of globalization and economic reform. David Roland-Holst and Finn Tarp use a new and detailed SAM for Vietnam constructed for the year 2000. The focus is on price transmission in a liberalizing economy as it goes from a closed administered price system to an open economy facing world prices. In many ways price transmission can be seen as the dual to the demand transmission examined by Pyatt and Round in Chapter 12. Alternative decompositions are presented and then quantified for Vietnam. A basic conclusion is that simply looking at direct effects of, say, equating domestic and world prices in a particular sector, is likely to be a very misleading guide to the true impacts once the economy wide repercussions are taken into account—"relying on intuition or rules of thumb alone in this context is very unlikely to achieve something approaching optimality." Moreover, decomposition based on path analysis, developed in the paper, highlights the sectors that are most crucial in terms of the economy wide price response.

Part IV of this volume addresses a broad range of issues in the institutional dimension of an economy wide perspective—all issues to which Erik Thorbecke has made significant contributions. A wide range of topics is considered, including agricultural policy, development assistance and debates on dualism. In Chapter 15 Christian Morrisson starts with the work of Engerman and Sokoloff on inequality, institutions and growth. The arguments in these papers, and in the subsequent literature, are examined with respect to the experiences of Ghana, Kenya and Senegal. Specifically, the paper assesses whether in the African context (i) inequality is related to factor endowments, (ii) institutional development depends on inequality and factor endowments and (iii) institutions in turn influence income distribution. These hypotheses are subjected to detailed country specific assessment and complex patterns are found which support some of these relations but not others, and at different time periods. The overall conclusion stresses the importance of context specificity in making claims about the causes or the consequences of institutional development and its links to inequality.

It is sometimes not fully appreciated that Erik Thorbecke has made major contributions to the debate on development assistance, including a comprehensive recent review of the relationship between how the evolution of views on development has interacted with the evolution of views on aid effectiveness and the institutions of aid delivery. In Chapter 16, Ravi Kanbur and Matti Tuomala pay tribute to this aspect of Erik's work by exploring optimal aid allocation when conditionality does not work, as indeed shown by the evidence. In other words, they assume that aid is simply a net addition to the recipient government's resources, the use of which cannot be influenced by the donor. Suppose the donor wanted to allocate aid between two recipients so as to maximize a welfare function that was sensitive to the wellbeing of the poorest. What should determine this aid allocation? Kanbur and Tuomala answer this question by modeling each recipient government's behavior as choosing an optimal non-linear income tax given the resources it gets from the outside. Apart from the volume of these resources, a number of other factors influence the progressivity of the recipient government's policies—including its inequality aversion and the structural inequality in that society. The calculations in the chapter show that poorer countries and those with greater inequality aversion should get a larger share of aid. By contrast, countries with greater structural inequality should get a smaller share of aid.

Chapter 17 continues the discussion of policy and institutions, but focuses on the agricultural sector. The author of this chapter, Per Pinstrup-Andersen, is Erik Thorbecke's successor in the H.E. Babcock Chair at Cornell University. His focus in this chapter is on nutrition goals, and the use of agricultural research and policy to achieve them. Needless to say, Erik has contributed much to the analysis of both nutrition and agricultural policy. The chapter addresses five questions: "1. How could agricultural research and policy improve nutrition? 2. Should nutrition goals guide agricultural research and policy? 3. What policy measures are likely to be effective? 4. Are nutrition goals best achieved through pre- or post-harvest changes? 5. Would consumer behavior enhance or reduce the intended effect?" Each of these questions is investigated and specific proposals are made to enhance the impact of agricultural research and policy on nutritional goals, but the general conclusion is that nutritional goals should be incorporated directly into the decision making processes in these policies.

At the same time, however, the impact of agricultural policies alone will be limited if determined in isolation from other interventions that affect determinants of nutritional outcomes such as contaminated water and sanitation.

The final chapter in this volume, Chapter 18, is perhaps the one with the broadest perspective. In it, Gus Ranis revisits one of the grand organizing themes of the development literature in the last fifty years—dualism. His question, "Is Dualism Worth Revisiting?", is given an affirmative answer. He sees dualism as a metaphor, encompassing the sociological dualism of Boeke as well as the conventional economic sense of the coexistence of sectors that are asymmetric in some aspect of organization. This latter sense, he argues, goes back to the physiocrats and their tableau économique, but its modern life obviously dates to Arthur Lewis's classic 1954 exposition on dualistic development. This framework has been developed and refined over the last fifty years, but, according to Ranis, "has been subjected to too much criticism over the past several decades," especially from the perspective of neoclassical economics. Ranis in turn subjects these critiques to a critique, and argues that while some of them are valid, there is still much to be said for the general metaphor of sectors that are asymmetric in crucial ways, as well as the specific insights generated by dual economy models.

This brings us to the end of our overview of this volume of essays in honor of Erik Thorbecke. The essays all address topics to which Erik has made, and continues to make, significant contributions in his own research and policy work. In their range and breadth, therefore, they show the extraordinary influence that Erik has had on the analysis of poverty, inequality and development. We look forward to his future contributions.

Chapter 2

ERIK THORBECKE: GROWTH AND ROOTS

Jacob Kol
Eramus University, kuijpers@few.eur.nl

1. GROWTH

'That's where they allowed him to grow'; this, in retrospect, was observed by Mrs. Charla Thorbecke about Iowa State University, where her husband began his career: assistant professor (1957-1960), associate professor (1960-1963), and professor (1963-1973), interrupted by two years at the Agency for International Development (1966-1968).

In 1957 Erik Thorbecke, then 28, received his Ph.D. with the thesis 'The Tendency Towards Regionalization in International Trade', published in 1960 by Martinus Nijhoff in The Hague, The Netherlands.

From the regionalization in world trade his interest in research went to the related but new and topical phenomenon of economic integration in Europe, the continent of his ancestry; it was studied from the perspective of the US and of the pattern of world trade (Thorbecke, 1961, 1963). Somewhat later Bela Balassa invited him to evaluate the Common Agricultural Policy (CAP) of the EEC (then, EU now). In cooperation with Emilio Pagoulatos a model was built to analyze the CAP's effects on trade and welfare. It was concluded that the CAP had increased protection of EEC agriculture, raised its degree of self-sufficiency, had led to a concomitant relative decline in extra-EEC imports and to considerable trade diversion; as a consequence the CAP had hindered the migration of labour out of agriculture into more profitable sectors and thus had reduced aggregate growth (Thorbecke, 1975).

2. A QUANTITATIVE APPROACH

Erik Thorbecke studied at the Netherlands School of Economics (now Erasmus University) from 1948 to 1951, where Professor Jan Tinbergen taught quantitative methods for economic research.

These lessons were applied when Erik Thorbecke was invited to contribute to a volume in honour of Jan Tinbergen, which was published in 1974. Together with A. Field a model of world trade was developed comprising 10 trading regions, with the Sino-Soviet bloc taken as exogenous; this was unique because till that time the most disaggregate model of world trade had distinguished only 3 regions. The model was demand oriented, of a short-term nature. The 10 regions were linked by import functions, specified differently for developed and developing regions. The model, containing 78 equations, was estimated and was shown to perform relatively accurately in explaining the changes in the endogenous variables over the sample period, 1953-1967.

Tinbergen's approach inspired many other quantitative publications in a number of areas; among others:

- The Theory of Quantitative Economic Policy with Application to Economic Growth Stabilization and Planning (1973)
- Planning Techniques for a Better Future (1976)
- A Multisectoral Framework for the Analysis of Labor Mobility and Development in LDCs (1988)
- A Multiplier Decomposition Method to Analyze Poverty Alleviation (1996)
- A Stochastic Social Accounting Matrix for Modeling (2003).

Erik Thorbecke did not confine himself to his study and lecture room. He went out to give advice to governments of many countries, among which Brazil, Indonesia, Kenya, Peru, the Philippines, and his own country, the US; he advised President L.B. Johnson on the world food problem (1967) and President W.J. Clinton on the GATT-round of negotiations (1994).

3. WILLEM JOHAN RUDOLF THORBECKE

When his father, Willem Johan Rudolf Thorbecke (1892-1989), died, Erik wrote 'Some Memories and Vignettes':

"My father was extremely quick witted and possessed a sharp sense of humour. He was a master of the repartee.

He was always very well dressed, well mannered and courteous. He was happiest when he could make you laugh, with him, or even at him.

He was an eternal optimist. To friends he could offer sound advice, with a light touch and good humour, emphasizing the positive elements of any situation, however bleak.

He had a keen intellectual curiosity, originality and independence of spirit. When in diplomatic service in China in the 1930's, he met with Teilhard de Chardin, and was much taken by the themes which would eventually culminate in 'The Phenomenon of Man' (1955).

My father was attracted by the search for a global, scientific and evolutionary humanism. He looked for universal, integrative forces and solutions. In particular, later as a political scientist and university professor, the themes of his books dealt with global integration, and also with improving East-West relations at a time, in the 1940's and 1950's, when this was not a particularly popular or even acceptable view.

Later, in the Hague, he met with Professor Jan Tinbergen and was much impressed by his attempt at global economic analysis and vision."

4. JOHAN RUDOLF THORBECKE

In the history of the Netherlands, Johan Rudolf Thorbecke (1798-1872) is a prominent prime-minister, if not the most important one. He designed a new constitution (1848), implemented in advance of his first term of office (1849-1853). His constitution has been described as an 'act of liberation' (Romein, 1979). It shifted power from the king and nobility to parliament. But it did more than that; Thorbecke insisted that the constitution should not be just a formal piece of paper, but a 'national force' for emancipation. And indeed, his constitution created the conditions for emancipation to continue and eventually to encompass all classes of society; this included the franchise, which was limited to wealthy men only in 1849, but evolved to include all adult citizens by 1919. He insisted that government should be truly public and open.

Earlier, in the 1820's and 30's, as a professor of diplomatic history, in Ghent and later in Leiden, he analyzed the changes that took place in the community of States in Europe, which provided a foundation for his thinking on a new constitution.

Thorbecke was three times prime-minister (1849-1853, 1862-1866, and 1871-1872). His legislation, other than the constitution, reflected the areas of additional interests:

- International trade; his legislation brought about the transition of the Netherlands to free trade and the abolition of slavery.
- Economic development; he foresaw an increasing role of the government, and promoted education and large scale infrastructure.

- Poverty alleviation; he was concerned that economic development through industrialization would be accompanied by poverty among the workers: 'It is not only the increase of national product that counts but equally its distribution'.

5. MEANDERING AND EQUILIBRIUM

The areas of Erik Thorbecke's interest and work are reflected in the themes of the conference in his honour 'Poverty, Inequality and Development', held at Cornell University, October 10-11, 2003; consequently, these themes are also represented in this volume.

In his closing address, Erik Thorbecke observed that he had 'meandered between topics' and – as this would imply a somewhat aimless movement – he added 'with some logic, I hope'. He observed furthermore that such a variety is not an option open to young academics nowadays; they have to specialize in a narrow field and stick to that. And: 'Interaction with colleagues is important; my research has evolved over time through interaction'.

In this perspective he then gave an overview of his work and career, which in 1952 started in the area of international economic relations, inspired by Professor J.B. Condliffe. In 1962, when advisor to the National Planning Institute of Peru, he was struck by the existing regional differences in income: 'It opened my eyes; and – henceforth – I wanted to work on development'. This – among others – resulted in a conference in Iowa and the volume: 'On the Theory and Design of Economic Development', edited with Irma Adelman.

Work in the World Employment Programme of the ILO started in 1972 and resulted in the Basic Needs Strategy. Being invited to Kenya to apply this approach caused a shift in interest again: to poverty; together with James Foster and Joel Greer a class of decomposable poverty measures was developed, still the standard today.

Meeting at the ILO with Richard Stone inspired research on the Social Accounting Matrix as a basis for development planning, leading – among others – to a World Bank publication in 1985.

In the 1980's interest rose in the role of institutions in economic development; a special issue of World Development explored this topic, edited by Erik Thorbecke together with, again, Irma Adelman.

The meandering between topics did not imply that interest in previous topics was lost; research in these areas continued as well.

Erik Thorbecke's list of publications counts 174 items now (reviews not included), with new studies still being added. The latest paper is with Henry Wan on East Asia's Development Model (May, 2004).

In his closing address Erik Thorbecke did not explicitly go into the internal logic of his meandering; but it can be found in his focus on economic development: in international economic relations, in economic integration, and above all for alleviating poverty; and on constructive policy making towards that end.

He did mention however another constant factor in his life, when he looked at Charla, his wife for nearly 49 years then, more than 50 years now, and said: 'a long-run equilibrium is a stable equilibrium'.

REFERENCES

Adelman, I. and E. Thorbecke (eds.) (1966), The Theory and Design of Economic Development, The Johns Hopkins Press, 1966.

Adelman, I. and E. Thorbecke (eds.) (1989), The Role of Institutions in Economic Development, World Development (special issue), September 1989.

Foster, J., J. Greer and E. Thorbecke (1984), A Class of Decomposable Poverty Measures, Econometrica, Vol. 52, 1984, pp. 761-766.

Fox, K.A., J.K. Sengupta and E. Thorbecke (1973), The Theory of Quantitative Economic Policy with Application to Economic Growth Stabilization and Planning, North-Holland, Amsterdam, 1973.

Jung, H.-S. and E. Thorbecke (1996), A Multiplier Decomposition Method to Analyze Poverty Alleviation, Journal of Development Economics, Vol. 48, 1996, pp. 253-277.

Pyatt, G. and E. Thorbecke (1976), Planning Techniques for a Better Future, ILO, Geneva, 1976.

Romein, J. (1979), Johan Rudolf Thorbecke (1798-1872), Ch. 27 in: J. and A. Romein (1979), Erflaters van onze beschaving ((Testators of our Civilization), Querido, Amsterdam, 1979.

Santiago, C.E. and E. Thorbecke (1988), A Multisectoral Framework for the Analysis of Labor Mobility and Development in LDCs: An Application to Postwar Puerto Rico, Economic Development and Cultural Change, Vol. 37, 1988, pp. 127-148.

Thorbecke, E. (1960), The Tendency Towards Regionalization in International Trade, Martinus Nijhoff, The Hague, 1960.

Thorbecke, E. (1961), The Attitude and Policy of the US with Reference to European Economic Integration, The American Review, September 1961.

Thorbecke, E. (1963), The Impact of the European Economic Community on the Pattern of World Trade, American Economic Review, May 1963, pp. 147-174.

Thorbecke, E. (1975), in cooperation with E. Pagoulatos, The Effects of European Economic Integration on Agriculture, Chapter 8 in: B. Balassa (ed.) (1975), European Economic Integration, North-Holland, Amsterdam, 1975.

Thorbecke, E. (1985), The Social Accounting Matrix and Consistency-Type Development Planning Models, in: G. Pyatt and J.I. Round (eds.) (1985), Social Accounting Matrices, A Basis for Planning, The World Bank, 1985.

Thorbecke, E. (1989), Willem Johan Rudolf Thorbecke (April 17, 1892 - May 5, 1989), Some Memories and Vignettes, Ithaca, May 8, 1989.

Thorbecke, E. (2003), Towards a Stochastic Social Accounting Matrix for Modeling, Economic Systems Research, Vol. 15, 2003.

Chapter 3

ON THE CONSISTENCY OF POVERTY LINES

Martin Ravallion and Michael Lokshin[i]
Development Research Group, World Bank, mravallion@worldbank.org and mlokshin@worldbank.org

1. INTRODUCTION

Poverty profiles — showing how a measure of poverty varies across sub-groups of a population — are widely used to inform policies for fighting poverty. A key ingredient is a set of poverty lines, to be used as deflators applied to sub-group specific distributions of income. Various methods are found in practice for setting poverty lines and the methodological choices made can matter greatly to the policy implications drawn. For example, a case study for Indonesia found virtually zero rank correlation between the regional poverty measures implied by two common methods of setting poverty lines (Ravallion and Bidani, 1994). This suggests that it is important to probe critically into the methods used to set poverty lines in practice.

Poverty lines are usually anchored to nutritional requirements for good health and normal activities. But there are many ways this can be done. There are two common methods of setting poverty lines in practice: the "Food-Energy Intake" (FEI) method and the "Cost-of-Basic Needs" (CBN) method.[ii] The FEI method finds the income or expenditure level at which pre-determined food-energy requirements are met in expectation within each sub-group. There is no explicit bundle of goods in the FEI method. The CBN method, by contrast, sets specific poverty bundles and costs them in each sub-group. The food bundles are typically anchored to nutritional requirements given prevailing diets, but allowances for non-food goods and services are also included.

The reviews alternative consistency criteria for poverty lines and tries to assess how well these two methods perform against those criteria. Section 2 discusses alternative theoretical foundations for defining the consistency of poverty lines. Section 3 then focuses on FEI poverty lines and argues that they are unlikely to be internally consistent for a reasonably broad concept of welfare. Section 4 turns to CBN poverty lines, and argues that these are potentially utility-consistent but whether they are in practice is an empirical question. Revealed-preference tests can be used to test consistency in terms of utility for given preferences. We then illustrate these tests using Russia's official poverty lines in Section 5. Conclusions can be found in Section 6.

2. CONSISTENCY OF POVERTY LINES IN THEORY

A poverty line can be defined as the money needed to achieve the minimum level of "well-being" that is required to not be deemed "poor." Thus everyone at the poverty line (no matter what sub-group they happen to belong) is deemed to be equally badly off, and all those below the line are worse off than all those above it. This much can be easily agreed. The more difficult question is what concept of well-being should serve as the anchor for poverty lines? For economists the obvious answer is "utility." A justification for utility consistent poverty lines can be found by applying standard welfare-economic principles to poverty measurement. These principles are that assessments of social welfare (including poverty measures) should depend solely on utilities, people with the same initial utility should be treated the same way, and social welfare should not be decreasing in any utility.

To formalize this approach to setting poverty lines, consider household i in sub-group j with characteristics x_{ij} (a vector).[iii] The household's preferences are represented by an interpersonally comparable utility function $u_j(q_{ij}, x_{ij})$. The household chooses its consumption vector q_{ij} to maximize utility. Notice that we allow the possibility that the same commodity bundle can yield different utility levels in different subgroups for households with the same characteristics. For example, a given bundle may yield a higher utility in a warm climate than a cold one, where more will be needed for clothing and energy.

The utility-consistent poverty line is the point on the consumer's expenditure function corresponding to a common reference utility level and prevailing prices. The consumer's expenditure function is $e_j(p_{ij}, x_{ij}, u)$, giving the minimum cost of utility u in sub-group j when facing the price vector p_{ij}. Let u_z denote the minimum utility level deemed to be needed to escape poverty. Consistency requires that this is a constant across all (*i*, *j*). The money metric of u_z defines a set of utility-consistent poverty lines:

$$z_{ij}^u = e_j(p_{ij}, x_{ij}, u_z) \text{ for all } (i, j) \tag{1}$$

When expenditure is deflated by such a poverty line one obtains a welfare metric with a number of desirable theoretical properties for policy analysis (Blackorby and Donaldson, 1987).[iv]

For economists, utility is the obvious anchor for setting poverty lines. However, it is not the only possible approach, and nor is it the approach that has had most influence on practices in applied work on poverty (as we will show in the following sections). Functioning-based concepts of well-being offer an alternative theoretical foundation for poverty measurement. Indeed, this can be viewed as an encompassing framework, for which utility consistency is a special case.

While versions of this approach go back a long way in philosophy and the social sciences, it can be characterized today in the terms of Amartya Sen's argument that "well-being" should be thought of in terms of a person's capabilities, i.e., the functionings ("beings and doings") that a person is able to achieve (Sen, 1985). By this view, poverty means not having an income sufficient to support specific normative functionings. Utility — as the attainment of personal satisfaction —can be viewed as one such functioning relevant to well-being (Sen, 1992, Chapter 3). But it is only one of the functionings that matter. Independently of utility, one might say that a person is better off if she is able to participate fully in social and economic activity, for example.

To formalize this approach, let a household's functionings be determined by the goods it consumes and its characteristics. The vector of actual functionings for household *i* in group *j* is:

$$f_{ij} = f_j(q_{ij}, x_{ij}) \tag{2}$$

where f_j is a vector-valued function. The quantities consumed are assumed to be utility maximizing, giving demand functions $q_{ij} = q_j(p_{ij}, y_{ij}, x_{ij})$ at total expenditure y_{ij}. One can also postulate that the household has preferences over functionings, for which $u_j(q_{ij}, x_{ij})$ is then a derived utility function, obtained by substituting (2) into the (primal) utility function defined over functionings (Ravallion, 1998).

Functioning-consistency requires that certain normative funtionings are reached at the poverty line in each sub-group. Let f_z denote the vector of critical functionings deemed to be needed to be not poor. (These are normative judgments, just as u_z is a normative judgment.) A commodity bundle, q_{ij}^c, is identified such that no functioning is below its critical value:

$$f_z \leq f_j(q_{ij}^c, x_{ij}) \tag{3}$$

There could well be more than one solution for q_{ij}^c satisfying (3). Each solution yields a set of functioning-consistent poverty lines $z_{ij}^c = p_{ij}q_{ij}^c$ when q_{ij}^c is valued at local prices. Two ways can be suggested for choosing a single functioning-consistent poverty line for each sub-group. The first possible way to resolve the indeterminacy of multiple solutions is to define z_{ij}^c as the minimum y such that:

$$f_z \leq f_j[q_j(p_{ij}, y, x_{ij}), x_{ij}] \tag{4}$$

Notice that one or more specific functionings will be decisive in determining z_{ij}^c, namely the functioning that is the last to reach its critical value as income rises. In this sense, the lowest priority functioning for the household will be decisive.

A second possible approach is to treat attainments as a random variable (i.e., with a probability distribution) and take a mean conditional on income and other identified covariates, including group membership. Then poverty lines are deemed to be functioning consistent if f_z is reached in expectation. This second approach is closer to current practices for an important class of methods for setting poverty lines, which we turn to in the next section.

3. THE FOOD-ENERGY-INTAKE METHOD

The FEI method can be interpreted as a special case of the functioning-based approach described above. The specialization is to focus on just one functioning, namely food-energy intake. The method finds the consumption expenditure or income level at which food energy intake is just sufficient to meet pre-determined food energy requirements for good health and normal activity levels. (Such caloric requirements are given in WHO, 1985, for example.) To deal with the fact that food energy intakes naturally vary at a given income level, the FEI method typically calculates an expected value of intake at given income. Figure 3-1 illustrates the method. The vertical axis is food-energy intake, plotted against income (or expenditure) on the horizontal axis. A line of "best fit" is indicated; this is the expected value of caloric intake at given income (i.e., the nonlinear regression function). By simply inverting this line, one finds the income z at which a person typically attains the stipulated food-energy requirement.[v] This method, or something similar, has been used often, including by Dandekar and Rath (1971), Osmani (1982), Greer and Thorbecke (1986), Paul (1989), Palmer-Jones and Sen (2001), and by numerous governmental statistics offices. It is probably the most common method found in practice in developing countries.

To explain the method more formally, let k denote food-energy intake, which is derived as the inner-product of the household's chosen consumption vector and a vector of calorific unit values. We take k to be a random variable. The stipulated requirement level is k^r which is taken to be fixed for given characteristics, such as age. As long as the expected value of food-energy intake conditional on total consumption expenditure, $E(k|y)$, is strictly increasing in y over an interval that includes k^r there will exist a FEI poverty line, z^{FEI}, defined implicitly by:

$$E(k|z^{FEI}) = k^r \tag{5}$$

Three points are notable. Firstly, the method is aiming to measure income poverty, rather than undernutrition. If one wanted to measure undernutrition, one would simply look at how many people had nutritional intakes $k \leq k^r$, ignoring incomes or consumption expenditures.

Secondly, the method is computationally simple. A common practice is to calculate the mean income or expenditure of a sub-sample of households whose estimated caloric intakes are approximately equal to the stipulated requirements. More sophisticated versions use regressions of the empirical relationship between food energy intakes and consumption expenditure. These can be readily used (numerically or explicitly) to calculate the FEI poverty line. The method avoids the need for price data; in fact, no explicit valuations are required.

Thirdly, the method automatically includes non-food consumption as long as one locates the total consumption expenditure at which a person typically attains the caloric requirement.

Can the FEI method assure that the resulting poverty lines will be consistent in terms of utility or capabilities more generally? To assess their utility consistency, consider first how FEI poverty lines respond to differences in relative prices, which can of course differ across the sub-groups (such as regions) being compared in the poverty profile and over time. For example, the prices of many non-food goods are likely to be lower relative to foods in urban than in rural areas. This will probably mean that the demand for food and (hence) food energy intake will be lower in urban than in rural areas, at any given real income. But this does not, of course, mean that urban households are poorer at a given expenditure level.

To see the problem more clearly, let there be two composite goods, "food" and "non-food" consumed in quantities q_0 and q_1 respectively, and let p denote the relative price of the non-food good. The utility-consistent poverty line is (simplifying notation) $z^u = e(p, u_z)$. By the envelope property, the derivative of z w.r.t p is simply the level of consumption of non-food goods for someone at the poverty line. As long as both goods are consumed, a higher relative price of non-food goods must mean a higher poverty line in terms of food.

However, this no longer holds using the FEI method to set the poverty line. Then one fixes instead the value of q_0 at the (unique) level needed to achieve the stipulated food-energy level. The corresponding FEI poverty line is z^{FEI} such that $q_0(p, z^{FEI})$ is the required food consumption, where $q_0(p, y)$ denotes the food demand function. The derivative of the FEI poverty line w.r.t. the price of non-food goods is now:

$$\frac{\partial z^{FEI}}{\partial p} = -\frac{q_{0p}(p, z^{FEI})}{q_{0y}(p, z^{FEI})} \tag{6}$$

where the subscripts "p" and "y" denote the partial derivatives w.r.t. those variables. It is reasonable to assume that non-food goods are normal ($q_{0y} > 0$). The sign of (6) will then depend on whether food and non-food goods are (uncompensated) substitutes ($q_{0p} > 0$) or complements ($q_{0p} < 0$). In the former case, the FEI poverty line will decrease with an increase in the price of non-food goods. A lower relative price of non-food goods in urban areas, for example, will perversely yield a higher poverty line using this method. The FEI poverty lines will then fail our consistency requirement since the consistent poverty lines must be increasing in all prices, given that this must hold for the consumer's expenditure function. Utility consistency would requite that food and non-food goods are complements.

There are other reasons to question the utility consistency of FEI poverty lines. Even if relative prices do not differ, the relationship between food energy intake and income will shift according to differences in tastes, activity levels and publicly-provided goods. There is nothing in the FEI method to guarantee that these differences are ones that would normally be considered relevant to assessing welfare. For example, tastes can differ across sub-groups even if relative prices do not. At given relative prices and real total expenditure, urban households may simply have more expensive food tastes than rural households; they eat more animal protein and less calories from starchy food staples, or simply eat out more often. Thus they pay more for each calorie, or (equivalently) food energy intake will be lower at any given real expenditure level. It is unclear why we would deem a person who chooses to buy fewer and more expensive calories as poorer than another person at the same real expenditure level. For these reasons, the real income at which an urban resident typically attains any given caloric requirement will tend to be higher than in rural areas. And this can hold even if the cost of living is no different between urban and rural areas.

Consider Figure 3-2, which gives a stylized food energy-income relationship for "urban" and "rural" areas. The urban poverty line is z_u while the rural line is z_r. However, the aforementioned concerns lead us to question whether the differential z_u/z_r could provide any reasonable approximation to the true differential in the cost of the same standard of living. The distribution of caloric intakes can readily vary between groups such that the regression function $E(k|y)$ also varies with the characteristics of those groups, and there is no reason to assume that $E(k|y)$ ranks welfare levels correctly at a given value of y. A differential in poverty lines can then appear, making the poverty profile utility inconsistent.

It is clear from these observations that one should then be wary of poverty lines generated by the FEI method if the aim is to reduce utility poverty; people at the poverty line in different sub-groups could well have very different levels of welfare defined as utility. Indeed, it is quite possible to find that the "richer" sector (by the agreed metric of utility) tends to spend so much more on each calorie that it is deemed to be the "poorer" sector. That has been found to be the case in studies of the properties of FEI poverty profiles for Indonesia (Ravallion and Bidani, 1994) and Bangladesh (Ravallion and Sen, 1996; Wodon, 1997).

Problems also arise in comparisons over time. Suppose that all prices increase, so the cost of a given utility must rise. There is nothing to guarantee that the FEI-based poverty line will increase. That will depend on how relative prices and tastes change; the price changes may well encourage people to consume cheaper calories, and so the FEI poverty line will fall. Wodon (1997) gives an example of this problem in data for Bangladesh. The FEI poverty line fell over time even though prices generally increased.

The potential utility inconsistencies in FEI poverty lines are worrying when there is mobility across the subgroups of the poverty profile, such as due to inter-regional migration. Suppose that, as the above discussion has suggested may well happen. The FEI poverty line has higher purchasing power in urban areas than rural areas. Consider someone just above the FEI poverty line in the rural sector who moves to the urban sector and obtains a job there generating a real gain less than the difference in poverty lines across the two sectors. Though that person is better off in that she can buy more of all goods, including food, the aggregate measure of poverty across the sectors will show an increase, as the migrant will now be deemed poor in the urban sector. Indeed, it is possible that a process of economic development through urban sector enlargement, in which none of the poor are any worse off, and at least some are better off, would result in a measured increase in poverty.

What about the functioning consistency of FEI poverty lines? By construction, the FEI lines are consistent with respect to one important functioning, namely reaching nutritional requirements. The issue is whether that constitutes a good basis for poverty comparisons. It might be if one deemed food-energy intake to be the sole functioning of interest. But there appears to be wide agreement that it is not, even among exponents of the FEI method. For if one deemed calories to be sufficient, none of this extra work would be necessary — all one would do is measure caloric shortfalls relative to requirements (all of which are already needed as data to implement this method of setting poverty lines). The FEI method acknowledges (at least implicitly) that meeting food-energy requirements is not enough.

To believe that FEI poverty lines are consistent for some broader set of functionings we must assume that meeting nutritional requirements has a low priority for people, for only then can we be sure that all other functionings have been reached once nutritional requirements have been reached. That is surely implausible on *a priori* grounds; if anything one would expect that food energy requirements had a relatively high priority.

In summary, a FEI-based poverty profile will not in general be utility consistent. Nor is functioning consistency likely to hold over a broader set of functionings. Next we turn to the main alternative method found in practice.

4. THE COST-OF-BASIC-NEEDS METHOD

The CBN method stipulates a consumption bundle deemed to be adequate for "basic consumption needs," and then estimates its cost for each of the subgroups being compared in the poverty profile. This is the approach of Rowntree in his seminal study of poverty in York, England, in 1899, and there have been numerous examples since, including the official poverty lines for the U.S.[vi] Some form of functioning consistency is assured by construction, since various valued functionings are essentially the starting point for defining "basic consumption needs." The poverty bundle is typically anchored to food-energy requirements consistently with common diets in the specific context. However, allowances for non-food goods are also included, to assure that basic non-nutritional functionings are assured. We give an example of how CBN poverty lines are constructed in Section 6, when we discuss Russia's poverty lines.

Superficially, the CBN method looks like a more promising route to utility-consistent poverty lines. The CBN poverty line can be written as the expenditure needed to achieve a specific bundle of goods. Similarly, the "ideal" utility-consistent poverty line in equation (1) can be written:

$$z_{ij}^{u} = p_{ij}q_{j}(p_{ij}, x_{ij}, u_{z}) \tag{7}$$

The CBN method will be utility consistent if the right bundle is used, corresponding to the relevant points on the utility-compensated (Hicksian) demand functions.

However, there is nothing to guarantee that the bundles of goods built into CBN poverty lines lie on the compensated demand functions, at the (common) reference level of utility (as in equation 7). Thus it is important to have some way of assessing a set of CBN poverty bundles. We explore one approach below, following Ravallion and Lokshin (2004).

A common problem in setting CBN poverty lines is missing data on non-food prices. A number of solutions have been proposed (as reviewed in Ravallion, 1998). The most common practice is to divide the food component by an estimate of the budget share devoted to food. For example, the widely used poverty line for the U.S. developed by Orshansky (1963) assumes a food share of one third, which was the average food share in the U.S. at the time. The total poverty line was set at three times the food poverty line.

However, the basis for choosing a food share is rarely transparent, and very different poverty lines can result, depending on the choice made. Why use the average food share, as in the Orshansky line? Whose food share should be used? Arguably a more appealing approach is to set an allowance for non-food goods that is consistent with demand behavior at (or in a region of) the food poverty line as proposed in Ravallion (1994). This will not be an issue in our empirical application (for which a complete set of goods is specified), but it may generate further concerns about consistency in other applications.

In practice, the most common application is likely to be the geographic poverty profile, so this is the case we focus on in the following exposition. Each geographic area (which could be a country) has its own poverty line, which is the cost in that area of a bundle of goods specific to that area.[vii]

It is convenient to change notation slightly such that $q_i = (q_i^1, ..., q_i^m)$ is the *m*-vector giving the CBN poverty bundle for region i=1,..n. (The bundle can also vary with household characteristics, but we ignore this to simplify notation.) The corresponding price vector is p_i and the poverty line in region *i* is $z_i = p_i q_i$. Let $r_i = (p_i^1 / z_i, ..., p_i^m / z_i)$ denote the vector of price relatives for region *i*, normalized by the poverty line, and let $P \equiv \{r_i, i = 1,...,n\}$ denote the set of all price relatives.

We define the *n*x*n* quantity-index matrix ***Q*** for which the *i*'th row and *j*'th column give the cost of *j*'s poverty bundle when valued at *i*'s price relatives:

$$Q_{ij} \equiv r_i q_j = \frac{p_i q_j}{p_i q_i} \tag{8}$$

We use the ***Q*** matrix to compare poverty bundles across regions; the higher Q_{ij} the higher the value of the poverty bundle for region *j* when judged by its cost in region *i*. The quantity index ranks poverty bundles across regions conditional on the price relatives.

We will say that the bundle for region *k* is "unconditionally higher" than the bundle for region *j* if $Q_{ik} \geq Q_{ij}$ for all r_i in ***P***. This means that all elements of the *j*'th column of ***Q*** are greater than the corresponding elements of the *k*'th column. There is no guarantee that such a ranking is possible; that is an empirical question.

To provide a summary statistic for the value of each region's poverty line we can calculate the simple mean quantity index formed by taking the column totals of the ***Q*** matrix; we write this index as

$$\overline{Q}_j = \sum_{i=1}^{n} Q_{ij} / n.$$

Finding that $\overline{Q}_j > \overline{Q}_k$ implies that bundle *j* dominates *k* at least partially (for some price relatives in ***P***), though (of course) not necessarily fully.

Can we decide whether a set of CBN poverty lines are utility consistent based on revealed preference theory? Consider, two regions, A and B, each of which has a poverty line, which is the cost in each region of pre-specified bundles of goods specific to each region. Our definition of consistency requires that these two bundles yield the same utility and are both utility-maximizing in their respective regions for someone at the poverty line.

If preferences are identical in the two regions, then there is a straightforward revealed preference test. This requires that the poverty line for A is no greater than the cost in region A of B's bundle, for otherwise the bundle in B is affordable when A was chosen, implying that A is preferred. Similarly, the region B poverty line cannot be greater than the cost in that region of the bundle for A. If this test fails than we can reject consistency for a broad class of possible preferences, though passing the test does not assure consistency for all possible preferences.

To outline the revealed preference test in more formal terms, assume that the (unknown) preferences over commodities of those living in region *i* can be represented by a utility function $u_i(.)$. (To simplify notation we treat households as homogeneous in all respects except their income and location, so we can drop the "*x*" for non-income characteristics from all functions, but allowing the function itself to vary by location.) Preferences are allowed to vary regionally due to (*inter alia*) differences in climate or differences in endowments of local public goods. We make the standard assumption that $u_i(.)$ traces out strictly convex indifference curves (though this can be weakened somewhat).

Ravallion and Lokshin (2004) propose the following two criteria for the utility consistency of CBN poverty lines relative to the preferences in region *i*:

$$z_i = e_i(p_i, u_z) \tag{9.1}$$

$$u_i(q_i) = u_i(q_j) = u_z \text{ for all } j \tag{9.2}$$

The testable implication of these two conditions is that $Q_{ij} \geq 1$ for all *j*. To see why, suppose instead that $Q_{ik} < 1$ for some region *k* i.e., $p_i q_k < p_i q_i$. Then the bundle q_k was affordable in region *i* with the expenditure required for obtaining q_i. However, for consistency, q_i is the utility-maximizing bundle for someone at the poverty line in region *i*; furthermore, given convex indifference curves, q_i is the unique such bundle. Then, q_i must have been strictly preferred to q_j ($u_i(q_i) > u_i(q_j)$), which contradicts utility consistency.

By repeating our test for successive rows of the ***Q*** matrix we can test consistency across the complete set of underlying (unknown) preferences. So the key testable implication of consistent poverty lines across the full set of preferences is that none of the elements of the ***Q*** matrix should be below unity.

A number of remarks can be made about our test.

(i) It is possible to find that $Q_{ij} \geq 1$ but $Q_{ji} < 1$. In other words, we may be unable to reject utility consistency between the bundles for regions *i* and *j* when assessed using *i*'s price relatives, yet we can reject it when using *j*'s. If we find that $Q_{ij} \geq 1$ but $Q_{ji} \geq 1$ we will say that the bundles *i* and *j* are <u>mutually utility consistent</u>.

(ii) Our test is necessary for utility consistency, but it is not sufficient. It is possible to find that $Q_{ij} \geq 1$ and yet bundles *i* and *j* do not yield the same utility when judged by *i*'s preferences. Figure 3-3 illustrates this point. Four bundles of two goods are identified. Point B represents the poverty bundle for region B, with the indifference curve indicated, while A, C and D are the bundles for three other regions. When assessed by region B's preferences, we can reject consistency between A and B; bundle A must be on a lower indifference curve than B. However, we cannot reject for C and D happen to be utility consistent with B; as drawn, C and B are consistent, but we do not of course know the actual indifference curves in practice.

(iii) Our test allows the possibility that preferences over commodities differ across the poverty profile, but it does so in a special way, namely that one compares the poverty bundles of different regions at a common utility function. The rejection of utility consistency could reflect heterogeneity in preferences.

(iv) This is a joint test of the two consistency requirements in (10.1) and (10.2), and if one fails to hold then the test loses all power to detect whether the other holds. For example, suppose that the bundle of goods on which a poverty line is based would not be chosen by someone at the poverty line income given the prevailing prices. Then it can still satisfy (10.2) even though our quantity index is less than unity.

5. CASE STUDY FOR RUSSIA

We shall now apply the revealed preference tests described above to Russia's official poverty lines.[viii] These were established under guidelines developed by the Ministry of Labor and Social Development (MLSD, 2000).

5.1 Russia's poverty lines

The poverty line is defined as the cost of specific baskets of goods and services that are deemed necessary for an individual to maintain health and a minimum activity levels, both personal and social, taking account of the geographic setting (notably climate). The actual compositions of goods and services that enter the poverty baskets are determined by local governments. An inter-ministry expert committee reviews the draft consumer baskets submitted by the local governments and provides recommendations to the Federal Government, which makes the final decision on the composition of the regional baskets.[ix] The expert committee evaluates the nutritional composition of every regional basket as well as the composition of the non-food components (VTsUZH, 2002).

The food baskets are defined based on nutritional requirements for calories, proteins, fats, and carbohydrates for various groups of people defined by age and gender. The baskets vary across the 16 geographical zones of Russia, to account for calorific differences by climatic zones and for regional differences in food consumption patterns. The caloric requirements for adult males, for example, range from 3030 kcal per day for the northern regions of Russia to 2638 kcal per day for the warmer zones. Norms for the consumption of proteins and carbohydrates can also vary substantially across regions. The final food poverty bundles comprise 34 items, which differ between regions.

Three zones for non-food goods and three zones for services/utility baskets are defined according to climatic conditions in Russia. The basket for non-food goods provides detailed quantities to be consumed by six groups of individuals. These groups are similar to the groups used in the construction of the food basket, except that separate baskets for non-food goods are defined for elderly men and women. The service basket consists of consumption norms for seven main utilities. While the food and non-food baskets are defined at the individual level, the service baskets are defined on a per capita basis.

The non-food bundles consist of a number of personal items and some consumer durables. The non-food goods include specific items of clothing, footwear, pens and notebooks. Goods for the household's collective use include furniture (table, chair, chest of drawers, mirror, etc.), appliances (TV, refrigerator, clocks,...), kitchen items (plates, pots and pans, silverware), as well as towels, sheets, blankets, and pillows. Every item in the non-food bundle has an approximate usage time that varies for different age-gender groups. For example, adult males aged 18 to 59 are supposed to use one coat for seven years, while the norm for male pensioners is 10 years. A blanket has a life-time of 20 years. Every prime age woman is entitled to five underwear with amortization period of 2.4 years and two bras every three years.

The services bundle includes allowances for housing, heating, electricity, hot and cold water, gas and transportation.[x] The norms for heating and electricity vary by zones. In the cold climate zones the per person heat consumption is equal to 8.0 Gcal (Giga calories) per year while in the warmer zones it is only 5.4 Gcal per year per person.

Price information on the items in the poverty baskets is collected quarterly by the Russian Central Statistical Agency ("GosComStat") in 203 cities and towns of Russia for 196 food and non-food items and services. The poverty lines for every geographical zone are calculated by multiplying the quantities of the items in the baskets by the corresponding prices in an appropriate city or town.

In order to construct a poverty line for a particular region the cost of the food basket corresponding to this region should be added to the regional costs of the non-food goods and services. We can define 23 geographical zones that correspond to the combinations of food, non-food goods and services (Ravallion and Lokshin, 2004). One hundred and thirty eight distinct baskets are specified as a combination of these geographical areas and the various demographic groups.

Ravallion and Lokshin (2004) give the detailed poverty lines for Russia by type of individual and location. They show that the poverty lines tend to decline from north to south. The household poverty line is determined by summing up the individual poverty lines of the household members. For our analysis we use the poverty lines for a typical household that consists of two parents (a male aged 18 to 59 and a female aged 18 to 54) and two children (one child 0 to 6 years old and one child 7 to 15 years old). We call this the "reference household."

Before we turn to our tests, it is worth reflecting on why we might expect inconsistencies in these poverty lines. Partial functioning consistency seems reasonably well assured, given that the lines are anchored to food-energy requirements specific to each geographic and demographic groups. Consistency in terms of other capabilities is less clear. The long list of essential non-food goods and services reflects perceptions of what is needed to maintain minimal activity levels in the specific setting, recognizing that this is more than a matter of adequate nutrition, but requires expenditures on clothing, housing, heating and transportation. Arguably there is a sense in which consistency with a reasonably broad set of capabilities for active participation in Russian society is built into this method of setting poverty lines.

However, no obvious attempts are made to assure utility consistency (in any explicit sense) of the poverty lines across regions. There can be random differences. But there are also likely to be systematic differences arising from two sources. Firstly, perceptions of what constitutes "poverty" will undoubtedly differ, with richer provinces tending to have higher real poverty lines (just as is found across countries; see Ravallion, 1994). (Clearly, this could generate functioning inconsistencies too.) Secondly, and probably working against the first factor, resource poor local governments in Russia may perceive an incentive to inflate their poverty lines to attract extra resources from the center. According to the Law on Social Protection any family or single person whose average per-capita income is below the regional poverty line is entitled to receive government social assistance. The Federal Government allocates funds for social protection based on the number of poor in the region. Therefore, the local governments have an incentive to inflate their baskets to secure a larger share of government transfers to the region. Furthermore, this incentive may well be stronger for poorer local government areas. On balance, we cannot predict which direction the bias might go.

5.2 Revealed preference tests

The Q matrix of Laspeyres quantity indices for the reference household is given in Table 3-1. (Ravallion and Lokshin, 2004, give the costs of the poverty baskets across the 23 zones for the reference household.) Comparing columns of the Q matrix, it is evident that the two most generous poverty bundles are those for zones 2 and 3, which make up Siberia. One of these dominates all other bundles, though 2 and 3 cannot be ranked unambiguously; for some price vectors, the zone 2 bundle dominates while for others it is zone 3. However, there can be no doubt which is the least generous bundle judged by the quantity index; the bundle for zone 20 is unconditionally lower than that for all other bundles, i.e., $Q_{i20} < Q_{ij}$ for all $j \neq 20$. Zone 20 is the small region of Kalmukia in the southwest.

Figure 3-4 gives the results of our revealed preference test based on the quantity matrix in Table 3-1. The elements of Q that are less than 1 (i.e., the test is not passing) are shaded. Overall, the test is passed for only 281 out of 529 elements of Q matrix.[xi] Strikingly, of the 253 distinct pairs of bundles, mutual utility consistency is rejected for all except six pairs, namely the pairs (10,17), (10,23), (11,9), (11,15), (23,13) and (23,17). Looking at the first row, we find that utility consistency at common preferences is rejected for all but two of the (i, j) combinations. Consistency is rejected for all regions when judged by region 3's preferences. Rejections tend to become less common as one moves down the table. The test comes very close in region 16, with only one narrow ($Q_{16,20} = 0.984$) rejection.

Zone 20 stands out as unusual in three respects. Firstly, as we have noted, it is the bundle with the lowest quantity index for all prices. Secondly, it is the only bundle that passes out test; judged by zone 20's preferences, we cannot reject consistency across all the bundles. Thirdly, the bundle for zone 20 accounts for more rejections than any other zone. Indeed, there is no zone for which consistency with zone 20 passes. Clearly these three observations are related. The low value of the zone 20 bundle makes it more likely to be utility consistent, and more likely to differ from the bundles elsewhere.

Why are our revealed preference criteria rejected so strongly? As we noted in the last section, the decentralized process generating Russia's regional poverty bundles may well yield utility inconsistencies. However, we cannot rule out geographic heterogeneity in preferences as an alternative explanation. Figure 3-5 maps the mean quantity indices ($\overline{Q}_j$). There is a marked north-south difference, which is clearly correlated (negatively) with

Table 3-1. [Matrix of Laspeyres quantity indices for the reference household]

		Baskets																						
Prices	No. zones test fails	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	20	1.000	1.132	1.123	0.982	0.778	0.825	0.890	0.792	0.848	0.750	0.851	0.803	0.740	0.953	0.852	0.736	0.744	0.879	0.781	0.720	0.782	0.762	0.753
2	22	0.896	1.000	0.998	0.883	0.697	0.740	0.788	0.704	0.755	0.666	0.753	0.718	0.662	0.849	0.762	0.657	0.666	0.785	0.693	0.641	0.697	0.674	0.664
3	21	0.897	1.001	1.000	0.881	0.699	0.741	0.792	0.704	0.757	0.668	0.756	0.718	0.663	0.853	0.763	0.659	0.667	0.789	0.694	0.644	0.700	0.677	0.666
4	19	1.020	1.153	1.147	1.000	0.795	0.841	0.909	0.806	0.866	0.767	0.869	0.819	0.754	0.975	0.869	0.751	0.759	0.898	0.796	0.736	0.799	0.777	0.768
5	7	1.292	1.470	1.470	1.278	1.000	1.066	1.154	1.021	1.095	0.974	1.103	1.048	0.961	1.244	1.101	0.951	0.961	1.138	1.014	0.941	1.017	0.992	0.984
6	12	1.212	1.379	1.379	1.199	0.938	1.000	1.083	0.958	1.027	0.913	1.035	0.984	0.902	1.167	1.033	0.892	0.902	1.068	0.951	0.883	0.954	0.930	0.923
7	17	1.119	1.273	1.273	1.107	0.866	0.924	1.000	0.885	0.949	0.843	0.956	0.908	0.833	1.077	0.954	0.824	0.833	0.986	0.879	0.815	0.881	0.859	0.853
8	10	1.273	1.428	1.431	1.260	0.986	1.054	1.122	1.000	1.070	0.950	1.074	1.032	0.948	1.216	1.085	0.933	0.948	1.119	0.988	0.920	0.998	0.963	0.952
9	13	1.189	1.334	1.337	1.178	0.921	0.984	1.049	0.934	1.000	0.888	1.003	0.964	0.886	1.136	1.014	0.872	0.886	1.045	0.924	0.860	0.932	0.900	0.889
10	3	1.352	1.509	1.507	1.329	1.050	1.112	1.187	1.058	1.135	1.000	1.132	1.077	0.995	1.278	1.146	0.988	1.001	1.183	1.041	0.967	1.051	1.017	1.000
11	13	1.191	1.332	1.330	1.171	0.925	0.981	1.047	0.933	1.002	0.883	1.000	0.951	0.878	1.129	1.011	0.872	0.883	1.044	0.919	0.853	0.927	0.897	0.882
12	11	1.253	1.397	1.396	1.230	0.974	1.031	1.101	0.981	1.054	0.929	1.050	1.000	0.922	1.185	1.063	0.917	0.928	1.098	0.965	0.896	0.974	0.941	0.926
13	2	1.346	1.505	1.509	1.333	1.043	1.113	1.185	1.057	1.131	1.003	1.131	1.089	1.000	1.280	1.146	0.986	1.001	1.182	1.043	0.971	1.052	1.016	1.002
14	18	1.059	1.184	1.180	1.039	0.823	0.871	0.930	0.831	0.891	0.787	0.889	0.846	0.779	1.000	0.899	0.775	0.785	0.927	0.817	0.757	0.823	0.796	0.784
15	14	1.173	1.333	1.333	1.160	0.908	0.968	1.047	0.927	0.994	0.884	1.002	0.952	0.873	1.129	1.000	0.864	0.873	1.033	0.920	0.854	0.923	0.900	0.892
16	1	1.359	1.537	1.539	1.346	1.052	1.123	1.206	1.071	1.150	1.019	1.154	1.100	1.011	1.304	1.158	1.000	1.012	1.196	1.063	0.984	1.066	1.036	1.022
17	3	1.344	1.510	1.515	1.329	1.040	1.110	1.189	1.057	1.131	1.003	1.135	1.088	0.999	1.281	1.145	0.985	1.000	1.180	1.046	0.972	1.053	1.021	1.009

Table 3-1 [Cont.]

Baskets																								
Prices	No. zones test fails	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
18	16	1.139	1.279	1.283	1.126	0.881	0.940	1.007	0.895	0.958	0.849	0.961	0.921	0.846	1.085	0.970	0.835	0.847	1.000	0.886	0.824	0.892	0.865	0.855
19	8	1.286	1.446	1.447	1.274	0.996	1.062	1.134	1.012	1.083	0.962	1.084	1.042	0.955	1.224	1.095	0.944	0.956	1.130	1.000	0.929	1.006	0.972	0.961
20	0	1.393	1.555	1.559	1.379	1.079	1.149	1.223	1.094	1.166	1.037	1.168	1.125	1.031	1.319	1.183	1.019	1.032	1.217	1.077	1.000	1.085	1.049	1.035
21	9	1.283	1.433	1.437	1.271	0.994	1.059	1.127	1.008	1.074	0.956	1.077	1.036	0.950	1.216	1.090	0.939	0.951	1.122	0.992	0.921	1.000	0.967	0.954
22	6	1.316	1.477	1.477	1.308	1.024	1.090	1.160	1.038	1.107	0.986	1.111	1.074	0.982	1.254	1.122	0.967	0.982	1.157	1.024	0.955	1.030	1.000	0.987
23	2	1.335	1.500	1.501	1.330	1.041	1.111	1.182	1.052	1.131	1.004	1.130	1.093	1.000	1.282	1.143	0.986	1.001	1.182	1.043	0.974	1.049	1.017	1.000
$\overline{Q}_j$		1.205	1.355	1.355	1.191	0.935	0.995	1.066	0.949	1.016	0.901	1.018	0.973	0.894	1.149	1.026	0.885	0.896	1.059	0.937	0.870	0.943	0.914	0.903

temperature; Figure 3-6 maps mean temperatures.[xii] The cooler the climate, the more generous the bundle as measured by the mean quality index. This suggests that the differences in the consumption bundles may well reflect differences in the commodities needed to reach the same utility level in different climates.

However, climate differences do not account for the violations of our revealed preference tests. By superimposing the temperature map (Figure 3-6) on the zones for which distinct poverty lines are identified (Figure 3-5) we can identify four distinct clusters of zones within a close range of temperatures, as identified in Table 3-2, which also give results of our revealed preference tests within each of these clusters. Again, rejections are indicated for about half the cases. Mutual utility consistency is rejected for every pair within each temperature band.

6. CONCLUSION

We have argued that possibly the most common method of setting poverty lines in practice — whereby the poverty line is the income at which pre-determined food-energy requirements are met in expectation in each subgroup — is unlikely to be utility consistent. Nor is this method likely to be consistent in terms of a broader set of functionings.

The poverty lines obtained by the main alternative method found in practice — the "cost-of-basic needs" method — have the potential to be utility consistent, and consistent for a broader set of normative functionings than reaching adequate nutritional status. Whether either holds in practice is a moot point. The specification of "basic consumption needs" is typically motivated by ideas of certain minimum functionings, notably (but not only) the ability to secure nutritional requirements. Their utility consistency is less obvious. In cases in which a complete set of basic consumption needs has been specified, we have shown that utility consistency for given preferences implies an empirical test, drawing on the theory of revealed preference.

As a case study, we have applied revealed preference tests to the official poverty lines for Russia. We find that we can generally reject utility consistency. Indeed, for only one region's price relatives do we find that our test passes. For all other region's prices, we reject consistency across at least one other regional bundle.

These rejections of our revealed preference tests may stem in part from underlying heterogeneity in preferences. The correlation we find across areas between the value of the Russian poverty bundles and mean temperature is suggestive of climatic differences in preferences, such that the same consumption does not yield the same utility in markedly different climates. Indeed, finding more generous bundles in colder climates is to be expected if the poverty lines are in fact utility consistent.

However, we still find numerous rejections of utility consistency when we control for climatic differences — by repeating our test for clusters of geographic areas within the same temperature band. The evidence of utility inconsistencies that we find in Russia's official poverty lines could well stem from the decentralized administrative process generating the poverty bundles. This may be less of a problem in settings in which the task of setting the normative bundles is more centralized.

Table 3-2. [Revealed preference tests for clusters of zone within common temperature bands]

Cluster 1: 8-10^0 Celsius zone

	Zones	
	20	22
20	1.000	1.049
22	0.955	1.000

Cluster 2: 2-4^0 Celsius zone

	Zones			
	5	15	16	18
5	1.000	1.101	0.951	1.138
15	0.908	1.000	0.864	1.033
16	1.052	1.158	1.000	1.196
18	0.881	0.970	0.835	1.000

Cluster 3: 0-2^0 Celsius zone

	Zones			
	7	8	9	13
7	1.000	0.885	0.949	0.833
8	1.122	1.000	1.070	0.948
9	1.049	0.934	1.000	0.886
13	1.185	1.057	1.131	1.000

Cluster 4: –4 - –2^0 Celsius zone

	Zones		
	10	11	14
10	1.000	1.132	1.278
11	0.883	1.000	1.129
14	0.787	0.889	1.000

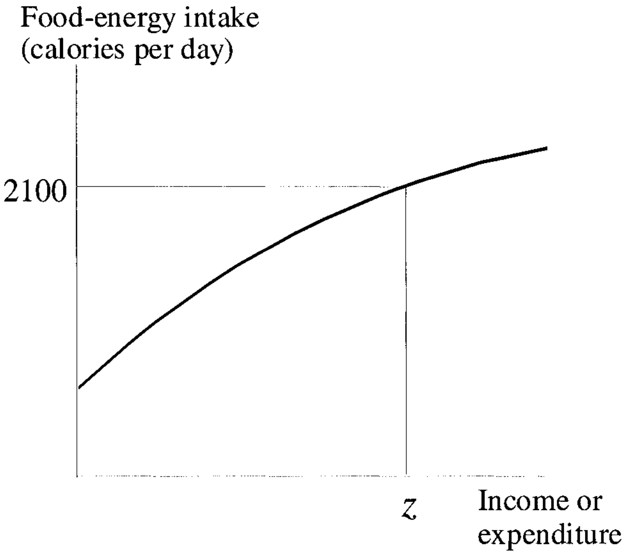

Figure 3-1. [The food-energy intake method of setting poverty lines]

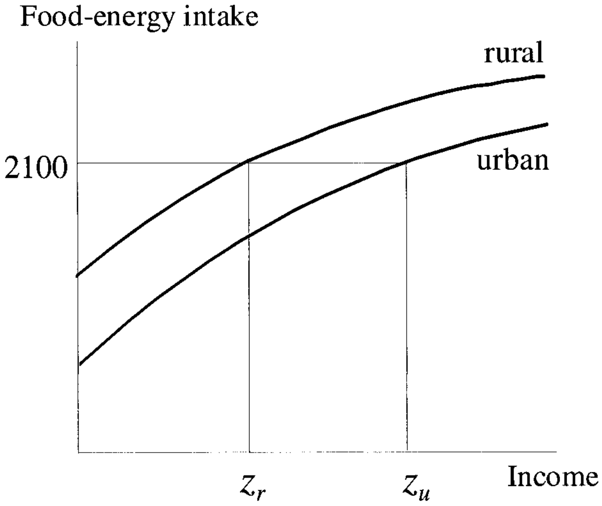

Figure 3-2. [Multiple poverty lines with the FEI method]

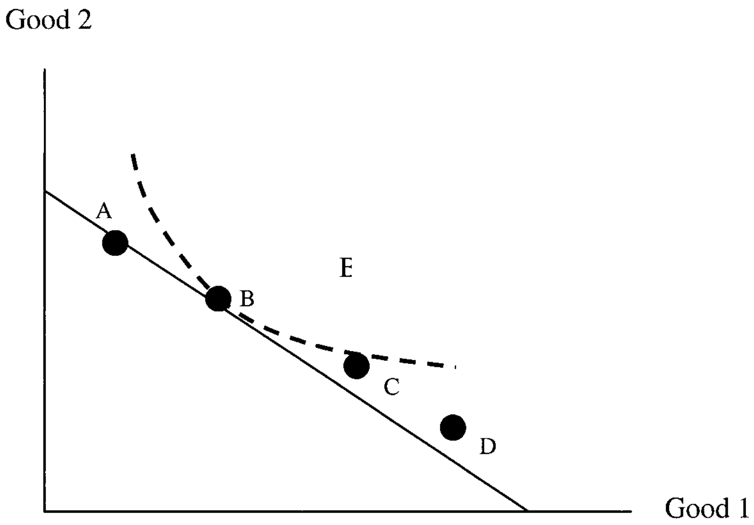

Figure 3-3. [Consistency test for four bundles]

Note: Consistency with bundle A is rejected for B but the test is inconclusive for C and D without knowing preferences.

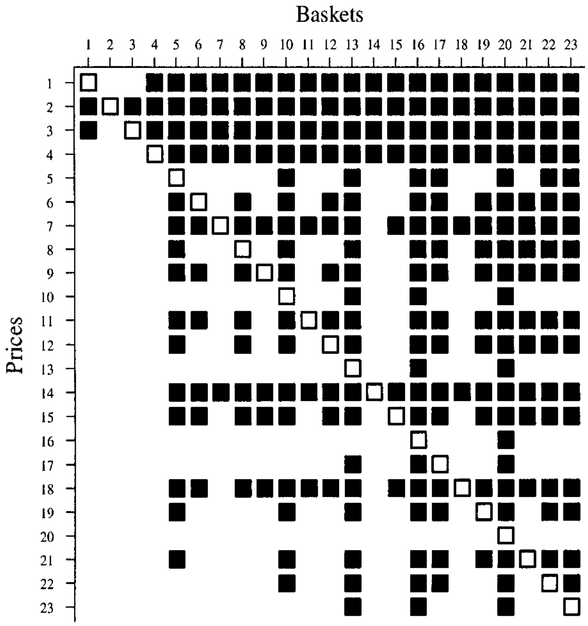

Figure 3-4. [The results of the consistency test for Russian regional poverty lines]

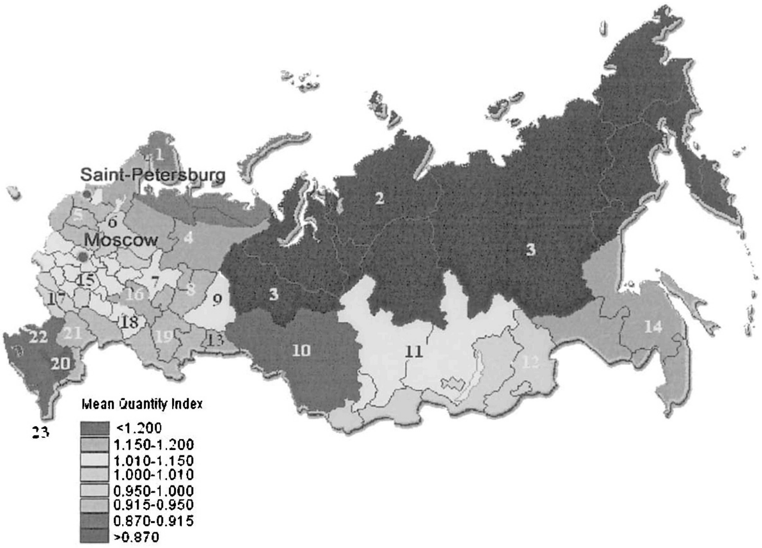

Figure 3-5. [Mean quantity index by zone]

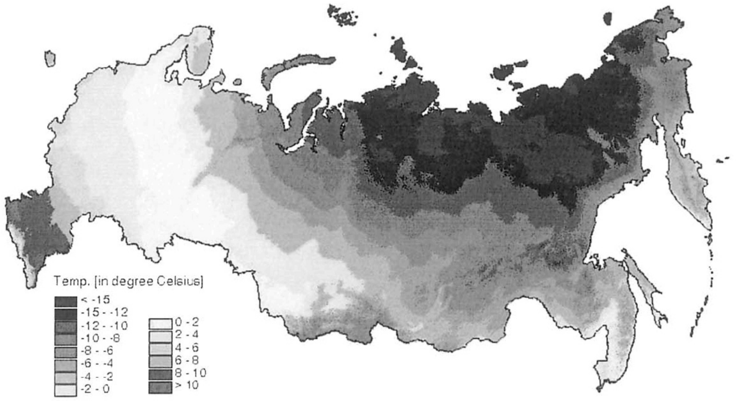

Figure 3-6. [Mean annual temperature in Russia]

Sources: Potsdam Institute for Climate Impact Research and Land Use Change Project IIASA -International Institute for Applied Systems Analysis, Austria (www.iiasa.ac.at/)

NOTES

[i] For comments the authors are grateful to Stefan Klonner and participants at the 2003 Cornell University Conference of "Poverty, Inequality and Development" in honor of Erik Thorbecke. These are the views of the authors, and need not reflect those of the World Bank or any affiliated organization. Correspondence: mravallion@worldbank.org, mlokshin@worldbank.org.

[ii] For an overview of alternative methods found in practice see Ravallion (1998). Note that we refer here to commonly used "objective" poverty lines. Subjective poverty lines can also be defined and measured based on perceived consumption adequacy and we believe that this approach has a number of attractions; see Pradhan and Ravallion (2000).

[iii] Ideally this would be the characteristics of individual rather than households. That is an important distinction, but not one that is implementable with the data normally available for measuring poverty.

[iv] Such poverty lines can also be used to construct true cost-of-living indices, by normalizing the poverty line by its value for some reference group (see, for example, Deaton and Muellbauer, 1980).

[v] Some versions of the FEI method regress (or graph) nutritional intake against consumption expenditure and invert the estimated function, while others avoid this step by simply regressing consumption expenditure on nutritional intake. These two methods need not give the same answer, though the difference is not germane to our present interest; either way the following points apply.

[vi] See Orshansky (1965) and Citro and Michael (1995).

[vii] For example, one way of setting the different bundles of goods is to base them on the actual consumption pattern in each region of a reference segment of the national population that is initially taken to be poor. Following the method described in Ravallion (1998) one can iterate until there is convergence such that the reference segment is in fact deemed to be in a neighborhood of the poverty line.

[viii] This section draws on Ravallion and Lokshin (2004), which provides greater detail.

[ix] The results of the latest 2001 review of the regional baskets indicate that out of 89 submitted proposals, 67 drafts attracted no criticism, while the remaining 22 drafts deviated in one way or another from the methodological recommendation.

[x] There is no allowance for health or education since by law (at least) these are free in Russia.

[xi] Consistency tests for the individual Q matrixes show different numbers of passing elements (Ravallion and Lokshin, 2004). The adult male matrix has 250 passes, while the matrices for adult females, children 0 to 7 and children 7 to 15 have 251, 247 and 248 respectively.

[xii] Given that the temperature map is at a much finer level, calculating a correlation coefficient would require considerable aggregation. From eye-balling the figures, the extent of the correlation is clearly high, however.

REFERENCES

Blackorby, C., and D. Donaldson. 1987. "Welfare Ratios and Distributionally Sensitive Cost-Benefit Analysis", *Journal of Public Economics* 34, pp. 265-290.

Citro, Constance F., and Michael, Robert T. 1995. *Measuring Poverty: A New Approach,*.Washington DC: National Academy Press.

Dandekar, V.M., and N. Rath. 1971. *Poverty in India.* Pune: Indian School of Political Economy.

Deaton, Angus and John Muellbauer. 1980. *Economics and Consumer Behavior,* Cambridge: Cambridge University Press.

Greer, J., and Erik Thorbecke. 1986. "A Methodology for Measuring Food Poverty Applied to Kenya," *Journal of Development Economics* 24, pp. 59-74.

Ministry of Labor and Social Development (MLSD). 2000. "Prozhitochnui Minimum v Rosiiskoi Federacii (The Subsistence Minimum in Russian Federation)." MLSD, Moscow, Russia.

Orshansky, Molly. 1963. "Children of the Poor," *Social Security Bulletin* 26, pp. 3-29.

Osmani, Siddiqur. 1982. *Economic Inequality and Group Welfare* Oxford: Oxford University Press.

Palmer-Jones, Richard and Kunal Sen. 2001. "On India's Poverty Puzzles and Statistics of Poverty," *Economic and Political Weekly,* January 20, pp. 211-217.

Paul, Satya. 1989. "A Model of Constructing the Poverty Line," *Journal of Development Economics* 30, pp. 129-144.

Pradhan, Menno and Martin Ravallion. 2000. "Measuring Poverty Using Qualitative Perceptions of Consumption Adequacy", *Review of Economics and Statistics,* 82(3), pp. 462-471.

Ravallion, Martin. 1994. *Poverty Comparisons,* Chur, Switzerland: Harwood Academic Press.

______________,. 1998. *Poverty Lines in Theory and Practice,* LSMS Paper 133, Washington DC, World Bank.

Ravallion, Martin and Benu Bidani. 1994. "How Robust is a Poverty Profile?" *World Bank Economic Review,* 8(1), pp. 75-102.

Ravallion, Martin and Michael Lokshin (2002), "Self-Rated Economic Welfare in Russia," *European Economic Review,* 46, pp. 1453-1473.

______________ and ________________. 2004. "Testing Poverty Lines," Development Research Group, World Bank.

Ravallion, Martin and Binayak Sen. 1996. "When Method Matters: Monitoring Poverty in Bangladesh," *Economic Development and Cultural Change,* 44, pp. 761-792.

Samuelson, Paul. 1938. "A Note on the Pure Theory of Consumer Behaviour" *Economica* 5, pp. 61-71.

Sen, Amartya. 1979. "The Welfare Basis of Real Income Comparisons: A Survey," *Journal of Economic Literature* 17, pp. 1-45.

___________. 1985. *Commodities and Capabilities,* Amsterdam: North-Holland.

___________. 1992. *Inequality Rexamined,* Oxford: Oxford University Press.

VTsUZH. 2002. "Living Standards of the Population of the Russian Federation. Legal Basis for Minimal Monetary Incomes." Part 1, Moscow, pp. 220-225.

Wodon, Quentin. 1997. "Food Energy Intake and Cost of Basic Needs: Measuring Poverty in Bangladesh," *Journal of Development Studies* 34, pp. 66-101.

World Health Organization (WHO). 1985. *Energy and Protein Requirements,* WHO, Technical Report Series 724, Geneva

Chapter 4

POVERTY INDICES

James E. Foster
Vanderbilt University, james.e.foster@vanderbilt.edu

1. INTRODUCTION

Most empirical and theoretical work has focused on poverty statistics drawn from income, expenditure or some other single dimensional variable. While an income-based approach can be limiting, it has proved to be extremely useful in practice and can be readily modified to account for variations in individual characteristics. There is also a well-developed axiomatic structure for measuring poverty in this context and a number of interesting and comprehensible functional forms for poverty indices to take. This paper presents an elementary overview of axiomatic literature on poverty indices – both to supplement the general discussion in Foster and Sen (1997) and to bring out a number of useful observations that have subsequently come to light.

The first section of the paper begins by defining the aggregation problem underlying poverty measurement and describing the "partial indices" that are the basic building blocks of poverty measurement: the frequency, average depth and dispersion of poverty. Section 2 provides the axiomatic framework for evaluating indices, including invariance axioms, continuity axioms, dominance axioms and subgroup axioms. Invariance axioms ensure that the poverty index ignores certain categories of information and thus emphasizes others. Continuity axioms guarantee that the index is a well-behaved function of incomes that has no unwarranted jumps. Dominance axioms ensure that an index appropriately reflects unambiguous improvements in the income distribution of the poor. Subgroup axioms require indices to give consistent judgments for subgroups and the overall population. The various implications of these axioms are discussed with reference to specific indices, including those of Sen, Clark-Hemming-Ulph, and Foster-Greer-Thorbecke. Recent work in inequality analysis has demonstrated the importance of so-called general means in the construction of measures of dispersion. We conclude by showing that these income standards also play a central role in the formation of poverty measures.

2. THE FRAMEWORK

In the standard Sen framework, the first step towards evaluating income poverty is the *identification* step, at which a poverty line z is chosen and the poor and non-poor groups are distinguished. Much has been written about the various approaches to determining this "minimum level of living," and we will not add to this discussion here.[i] The second step is the *aggregation* step, at which the data are put together to form an overall picture of poverty. This is usually done with the help of a *poverty index* or *measure*, which is a function P(x;z) of the income distribution $x = (x_1,\ldots,x_n)$ and the poverty line z. Sometimes the income distribution is represented by a cumulative distribution function F, in which case the poverty index is a function P(F;z).

Several questions must be answered at the aggregation step. Which aspects of the distribution are considered relevant in measuring the extent of poverty? How should measured poverty change in response to basic changes in the distribution? Can we obtain a single overall measure of poverty, or must we be content with a number of "partial" indices, each of which focuses on a specific aspect or dimension of poverty? We will now try to provide answers to these questions.

2.1 Partial indices

A crude but important indicator of the extent of poverty is the number of poor, which we denote by q = q(x;z), or its per capita version commonly called the *headcount ratio* H(x;z) = q/n, where n = n(x) is the population size of x. In the case of distribution functions, the headcount ratio is just H(F;z) = F(z), the distribution function evaluated at the poverty line – yielding the proportion of the population with incomes on or below the poverty line.[ii] Note that H provides a concrete answer to the question, "How widespread or prevalent is the condition of poverty?" But in order to provide a sharp answer to this question, it by necessity ignores other aspects of poverty, such as the depth or distribution of poverty. It is a key "partial index" of poverty, but not a very credible overall measure.

Suppose the headcount ratio were adopted as the unique criterion for allocating a limited pool of extra resources.[iii] The goal would then be to lift as many of the poor above the poverty line, which would require focusing only on those nearest the poverty line and ignoring the poorest of the poor. Indeed, if income could be extracted from the poorest and given to the richest poor person, thereby lifting the latter above the poverty line, it would be regarded by the index as a favorable outcome. This kind of "lifeboat logic" is hardly surprising, given the headcount ratio's extreme reliance on the poverty line as a meaningful and unique boundary between poor and nonpoor. If escaping from poverty truly required a specific, indivisible level of income, and if the particular levels of income held by those below the line were irrelevant, then the use of the headcount ratio might be justified. However, if the poverty line is viewed as a potentially arbitrary selection from a continuous range of potential lines, and gradations of income are important even among the poor, the recommendation "cross the line at all cost" carries limited weight. The specific incomes of the poor, and not just their number, must be taken into account by the index in order to produce reasonable policy recommendations.

Consider a distribution in which all the poor have incomes that are right next to the poverty line and a second distribution with the same q and n (and hence H) for which all poor incomes are extremely low. One could capture the obvious and important differences between these distributions with the help of a *gap* measure – based on the aggregate, average, or per capita shortfall of the poor incomes from the poverty line. These measures provide information on an additional aspect of poverty: its depth. The gap measures are often used as aggregate poverty indicators, but like the indices based on poverty counts, they are best seen as partial indices.

The *income gap ratio* $I = (z - \mu_p)/z$, for example, is the average shortfall of the poor (the poverty line minus the mean poor income μ_p) expressed as a share of the "maximum" shortfall z. Using the income gap ratio as the criterion for allocating resources would lead to policy recommendations that are in some sense opposite to the headcount ratio's, but still counterintuitive: It does not matter who among the poor receives the extra income, so long as no one is allowed to escape! Apart from this restriction against lowering the number of the poor, the income gap ratio has no implications for the pattern of transfers. Like all gap and counting measures, it is supremely indifferent to the distribution of income among the poor. In providing a conclusive answer to the question, "What is the average depth of poverty?" it neglects other important aspects of poverty. But in combination with other partial indices, it too can be extremely useful in poverty evaluations.

2.2 Inequality in poverty?

Sen (1976), Watts (1969), and others have argued that an overall evaluation of poverty should take into account the distribution of income among the poor. There are several compelling reasons why it should do so. The first is based on the traditional economic view that the marginal utility of income – and hence marginal deprivation from the lack of income – is higher at low incomes. A second line of argument draws from a sociological perspective that relative deprivation is more acute at lower incomes.[iv] A third argues that, at a descriptive level, there is a significant difference between a distribution x in which all poor incomes are located at their mean μ_p and another distribution y having identical H and I in which the poor incomes range from extremely low to right up against the poverty line. The partial indices H and I are sufficient to capture the key aspects of the distribution x. However, an additional partial index measuring inequality among the poor is helpful in evaluating the distribution y and differentiating it from x. Information on inequality can be particularly important when considering policies for reducing poverty – where it helps to know whether there is a homogeneous group of poor or whether they are wide-ranging in their circumstances.

3. AXIOMS

How should these various dimensions be brought together in an overall measure of poverty? Following Sen (1976), there have been many functional forms proposed in the literature. Consider, for example, the *Sen index* S defined as

$$S = H(I + (1-I)G_p) \tag{1}$$

where G_p is the Gini inequality index among the poor; or the measure P_2, drawn from the *Foster-Greer-Thorbecke (FGT)* class of indices and having the formula

$$P_2 = H(I^2 + (1-I)^2C_p^2) \tag{2}$$

where C_p is the coefficient of variation among the poor.[v] But why adopt these specific formulas? A key step towards justifying a particular measure of poverty is identifying the properties it satisfies. We now consider the various categories of axioms that are commonly required for poverty measures.

3.1 Invariance Axioms

Our initial collection of axioms requires the poverty measure to ignore certain aspects of the distribution and to focus on others. The first ensures that a reordering of incomes leaves the poverty value unchanged; the second ensures that poverty is measured on a per capita basis, to allow comparability across populations of different sizes. We say that x is obtained from y by a *permutation of incomes* if x = Py for some permutation matrix (i.e., a square matrix having a single "1" in each row and column, and "0" for all remaining entries). We say that x is obtained from y by a *replication of incomes* if x = (y,y,...,y), where x is constructed by "cloning" the original distribution y a finite number of times.

Symmetry: If x is obtained from y by a permutation of incomes, then P(x;z) = P(y;z).

Replication Invariance: If x is obtained from y by a replication of incomes, then P(x;z) = P(y;z).

These axioms together imply that any two income distributions having the same cumulative distribution function F must have the same level of measured poverty.

The next axiom is a defining characteristic that ensures that a poverty measure is independent of the incomes of the nonpoor. We say that x is obtained from y by *an increment to a nonpoor person* if $x_i > y_i$ for some i satisfying $y_i > z$, while $x_j = y_j$ for all $j \neq i$.

Focus: If x is obtained from y by an increment to a nonpoor person, then P(x;z) = P(y;z).

The focus axiom says that changing nonpoor incomes without altering the number of the poor leaves the poverty value unchanged. This axiom defines one of the key aspects of a measure of poverty, as opposed to inequality or other characteristics of the distribution: it depends on the incomes of the poor while ignoring the incomes of the nonpoor (but not necessarily their number). When combined with symmetry and replication invariance, it ensures that any two income distributions having the same distribution function over the income range [0, z] must have the same poverty value.

Until now, we have not considered properties that compare distributions at different poverty lines. The presumption has been that when comparisons are made, they are taken relative to the same poverty line, e.g., $P(x;z)$ vs. $P(y;z)$. But the poverty line may change over time and space due to, say, a reevaluation of the cost of the bundle of basic goods. The next property ensures that the poverty measure is *relative* in the sense that doubling the poverty line and all incomes leaves the level of poverty unchanged. More precisely, we say that $(x';z')$ is obtained from $(x;z)$ by *a proportional change* if $(x';z') = (\alpha x;\alpha z)$ for some $\alpha > 0$.

Scale Invariance: If $(x';z')$ is obtained from $(x;z)$ by a proportional change, then $P(x';z') = P(x;z)$.

For measures satisfying scale invariance, comparability at different poverty lines is obtained by allowing the poverty line z itself to be the unit of measure. The technical implication is that the poverty index can be expressed as a function of the income distribution *normalized by the poverty line* (in contrast to normalization by the distribution's mean, which is commonly used in inequality analysis). In conjunction with the above axioms, it ensures that whenever two income distributions have normalized income distributions whose cumulative distribution functions are identical on the interval [0, 1], they must have the same poverty value. Thus, for example, the distribution (2, 12) evaluated at a poverty line of 8 has the same level of poverty as the distribution (6, 6, 30, 40) evaluated at a poverty line of 24.

3.2 Dominance Axioms

The invariance axioms are useful in identifying the information that a poverty measure can use in arriving at the overall level of poverty, but leave the shape of the function unrestricted. The next series of axioms restricts the shape of the measure by identifying transformations that appear to have a certain effect on poverty, and then requiring the poverty values to follow this intuition. We begin with the definitions of two basic transformations. First, we say that x is obtained from y by *a decrement among the poor* if for some i with $y_i \leq z$ we have $x_i < y_i$, while $x_j = y_j$ for all $j \neq i$. In other words, x and y are identical except that one poor income is lower in x than in y. Now suppose x and y are two distributions with the same mean. We say that x is obtained from y by *a progressive transfer among the poor* if for some i and j with $y_i < y_j \leq z$ we have $y_i < x_i \leq x_j < y_j$, while $x_k = y_k$ for all $k \neq i,j$. In other words, a poor person transfers a given amount to an even poorer person. Consider the following two axioms proposed by Sen (1976).

Monotonicity: If x is obtained from y by a decrement among the poor, then $P(x;z) > P(y;z)$.

Transfer: If x is obtained from y by a progressive transfer among the poor, then $P(x;z) < P(y;z)$.

A decrement among the poor represents an unambiguous increase in a poor individual's poverty level; monotonicity requires the poverty measure to reflect this increase at the aggregate level. In terms of the three partial indices, the frequency is unchanged, the average depth rises, and inequality can go either way as a result of the decrement. For transformations of this sort where frequency is unchanged, average intensity rises and inequality can go either way, the monotonicity axiom requires overall poverty to follow the conclusions of the second partial index (average depth). The transfer axiom likewise requires a particular transformation to change the aggregate level of poverty. However, since the underlying transformation entails a decrease in poverty for one poor person (the poorer of the two) and an increase for the other (the richer), the axiom goes beyond monotonicity (which involves no such trade off). The axiom requires greater weight to be placed on the poorer of the two, with the decrease in the poorer person's poverty more than making up for the simultaneous increase in the richer person's poverty. Put differently, for transformations of this sort where frequency and average depth are unchanged and inequality among the poor falls, the transfer axiom requires overall poverty to follow the third partial index (inequality among the poor).

The monotonicity axiom ensures that the poverty level must fall whenever a poor income rises but stays poor. A reasonable case may be made for expanding this axiom to address the situation where the poor income rises *above* the poverty line. Note, however, that if the poor income is originally on the poverty line z, it is not immediately clear why poverty must *strictly* fall as a result of the increased level of income. Once again, this depends upon one's conception of the poverty line. If the line is indeed the great divider, and the size of the poor person's shortfall is irrelevant, then a small increase in the income of a poor person with income z will be regarded as a discrete improvement in poverty. But if the line is acknowledged to be an arbitrary standard, with the extent of individual poverty decreasing to zero as the income rises to z, then an increment to a person with an income of z may leave overall poverty unaffected. Our extension of monotonicity will adopt the less exacting point of view that increasing a poor person's income should not increase poverty.

A related issue is whether we should broaden the coverage of the transfer axiom to allow transfers that alter the number of the poor. Sen's original formulation of the transfer axiom required poverty to rise in the case of a regressive transfer among the poor – even when the recipient rises out of poverty as a result. We present this stronger axiom below. But first we must define the transformations associated with the two axioms.

Let x and y be income distributions with same population size. We say that x is obtained from y by *an increment among the poor* if for some i with $y_i \leq z$ we have $x_i > y_i$, while $x_j = y_j$ for all $j \neq i$. In other words, x and y are identical except that a poor income in y is now higher in x. Now suppose x and y are two distributions with the same mean. We say that x is obtained from y by *a regressive transfer among the poor* if for some i and j with $y_i \leq y_j \leq z$ we have $x_i < y_i \leq y_j < x_j$, while $x_k = y_k$ for all $k \neq i,j$. In other words, a transfer has been made from a poor person i to a richer poor person j, with the latter becoming better off at the expense of the former. Unlike the progressive transfer used earlier, the number of the poor can change as a result of a regressive transfer. We have the following stronger versions of the two dominance axioms.

Strong Monotonicity: P satisfies monotonicity; in addition, if x is obtained from y by an increment among the poor, then $P(x;z) \leq P(y;z)$.

Strong Transfer: If x is obtained from y by a regressive transfer among the poor, then $P(x;z) > P(y;z)$.

The strong monotonicity axiom adds one additional requirement to the monotonicity axiom: that if the frequency of poverty falls as a result of an increase in the income of a poor person, the overall poverty level should not rise – and this must be true irrespective of the impact on the other two partial indices. The income gap ratio I itself is clearly one example of a measure that violates this stronger form of monotonicity since the average gap can rise when a poor income crosses the poverty line. (It does satisfy the weaker version of monotonicity.) The stronger transfer axiom immediately implies the original transfer axiom, since any progressive transfer among the poor is the converse of a regressive transfer among the poor in which the number of the poor is unchanged. However, the stronger axiom applies to regressive transfers that alter the number (and average depth) of poverty as well – requiring the poverty index to rise. As we will see below, this can cause trouble for poverty measures that are not continuous – like Sen's original poverty measure, the headcount ratio H and the income gap I – but for continuous measures the stronger axioms will be shown to be no more exacting than the original ones.[vi]

The monotonicity and transfer properties are called dominance properties in part because of their connection to the traditional properties of first and second order (stochastic) dominance on a range of incomes.[vii] A decrement among the poor always results in a pair of normalized (by the poverty line) distributions that can be ranked by first order dominance over the range [0,1). It can be shown that any continuous poverty index satisfying the invariance axioms and monotonicity is always consistent with first order dominance rankings over the normalized distributions for the range [0,1) – with the first order dominating distribution having less poverty according to *any* such poverty index. Similarly, a progressive transfer among the poor always leads to normalized distributions that can be ranked by second order dominance on [0,1). Any continuous measure satisfying the invariance axioms and both dominance axioms must be consistent with second order dominance over the normalized distributions for the range [0,1). The second order dominating distribution necessarily has lower poverty according to *any* such poverty index.[viii]

3.3 Continuity Axioms

Each of the above axioms places an ordinal restriction on the poverty measure either by requiring the value to be unchanged for various transformations or by requiring the value to change in a particular direction. The next axioms are of a somewhat different type in that they also place restrictions on the cardinal form of the measure. Consider a distribution x and a sequence x^k of distributions, all having the same population size, and a fixed poverty line z. We say that sequence x^k has a *fixed nonpoor income distribution* if whenever a person i is not poor in x^k for some k, then i has the same (nonpoor) income in x^k for all k. The first axiom focuses entirely on the behavior of P when only the incomes of the poor are varied, while the second allows all incomes to change.

Restricted continuity: For any sequence x^k having a fixed nonpoor income distribution, if x^k converges to x, then $P(x^k;z)$ converges to P(x;z).

Continuity: For any sequence x^k, if x^k converges to x, then $P(x^k;z)$ converges to P(x;z).

In other words, under restricted continuity P is a continuous function of poor incomes, while under continuity P is a continuous function of all incomes.

Restricted continuity simply requires that the poverty value does not abruptly jump as poor incomes are altered over the range [0,z]. It is satisfied by all of the usual measures. The stronger continuity axiom, which applies over all incomes (and specifically when a poor income rises above the poverty line z), rules out a rich class of measures that are especially sensitive to the headcount ratio, and hence may discontinuously jump when an income crosses the poverty line. H and I both violate continuity while satisfying the restricted form. The Sen measure S also violates the stronger axiom due to its reliance on a weighting system that depends crucially on the number of the poor (as described below). The FGT index P_2 is fully continuous. The product HI is also continuous even though each of its components is not.

There is a useful result linking the stronger and weaker versions of the dominance axioms in the presence of continuity (see Donaldson and Weymark, 1986, for related results).

Result: Suppose that P is a continuous poverty measure satisfying the invariance axioms. Then P satisfies monotonicity and the transfer axiom if and only if it satisfies strong monotonicity and the strong transfer axiom.

The proof is as follows: Let P satisfy the invariance axioms, continuity, monotonicity and the transfer axiom. Suppose that x is obtained from y by an increment among the poor (say involving person i). If $y_i < x_i \leq z$ then y is obtained from x by a decrement among the poor and $P(y;z) > P(x;z)$. If $z = y_i < x_i$ then by continuity and the focus axiom, $P(y;z) = P(x;z)$. If $y_i < z < x_i$ then where x' is the distribution with $x_i' = z$ and $x_j' = x_j$ for all $j \neq i$, we have $P(y;z) > P(x';z) = P(x;z)$ by the above. Thus $P(y;z) \geq P(x;z)$, and P satisfies strong monotonicity. Now, let x be obtained from y by a regressive transfer among the poor (say from i to j). If $y_j < x_j \leq z$ then y is obtained from x by a progressive transfer among the poor and $P(y;z) < P(x;z)$. If $z = y_j < x_j$ then by monotonicity, continuity and the focus axiom, $P(y;z) < P(x;z)$. If $y_j < z < x_j$ then where y' is the distribution with $y_j' = z$ and $y_i' = y_i-(z-y_j)$ while $y_k' = y_k$ for all $k \neq i,j$, we have $P(y;z) < P(y';z) < P(x;z)$, by the above. Thus $P(y;z) < P(x;z)$, and P satisfies the strong transfer axiom. The converse implications are obvious.

For continuous poverty measures, then, there is no distinction between the weaker dominance axioms and the stronger dominance axioms.

A key step of the above proof was to note that for a continuous measure P satisfying the focus axiom, poverty is the same whether a person has the poverty line income or any income above it. This may be stated more generally with the help of a definition: For any x, define the *censored distribution* $x^* = x^*(z)$ to be the distribution in which all incomes above z are replaced by z and the remaining incomes are unchanged. Then for any continuous poverty index P satisfying the focus axiom, $P(x^*;z) = P(x;z)$.

The censored distribution can also be used to create continuous measures from those satisfying restricted continuity. Let $P(x;z)$ be a poverty measure satisfying the invariance axioms, monotonicity, the transfer axiom and restricted continuity. As in Foster and Sen (1997), define the *continuous version* P* of P as follows:

$$P^*(x;z) = P(x^*;z) \text{ for all distributions } x. \tag{3}$$

In other words, P is applied to the censored distribution x* rather than to x itself. Since all incomes in x* are less than or equal to z, this has the effect of extending the restricted continuity of P to full continuity for P*. In addition, it does so without violating any of the basic axioms satisfied by P and, by the above result, it strengthens the dominance axioms to their stronger versions. Note that while the continuous version of the headcount ratio is H* = 1, and hence is not useful as a poverty measure, the transformation converts the income gap ratio I into

$$I^* = HI = \frac{1}{n}\sum_{i=1}^{n}(z - x_i^*)/z,$$

or the *per capita poverty gap*, which has an important role in both the theory and practice of poverty measurement. We will revisit this process when we discuss specific measures in Section 4.

3.4 Subgroup Axioms

The next two axioms require a link to exist between overall poverty and the poverty levels in population subgroups. Decomposability requires overall poverty to be the weighted average of the subgroup levels, where the weights are the group population shares. Subgroup consistency requires a decrease in one subgroup's poverty to be reflected in the overall poverty level. Both are important in applied poverty analysis.

Suppose that a population of size n is divided into several (say m) subgroups according to some characteristic of interest. Let y^j be the distribution and n^j be the population size of subgroup j = 1,...,m, so that $y = (y^1, y^2, \ldots, y^m)$ is the overall distribution. We say that P is *decomposable* (or *additively decomposable in population subgroups*) if for every such population and partition the following is true:

$$P(y;z) = \sum_{j=1}^{m}\frac{n^j}{n}P(y^j;z) \tag{4}$$

What is the practical import of decomposability? A main implication is that it allows the construction of consistent "poverty profiles" (see Anand, 1977, for example), which show how overall poverty may be attributed to various population subgroups. Actually there are two conceptually distinct objectives of subgroup analysis. In the first, the aim is to identify subgroups where poverty is particularly high as measured absolutely by $P(x^j;z)$ or relatively by $P(x^j;z)/P(x;z)$. This can be especially useful in targeting the poor to achieve a maximal reduction in overall poverty.[ix] If such comparisons are the only goal of the analysis, there may not be a need for the poverty measure to be decomposable, since one can easily evaluate and compare poverty values across different groups using any poverty measure. However, if the aim includes a consistent reckoning of subgroup contributions to total poverty – say, as measured by $(n^j/n)P(x^j;z)/P(x;z)$ – a nondecomposable index is problematic, since the sum of the contributions can exceed or fall short of 1. With decomposable measures, the contributions always sum to 1.

What theoretical implications does decomposability have for P? It can be shown that decomposable measures must have the form

$$P^f(x;z) = \frac{1}{n}\sum_{i=1}^{n} f(x_i;z) \qquad (5)$$

for some *individual poverty function* f of the income and poverty line.[x] In other words, overall poverty is the average individual poverty level. The other axioms satisfied by P are reflected in the shape and characteristics of f.

Now let us consider another type of subgroup axiom for poverty measures. Suppose a given population subgroup achieves a reduction in poverty while the rest of the population has a poverty level that is unchanged. Barring changes in population sizes – as might occur with migration – we should expect poverty to decline overall. Such is the requirement of the following axiom[xi] investigated by Foster and Shorrocks (1991):

Subgroup Consistency: A poverty index $P(x;z)$ is subgroup consistent if for every poverty line z and any distributions x, x', y, and y' for which $n(x) = n(x')$ and $n(y) = n(y')$ we have $P(x,y;z) > P(x',y';z)$ whenever $P(x;z) > P(x';z)$ and $P(y;z) = P(y';z)$ both hold.

Apart from sheer intuition, the justification of subgroup consistency rests on practical concerns of targeting and coherent regional poverty programs. When a poverty measure is not subgroup consistent, a successful regional campaign against poverty may lead to a perceived worsening of poverty for the country as a whole. Note that the reach of the axiom is not limited to the two-group case: the axiom can be applied to any number of subgroups by repeated application to pairs. Also it is clear that subgroup consistency is closely related to decomposability. If, for a decomposable measure, poverty rises in one subgroup and is unchanged in the second, then since overall poverty is a weighted average of subgroup poverty levels (with population share weights), it clearly must rise. Thus decomposability implies subgroup consistency – but the converse implication is not true.[xii] However, under a natural set of assumptions (see Foster and Shorrocks (1991)) it *is* true that any continuous subgroup consistent index is a *monotonic transformation* of some decomposable measure. We will see the implications of this when we discuss individual indices below.

4. INDICES

We now turn to a discussion of various poverty indices that have been proposed in the literature. Given the rather sizeable lists of contributions in this area, we will not try to be exhaustive. Instead we focus on three main families of indices (from which nearly all the common measures can be derived) and refer the reader to the surveys of Foster (1984), Donaldson and Weymark (1986), Seidle (1988), Chakravarty (1990), and Zheng (1997) for more detailed discussions.

4.1 The Sen index

Sen's (1976) influential paper on poverty measurement set the stage for the subsequent literature through its presentation of a rich axiomatic framework and a new, "distribution sensitive" index of poverty. Recall from (1) in Section 3 that the Sen index can be written in terms of the partial indices H, I and G_p as $S = H(I + (1\text{-}I)G_p)$. Sen provides an alternate definition based on a function $r_i = r_i(x;z)$ called the *ranking of i among the poor* that can be constructed as follows: One poor person with the lowest income is assigned a rank of $r_i = q$; of the remaining poor persons, one individual with the lowest income is assigned a rank of $r_i = q\text{-}1$; and so on until a poor person with the highest income receives the rank of $r_i = 1$; the remaining persons can be assigned ranks of $r_i = 0$, although the precise value of r_i for the nonpoor will not affect the formula. Where x_i^* is person i's censored income, let the *income gap* of person i be defined by $g_i^* = z - x_i^*$ (or, equivalently, $g_i^* = z - x_i$ for all poor i, and $g_i^* = 0$ for all nonpoor i).

The *Sen measure* can be defined as

$$S(x;z) = \frac{1}{n} \sum_{i=1}^{n} \frac{(2r_i - 1)}{q} \frac{g_i^*}{z} \tag{6}$$

where n = n(x) is the population size of x while q = q(x;z) is the number of poor.[xiii] S is a weighted sum of the normalized gaps, where the weights depend on i's rank among the poor and thus are higher for poor persons with larger income gaps. (The weighted normalized gaps for nonpoor persons are obviously zero.) In his original presentation, Sen (1976) invoked the literature on relative deprivation in his justification of rank-based weighting and derived the final form of S within a general additive framework.

That S satisfies the invariance axioms and restricted continuity follows immediately from the equivalent definition (1) and the fact that the partial indices H, I and G_p satisfy these properties. The dominance axioms can be evaluated with the help of definition (6). For example, let x be obtained from y by a decrement among the poor. This will increase the gap of one poor person and may alter the ranking. Holding the ranking fixed, an increased gap increases (6). Shifting to the final ranking (so that rank-based weights are perfectly correlated with the gaps) will only raise (6) even further. Thus, monotonicity is satisfied by S. A similar argument shows that S satisfies the transfer axiom. Now, what about the stronger versions of the dominance axioms and continuity? Consider an increment to a person with rank 1 on the poverty line. The increment leaves all normalized income gaps unchanged, but lowers every poor person's rank by 1. This, in turn, lowers each poor person's weight from $(2r_i\text{-}1)/q$ to $(2(r_i\text{-}1)\text{-}1)/(q\text{-}1)$ and hence S must strictly fall (unless each poor person is on the line, in which case S is unchanged). Thus S violates continuity, and we cannot apply our equivalence result to obtain the stronger dominance axioms. Yet strong monotonicity does indeed hold for S. To see this note that any arbitrary increment among the poor is either an increment of the type described above (where the person is initially on the poverty line), or the converse of a decrement among the poor, or a combination of the two. It follows that S must fall or stay the same and hence strong monotonicity is satisfied. The strong transfer axiom, however, is not satisfied by S. A regressive transfer among the poor can lift the recipient above the poverty line and lower the number of the poor. With q lower, the discontinuous drop via the change in weights may override the increased shortfall of the giver of the transfer. For example, if the poverty line is 30 and the distribution of income among the poor is (5,5,10,15,26), then *any* regressive transfer that changes the number of the poor will lead to a lower value for S.

The Sen measure also violates both decomposability and subgroup consistency. To see this, set z = 10 and note that the Sen poverty levels of the group distributions $y^1 = (4,5,9,10,20)$ and $y^2 = (1,9,10,10,20)$ are 0.350 and 0.340, respectively. When the populations are combined and the weights on the normalized gaps are appropriately altered, the overall value of the Sen measure is seen to be 0.355 – which could never be obtained as a weighted average of subgroup poverty levels. Put differently, the percentage contribution of subgroup 1 is less than 49.5% while subgroup 2 accounts for less than 48%, leaving more than 2.5% of total poverty unaccounted for. It turns out that the Sen measure has the interesting property[xiv] – inherited from the Gini coefficient – that the weighted average of the subgroup poverty values never exceeds and typically lies below the poverty value of the merged population, i.e.,

$$P(y;z) \geq \sum_{j=1}^{m} \frac{n^j}{n} P(y^j;z) \tag{7}$$

This implies that as a finer profile of poverty is constructed, less aggregate poverty is accounted for by the subgroup poverty levels. In the extreme case where each subgroup is made up of a single individual, the subgroup poverty level becomes g_i^*/z and the weighted average of subgroup poverty levels is clearly $HI = (1/n)\Sigma_i\, g_i^*/z$, the per capita poverty gap measure. Decomposing $S = HI + H(1\text{-}I)G_p$ by subgroup removes a portion of the distribution sensitive term $H(1\text{-}I)G_p$, and eventually eliminates it altogether.

The above example also leads to a violation of subgroup consistency. Recall that $S(y^1;10) = 0.350$, $S(y^2;10) = 0.340$, and $S(y^1,y^2;10) = 0.355$, and hence by replication invariance $S(y^1,y^1;z) = 0.350$. When the distribution in subgroup 1 is fixed at y^1, and the distribution in subgroup 2 changes from y^2 to y^1, poverty in subgroup 2 rises from 0.340 to 0.350 while overall poverty *falls* from 0.355 to 0.350. This violation arises due to the link with the Gini coefficient, which is known to violate subgroup consistency.[xv]

4.2 The continuous Sen index

It was the Sen measure's violation of continuity and the strong transfer axiom that led Thon (1979) to construct a modified version that satisfies both. Rather than using $r_i(x;z)$, the ranking of i among the poor, he constructed a ranking function $R_i = R_i(x)$ which assigned every person their rank in the overall distribution (with the person with rank $R_i = 1$ having the highest income and the person with rank $R_i = n$ having the lowest income). The modified Sen measure presented by Thon, and later discussed at some length by Shorrocks (1995), is defined as[xvi]

$$S^*(x;z) = \frac{1}{n}\sum_{i=1}^{n}\frac{(2R_i-1)}{n}\frac{g_i^*}{z} = HI + (1-HI)G^* \tag{8}$$

where G* is the Gini coefficient applied to the censored distribution. It is an easy matter to show that this measure is the continuous version of the Sen measure, $S^*(x;z) = S(x^*;z)$, and hence it inherits the invariance axioms, restricted continuity, and the dominance axioms (apart from strong transfer) from S; it gains continuity and the strong transfer axiom via the definition of the continuous version. The key difference between S* and the original Sen index is, of course, the weights on the shortfalls. S has weights $(2r_i-1)/q$ that vary with q; S* has weights $(2R_i-1)/n$ which are independent of the number of the poor. Consequently, S* does not abruptly change when a poor income crosses the poverty line. As for the subgroup axioms, S* still suffers from the types of violations apparent with S, with the culprit remaining the Gini coefficient. Overall poverty will in general exceed the population share weighted sum of subgroup levels; and a decrease in a subgroup's poverty level (with unchanged poverty elsewhere) can lead to an increase in overall measured poverty.

4.3 The CHU indices

In a 1981 paper, Clark, Hemming and Ulph took an entirely different approach to constructing a distribution sensitive poverty index with the help of Atkinson's well-known class of inequality measures and the censored distribution. The *Clark, Hemming and Ulph (CHU)* parametric family of poverty measures is defined by:

$$C_\beta(x;z) = \begin{cases} 1 - [\frac{1}{n}\sum_{i=1}^{n}(x_i^*/z)^\beta]^{1/\beta} & \text{for } \beta \leq 1 \text{ and } \beta \neq 0 \\ 1 - [\prod_{i=1}^{n}(x_i^*/z)]^{1/n} & \text{for } \beta = 0 \end{cases} \tag{9}$$

In words, C_β is the departure from unity of the "representative" income associated with the normalized censored distribution, using the functional form employed by Atkinson (1970).

Each CHU index clearly satisfies the invariance axioms and continuity. It achieves a minimum of 0 when all the incomes are at z or higher. For parametric values $\beta > 0$, the measure achieves a maximum of 1 when *all* incomes are zero. For values $\beta \leq 0$ the index rises to 1 as *any* income falls to 0; to satisfy monotonicity and the transfer axiom, the domain must be restricted to strictly positive income distributions. With this restriction, and apart from the parameter value $\beta = 1$ where the measure becomes HI, the members of this family satisfy all four dominance axioms. The one axiom not satisfied by the CHU family of indices for $\beta < 1$ is decomposability. While the functional form does have a fundamentally additive structure, as required by decomposability, the additive term is raised to a power. This makes it impossible for the aggregate poverty level to be expressed as a weighted average of subgroup poverty levels. Nonetheless, the family *does* satisfy subgroup consistency, and thus is able to evaluate subgroup and overall poverty levels in a coherent manner.[xvii]

One question that arises for this family is: Why should this particular formula be used? To understand the motivation for this formula, we return to the definition of the Sen measure and a simple observation first made by Anand (1977) that S is the product of the headcount ratio and an alternative income gap ratio using the "equivalent poor income" $\mu_p(1\text{-}G_p)$ – or the mean income among the poor discounted by inequality among the poor – in place of μ_p. Hence,

$$S(x;z) = H(1 - \mu_p(1\text{-}G_p)/z). \qquad (10)$$

Anand went on to suggest the use of different inequality measures in place of G_p, an approach that was later explored by Blackorby and Donaldson (1980) for a variety of relative inequality measures R_p among the poor. There are some general caveats associated with the use of this transformation to obtain "welfare" or "equivalent income" functions from inequality measures, and they apply here as well. For example, if R_p takes values above unity (e.g., the coefficient of variation), then $\mu_p(1\text{-}R_p)$ may not be increasing in (poor) incomes,[xviii] and hence the monotonicity axiom may or may not be satisfied. Apart from these minor difficulties, this approach to constructing distribution sensitive poverty indices represents a nice generalization of the Sen measure.

As noted by Chakravarty (1990), the resulting measures $H(1 - \mu_p(1\text{-}R_p)/z)$ share the Sen measure's sensitivity to the number of the poor and its violation of continuity (and the strong transfer axiom) at the poverty line. However, by applying the measure to the censored distribution x*, rather than the distribution x, we can obtain the "continuous versions" of the Blackorby-Donaldson indices. Evaluating the components of the index at the censored distribution, H becomes 1, the mean among the poor becomes the mean of the censored distribution, μ^*, and R_p becomes the inequality in the censored distribution, R*. The resulting class of measures is therefore

$$P(x;z) = 1 - \mu^*(1\text{-}R^*)/z, \qquad (11)$$

which is the general form suggested by Clark, Hemming and Ulph (1981) and later thoroughly explored by Chakravarty (1983a). For appropriate choice of R, we obtain a relative poverty measure satisfying the invariance axioms, continuity and the strong dominance axioms. When the Gini coefficient is used we obtain the continuous Sen measure S* = HI + (1 - HI)G* described above. In the case where Atkinson's (1970) parametric family of inequality measures is used, we obtain the CHU family C_β defined above.

The CHU indices have some other special features that deserve mention. First, β has an interesting interpretation as a measure of "aversion to inequality in poverty," with β = 1 at one extreme where inequality does not affect the index, and β approaching -∞ at the other where the largest shortfall is all that matters. Second, while C_β itself is not decomposable, we shall soon discover an interesting monotonic transformation that is.

4.4 The decomposable CHU indices

In a much overlooked paper, Harold Watts (1969) presented the following interesting distribution sensitive poverty measure:

$$P_W(x;z) = \frac{1}{n}\sum_{i=1}^{n}(\ln z - \ln x_i^*). \qquad (12)$$

It is easy to see that, so long as incomes are strictly positive, this measure satisfies each of the invariance and dominance axioms and is fully continuous. It is clearly decomposable, given its additive form, and hence subgroup consistent.

One interesting aspect of the Watts index is its interpretation as a measure of the per capita *utility* shortfall $(1/n)\Sigma_i\,[u(z) - u(x_i^*)]$, when the Bernoulli utility function u(s) = ln(s) is used.[xix] For arbitrary u, one would not expect a poverty measure of this form to satisfy the scale invariance property; for example, with $u(s) = s^{1/2}$, doubling the poverty line and all incomes increases poverty by a factor of $2^{1/2}$. By the special properties of logarithmic utility, though, $u(z) - u(x_i^*) = u(z/x_i^*)$, and the poverty measure clearly satisfies the invariance axioms.[xx] Note that the Watts index is similar to the CHU indices with nonnegative parameters in that P_W is overwhelmed by the presence of a single income close to zero.

Chakravarty (1983b) and Hagenaars (1987) adopt a related general functional form for poverty measures

$$P_D(x;z) = \frac{1}{n}\sum_{i=1}^{n} \frac{u(z)-u(x_i^*)}{u(z)} \tag{13}$$

in which the utility shortfall is expressed as a share of u(z). Hagenaars calls this the Dalton form of poverty measure, drawing on its analogy to Dalton's (1920) inequality measure. Decomposability of P_D follows directly from this functional form – regardless of how u(s) is specified. However, once again we find that scale invariance places important restrictions on u(s).[xxi] Chakravarty (1983b) shows that if a utility function u(s) defined for all $s \geq 0$ yields a poverty measure satisfying the invariance axioms then u(s) is a positive multiple of s^β where $0 < \beta < 1$. This leads to the Chakravarty (1983b) index

$$P_C(x;z) = \frac{1}{n}\sum_{i=1}^{n}[1-(x_i^*/z)^\beta] \qquad \text{for } 0<\beta<1, \tag{14}$$

which he observes is decomposable. Each member of this class satisfies all of the invariance, dominance, continuity and subgroup axioms.

Now recall that the CHU measures satisfy all the axioms but decomposability, so that by the characterization result of Foster and Shorrocks (1991) it must be a monotonic transformation of some decomposable index. In fact the Watts index along with Chakravarty's decomposable family, appropriately extended, will do the trick:

$$D_\beta(x;z) = \begin{cases} \dfrac{1}{\beta n}\sum_{i=1}^{n}[1-(x_i^*/z)^\beta] & \text{for } \beta\leq 1 \text{ and } \beta\neq 0 \\ \dfrac{1}{n}\sum_{i=1}^{n}(\ln z - \ln x_i^*) & \text{for } \beta=0 \end{cases} \tag{15}$$

This family of transformed CHU measures satisfies all the axioms, including decomposability. Note that the Watts index (obtained when $\beta = 0$) is the continuous limit of the remaining indices as β tends to 0.[xxii]

4.5 The FGT indices

Return for a moment to the functional form (6) of the Sen measure. The weight on each poor person's normalized income gap is a function of the person's ranking, hence it depends crucially on the incomes of others, not just the person in question. Foster, Greer, and Thorbecke (1984) turn to a simpler form of weighting which depends on the poor person's income and the poverty line. More specifically, they weight the normalized shortfalls by the normalized shortfall itself, raised to a power. The *Foster-Greer-Thorbecke (FGT)* poverty indices are defined as:

$$P_\alpha(x;z) = \frac{1}{n} \sum_{i=1}^{n} (g_i^*/z)^\alpha \quad \text{for } \alpha > 0 \qquad (16)$$

and $P_0 = H$. All indices with $\alpha > 0$ satisfy continuity, monitonicity and strong monitonicity. The index obtained when $\alpha = 1$ is the per capita poverty gap HI, which just violates the transfer axiom. All indices with α beyond 1 satisfy the transfer axiom and, since they are continuous, the strong transfer axiom as well. The invariance axioms and restricted continuity are satisfied by all members of the family.

The measure

$$P_2 = \frac{1}{n} \sum_{i=1}^{n} (g_i^*/z)^2$$

obtained when $\alpha = 2$ has a particularly straightforward structure in which the normalized shortfalls are weighted by the normalized shortfalls themselves, and averaged across the population.[xxiii] This has the effect of emphasizing the shortfalls of the poorest persons. It is easy to show that

$$P_2 = H(I^2 + (1-I)^2 C_p^2),$$

which is the expression involving H, I and the (squared) coefficient of variation among the poor given in (2) above. This relationship with the coefficient of variation ensures that P_2 has an income-based "transfer neutrality" property: the effect of a given-sized progressive transfer between two poor persons a given "income-distance" apart is the same regardless of the absolute levels of income. It is also easy to verify that for α beyond 2 the poverty measures are "transfer sensitive" in that they stress transfers at lower income levels.[xxiv] As with the CHU indices, the parameter α has an interpretation as an indicator of degree to which the distribution of poverty matters.

A quick calculation verifies that each member of this class is decomposable. This was one of the major reasons the class was developed and has why it has been so well received by other researchers interested in evaluating poverty by subgroups. Of course, as a decomposable index, each also satisfies subgroup consistency.

5. SOME NEW OLD INDICES

In the foregoing we have presented the general axiomatic framework for poverty measurement and explored how various indices fit within that structure. We have seen how the literature on poverty measurement has made use of techniques and tools drawn from the inequality measurement literature. We now present an additional example of this type of "technology transfer."

In a recent paper on decomposable inequality measures, Foster and Shneyerov (1999) noted the fundamental role that the *general means*, defined as

$$\mu_\alpha(x) = \begin{cases} (\frac{1}{n}\sum_{i=1}^{n} x_i^\alpha)^{1/\alpha} & \text{for } \alpha \neq 0 \\ \prod_{i=1}^{n} x_i^{1/n} & \text{for } \alpha = 0 \end{cases} \qquad (17)$$

play in inequality measurement.[xxv] The general mean μ_α becomes the *arithmetic* mean for $\alpha = 1$, the *geometric* mean for $\alpha = 0$, and the *harmonic* mean for $\alpha = -1$. General means with higher α place greater emphasis on higher incomes; a lower α emphasizes lower incomes. Of course, for $\alpha < 1$, the general means are simply the special form of *equally distributed equivalent income functions* used by Atkinson (1970) to construct his parametric family of inequality measures, namely, $I = 1 - (\mu_\alpha/\mu)$. The so called *generalized entropy* measures can similarly be defined using the entire range of general means as follows:

$$I_\alpha = \frac{1}{\alpha(1-\alpha)}[1 - (\mu_\alpha/\mu)^\alpha] \qquad \text{for } \alpha \neq 0{,}1. \qquad (18)$$

The two Theil measures are obtained as continuous limits at $\alpha = 0$ and 1.

Foster and Shneyerov (1999) observed that virtually every inequality measure (with the exception of the Gini coefficient) is a function of a ratio of two general means, or the limit of functions of a ratio of general means. Can a comparable statement be made in the context of poverty measurement? The ratio typically entails a comparison between a general mean and the arithmetic mean, which in that context is viewed as an income standard. In poverty measurement, the role of income standard is typically assumed by the poverty line.[xxvi] Can meaningful poverty indices be constructed using the ratio of a general mean and the poverty line?

Consider the per capita poverty gap measure HI = $(1/n)\Sigma_i\, g_i{}^*/z$, which indicates the average income gap across the entire population. This measure can be written as:

$$\mathrm{HI} = \mu(g^*/z) = 1 - \mu(x^*/z)$$

In other words it is the mean of the normalized gaps or the gap associated with the normalized mean. Suppose we were to use a general mean in each of the above expressions, or perhaps a general mean to the power α. The resulting poverty measures would necessarily be sensitive to inequality, but restrictions on parameter values would be needed to ensure that this sensitivity is consistent with the transfer axiom. The four measures obtained are given as follows:

$$A_\alpha(x;z) = \mu_\alpha(g^*/z) \qquad B_\alpha(x;z) = [\mu_\alpha(g^*/z)]^\alpha \qquad \text{for } \alpha > 1$$

$$C_\beta(x;z) = 1 - \mu_\beta(x^*/z) \qquad D_\beta(x;z) = 1 - [\mu_\beta(x^*/z)]^\beta \qquad \text{for } \beta < 1$$

It is clear that the resulting measures satisfy the invariance, continuity, and dominance axioms. The aggregation properties of the general means ensure that each is subgroup consistent and, indeed, the two measures on the right are fully decomposable. Note, thought, that this is all very well known, since B_α, C_β and D_β are essentially the FGT, the CHU and the decomposable CHU measures (apart from the Watts index), respectively. The remaining measure

$$A_\alpha(x;z) = (P_\alpha)^{1/\alpha} = [\frac{1}{n}\sum_{i=1}^{n}(g_i^*/z)^\alpha]^{1/\alpha}$$

is a nondecomposable version of the FGT index whose cardinal interpretation is quite nice and perhaps even easier to understand than FGT itself.[xxvii] The general means play a role in poverty measurement not dissimilar to their role in inequality measurement: All the main poverty measures (excluding the Sen-type measures) are functions of the ratio of a general mean to the poverty line, or a limit of such a functions, where the general means are applied to vectors of income gaps or censored incomes.

NOTES

[i] See Foster and Szekely (2004) for a recent contribution to the identification step that provides a methodology for constructing a "hybrid" poverty line for use across space and time.

[ii] Note that we are defining as poor someone who falls on or below the poverty line, as is done in Sen (1976). One might alternatively use a stronger notion of the poor as anyone whose income falls below the poverty line. See, for example, Donaldson and Weymark (1986) or Atkinson (1987).

[iii] See Bourguignon and Fields (1990) for exercises of this type.

[iv] See Sen (1976).

[v] See Sen (1997) for the definitions of these inequality measures.

[vi] The link between continuity and the strong transfer axiom is discussed in Foster (1984) and Donaldson and Weymark (1986).

[vii] If F is one cumulative distribution function and G is another and L is a range of incomes, we say that F first order dominates G on L if $F(s) \leq G(s)$ for all s in L with < for some s. The second order dominance relation uses integrals of F and G up to the given s. See Bawa (1975), Foster and Shorrocks (1988ab), or Foster and Sen (1997).

[viii] See Foster (1984) and Atkinson (1987) for further discussions.

[ix] See Kanbur (1987) and Ravallion (1994) for discussions of targeting. Ravallion also argues that this first type of analysis is key for policy while the second type (which relies crucially on decomposablity) is less useful. To the extent that he is right, the property of decomposability becomes less a necessity and more a convenience in the analysis of poverty profiles.

[x] See Foster and Shorrocks (1991).

[xi] See also the "subgroup monotonicity" axiom presented in Foster, Greer and Thorbecke (1984)

[xii] Examples are found below and in Foster and Shorrocks (1991, p. 692).

[xiii] The verification that (1) and (6) are the same index follows from Sen (1976). Note that this is the replication invariant version of the Sen measure.

[xiv] To see this, we note that the right hand side of (7) reduces to $HI + (1/(nz))[\Sigma_{j=1}^{m} Y^j_p G^j_p]$, where Y^j_p is the sum of the poor incomes in group j and G^j_p is the Gini coefficient among the poor. Now applying inequality (b) from Zagier (1983) we find that this is less than or equal to $HI + (1/(nz))Y_p G_p$, which is the Sen index over the entire distribution. Note that equality holds if all y^j are identical or are replications of the same distribution.

[xv] See, for example, the discussion of the Gini coefficient in Foster and Sen (1997).

[xvi] As with the original Sen index, I am presenting the replication invariant version of this measure.

[xvii] See the verification given in Foster and Shorrocks (1991, p. 692).

[xviii] See Foster (1984) for specific examples. Given this difficulty, it may be preferable to specify a welfare function with the appropriate properties directly.

[xix] Alternatively, P_W is seen as the difference between the welfare level of the distribution in which everyone has z and the welfare level of the censored distribution: $P_W(x;z) = W(z,...,z) - W(x^*(z))$, where (z,...,z) has n(x) entries.

[xx] Zheng (1993) verified that $u(s) = \ln(s)$ is the only function that will yield a scale invariant poverty measure when substituted into the utility shortfall form.

[xxi] See also Foster and Jin (1998). Note that several of the specific examples discussed by Hagenaars (1987) either violate scale invariance or involve a utility function that apparently depends on z.

[xxii] See Atkinson (1987, p. 755).

[xxiii] Kundu (1981) independently discovered P_2 and verified several of its properties.

[xxiv] Shorrocks and Foster (1987) provide the precise definition of transfer sensitivity. Kakwani (1980) was first to discuss this property with respect to poverty measurement.

[xxv] Foster and Szekely (2003) recently used general means to evaluate the impact of growth on the incomes of the poor.

[xxvi] See Foster (1984).

[xxvii] One difficulty of the FGT indices has been that the values of the measures become very small as α rises. This will not occur with A_α, which uses the power $1/\alpha$ to transform the average of the terms $(g_i{*}/z)^\alpha$ into a number similar in magnitude to the original $(g_i{*}/z)$ terms.

REFERENCES

Anand, S., "Aspects of Poverty in Malaysia," *Review of Income and Wealth*, March 1977, 23(1), pp. 1-16.

Atkinson, A. B. (1970): "On the Measurement of Inequality," *Journal of Economic Theory*, 2, pp. 244-263.

Atkinson, A. B., "On the Measurement of Poverty," *Econometrica*, 55, 1987, pp. 749-764.

Bawa, V., "Optimal Rules for Ordering Uncertain Prospects," *Journal of Financial Economics*, 2, 1975, pp. 95-121.

Besley, Timothy and Ravi Kanbur, "Food Subsidies and Poverty Reduction," *Economic Journal*, 98, 1988, pp. 701-719.

Blackorby, C., and D. Donaldson, "Ethical Indices for the Measurement of Poverty," *Econometrica*, Vol. 48, 1980, pp. 1053-1060.

Bourguignon, F., and G. Fields, "Poverty Measures and Anti-Poverty Policy," *Recherches Economiques de Louvain*, Vol. 56, 1990.

Chakravarty, S. R., "Ethically Flexible Measures of Poverty," *Canadian Journal of Economics*, 16, 1983a, pp. 74-85.

Chakravarty, S. R., "A New Index of Poverty," *Mathematical Social Sciences*, 6, 1983b.

Chakravarty, S. R., *Ethical Social Index Numbers*, Berlin: Springer-Verlag, 1990.

Clark, S., R. Hemming and D. Ulph, "On Indices for the Measurement of Poverty," *Economic Journal*, 91, 1981, pp. 515-526.

Dalton, Hugh, "The Measurement of the Inequality of Incomes," Economic Journal, 91, 1920, pp. 348-361.

Donaldson, D., and Weymark, J.A., "Properties of Fixed Population Poverty Indices," *International Economic Review*, Vol. 27, 1986, pp. 667-688.

Foster, James E., "On Economic Poverty: A Survey of Aggregate Measures," in *Advances in Econometrics*, Volume 3, (R. L. Basmann and G. F. Rhodes, Jr., eds.) 1984, pp. 215-251.

Foster, James E., J. Greer and E. Thorbecke, "A Class of Decomposable Poverty Measures," *Econometrica*, 52, May 1984, pp. 761-766.

Foster, James E. and Yong Jin, "Poverty Orderings for the Dalton Utility-Gap Measures" in *The Distribution of Welfare and Household Production: International Perspectives* (S. Jenkins, A. Kapteyn, B. van Praag, eds.), Cambridge University Press, 1998.

Foster, James E. and A. K. Sen, 1997, "*On Economic Inequality* After a Quarter Century," in Sen (1997).

Foster, James E. and A. F. Shorrocks, "Poverty Orderings," *Econometrica* 52, 1988a, pp. 761-765.

Foster, James E. and A. F. Shorrocks, "Poverty Orderings and Welfare Dominance," *Social Choice and Welfare*, 5, 1988b, pp. 179-198.

Foster, James E. and A. F. Shorrocks, "Subgroup Consistent Poverty Indices," *Econometrica* 59, 1991, pp. 687-709.

Foster, James E., and Artyom Shneyerov, "A General Class of Additively Decomposable Inequality Measures," *Economic Theory*, Volume 14, 1999, pp. 89-111.

Foster, James E., and Miguel Szekely, "Is Economic Growth Good for the Poor? Tracking Low Incomes Using General Means," mimeo, Department of Economics, Vanderbilt University, 2003.

Foster, James E., and Miguel Szekely, "Poverty Lines across Space and Time," mimeo, Department of Economics, Vanderbilt University, 2004.

Greer, Joel and Erik Thorbecke, "Food Poverty Profile Applied to Kenyan Smallholders," *Economic_Development and Cultural Change*, October 1986a, pp. 115-141.

Greer, Joel and Erik Thorbecke, "A Methodology for Measuring Food Poverty Applied to Kenya," *Journal of Development Economics*, Vol. 24 (1), November 1986b, pp. 59-74.

Hagenaars, Aldi J. M., *The Perception of Poverty*, Amsterdam: North-Holland, 1986.

Hagenaars, Aldi J. M., "A Class of Poverty Measures," *International Economic Review*, 28, 1987, pp. 583-607.

Howes, Stephen, "Mixed Dominance: A New Criterion for Poverty Analysis," Distributional Analysis Research Programme, No. 3, London School of Economics, London, 1993a.

Howes, Stephen, "Robust Distributional Analysis," Mimeo, London School of Economics, London, 1993b.

Kakwani, N., "On a Class of Poverty Measures," *Econometrica*, Vol. 48, 1980, pp. 437-446.

Kanbur, S. M. Ravi, "Measurement and Alleviation of Poverty: With an Application to the Effects of Macroeconomic Adjustment," *International Monetary Fund Staff Papers*, Vol. 34 (1), March 1987, pp. 60-85.

Kundu, A., "Measurement of Poverty – Some Conceptual Issues," *Anvesak*, Vol. 11, 1981.

Ravallion, Martin, *Poverty Comparisons*, Switzerland: Harwood Academic Publishers, 1994.

Seidl, C., "Poverty Measurement: A Survey," in Bos D., Rose M., and Seidl C. (Eds.) *Welfare and Efficiency in Public Economics*, Heidelberg: Springer-Verlag, 1988.

Sen, Amartya K., "Poverty: An Ordinal Approach to Measurement," *Econometrica*, 44, 1976, pp. 219-231.

Sen, Amartya K., *Poverty and Famines: An Essay on Entitlement and Deprivation*, Oxford: Oxford University Press, 1981.

Sen, Amartya, *On Economic Inequality*, Oxford: Clarendon Press, 1997.

Shorrocks, Anthony F. and James E. Foster, "Transfer Sensitive Inequality Measures," *Review of Economic Studies*, Vol. 54, July 1987, pp. 485-497.

Shorrocks, Anthony F., "Revisiting the Sen Poverty Index," *Econometrica*, Vol. 63, 1995.

Thon, D., "On Measuring Poverty," *Review of Income and Wealth*, Vol. 25, 1979, pp. 429-439.

Watts, H. W., "An Economic Definition of Poverty," in D. P. Moynihan (ed.), *On Understanding Poverty*, New York: Basic Books, 1969.

Zagier, Don, "Inequalities for the Gini Coefficient of Composite Populations," Vol. 12(2), October 1983, pp. 103-18.

Zheng, Buhong, "Aggregate Poverty Measures," *Journal of Economic Surveys*, Vol. 11(2), June 1997, pp. 123-62.

Zheng, Buhong, "An Axiomatic Characterization of the Watts Poverty Index," *Economics Letters*, Vol. 42, 1993, pp. 81-86.

Chapter 5

SHOULD POVERTY AND INEQUALITY MEASURES BE COMBINED?*

Gary S. Fields
Cornell University,gsf2@cornell.edu

1. INTRODUCTION

It is a great pleasure for me to present a paper at a conference in Erik Thorbecke's honor. Erik was instrumental in attracting me to Cornell. In the years during which we have been colleagues, Erik and I offered the graduate economic development sequence together. We have had many stimulating discussions, none more so than when Foster, Greer, and Thorbecke's P_α poverty index was being developed in Uris Hall. Erik and I have talked often about poverty, not only how to measure it but how to combat it.

As the title of my paper indicates, the purpose of today's presentation is to analyze whether poverty and inequality measures should be blended into a single index. This question was first raised to me by Erik, so it is very fitting that an answer be presented at a conference honoring his distinguished career.

The first step in answering Erik's question is to be clear on what one means by "poverty" and "inequality." For this conference, there can be only one way of measuring poverty: the P_α (or *FGT*) class of indices (Foster, Greer, and Thorbecke, 1984). "Inequality" will be measured by Lorenz curves and a particular Lorenz-consistent inequality index.

The next step is to ask what purpose would be served by blending poverty and inequality measures. Researchers who work with poverty lines and poverty measures judge it valuable to assess deprivation vis-à-vis a poverty line. Deprivation profiles and deprivation indices have been analyzed systematically by Jenkins and Lambert (1997) and Shorrocks (1998). As with many other poverty measures, the P_α class gauges deprivation by fixing a poverty line in real terms and calculating poverty based only on the incomes of people below that line. This is what Amartya Sen dubbed the "focus axiom" in justifying his own poverty measure (Sen, 1976). One critique of this exclusive focus on those at or below a fixed real poverty line is that it is *too* focused, because poverty has a relative as well as an absolute component. That is, *i's* poverty can change not because *i's* real income changes but because *others'* real incomes change. Such "relative poverty" notions are described in Fields (2001). This motivation for combining poverty and inequality may be called the "relative poverty" rationale.

With this as motivation, I now proceed to analyze whether poverty and inequality should be blended into a single index. The first step is to ask, can it be done, to which the answer is "yes, it is possible." The second step is to ask, is it desirable to do it? My answer is, "for the most part, no."

2. NOTATION AND TERMINOLOGY

The P_α poverty measure applies to deficits of any variable among any units of observation. For ease of exposition, I will refer here to "incomes" received by "persons." Incomes among the n persons in the population are denoted by $z_1,\ldots,z_n$ and are assumed to be ordered from lowest income to highest.

The P_α measure was designed to be an index of absolute poverty and is usually thought of in that way. Absolute poverty holds a poverty line constant in real dollars, pesos, or rupees. The poverty line therefore varies with inflation and only with inflation; it is therefore invariant with respect to other changes in the economy, in particular, to economic growth or economic decline. This fixed real poverty line will be denoted here by $\underline{z}$. (It is straightforward to extend the analysis to allow for persons to be of different types and for poverty lines to vary with needs, but I shall not pursue that extension here.)

The P_α index is a function of the normalized poverty deficits of individuals. The normalized poverty deficit of the i´th individual is

$$x_i = \frac{\underline{z} - z_i}{\underline{z}} \text{ if } \frac{\underline{z} - z_i}{\underline{z}} \geq 0, \qquad (1)$$

$$= 0 \text{ if } \frac{\underline{z} - z_i}{\underline{z}} < 0.$$

The P_α index takes each of the normalized deficits, raises them to some power α, and then averages. Specifically, the P_α index, is

$$\mathrm{P}_\alpha \equiv \frac{1}{n} \sum_i x_i^\alpha . \tag{2}$$

3. BLENDING POVERTY AND INEQUALITY

The natural way of combining poverty and inequality would be to combine measures of each into a blended index. The *BLEND* concept is an amalgam of poverty and inequality, and as such can be thought of as an indicator of economic ill-being. Accordingly, any b(.) function is assumed to be increasing in both its poverty and its inequality arguments. Assuming differentiability, such a blend function would be of the form

$$BLEND = b(POV, INEQ),\ b_1>0,\ b_2>0. \tag{3}$$

An example of such a function is

$$BLEND = [w * POV] + [(1 - w) * INEQ], \tag{4}$$

with the weight *w* being chosen by the analyst. For the poverty measure, I shall use the P_α index, and for the inequality measure, the ratio of high incomes to low incomes.

Functions such as (3) and (4) show that it is *possible* to combine poverty and inequality by blending them into a single index. The next question we turn to is whether we *want* to.

4. FOUR STYLIZED GROWTH TYPES

In order to be able to gauge the suitability of blend measures of the form *BLEND = b(POV, INEQ)*, it is useful to choose some stylized types of economic growth and ask how this class of measures behaves in each. I do this based on stylized dualistic development models of the type first formulated in Fields (1979). To sharpen the arguments, the analysis is carried out in a world of two incomes, z_p for the poorer persons and z_r ($> z_p$) for the richer ones. In all cases, the poverty line $\underline{z}$ is assumed to be between both the original and the final values of z_p and z_r.

Four cases of economic growth are considered:

Case I: Uniform percentage enrichment of everyone, raising both z_p and z_r by the same multiple $m > 1$.

Case II: Poorer sector enrichment, which involves raising z_p only, holding z_r constant.

Case III: Richer sector enrichment, which involves raising z_r only, holding z_p constant.

Case IV: Disequalizing growth, which involves raising both z_p and z_r and raising z_p by a smaller percentage than z_r.

Let us now see how the *BLEND* class performs in each of these stylized growth types.

5. APPLYING *BLEND* TO STYLIZED TYPES OF ECONOMIC GROWTH

Consider an index belonging to the class of *BLEND* measures *BLEND* = *b(POV, INEQ)*, $b_1>0$, $b_2>0$. Any such index i) rises when *POV* and *INEQ* both rise or when one rises and the other stays the same, ii) falls when *POV* and *INEQ* both fall or when one falls and the other stays the same, and iii) changes ambiguously when *POV* changes in one direction and *INEQ* changes in the other.

Taking the P_α index

$$P_\alpha \equiv \frac{1}{n}\sum_i x_i^\alpha$$

as our measure of *POV*, we find that P_α falls when some or all of the incomes below the poverty line rise; this happens in Cases I, II, and IV. Taking the rich/poor income ratio z_r / z_p as our measure of inequality, we find that *INEQ* is constant in Case I, falls in Case II, and rises in Cases III and IV.

Looking now at the four growth cases one by one:

Case I involves a uniform percentage enrichment of everyone, raising z_p and z_r by the same multiple $m > 1$. We would expect that economic ill-being would fall if all incomes rise by the same percentage, and indeed both approaches agree with this expectation.

Case II entails poorer sector enrichment, in which only z_p is raised, holding z_r constant. Increases in incomes of the poorer persons, holding incomes of the richer persons constant, lowers poverty and lowers inequality. Because *BLEND* is increasing in both of these arguments, the *BLEND* approach therefore concludes that economic ill-being falls when poorer sector enrichment takes place.

In Case III, the richer people get richer. Because they are above the poverty line to begin with, *POV* does not change. However, inequality as measured by z_r / z_p increases. Constant *POV* and rising *INEQ* imply that economic ill-being *increases* with this type of growth.

Finally, in Case IV, everyone gets richer but those with z_r enjoy larger percentage gains than those with z_p do. Because the poor are getting richer, *POV* falls. And because the richer are getting larger percentage gains, *INEQ* as measured by the ratio z_r / z_p increases. *BLEND* is therefore ambiguous.

In summary, ill-being as gauged by *BLEND* falls in Cases I and II, rises in Case III, and changes ambiguously in Case IV. These results are summarized in Table 5-1.

6. DISCUSSION OF RESULTS

6.1 Discrepancy between *BLEND* results and *POV* results

The discrepancies that arise between the *BLEND* results and the *POV* results in Cases III and IV come about for the following reason. *BLEND* assigns a *positive* change in ill-being to an increase in *INEQ*. On the other hand, the *POV* approach in general and the P_α approach in particular are concerned only with the incomes of the poor, and so give no weight, positive or negative, to the income gains of the non-poor. When z_p rises, *POV* falls and so too does ill-being.

Which is the "better" way to look at distributional change? The choice is best made axiomatically. The interested reader is invited to contrast the POV axioms in Chakraborty, Pattanaik, and Xu (2002) and Foster (this volume) with the *INEQ* axioms in Fields and Fei (1978) and Foster and Sen (1997). You might also try to axiomatize *BLEND*. I have been singularly unsuccessful in getting anywhere with it; I think this is because it is unclear to me what the primitive concept of *BLEND* is.

6.2 Contrast with welfare dominance results

The *BLEND* and *POV* results differ from the results that would be obtained by comparing the distributions using dominance methods. In all four of these cases, applying the methods devised for welfare dominance (Hadar and Russell, 1969; Saposnik, 1981) and for poverty dominance (Atkinson, 1987; Ravallion, 1994), we find that the new income-distribution first-order-dominates the old. Therefore, the class of economic well-being functions of the form

$W=W(z_1, z_2, \ldots z_n)$, W(.) increasing in all z_i

shows *higher* economic well-being when any of these types of economic growth takes place. The unambiguous improvements recorded in Cases III and IV using dominance methods are at odds with the results using *BLEND*. As is well-known (e.g., Shorrocks, 1983) the different judgments come about because of the way the different methods evaluate the income gains of the richer people.

6.3 How does using *BLEND* compare with using its components but not blending them?

Let us now evaluate the usefulness of extending the analysis beyond P_α by using *BLEND*. The advantage is that which Erik Thorbecke first suggested: that although the P_α and other *POV* measures pay no attention to incomes of people above the poverty line, *BLEND* does. In particular, *BLEND* takes account of the incomes of the non-poor by looking at inequality of the income distribution and then combining *POV* and *INEQ* into a single number. An example of how to do this is to choose a particular *BLEND* function – for example,

$$BLEND = [w * POV] + [(1-w) * INEQ], \tag{4}$$

with the weight *w* being chosen by the analyst – and calculating just *BLEND*.

Anyone who only calculates *BLEND* will miss entirely what happens with each of the components. To me, this is a serious omission, because I would most assuredly want to know what happens to *POV* and *INEQ* *separately*, not just the blend of the two.

There is, however, one distinct advantage of using a particular *BLEND* function such as (4) with particular weights *w*. It is that the analyst may wish to use the calculated values to *decide* what s/he thinks of the overall change in the income distribution – in particular, whether the changes that take place have made things better or worse. Of course, for *BLEND* calculations to be useful, the analyst must be extremely careful about which *POV* and *INEQ* measures to use and which weights to choose.

7. CONCLUSION

This paper set out to answer two questions. The first was whether poverty in general and the P_α measure in particular can be extended to include inequality considerations. The answer has been "yes." The second question was, given that the answer to the first question is "yes," is it desirable to extend the *POV* measure in this way? The analysis was carried out using the P_α measure and also *INEQ* and applying them to four stylized growth types. My summary judgment is that poverty and inequality measures can be combined but anyone who does so should be very aware of the limitations involved.

Table 5-1. [Changes in Four Dualistic Growth Typologies]

Growth Type	*BLEND* and its Components			Dominance
	POV	*INEQ*	*BLEND* ≡ *b(POV, INEQ)*	*First order dominance*
Case I. Uniform percentage increases.	↓	→	↓	↑
Case II. Increase in incomes only of the poorer.	↓	↓	↓	↑
Case III. Increase in incomes only of the richer.	→	↑	↑	↑
Case IV. Increase in all incomes with larger percentage increases for the richer.	↓	↑	↑ or ↓	↑

ACKNOWLEDGEMENTS

* I am grateful to Robert Duval Hernandez, James Foster, Dhushyanth Raju, Rumki Saha, and Maria Laura Sánchez Puerta for helpful comments on earlier drafts of this paper.

REFERENCES

Atkinson, Anthony B., "On the Measurement of Poverty," Econometrica, 1987.

Bourguignon, François and Gary S. Fields, "Discontinuous Losses from Poverty, Generalized P_α Measures, and Optimal Transfers to the Poor," Journal of Public Economics, January, 1997.

Chakraborty, Achin, Prasanta K. Pattanaik, and Yongsheng Xu, "On the Structure of Some Measures of Deprivation," Working Paper, July, 2002.

Fields, Gary S., "A Welfare Economic Approach to Growth and Distribution in the Dual Economy," Quarterly Journal of Economics, August, 1979.

Fields, Gary S., Distribution and Development: A New Look at the Developing World. (Cambridge, MA: MIT Press and the Russell Sage Foundation, 2001).

Fields, Gary S. and John C.H. Fei, "On Inequality Comparisons," Econometrica, March, 1978.

Foster, James, "Poverty Indices," Chapter 3 in this volume.

Foster, James, Joel Greer, and Erik Thorbecke, "A Class of Decomposable Poverty Measures," Econometrica, May, 1984.

Foster, James and Amartya K. Sen, "On Economic Inequality After a Quarter Century," in Amartya K. Sen, On Economic Inequality, Expanded Edition. (New York: Oxford, 1997).

Hadar, Josef and William Russell, "Rules for Ordering Uncertain Prospects," American Economic Review, 1969.

Jenkins, Stephen and Peter Lambert, "Three I's of Poverty Curves, with an Analysis of UK Poverty Trends," Oxford Economic Papers, July, 1997.

Ravallion, Martin, Poverty Comparisons. (Chur, Switzerland: Harwood Academic Publishers, 1984).

Saposnik, R., "Rank Dominance in Income Distributions," Public Choice, 1981.

Sen, Amartya K., "Poverty: An Ordinal Approach to Measurement," Econometrica, 1976.

Shorrocks, Anthony F., "Ranking Income Distributions," Economica, 1983.

Shorrocks, Anthony F., "Deprivation Profiles and Deprivation Indices," in Stephen P. Jenkins, Arie Kapteyn, and Bernard M.S. van Praag, eds., The Distribution of Welfare and Household Production. (Cambridge: Cambridge University Press, 1998).

Chapter 6

EQUALITY OF WHAT? EVIDENCE FROM INDIA

David E. Sahn
Cornell University, david.sahn@cornell.edu

1. INTRODUCTION

Erik Thorbecke's work conceptualizing and measuring poverty is among his numerous outstanding contributions to the field of economics. While his seminal work on the now famous Foster-Greer-Thorbecke (FGT) measures has encouraged a generation of economists with a focus on poverty alleviation, I take his inspiration to argue the need for according greater attention to the multiple dimensions of poverty and welfare. While income poverty remains the predominant means indicator for assessing well-being and social progress, there has been increased emphasis placed on defining poverty in terms of the lack of such basic functionings as hunger and malnutrition (Sen 1985, 1987). There are compelling reasons to adopt non-welfarist indicators that are direct measures of well-being and have powerful implications for the individual, household, and community.[i] To borrow Sen's terminology, while income and expenditures are *instrumentally* important, or a means to an end, other measures of well-being, such as nutritional status are *intrinsically* important. Furthermore, the mapping of individuals' functionings to incomes is not constant and involves systematic differences across individuals. Thus, making interpersonal comparisons that take into account their non-income circumstances is an important part of the assessment of inequality.

The emphasis on non-income measures of well-being is now reasonably well established in the policy arena,[ii] and has more recently seen various innovations designed to theoretically and empirically explore the concept of multi-dimensionality in living standards.[iii] Far less has been done, however, in the domain of multiple dimensions of inequality, which still is almost exclusively measured and analyzed in terms of income or expenditure.[iv] This focus on income (or expenditure) inequality ignores the fact that inequality in well-being may have different faces than income alone. Instead, functionings and capabilities, such as health and nutrition, may also be distributionally progressive or regressive. Like inequality in incomes, there is also an important question of whether inequality matters for non-income dimensions of well-being, such as health and education. And specifically, to the extent that these non-income measures are included in a concave social welfare function, and individual utility is a function of not only levels of, but relative well-being, there is ample motivation to focus on dimensions of inequality beyond income exclusively.

Exploring the non-income dimensions of inequality also helps concentrate attention on the fact that, again similar to income, improvements in well-being, such as health or education, may not be equally distributed. Put differently, the same level of average health in a population – for example, the number of individuals failing to achieve a basic level of health (say, a rate of growth in children that is consistent with good health) – can differ depending on how health outcomes are distributed. And, a change in a given average level of health can be distributed differently, thereby affecting the extent to which, for example, investment in health care infrastructure will affect the share of the population who are malnourished or unhealthy.

In this paper, I present an approach to measuring an important dimension of inequality: health. Using data from India, I make comparisons, both spatially between states, and inter-temporally, to both illustrate the methods for measuring and decomposing health inequality, while providing some interesting empirical findings that are compared to results presented elsewhere on income inequality. More specifically, after discussing the methods employed in Section 2 and data used in Section 3, I first present inequality measures for health by state, and by year, employing statistical tests of both cardinal inequality measures as well as tests of stochastic dominance to order inequality spatially and temporally in India. Following this, I conduct several decompositions. This includes decomposing the inter-temporal changes in nutrition poverty into the effect of changes in the mean heights versus changes in the distribution; and also examining the extent to which malnutrition would be reduced in the various states if the distribution of heights corresponded to the patterns that exists in Kerala. Finally, using estimates of income inequality from the same states and for the same year, I examine the extent to which the story on non-income poverty and income poverty are the same, thus addressing whether "Inequality of what?" is an important question.

1.1 Defining Health Inequality

Health inequality has increasingly been the subject of attention by both policy makers and researchers, particularly by empirical economists and sociologists. Two major approaches to defining health inequality are found in the literature. The first, representing the vast majority of the work in this area, involves examining inequality in health across a variety of dimensions of social and economic stratification. These include income, race, ethnicity, location, and gender. Making comparisons of health across populations with different social and economic characteristics is often referred to in the literature as following the so-called "gradient" or "socioeconomic" approach to health inequality. Much of the motivation for this work on the gradient approach to health inequality arises out of fundamental concerns over social and economic justice. The roots of the gradient will often arise from various types of discrimination, prejudice, and other legal, social, and economic norms that may contribute to stratification and fragmentation, and subsequently inequality in access to material resources and various correlated welfare outcomes.

The second approach, which has been referred to as the univariate approach to measuring pure health inequality, involves making comparisons of cardinal or scalar indicators of health inequality and distributions of health, regardless of whether health is correlated with welfare measured along other dimensions. This univariate approach measures pure inequalities in health (or other non-income measures of well-being, such as education), quite analogous to what is done for income distribution[v] – but rather than ordering individuals along the x-axis by income (or expenditures) and drawing Lorenz curves using cumulative income, health is used instead to both order individuals from the least to most healthy, and describe the distribution of health.[vi]

Put more formally, consider a social welfare function, S, which as formulated by Atkinson (1970) can be defined in terms of the additively separable utilities across persons i. Instead of individual utility being measured in terms of income alone, however, I define the utility function with two arguments, say, income, y, and health, h, where utility is an increasing and concave function of both these arguments, such that:

$$S = \frac{1}{N}\sum_{i=1}^{N} U(h_i, y_i)$$

The gradient or socioeconomic inequality approach assumes that the two factors, income, y, and health, h, are substitutes in the social welfare function, $U''hy < 0$, while the univariate approach implies $U''hy = 0$, or $U(h) + U(y) = U(h,y)$. The implication of the view that the two factors are substitutes is that a worsening in health status of a population can be compensated in part by poor health being less concentrated among the income poor. Likewise, the substitutability implies the social welfare function is greater if poor health care is concentrated less among the poor, even if the distribution of pure health inequality worsens. Pure inequality, in contrast, assumes that the two factors are independent.

The preponderance of the literature on health inequality is focused on the gradient, and the socioeconomic correlates of poor health. Findings generally reveal a positive correlation between health and many indicators of socioeconomic status or various measures of social stratification.[vii] We know from the production function literature, however, that the correlation between health and other social indicators, including income and expenditures, is sufficiently weak that in any given sample, income and the measures of social stratification usually predict only a small portion of the variation in health status. This reflects that many other observable and unobservable factors are important in explaining health status. A wide variety of social and economic circumstances and behaviors, that are not captured by the gradient, matter. These include the psychological state of the primary care-giver, weaning and other feeding practices, the social norms and behaviors that govern sexual transmission of diseases, and the natural occurrence of trace minerals and vitamins available in soils and foods. Community factors often matter, too. The availability and quality of the health care system and related public health measures such as water and sanitation, vaccination coverage, etc., have all been shown to be of equal or greater importance in determining child health than income (Pitt and Rosenzweig 1986; Thomas, Lavy, and Strauss 1996; Sahn and Alderman 1997; Haughton 1997).

By implication, reducing income inequality, or the health/income gradient, will not necessarily be an effective way of dealing with pure or univariate inequality in health – that is, inequality of health ordered by the health status of the population[viii] – since the gradient fails to capture the fact that a wide range of factors beyond income (or whatever measure of social stratification is being employed) influence health. This is especially the case because income distributions have a long right tail, an issue I will come back to in more detail in the final section of this paper.

The conceptual weakness of the widespread practice of measuring health inequality based on the gradient approach (relative to income of other measures of social stratification) can be illustrated by drawing upon an example from Deaton (2003) where he considers two populations, A and B, with equal levels of average health and equal levels of pure health inequality. Assume that in population A there is a strong correlation between health and income, and in B, the correlation between health and income is weak, and thus, the gradient is quite flat. I would argue that it is not always appropriate to adopt the widely held view that health inequality in population A is a more serious public policy problem than in population B, owing to the stronger correlation of health with income (or some other measure of social stratification) in the case of the former.

One implication of this substitutability, implied by the gradient approach, is as follows: Consider the case of the Titanic going down. Social welfare would be improved if more of those traveling steerage were saved at the expense of those wealthier passengers traveling on the upper decks.[ix] Naturally, the shape of the utility function will determine the extent that these considerations apply. But regardless, I remain troubled by the prospect of justifying a worsening of health status of a part of the population in the name of reducing the gradient; or put another way, I start from the agnostic view that that cross derivative should be zero, at least until I have greater basis to suggest otherwise and know exactly what the tradeoffs should be. In doing so, however, I also acknowledge what I consider the most important limitation of this approach: that the independence of the two factors makes it difficult to rank inequality in society, especially when the ordering of inequality differs by factors.

Thus, I have argued elsewhere for a greater focus on the development of a measure of pure health inequality, given that it has a number of compelling motivations, both relative to the gradient approach and relative to the standard univariate income approach to measuring inequality that has been at the core of modern economics.

2. DATA

The data used in this paper are the 1992-1993 and 1998-1999 National Family Health Surveys of India. The surveys are designed to be representative at the state and regional level and form a comprehensive database on health and demographic characteristics of the population in India. Consequently, these are large surveys. For example, the 1998-99 survey collected data from 91,196 households, and over 30,000 children up to three years of age were included in the survey. In addition to the anthropometric data used in this paper, the surveys contain data on fertility, family planning practices, health surveys, health status, education, and other measures of the standard of living. Funding from the survey comes from the U.S. Agency for International Development through ORC Macro. The International Institute for Population Sciences was the agency selected by the Ministry of Health and Family Welfare of India as the nodal agency for this data collection effort.

3. METHODS

3.1 Choice of Health Indicator

The measure of health inequality used in this paper is based on the height of children up to 36 months old. Indeed, there are many dimensions of health, and no single indicator of health status captures all aspects of health. The growth of children, however, has some desirable properties and characteristics for the purposes of this paper.

First, there is an abundance of good quality and comparable anthropometric data on growth of pre-school age children in India (and elsewhere). Second, growth of children is widely acknowledged as an excellent and objective indicator of general health status of children (Cole and Parkin 1977; Mata 1978; Tanner 1981; Mosley and Chen 1984; WHO 1995; Martorell et al. 1975, Beaton et al. 1990; Strauss and Thomas 1995; Behrman and Deololikar 1988), capturing various dimensions of the health of individuals and the developmental and socio-economic environment in which they grow. Third, measuring growth of children is not susceptible to self-reporting bias, and the errors in measurement are unlikely to be correlated with socioeconomic characteristics. Fourth, unlike data on adult anthropometrics, we need not worry about the genetic (racial or ethnic) makeup of the population influencing our comparisons, since it is now well established that the distributions of heights of healthy children among populations are comparable around the world (Habicht et al. 1974, Bustos et al. 2001).

There are also serious methodological and conceptual limitations among the competing options for measuring inter-personal variability in health, including life expectancy, mortality, and morbidity. In the case of life expectancy, one approach would be to examine the number of years a cohort, say, born in 1980, would be projected to live. However, the life tables used to calculate life expectancy are based on data collected at a given point in time, but that time does not correspond to the future experiences of those presently alive (Deaton 2003). A further problem with this approach is that it does not indicate anything about the "healthiness" of the life spans observed, and the various techniques for doing so are riddled with methodological challenges that make them inappropriate for my purposes. As for the use of infant or child mortality, it is plagued by the fact that the rarity of death implies that very large samples are required (Mosley and Chen 1984); in addition, one cannot study inequality with a discrete variable. The alternative of using a predicted probability of death is not feasible since the variation in predicted mortality will be substantially less than the true variation in mortality probability. And, as for morbidity, not only is mis-measurement of self-reported illness well documented for both specific ailments and general health status (Kroeger 1985; Hill and Mamdani 1989; Over et al. 1992; Schultz and Tansel 1997), but such bias in reporting is not random, further precluding use of such indicators to characterize the distribution of health.

I therefore use anthropometric data, specifically children's growth, to assess health inequality. In that regard, the standard procedure for analyzing anthropometric data, such as height, is to transform it into age and gender standardized z-scores, or standard deviation scores, to assess the extent to which a child is above/below the median of the distribution of healthy children. The limiting distribution of the z-score is standard normal, so a child who is below –2 z-scores has only a very low probability of being of normal height. Thus, the World Health Organization (1983) takes –2 z-scores to be the cut-off point or nutrition poverty line, below which a child is judged to be stunted or "nutrition poor."

Since z-scores can be negative, and typically are for most poor children, they do not work well for my purpose of defining distribution statistics, such as the Gini or log-variance. Thus, rather than use z-scores, I use "standardized heights." Each child's height is transformed to the height for a reference age and gender. In this paper, I arbitrarily use 24-month-old girls, although the results are not sensitive to the choice of age/gender for standardization. The standardized height measure is constructed such that a child's position in the distribution, in terms of percentiles, is the same for actual height in the actual age/sex group and the transformed height in the reference group WHO distribution. More specifically,

$$H = F_{\bar{a},\bar{g}}^{-1}(F_{a,g}(h))$$

where F is the distribution function of heights in the WHO population for age/sex group defined by a (age) and g (gender), h is the actual height, $\bar{a}$ = 24 months, $\bar{g}$ = female, and H is standardized height.

3.2 Comparing Inequality

Using the standardized heights, I summarize the distributions using the extended Gini coefficient,[x] where:

$$G(v) = -v * Cov\{h,[1-F(h)]^{(v-1)}\} / \bar{h} \quad v>1$$

In the case of my health inequality measure, h measures individuals' health (standardized height); F(h) is the cumulative density function of the welfare ordering; $\bar{h}$ is mean welfare (standardized height); and v is a parameter that affects the weighting of each point on the Lorenz curve. I thereafter make comparisons of the Ginis based on the computation of standard errors to order inequality among the states and across time. In addition, I calculate the log variance measure of inequality to later make comparisons with the income inequality data that use that indicator.

While the Ginis are useful for making comparisons of the degree of progressivity in the health indicator, such cardinal indicators lack the rigorous foundation in welfare analysis of ordinal measures that compare the entire distribution of income or health. Therefore, I also employ tests of stochastic dominance that are very general and provide an ordering of progressivity for any anonymous social welfare function that favors equality.

Because the Lorenz curves are constructed from sample data, comparisons between them need to be statistical.[xi] Davidson and Duclos (2000) derive distribution-free standard errors for the difference between two concentration curves, which I use to establish a confidence interval around the estimated Lorenz curves. Thereafter, I test for significant differences between them.

More specifically, I apply statistical tests to the difference between the ordinates of Lorenz curves at nine abscissa (i.e., 0.1, 0.2 ... 0.9) and reject the null hypothesis of non-dominance when any one of the ordinates does not differ statistically in the direction of dominance.[xii] The failure to reject the null of non-dominance using the above procedure implies that we cannot say anything about the relative progressivity of one distribution versus the other under a broad class of social welfare function. This is indeed a demanding criterion, especially in light of the low power of the test. I therefore also show results where I allow the bottom or top ordinate pair not to be significantly different, although, the direction of difference must be the same as with the other statistically significant pairs.

3.3 Decompositions

I employ widely used methods proposed by Datt and Ravallion (1992) and Kakwani (1997), as adapted elsewhere by Sahn and Younger (2005), to decompose the total change in nutrition poverty headcount index and the nutrition poverty gap index into the impact of changes in the mean versus changes in the distribution of heights. Then I further employ these decomposition techniques to determine to what extent the nutrition problem in Indian states would be mitigated if the distribution of height followed the highly progressive pattern observed in Kerala.

The point of departure for the decomposition exercise is the estimation of the nutrition poverty indexes, following the work of Sahn and Stifel (2002) that applies the FGT-type poverty measures to the nutrition data. This class of measures can be written as:

$$M_\alpha = \tfrac{1}{N}\sum_{i=1}^{N}(z - y_i)^\alpha I(y_i \le z),$$

where y_i is an independent observation of my welfare indicator (standardized height) from a sample of size *N*, *z* is the poverty line – equal to the 2.27th percentile of the cumulative distribution (approximately equal to -2 z-scores) – and $I(\cdot)$ is an indicator function that takes on a value of one if its argument is true, and zero otherwise.[xiii] When α is 0, 1, and 2, we have, respectively: the *prevalence* of nutrition poverty index, or the percentage of the population who are malnourished; the nutrition *gap* index, or the mean distance below the poverty line (and where the mean is defined over the entire population); and the nutrition *severity* index (or the squared nutrition gap) defined as the mean squared nutrition gap.[xiv]

The basic approach to decomposing changes over time, or differences between states, can be captured using a class of measures that are fully characterized by the cut-off line (*z*), the mean of the distribution (μ), and the Lorenz curve (*L*). For date *t*, the nutrition poverty measure can be written as

$$N_t = N(z, \mu_t, L_t).$$

A change in nutrition poverty between period *t* and *t*+*n* can then be decomposed as follows:

$$N_{t+n} - N_t = \underbrace{G(t,t+n;r)}_{\text{growth component}} + \underbrace{D(t,t+n;r)}_{\text{redistribution component}} + \underbrace{R(t,t+n;r)}_{\text{residual}}$$

The component *G()* is defined as the change in nutrition poverty due to a change in the mean of the distribution, while holding the Lorenz curve constant at that of the reference year *r*:

$$G(t,t+n;r) \equiv N(z, \mu_{t+n}, L_r) - N(z, \mu_t, L_r).$$

Similarly, the redistribution component, *D()*, is defined as the change in the Lorenz curve while keeping the mean of the distribution constant at that of the reference year *r*:

$$D(t,t+n;r) \equiv N(z,\mu_r,L_{t+n}) - N(z,\mu_r,L_t).$$

The presence of the residual *R()* occurs whenever a change in the poverty measure due to changes in the mean (distribution) also depends on the precise distribution (mean).

Kakwani (1997) has argued for an axiomatic approach to the decomposition problem, in which the growth and inequality contributions to changes in overall poverty should be of the same magnitude and of opposite sign when going from the base to terminal year, and vice versa. This has led to a widely employed approach that basically averages the two growth components – a change in the mean income when the distribution is fixed at $t = 0$, and the other when $t = 1$. A similar averaging is applied to the two distribution components. This practice has been adopted widely (McCulloch et al. 2000; Dhongde 2002; Shorrocks and Kolenikov 2001; Christiaensen, Demery and Paternostro 2000), not only because it is consistent with the axiomatic properties proposed by Kakwani, but also because it has the additional advantage of dealing with the residual which is difficult to interpret.

4. RESULTS

4.1 Inequality comparisons across states

In Figures 6-1 and 6-2, I present the difference between the Lorenz curves and the 45-degree line for the various Indian states using the data from 1999.[xv] The visual results are suggestive of the significant differences in the degree of inequality. For example, the results in Figure 6-1 would suggest that health inequality is a more serious problem in Bihar than Maharashtra; similarly, the curves in Figure 6-2 suggest that inequality is less of a problem in Arunachal Pradesh than in most of the other states shown.

The results of the health inequality Gini coefficients are found in Table 6-1 for each state in 1999. As expected, the inequality parameters closely reflect the story depicted in the figures. I also rank the Ginis by assessing whether the differences between them are statistically significant. The results of the statistical comparison are found in the off-diagonal matrix in Table 6-1, where an "X" signifies that the Ginis are significantly different from each other – or more specifically, that the Gini in the row is statistically lower than the Gini in the column. There are cases, for example, where Andrah Pradesh has a lower Gini than West Bengal according to the first column of Table 6-1, but statistically, one cannot reject the null, that inequality is the same in the two states. Similarly, although the Gini in Orissa of 0.367 is much lower than 0.391 in Arunachal Pradesh, statistically they do not differ from each other.

While the ranking of Ginis or other inequality parameters are informative, as discussed above in the methods section, such cardinal measures do not necessarily imply an ordering of inequality for each and every anonymous welfare function. Thus, a far more rigorous and demanding approach to comparing distributions involves relying on tests of stochastic dominance. The results of these inter-state comparisons are found in Table 6-2. The letter "D" denotes the rejecting of the null of non-dominance between the pairs of state-specific health distributions, whereby the row dominates the distribution of the state shown in the column. As expected, given the generality of the dominance tests, I fail to reject the null quite frequently. This implies a fair amount of ambiguity when using these tests to rank the inequality of health outcomes among the states. When I test dominance over a slightly more restricted range of test points – allowing either the top or bottom test point to be insignificant – I get far more results, as shown the lower case "d" in Table 6-2.

Because of the property of my health inequality measure that implies that most inequality will be found in the left hand side tail, rather than the right tail as in income inequality, I would expect there to be a negative correlation between health and health inequality.[xvi] The question emerges, however, as to whether the inequality measure itself is a close transformation of the nutrition poverty headcount value, which is reported in Table 6-5. There is a negative and significant correlation between health inequality and health itself across the states, although the strength of this correlation is sensitive to the inclusion of a couple of individual states. For example, the correlation coefficient in 1993 is 0.55 and significant at the 10 percent level. When I eliminate Kerala, however, the correlation falls markedly and is no longer statistically significant, even at the 10 percent level. Of equal importance is that there are a number of cases that have nearly identical inequality parameters, but levels of stunting that vary widely. For example, in 1999, the levels of health inequality in Karnataka and Haryana are 3.89 and 3.87, respectively. However, the level of stunting in the former is 36 percent, and in the latter, 50 percent. Similarly, Assam and Rajasthan have nearly the same level of stunting, but inequality parameters of 3.72 and 5.21, respectively. This puts the rankings for Assam and Rajasthan at 9 and 19 out of 20, respectively. Thus, it is clear that a state's overall health status, as measured by the nutrition headcount indicator, is not a good predictor of health inequality, and vice versa.

How do the results of health inequality compare to traditional income inequality measures? To answer that question, I rely on a recent paper of Deaton and Dreze (2002) where they present inequality measures, by state, for the same time periods as the DHS data I use in this paper. Since they compute log variance as their measure of inequality, I do the same for standardized heights (Table 6-3). The ranks of the income and health inequality measures are shown, as are the correlation coefficients. The results indicate a negative and significant correlation. For example, the Spearman rank correlation between the log variance of health and income was –0.24 in 1993, and jumps up to –0.49 in 1999. This reflects, for example, cases such as Bihar, where health inequality is quite high, while income is among the most equitable of all the Indian states; or conversely, the case of Kerala, where health inequality is low, and income inequality is high.

4.2 Inter-temporal comparisons

Similar comparisons of health inequality can be made inter-temporally. In Table 6-4, I present the health Ginis for the various states for which there are data from 1993 and 1999. While there is considerable shifting of the rankings between 1993 and 1999, states with low inequality include Goa, Kerala, and Punjab, while high health inequality states include Arunachal Pradesh, Bihar, and Rajasthan.

Those states where the changes in Ginis are statistically significant are shown in bold. Out of the 20 states, there are statistically significant changes in 12. Among those, seven witness a worsening of inequality, and five, an improvement. In the final column, a "D" is entered where I can reject the null of non-dominance. The number of statistically significant changes drops to nine, with the differences for Goa, Punjab, and Tripura not proving significant for the more general test of dominance.

4.3 Decompositions

4.3.1 Means versus distribution

It is clear from the results that there are differences in the distribution of health among the Indian states, whether I use cardinal measures such as the Gini and the log variance, or the more general dominance criteria. However, as alluded to above, the question remains as to how much practical importance there is to the distributional component of health, versus the mean level. The role of changes in the mean level of welfare, versus the distribution, has been a source of considerable attention in the income literature. And while the evidence generally points to changes in mean levels of income as being the key to poverty alleviation, rather than the distribution of income, it is also clear that the income distribution affects the extent to which a given amount of growth reduces poverty (Ravallion 1995; Chen and Ravallion 2001; World Bank 2000).

To explore a similar question in the context of my health indicator, I next turn to the decomposition of changes in child health over time into the effects of a shift in the means versus the underlying distribution of heights. Table 6-5 presents the results of this decomposition, with the third column of data showing the actual change in the nutrition headcount index, followed by an examination of the contribution of the overall change in mean heights, assuming that the health distribution remains the same, and the distribution of heights, assuming that the mean levels remain constant, to this difference. I present results using both the Datt-Ravallion and Kakwani decompositions.

The results indicate that the effect of shifts in the mean are in general of a larger magnitude than the redistribution effect in explaining changes over time among the 19 states for which there are data. Similarly, any large movements over time, such as observed in Haryana and Rajasthan, are primarily explained by a shift in the mean, rather than redistribution. This is broadly consistent with what has been observed elsewhere for explaining changes in income poverty (Datt and Ravallion 1992; Kakwani 1997;

Balisacan 2000; Bigsten et al. 2003), as well as what I have found elsewhere for nutrition poverty in other countries (Sahn and Younger 2005).

Even when the shift in the mean drives the results, there are important redistribution effects that impact the degree of overall change. For example, the shift in the mean level of height in Karnataka would have resulted in a 5.55 percent decline in malnutrition, holding the distribution of heights constant. However, the opposite signed distribution effect mitigates the drop in nutrition poverty, which was only 3.85 percent. In Kerala, where the decline in the nutrition headcount index was only 3.12 percent, the overall decline attributable to the shift in the mean was only 1.72 percent. However, the improvement in the distribution of health contributed to a nearly comparable overall improvement as observed in Karnataka. In other cases, such as Goa, the nearly 7 percent decline in the headcount index that was attributable to the shift in the mean was nearly wiped out by the 4.93 increase due to a worsening of the distribution of heights. A similar story applies to the inter-temporal results from Tripura.

Table 6-6 shows results of the decompositions for the nutrition poverty gap measure. In this case, the redistribution component takes on a relatively greater importance than shifts in the means. And in many cases, such as in Maharashtra and Punjab, the redistribution component is around three times greater in magnitude than the effect of mean shifts.

While it is clear that distribution and changes therein matter over time in explaining the extent to which there are secular improvements in health status, there is a related question that I also address: what would nutrition poverty indicators look like among the various Indian states if the distribution of heights was the same as in Kerala in 1999, keeping the mean level of heights in the states at a level that was observed in 1999. The results, found in Table 6-7, suggest that in some states, a distribution of income comparable to Kerala would result in a large decline in nutrition headcount measures, while in other states this is not the case. For example, consider Rajasthan and West Bengal, both with roughly the same level of malnutrition, 43 and 42 percent, respectively. If the distribution of child heights in Rajasthan mirrored that in Kerala, the level of malnutrition would be 32 percent. Contrast this, however, with West Bengal, where the level of malnutrition would fall only a trivial amount, to 40 percent, if the height distribution was the same as that observed in Kerala. Similarly, take the cases of Andrah Pradesh and Punjab, both with 39 percent of the children falling below the nutrition poverty line. Redistribution of the heights to be consistent with the degree of inequality in Kerala, while maintaining the mean at the existing level, would result in no change in malnutrition in Andrah, but a decline from 40 to 35 percent in the case of Punjab.

A similar story is observed in Table 6-7, which reports comparable results for the nutrition gap measure. While the magnitude of the drop in the index is small, proportionally, it is quite large in many instances. For example, the decline from 3.3 to 2.6 in Uttar Pradesh is around 25 percent of the first period level, while the decline is Bihar is nearly a third of the original value.

5. DISCUSSION

In this paper, I explore the spatial and temporal patterns of health inequality among the states of India. I am motivated by the fact that there is a paucity of research on the non-income dimensions of inequality, and the fact that health inequality, in particular, may be considered of equal importance to public policy as more traditional measures of income inequality.

The development of cardinal indicators of health inequality, based on child growth, as well as ordinal comparisons of distributions using tests of stochastic dominance, show that there are marked differences among the Indian states. Likewise, I find that there are significant changes in health inequality when making inter-temporal comparisons for individual states.

While there are statistically significant differences in the distributions and inequality measures, their economic and social significance are not clear by looking at the numbers. Therefore, I address the question of whether reducing health inequality matters, even without improvements in mean health. I decompose the inter-temporal changes in nutrition poverty into the effect of changes in the average heights versus changes in the distribution. The results suggest that the changes in the distribution of height are less important than the changes in the mean values, when explaining the evolution of the nutrition poverty index over time in India. However, this does not apply to the nutrition gap index, where the distribution component is generally of equal importance as the change in the mean in terms of explaining the changes in nutrition poverty. But perhaps of greater importance is that I find that the level of stunted growth would be reduced among the various Indian states in a rather marked fashion if the distribution of heights corresponded to the pattern that exists in Kerala, where health of children is relatively equally distributed. Thus, there is ample scope for redistributive policies to raise health, even without an overall improvement in the mean.

Of course, any discussion of redistributive policies, when applied to health, implies that health outcomes can be redistributed in a manner analogous to income. While income can be redistributed in the short-term through measures that are now well understood, in practice, redistribution of health will likely be achieved only over the medium to long term through redistribution of services and the other determinants of health, such as access to health services, nutrients, and so forth.

Another finding that deserves further consideration is the difference between the results for my health inequality indicator and income inequality. Why is this the case? Part of the reason could be simply that the underlying factors that determine income inequality are different from those that contribute to health inequality. For example, the income inequality is undoubtedly explained by the nature of the labor market, the role of non-earned incomes, including the large flows of remittances from overseas workers, and the distribution of productive assets. While these factors may also directly affect health inequality, perhaps of greater importance is the long-term commitment of the state to promoting social welfare and equality through investments in services and institutions that affect health directly. Beyond the different determinants of health and income equality, however, there are properties of the indicators themselves that contribute to the propensity for different outcomes. Specifically, the long rightward tail of income distributions implies that inequality can be unduly influenced by the welfare of the rich. The height distribution does not have a rightward skew; measuring height inequality does not suffer from this problem. If anything, the opposite is true, a fact that may in part reflect the relationship between income and health, particularly when measured by nutrition indicators. The income-health relationship is a strongly concave function. As incomes improve from low levels, the associated improvements in health taper off quickly. Even modest improvements in a population's welfare can compress its height distribution significantly, precisely because height has a genetic upper bound while income does not. Since most policy analysts care, rightly, about the poor, I would argue this is yet another reason to consider the use of non-income dimensions, such as health inequality, in considering the relative distribution of welfare in a population.

It is finally important to emphasize once again that I am looking at only one dimension of health, that captured by the growth of young children. Inequality in health defined along other dimensions may look quite different, and in fact, may be far less stable over time than child health inequality. One obvious example would be if we could measure health inequality based on life expectancy, which incorporates the entire population in these estimates. The impact of a disease such as HIV/AIDS would be dramatic and may contribute to a very different looking health distribution than I have observed with child heights, as well as to a distribution of health that changes markedly over short periods as the epidemic grows. Clearly, considering alternative health indicators, and the processes that contribute to pure inequities in health, is an important area for future research.

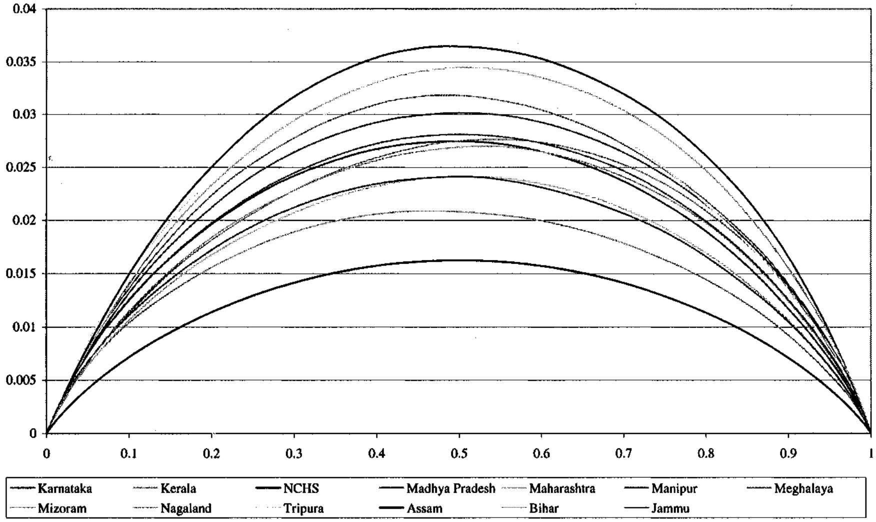

Figure 6-1. [Standardized heights (1999) Distance of Lorenz curves from the 45 degree line]

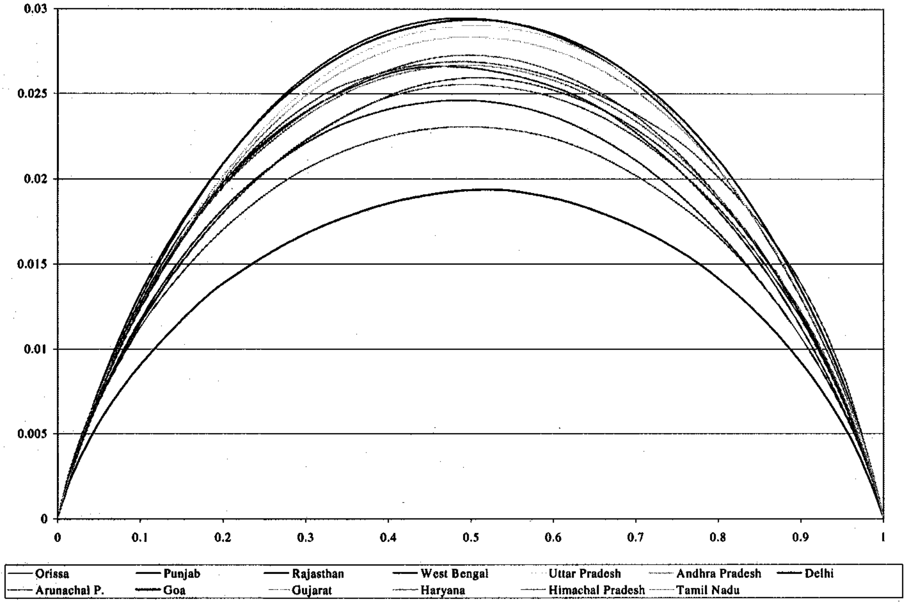

Figure 6-2. [Standardized heights (1999) Distance of Lorenz curves from the 45 degree line]

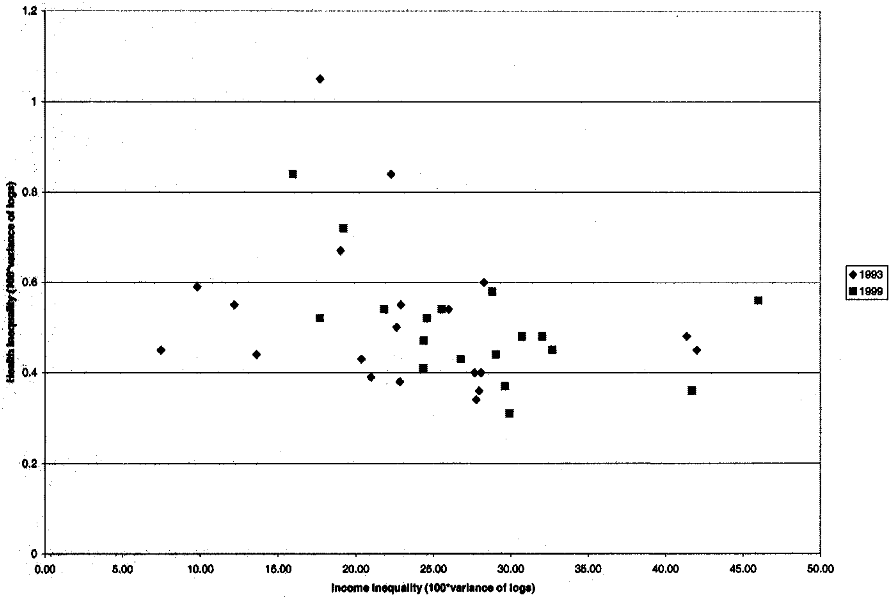

Figure 6-3. [Health and income inequality for Indian states, 1993 and 1999]

Table 6-1. [Health inequality ordering of Indian states based on the Gini coefficient (1999)]

State	Gini Co-efficient*	Kerala	Andhra	Maharashtra	Manipur	Sikkim	W. Bengal	Orissa	Himachal P.	Tamil Nadu	Punjab	Mizoram	Haryana	Karnataka	Arunachal P.	Nagaland	Jammu & K	Gujarat	Uttar P.	Rajasthan	Delhi	Madhya P.	Meghalaya	Tripura	Bihar	Assam
Goa	0.2800		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Kerala	0.3032		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Andhra P.	0.3350							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Maharashtra	0.3384							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Manipur	0.3411							X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Sikkim	0.3494											X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
W. Bengel	0.3552										X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Orissa	0.3670																X	X	X	X	X	X	X	X	X	X
Himachal P.	0.3685																X	X	X	X	X	X	X	X	X	X
Tamil Nadu	0.3737																X	X	X	X	X	X	X	X	X	X
Punjab	0.3792																	X	X	X	X	X	X	X	X	X
Mizoram	0.3859																		X	X	X	X	X	X	X	X
Haryana	0.3873																		X	X	X	X	X	X	X	X
Karnataka	0.3885																		X	X	X	X	X	X	X	X
Arunachal P.	0.3913																					X	X	X	X	X
Nagaland	0.3924																					X	X	X	X	X
Jammu & K	0.4022																					X	X	X	X	X
Gujarat	0.4052																					X	X	X	X	X
Uttar P.	0.4122																					X	X	X	X	X
Rajasthan	0.4149																							X	X	X
Delhi	0.4205																								X	X
Madhya P.	0.4284																								X	X

Table 6-1 [Cont.]

State	Gini Co-efficient*	Kerala	Andhra	Maharashtra	Manipur	Sikkim	W. Bengal	Orissa	Himachal P.	Tamil Nadu	Punjab	Mizoram	Haryana	Karnataka	Arunachal P.	Nagaland	Jammu & K	Gujarat	Uttar P.	Rajasthan	Delhi	Madhya P.	Meghalaya	Tripura	Bihar	Assam
Meghalaya	0.4425																								X	X
Tripura	0.4572																									X
Bihar	0.4829																									X
Assam	0.5217																									

* Multiplied by 10

“X” in cell (Row entry, Column entry) indicates that the Gini for the Row entry is statistically less than the Gini for the Column entry.

Table 6-2. [Dominance table of health inequality for Indian states (1999)]

State	Kerala	Andhra	Maharashtra	Manipur	W. Bengal	Orissa	Himachal P.	Tamil Nadu	Punjab	Mizoram	Haryana	Karnataka	Arunachal P.	Nagaland	Jammu & K	Gujarat	Uttar P.	Rajasthan	Delhi	Madhya P.	Meghalaya	Tripura	Bihar	Assam
Goa				d	D	d	d	D	D	D	D	D	d	D	D	D	D	D	D	D	D	D	D	D
Kerala					d		d	d	d		D	d	d		d	d	d	d	D	D	D	D	D	D
Andhra P.					d			d	d	d	D	d		d	D	d	d	D	D	D	D	D	D	D
Maharashtra						D			d		D	D		D	D	D	D	D	D	D	D	D	D	D
Manipur									d		d	d			d	d	d	d	D	D	D	D	D	D
W. Bengel																		d	D	d	d	d	D	D
Orissa																d	D	d		D	d	d	D	D
Himachal P.																d	d	d	d	D		D	D	D
Tamil Nadu																		d	d	D	D	d	D	D
Punjab																						D	D	D
Mizoram																				d		d	D	D
Haryana																						d	D	D
Karnataka																				d		d	D	D
Arunachal P.																							d	D
Nagaland																							d	D
Jammu & K																							D	D
Gujarat																							D	D
Uttar P.																							D	D
Rajasthan																							D	D
Delhi																								D
Madhya P.																							D	D

Table 6-2 [Cont.]

State	Kerala	Andhra	Maharashtra	Manipur	W. Bengal	Orissa	Himachal P.	Tamil Nadu	Punjab	Mizoram	Haryana	Karnataka	Arunachal P.	Nagaland	Jammu & K	Gujarat	Uttar P.	Rajasthan	Delhi	Madhya P.	Meghalaya	Tripura	Bihar	Assam
Meghalaya																								D
Tripura																								
Bihar																								

"D" in cell (Row entry, Column entry) indicates that the Row entry has first-order dominance over the Column entry.

"d" in cell (Row entry, Column entry) indicates that the Row entry has first-order dominance over the Column entry except for one ordinate.

Table 6-3. [Health and income inequality for Indian states, 1993 and 1999]

	Income inequality (100*variance of logs)				Height inequality (100*variance of logs)			
State	**1993**	*Rank*	**1999**	*Rank*	**1993**	*Rank*	**1999**	*Rank*
Andhra Pradesh	27.48	*17*	29.62	*12*			0.37	*4*
Arunachal Pradesh (AR)	28.27	*22*			0.6	*17*	0.51	*15*
Assam (AS)	13.62	*4*	15.97	*1*	0.44	*8*	0.84	*25*
Bihar (BI)	19.02	*6*	19.21	*3*	0.67	*18*	0.72	*24*
Goa (GO)	28.08	*21*			0.4	*5*	0.26	*1*
Gujarat (GU)	22.63	*10*	24.59	*7*	0.5	*12*	0.52	*16*
Haryana (HA)	27.95	*20*	30.73	*14*	0.36	*2*	0.48	*12*
Himachal Pradesh	25.73	*14*	29.04	*11*			0.44	*8*
Jammu & Kashmir (JK)	20.37	*7*	17.69	*2*	0.43	*7*	0.52	*16*
Karnataka (KA)	27.67	*18*	32.04	*15*	0.4	*5*	0.48	*12*
Kerala (KE)	27.77	*19*	29.92	*13*	0.34	*1*	0.31	*2*
Madhya Pradesh	27.38	*16*	28.81	*10*			0.58	*21*
Maharashtra (MA)	41.38	*24*	41.71	*17*	0.48	*11*	0.36	*3*
Manipur (MN)	7.48	*1*			0.45	*9*	0.38	*5*
Meghalaya (ME)	17.72	*5*			1.05	*20*	0.62	*22*
Mizoram (MI)	12.20	*3*			0.55	*14*	0.47	*10*
Nagaland (NA)	9.82	*2*			0.59	*16*	0.49	*14*
New Delhi (DE)	42.03	*25*	46.00	*18*	0.45	*9*	0.56	*20*
Orissa (OR)	22.91	*12*	26.77	*9*	0.55	*14*	0.43	*7*
Punjab (PU)	22.84	*11*	24.38	*6*	0.38	*3*	0.47	*10*
Rajasthan (RA)	22.27	*9*	21.83	*4*	0.84	*19*	0.54	*18*
Tamil Nadu	34.43	*23*	32.69	*16*			0.45	*9*
Tripura (TR)	20.98	*8*			0.39	*4*	0.67	*23*
Uttar Pradesh (UP)	26.00	*15*	25.55	*8*	0.54	*13*	0.54	*18*
West Bengal	24.80	*13*	24.36	*5*			0.41	*6*

Table 6-4. [Changes in health inequality in Indian states]

	1993	**1999**	
State	**Gini***	**Gini***	**Dominance**
Assam	**0.37102**	**0.52171**	**D**
Bihar	**0.46492**	**0.48287**	
Goa	**0.34752**	**0.28002**	
Gujarat	0.39728	0.40518	
Haryana	**0.33230**	**0.38735**	**D**
Jammu	**0.36354**	**0.40221**	**D**
Karnataka	**0.35077**	**0.38847**	**D**
Kerala	**0.32490**	**0.30317**	
Maharashtra	**0.38974**	**0.33842**	**D**
Manipur	**0.37055**	**0.34108**	
Meghalaya	**0.49090**	**0.44253**	
Mizoram	0.40645	0.38589	
Nagaland	**0.42956**	**0.39239**	
Orissa	**0.41629**	**0.36702**	**D**
Punjab	**0.34259**	**0.37919**	
Rahasthan	**0.52057**	**0.41495**	**D**
Uttar Pradesh	0.41485	0.41224	
New Delhi	**0.37517**	**0.42045**	**D**
Arunachal	**0.42740**	**0.39128**	
Tripura	**0.34658**	**0.45719**	

* Values for Gini coefficients have been multiplied by 10.

Note: Rows in bold indicate that the Gini coefficients are statistically different in the two years.

Table 6-5. [Decomposition of changes in nutrition headcount index in India, 1993 to 1999]

	Nutrition Poverty Headcount Index			Ravallion-Datt			Kakwani	
Country (DHS years)	First Survey	Last Survey	Change	Change in Mean	Redistribution	Residual	Change in Mean	Redistribution
Arunachal Pradesh	49.398	26.999	22.399	18.072	4.638	-0.311	17.9169	4.4821
Assam	50.367	50.635	-0.268	-1.940	2.448	-0.776	-2.3280	2.0598
Bihar	55.641	53.866	1.774	2.262	-0.732	0.244	2.3843	-0.6100
Delhi	39.641	36.982	2.659	3.382	-0.659	-0.063	3.3501	-0.6911
Goa	30.000	18.090	11.910	7.102	5.061	-0.253	6.9757	4.9346
Gujarat	44.145	43.906	0.239	1.298	-0.691	-0.368	1.1136	-0.8748
Haryana	42.988	50.161	-7.173	-7.234	2.260	-2.199	-8.3335	1.1608
Jammu	38.205	38.965	-0.760	1.836	-3.691	1.095	2.3835	-3.1435
Karnataka	40.432	36.578	3.854	5.257	-1.990	0.588	5.5508	-1.6966
Kerala	25.194	22.073	3.121	1.594	1.280	0.247	1.7174	1.4033
Maharashtra	41.003	39.949	1.055	1.293	-0.502	0.264	1.4249	-0.3703
Manipur	24.620	31.437	-6.817	-6.047	0.478	1.248	-6.6708	-0.1458
Mizoram	36.704	34.742	1.962	6.019	-2.532	-1.525	5.2568	-3.2945
Nagaland	29.204	32.579	-3.375	-1.652	-0.721	-1.003	-2.1534	-1.2221
Orissa	45.237	44.028	1.209	1.256	0.117	-0.164	1.1738	0.0354
Punjab	38.152	39.232	-1.079	-1.707	0.334	0.294	-1.5602	0.4808
Rajasthan	42.959	52.132	-9.174	-12.563	1.665	1.724	-11.7008	2.5273
Tripura	41.429	40.498	0.931	4.353	-3.831	0.409	4.5573	-3.6268
Uttar Pradesh	54.387	55.642	-1.255	-0.505	-0.845	0.095	-0.4574	-0.7975

Table 6-6. [Decompositions of changes in nutrition poverty gap index in India, 1993 to 1999]

	Nutrition Poverty Headcount Index			Ravallion-Datt			Kakwani	
Country (DHS years)	First Survey	Last Survey	Change	Change in Mean	Redistribution	Residual	Change in Mean	Redistribution
Arunachal Pradesh	2.814	1.289	1.525	1.315	0.396	-0.186	1.2224	0.3026
Assam	2.527	3.845	-1.318	-0.218	-1.084	-0.016	-0.2263	-1.0921
Bihar	3.893	3.742	0.151	0.288	-0.135	-0.002	0.2869	-0.1361
Delhi	19.944	1.999	-0.055	0.243	-0.298	0.000	0.2437	-0.2983
Goa	1.284	0.563	0.721	0.263	0.368	0.090	0.3083	0.4126
Gujarat	2.434	2.425	0.009	0.057	-0.047	-0.001	0.0566	-0.0478
Haryana	1.753	2.799	-1.046	-0.640	-0.392	-0.014	-0.6467	-0.3991
Jammu	1.822	1.958	-0.136	0.133	-0.260	-0.009	0.1286	-0.2649
Karnataka	1.917	1.801	0.116	0.392	-0.268	-0.008	0.3881	-0.2720
Kerala	0.953	0.836	0.117	0.059	0.054	0.004	0.0610	0.0564
Maharashtra	2.068	1.606	0.462	0.110	0.354	-0.002	0.1092	0.3531
Manipur	1.152	1.325	-0.173	-0.311	0.135	0.003	-0.3093	0.1365
Mizoram	2.001	1.537	0.464	0.320	0.177	-0.032	0.3038	0.1605
Nagaland	1.506	1.282	0.224	-0.111	0.330	0.006	-0.1084	0.3329
Orissa	2.449	1.992	0.456	0.119	0.339	-0.001	0.1181	0.3382
Punjab	1.580	2.002	-0.422	-0.114	-0.305	-0.002	-0.1150	-0.3066
Rajasthan	2.876	3.196	-0.319	-1.004	0.744	-0.059	-1.0340	0.7148
Tripura	2.015	2.598	-0.583	0.233	-0.801	-0.015	0.2258	-0.8089
Uttar Pradesh	3.289	3.308	-0.019	-0.055	0.035	0.001	-0.0547	0.0356

Table 6-7. [Nutrition headcount and gap index adjusted for the Kerala health distribution]

State	Nutrition Headcount Index		Nutrition Gap Index	
	Actual	Adjusted for Kerala Distribution	Actual	Adjusted for Kerala Redistribution
Andhra	38.754	37.1329	1.730	1.565
Arunachal Pradesh	26.999	21.6246	1.289	0.802
Assam	50.635	43.4360	3.845	2.375
Bihar	53.866	46.9979	3.742	2.546
Delhi	36.982	30.3432	1.999	1.295
Goa	18.090	20.0465	0.563	0.741
Gujarat	43.906	37.9591	2.425	1.795
Haryana	50.161	46.7668	2.799	2.253
Himachal Pradesh	41.406	36.8148	2.007	1.635
Jammu	38.965	32.1356	1.958	1.374
Karnataka	36.578	30.7776	1.801	1.276
Maharashtra	39.949	35.5311	1.606	1.448
Manipur	41.003	35.4690	2.068	1.559
Meghalaya	45.225	37.8980	2.794	1.852
Mizoram	34.742	28.5935	1.537	1.131
Madhya Pradesh	51.225	45.9078	3.124	2.309
Nagaland	32.579	23.3955	1.282	0.916
Orissa	44.028	38.1375	1.992	1.648
Punjab	39.232	35.4197	2.002	1.496
Rajasthan	42.959	32.4092	2.876	1.513
Tamil Nadu	29.493	25.5410	1.373	0.992
Tripura	40.498	34.5072	2.598	1.573
Uttar Pradesh	55.642	51.4935	3.308	2.611
West Bengal	41.812	40.0159	2.108	1.776

NOTES

[i] Most obvious is the ability to achieve success in school and undertake productive work.

[ii] See, for example, the recent World Development Report on poverty (World Bank 2000).

[iii] See for example, Duclos, Sahn and Younger 2003; Atkinson 2003; and Bourguignon and Chakravarty 2003.

[iv] Atkinson and Bourguignon (1982) examine inequality across two dimensions, income and life expectancy.

[v] This dichotomy between the univariate and gradient approach is discussed in some detail in Wagstaff and van Doorslaer (2002) who focus on the empirical differences, as well as in the recent paper by Bommier and Stecklov 2002. Empirical approaches to measuring health inequality are developed by Le Grand 1987, Gakidou, Murray and Frenk 2000, and Pradhan, Sahn and Younger 2003.

[vi] This univariate approach to measuring non-income inequality has also been applied to education. See Thomas, Wang and Fan (2000) and Sahn and Stifel (2002).

[vii] This vast literature is both from developing and developed countries. See, for example, Wagstaff et al. 1991; Contoyannis and Forster 1999; Preston and Taubman 1994; van Doorslaer et al. 1997; Mackenbach and Kunst 1997; Navarro 1998; Hummer, Rogers and Eberstein 1998; Glied and Lleras-Muney 2003.

[viii] Wolfson and Rowe (2001) and Wagstaff and van Doorslaer (2002) use the term univariate to distinguish this approach from one that examines health inequality based on correlations with income or other indicators of social stratification.

[ix] Erik Thorbecke pointed out to me that someone from the Chicago school may argue just the opposite – that saving the life of someone in the upper deck would result in more investment, earnings, and greater economic growth (and, thereby, may even help the poor in the long term).

[x] See Yitzhaki (1983) for a discussion of the extended Gini coefficient.

[xi] It is not unusual that findings regarding dominance are not based on statistical tests of differences in concentration curves. See for example, Jenkins and Lambert (1993).

[xii] Howes (1996) shows that I can only be sure that the probability of type I error is no more than the critical value if I reject the null hypothesis in the case that the difference in the ordinates of the two curves is non-zero for *every* ordinate tested and, obviously, that the difference be of the same sign. This decision rule is clearly less likely than the more common one to reject the null in favor of dominance. In practice, I have found elsewhere (Sahn and Younger 2000) that it leads us to accept the null quite often, limiting what we conclude about the relative ordering of concentration or Lorenz curves. However, bounding the size of the test at the risk of low power is consistent with standard econometric practice, and we follow it here. Of course, failure to reject the null leaves us with an indeterminate result, unless I can establish that the two Lorenz curves cross, something shown by two significant differences in ordinates of opposite signs.

[xiii] The FGT measure is typically defined as,

$$P_\alpha = \tfrac{1}{N}\sum_{i=1}^{N}\left(\frac{z-y_i}{z}\right)^{\alpha} I(y_i \le z),$$

where the individual's poverty gap is expressed as a proportion of the poverty line. This creates a unit free measure that is comparable across populations. The measure I present in the text does not follow this convention because (a) z-scores are already standardized across populations, and (b) the absolute gap (i.e., $z - y_i$) has a meaningful interpretation – it is the number of standard deviations that a child's z-score falls below the poverty line.

[xiv] In the results section, I do not present the severity index because the magnitude of changes is so small that the decomposition is of little interest.

[xv] I choose to present the differences, rather than the Lorenz curves themselves, since visually, it is easier to depict the differences among states with this approach.

[xvi] An anonymous referee pointed out that this negative correlation would also be found in the case of income inequality on the downward sloping part of the Kuznet's curve. In the income case, however, the negative correlation is attributable to factors such as redistributive policies in better off societies, while the in case of my health inequality

measure, I would expect this negative correlation to be across the spectrum of levels of health due to the innate properties of the measure discussed above.

REFERENCES

Atkinson, Anthony. 2003. Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches. *Journal of Economic Inequality.* 1:1, pp. 51-65.

Atkinson, Anthony and François Bourguignon. 1982. "The Comparison of Multi-Dimensional Distributions of Economic Status" in *Social Justice and Public Policy.* London: Harvester Wheatsheaf, Chapter 2.

Balisacan, Arsenio. 2000. Growth, Redistribution and Poverty: Is the Philippines an exception to the standard Asian story? *Journal of the Asian Pacific Economy.* 5:1/2, pp. 125-140.

Beaton, George, A. Kelly, John Kevany, Reynaldo Martorell, and J. Mason. 1990. Appropriate Uses of Anthropometric Indices in Children: A Report Based on an ACC/SCN Workshop. United Nations Administrative Committee on Coordination/Subcommittee on Nutrition ACC/SCN State-of-the-Art Series, Nutrition Policy Discussion Paper No. 7, New York.

Behrman, Jere and Anil Deolalikar. 1988. "Health and Nutrition" in *Handbook of Development Economics, Vol. 1.* Hollis Chenery and T.N. Srinivasan, eds. Amsterdam: North-Holland Press, pp. 631-711.

Bigsten, Arne, Bereket Kebede, Abebe Shimeles, and Mekonen Taddesse. 2003. Growth and Poverty Reduction in Ethiopia: Evidence from Household Panel Surveys. *World Development.* 31:1, pp. 87-106.

Bommier, Antoine and Guy Stecklov. 2002. Defining Health Inequality: Why Rawls Succeeds Where Social Welfare Theory Fails. *Journal of Health Economics.* 21:3, pp. 497-513.

Bourguignon, François and Satya Chakravarty. 2003. The Measurement of Multidimensional Poverty. *Journal of Economic Inequality.* 1:1, pp. 25-49.

Bustos, Patricia, Hugo Amigo Sergio Muñoz. and Reynaldo Martorell. 2001. Growth in Indigenous and Nonindigenous Chilean Schoolchildren from 3 Poverty Strata. *American Journal of Public Health.* 91:10, pp. 1645-1649.

Chen, Shaohua and Martin Ravallion. 2001. How Did the World's Poorest Fare in the 1990s. *Review of Income and Wealth.* 47:3, pp. 283-300.

Christiaensen, Luc, Lionel Demery, and Stefano Paternostro. 2002. Growth, Distribution and Poverty in Africa. Messages from the 1990s. Washington, D.C.: World Bank Working Paper 2810.

Cole, T. J. and J. M. Parkin. 1977. Infection and its Effect on Growth of Young Children: A Comparison of the Gambia and Uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* 71, pp. 196-198.

Contoyannis, Paul. and Martin Forster. 1999. The Distribution of Health and Income: A Theoretical Framework. *Journal of Health Economics.* 18, pp. 605-622.

Datt, Gaurav and Martin Ravallion. 1992. Growth and Redistribution Components of Changes in Poverty Measures: A Decomposition with Applications to Brazil and India in the 1980s. *Journal of Development Economics.* 38:2, pp. 275-295.

Davidson, Russell, and Jean-Yves Duclos. 2000. Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica.* 68, pp. 435-1465.

Deaton, Angus. 2003. Health, Inequality, and Economic Development. *Journal of Economic Literature.* 41:1, pp. 113-158.

Deaton, Angus, and Jean Dreze. 2002. Poverty and Inequality in India: A Re-Examination. *Economic and Political Weekly.* September 7, 2002, pp. 3729-3748.

Dhongde, Shatakshee. 2002. Measuring the Impact of Growth and Income Distribution on Poverty in India, mimeo, Department of Economics, University of California, Riverside.

Duclos, Jean-Yves, David E. Sahn, and Stephen. D. Younger. 2003. Robust Multidimensional Poverty Comparisons, Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi (CIRPÉE) Cahier de recherche/Working Paper 03-04, Québec, Canada: Université Laval.

Gakidou, Emmanuela, Christopher Murray, and Julio Frenk. 2000. Defining and Measuring Health Inequality: An Approach Based on the Distribution of Health Expectancy [Review]. *Bulletin of the World Health Organization.* 78:1, pp. 42-54.

Glied, Sherry and Adriana Lleras-Muney. 2003. Health Inequality, Education and Medical Innovation. NBER Working Paper No. w9738 (June), National Bureau of Economic Research, Cambridge, MA.

Habicht, J.-P., R. Martorell, C. Yarbrough, R.M. Malina, and R.E. Klein. 1974. Height and Weight Standards for Preschool Children. How Relevant Are Ethnic Differences in Growth Potential? *Lancet.* I, pp. 611-614.

Haughton, Jonathan. 1997. Falling Fertility in Vietnam. *Population Studies.* 51:2, pp. 203-211.

Hill, Ann and Masuma Mamdani. 1989. Operational Guidelines for Measuring Health through Household Surveys, mimeo. Centre for Population Studies, School of Hygiene and Tropical Medicine. University of London.

Howes, Stephen. 1996. A New Test for Inferring Dominance from Sample Data, draft.

Hummer, Robert A., Richard G. Rogers, and Isaac. W. Eberstein. 1998. Sociodemographic Differentials in Adult Mortality: A Review of Analytic Approaches. *Population and Development Review.* 24:3, pp. 553-578.

Jenkins, Stephen. P. and Peter J. Lambert. 1993. Ranking Income Distributions When Needs Differ. *Review of Income and Wealth.* 39:4, pp. 337-356.

Kakwani, Nanak. 1997. On Measuring Growth and Inequality Components of Changes in Poverty with Application to Thailand. mimeo, School of Economics, The University of New South Wales, Sydney.

Kroeger, Axel. 1985. Response Errors and Other Problems of Health Interview Surveys in Developing Countries. *World Health Statistics Quarterly.* 38:1, pp. 15-37.

Le Grand, Julian. 1987. Inequalities in Health: Some International Comparisons. *European Economic Review.* 31, pp. 182-191.

Mackenbach, Johan. P. and Anton. E. Kunst. 1997. Measuring the Magnitude of Socioeconomic Inequalities in Health: An Overview of Available Measures Illustrated with Two Examples from Europe. *Social Science and Medicine.* 44, pp. 757-71.

Martorell, R., Habicht, J.-P., Yarbrough, C., Lechtig, A., Klein, R. E. and K. A. Western. 1975. Acute Morbidity and Physical Growth in Rural Guatemalan Children. *American Journal of Diseases in Childhood.* 129, pp. 1296-1301.

Mata, Leonardo. 1978. *The Children of Santa Maria Cauque: A Prospective Field Study of Health and Growth.* Cambridge, MA: MIT Press.

McCulloch, Neil, Milasoa Cherel-Robson, and Bob Baluch. 2000. Growth, Inequality and Poverty in Mauritania 1987-96, mimeo, Institute of Development Studies, University of Sussex, Brighton.

Mosley, W. Henry and Lincoln C. Chen. 1984. An Analytical Framework for the Study of Child Survival in Developing Countries. *Population and Development Review.* 10:0, pp. 25-45.

Navarro, Vicente. 1998. A Historical Review (1965-1997) of Studies on Class, Health, and Quality of Life: A Personal Account. *International Journal of Health Services.* 28:3, pp. 389-406.

Over, Mead, Randall P. Ellis, Joyce H. Huber, and Orville Solon. 1992. "The Consequences of Adult Ill-health" in *The Health of Adults in the Developing World.* R.G. Feachem, T. Kjellstrom, C. J. L. Murray, M. Over, and M. A. Phillips, eds. Oxford University Press for World Bank, New York, Toronto, and Melbourne, pp. 161-207.

Pitt, Mark. M. and Mark R. Rosenzweig. 1986. "Agricultural Prices, Food Consumption and the Health and Productivity of Farmers" in *Agricultural and Household Models: Extensions, Applications and Policy.* I. J. Singh, L. Squire, and J. Strauss eds., World Bank, Washington.

Pradhan, Menno, David E. Sahn, and Stephen D. Younger. 2003. Decomposing World Health Inequality. *Journal of Health Economics.* 22:2, pp. 271-293.

Preston, Samuel H. and Paul Taubman. 1994. "Socioeconomic Differences in Adult Mortality and Health Status" in *Demography of Aging.* L. G. Martin and S. H. Preston, eds. Washington, DC: National Academy Press, pp. 279-318.

Ravallion, Martin, 1995. Growth and Poverty: Evidence for Developing Countries in the 1980s. *Economics Letters.* 48, pp. 411-417.

Sahn, David E. and Harold Alderman. 1997. On the Determinants of Nutrition in Mozambique: The Importance of Age-Specific Effects. *World Development.* 25:4, pp. 577-588.

Sahn, David E. and David C. Stifel. 2002. Parental Preferences for Nutrition of Boys and Girls: Evidence from Africa. *Journal of Development Studies.* 39:1, pp. 21-45.

Sahn, David E. and Stephen D. Younger, 2000. Expenditure Incidence in Africa: Microeconomic Evidence. *Fiscal Studies.* 21:3, pp. 329-347.

Sahn, David E. and Stephen D. Younger. Forthcoming, 2005. Improvements in Children's Health: Does Inequality Matter? *Journal of Economic Inequality.*

Schultz, T. Paul and Aysit Tansel. 1997. Wage and Labor Supply Effects of Illness in Côte D'Ivoire and Ghana: Instrumental Variable Estimates for Days Disabled. *Journal of Development Economics.* 53:2, pp. 251-286.

Sen, Amartya. 1985. *Commodities and Capabilities.* North Holland, Amsterdam.

Sen, Amartya. 1987. "The Standard of Living: Lecture II, Lives and Capabilities" in *The Standard of Living.* G. Hawthorn, ed. Cambridge University Press, Cambridge, pp. 20-38.

Shorrocks, Anthony and Stanislav Kolenikov. 2001. Poverty Trends in Russia during the Transition, mimeo, World Institute of Development Research, Helsinki and University of North Carolina.

Strauss, John and Duncan Thomas. 1995. "Empirical Modeling of Household and Family Decisions" in *Handbook of Development Economics*, Vol. IIIA. J. Behrman, and T. N. Srinivasan, eds. North-Holland, Amsterdam, pp. 1883-2023.

Tanner, James M. 1981. *A History of the Study of Human Growth.* New York: Cambridge University Press.

Thomas, Duncan, Lavy, Victor, and John Strauss. 1996. Public Policy and Anthropometric Outcomes in the Côte D'Ivoire. *Journal of Public Economics.* 61:2, pp. 155-192.

Thomas, Vinod, Yan Wang, and Xibo Fan. 2000. Measuring Education Inequality: Gini Coefficients of Education. Education Policy Research Working Paper #2525, World Bank, Washington, DC.

Van Doorslaer, E., A. Wagstaff, H. Bleichrodt, S. Calonge, Ulf-G. Gerdtham, Michael Gerfin, Jose Geurts, Lorna Gross, Unto Häkkinen, R. Leu, O. O'Donnell, C. Propper, F. Puffer, M. Rodriguez, G. Sundberg, and O. Winkelhake, 1997. Income-related Inequalities in Health: Some International Comparisons. *Journal of Health Economics.* 16:1, pp. 93-112.

Wagstaff, Adam and Eddy van Doorslaer. 2002. Overall vs. Socioeconomic Health Inequality: A Measurement Framework and Two Empirical Illustrations, unpublished manuscript.

Wagstaff, Adam, Pierella Paci, and Eddy van Doorslaer. 1991. On the Measurement of Inequalities in Health. *Social Science and Medicine.* 33, pp. 545-557.

WHO (World Health Organization), 1995. *Physical Status: The Use and Interpretation of Anthropometry*: Report of WHO Expert Committee. WHO, Geneva.

WHO (World Health Organization). 1983. *Measuring Change in Nutritional Status: Guidelines for Assessing the Nutritional Impact of Supplementary Feeding Programmes for Vulnerable Groups.* WHO, Geneva.

Wolfson, Michael and Geoff Rowe. 2001. On Measuring Inequalities in Health. *Bulletin of the World Health Organization.* 79:6, pp. 553-560.

World Bank. 2000. World Development Report 2000/2001: *Attacking Poverty.* London: Oxford University Press.

Yitzhaki, Shlomo. 1983. On an Extension of the Gini Inequality Index. *International Economic Review.* 243, pp. 617-628.

Chapter 7

HOUSEHOLD INVESTMENTS IN EDUCATION AND INCOME INEQUALITY AT THE COMMUNITY LEVEL: EVIDENCE FROM INDONESIA

Jeffrey B. Nugent[1]and Shailender Swaminathan[2]
[1]*University of Southern California, nugent@usc.edu and* [2]*University of Alabama Birmingham, SSwaminathan@ms.soph.uab.edu*

1. INTRODUCTION

This paper is very much inspired by Erik Thorbecke's work. Two of Erik Thorbecke's many areas of important contributions to economics have been to the measurement, determinants and effects of income inequality and the analysis of the Indonesian economy. Indeed, a not insignificant part of his work on each of these subjects has been at the intersection of these two fields of scholarship. Among such contributions are: (1) Khan and Thorbecke (1988) which applied structural path analysis and SAM tables to investigate the effects of changes in sectoral demand patterns, technology and energy sources on Indonesian employment and income distribution; (2) Thorbecke (1988) which examined the effects of various possible policy reforms in Indonesia on agricultural efficiency and distribution; (3) Thorbecke (1992) which applied various kinds of SAM and CGE models to examine the distributional effects of Indonesia's structural adjustment and Indonesia's budgetary retrenchment in the 1980's, and (4) Thorbecke and van der Pluijm (1993) which focused more on the effects of policy interventions in agriculture and rural institutions on poverty and income distribution.

In most of these analyses, the effects of income distribution were also captured through their influence on household expenditure and labor supply patterns. Several of these works credited devaluation and increased expenditures on health and education during the 1980's as important means of reducing poverty and income inequality in Indonesia in spite of the structural adjustment and stabilization policies that elsewhere have often had less positive effects.

Although not focused on Indonesia, Thorbecke and Charumilind (2002) provide a very comprehensive overview of the theoretical arguments from several disciplines on the direction and magnitude of the effects of income inequality on savings, investment and growth. These authors also assess the existing empirical evidence from around the world, especially that pertaining to investments in health and education that they feel deserve more attention from economists. Much of this empirical evidence is of the international cross-section type that fails to distinguish sharply between the effects of unobserved cross-country differences and those of income inequality. Therefore, in this study we focus on the effects of variations in income inequality and household investments in education across local communities within a single country where history and national policies are held constant and the effects of numerous other community and household characteristics can be controlled for.

In particular, in exploring Thorbecke's theme of the effects of income inequality on household investments in education, we make use of micro-level data for Indonesia from the Indonesian Family Life Survey (IFLS). Indonesia is deemed appropriate as a case study for several reasons. First, it is a large country characterized by considerable income inequality. Second, it is a country where educational investments are still relatively low. Third, there is considerable variability in both inequality and educational investments across households and communities. Fourth it allows us to take advantage an excellent micro dataset, the Indonesian Family Life Survey (IFLS). Despite making the second greatest increase in the world since 1975, with a score of 0.684 in 2001, Indonesia remains ranked 112 on the UNDP's Human Development Index (HDI), unusually low for an oil exporting country.[i] Moreover, it was ranked quite a bit lower at the times the data used in this study were collected (1993 and 1997).

Such an analysis is also relevant to the ongoing debate over the net direction of the effects of income inequality on growth. Theoretically, these effects can be either positive or negative and can operate through various transmission mechanisms. We focus on effects operating through the quality and quantity of public goods and labor markets. In this way, our results can be of help in identifying useful policy interventions.

Section 2 surveys some of the relevant literature on the effects of inequality on education with special reference to the transmission mechanisms. Section 3 identifies the IFLS data used in the study and its use in getting at inequality effects on investments in education. Section 4 contains our empirical analysis and Section 5 concludes.

2. INEQUALITY EFFECTS ON HUMAN CAPITAL INVESTMENTS

As already mentioned, the overall relationship between income inequality and growth can be either positive or negative and most such effects operate through investments of various sorts. For example, in the absence of perfect capital and insurance markets but the presence of adverse selection and moral hazard, negative effects on investment may arise from binding financial constraints (Birdsall, Pinckney and Sabot 1999). But under certain conditions, such as non-linearity in savings and investment functions, and people with higher incomes being less risk averse than those with lower incomes, savings and investment rates can also be higher where incomes are more unequally distributed (Kuznets 1955). But then again, under certain circumstances, inequality can reduce growth by reducing investment. Yet, to the extent that the inequality arises from artificial rents, inequality can also produce growth-retarding, rent-seeking behavior (Acemoglu 1995). Inequality can also undermine trust and thereby social capital and can stimulate social tensions, social conflict, and political and policy instability, all of which are likely to be negatively related to property rights, investment and growth.[ii] In democratic contexts, moreover, income inequality may increase the demand for inequality and growth-reducing transfers financed by distortionary taxes.[iii]

Virtually all of these general arguments can apply to investments in human capital. As Thorbecke and Charumilind (2002) have pointed out, economists have tended to pay much less attention to the effects of inequality on these forms of investment. Indeed, the few exceptions such as the recent and very comprehensive survey by Deaton (2003) in the case of health have suggested that the apparent relation between income inequality and such investment disappears once non-linearities in income and perhaps other controls are introduced. Yet, most of these studies have been confined to developed countries where inequality is relatively low and capital and insurance markets well developed. In developing countries, how might these effects operate?

At very low levels of income where very few have any opportunity to invest in human capital, high inequality could make for larger such investments than might be expected on the basis of average income alone. But on the other hand, greater inequality would imply larger percentages of poor people who, because of credit constraints, would be unable to make private investments in human capital. This can have especially deleterious effects on growth because in poor countries the private and social rates of return to human capital are likely to be higher than in developed countries (Psacharopoulos 1985; Psacharopoulos and Patrinos 2002).

Naturally, the way in which education or health is financed can make a big difference and as a result can introduce political economy effects (Benabou 1996a, 1996b). These, in turn, may differ between democratic and non-democratic settings. In the former, the median voter theorem might suggest that inequality would imply that the median voter's income is well below the mean, quite possibly inducing a shift from private toward public provision of education and health services, the effect of which could be to increase human capital formation. In a non-democratic setting, however, the educated and healthy rich would likely have much stronger access to power and thus prefer to avoid high taxation by limiting the financing of human capital to private sources or alternatively focusing publicly supplied human capital on expensive items (like higher education) that would primarily benefit the rich.

At the same time, because of the greater role of public provision of schools and infrastructure, at least as complements to private investments, the local provision of public goods can exert very important influences over household investments in human capital. As the theory of collective action and local public goods provision has suggested, many different group characteristics can influence the strength of collective action and hence the extent to which a given community or group is likely to be successful in producing human capital-inducing public goods. The effects of income inequality in this context are somewhat controversial.[iv] Yet, a majority of analysts would probably expect income inequality, like ethnic diversity, to weaken trust, altruism and willingness to cooperate, thereby lowering the production of local public goods (either quantitatively or qualitatively), and thereby raising the private cost (especially to the poor) of investments in human capital and lowering investments in human capital (Alesina, Baqir and Easterly 1999; Alesina and LaFerrara 2000).[v]

Among non-economists such as Wilkinson (1996, 2000) but also relative income and relative deprivation theory advocates among economists, there is the notion that inequality can trigger psychological feelings of resentment and despair, that can lead to lower investments or even disinvestments in human capital (especially health). Being subjective in nature, such effects, however, are very difficult to test and in any case may apply to people who are older than most candidates for investments in education.

Largely overlooked, however, would seem to be the effects of income inequality on the character of the labor and capital markets. For example, income inequality could interact with other characteristics of these markets, such as the availability of credit for such investments and entrepreneurial start-ups, and wage rates for higher levels of human capital, thereby affecting the rate of return on human capital investments, be it on education or health.

One such mechanism that would apply to communities with relatively closed labor markets arises from the fact that greater income inequality would imply fewer employers of human capital in the community. As a result, these fewer employers might be able to exercise more monopsony power in the market for better educated or healthier labor than they would in a community in which employers would be greater in number and less concentrated. The wage premium for higher skilled and educated workers would be lower, thereby lowering the rate of return on such labor.

Another means of obtaining a similar effect on the rate of return and hence willingness to invest in human capital could be the poorer quality of publicly supplied educational services. The source of such differences presumably would lie in the local public goods provision and hence the same considerations mentioned above.

For rural residents, the returns to education are often realized when the individual makes a transition from farming or agricultural labor (where the returns to education are low) into entrepreneurial and self-employment activities (where the returns are higher). For any given level of community income, greater income inequality may imply greater difficulty in finding financial support for entrepreneurial startups and hence in the ability to make these transitions. The result, once again, may be lower returns to education.

Another connection between income inequality and education investments derives from the trade-off between numbers of children and education. Recently, de la Croix and Doepke (2003) have argued that where poor families decide to have more children but invest less in educating them than rich families, a mean-preserving increase in inequality implies that more weight would be given to those members of the community who provide little education to their children.

In the empirical literature on education, income distributional effects have tended to receive little attention. But, empirical studies on the closely related effects of inequality on health have been expanding rapidly, especially so for infant and child mortality. The vast majority of empirical studies on this subject have shown a simple positive relationship between income inequality and infant mortality. There is considerable controversy, however, as to how robust this relationship is to the addition of various controls for other factors influencing infant mortality. While some of the earlier studies[vi] showed the results to be quite robust in a number of respects, increasingly others[vii] have shown that the inclusion of controls especially for household income (in a non-linear way) and parents' education tends to eliminate the income inequality effect, at least in the United States. Yet, since some of these controls could themselves be affected by income distribution, even the disappearance of a direct effect with the introduction of controls does not necessarily indicate the absence of an effect. Indeed, as Deaton (2002, 2003) has pointed out, the controls could be interpreted as channels through which the effects of income inequality are realized. But, just as there is little consensus on the robustness of the results to different specifications, there is equally little agreement on the channels or links through which the inequality effects are realized.

Virtually all such studies are based on cross-section data, either across countries or across broad regions or states within a country. As such they are subject to numerous possible criticisms. Among these are (1) the high level of aggregation (that may make it impossible to distinguish the effect of inequality per se from that of non-linearity in income), (2) the omitted variable bias that may bias the results either positively or negatively, (3) the use of poor and incomparable measures of income inequality, sometimes measured from income data and sometimes from expenditure data that introduce measurement error bias, and (4) potential endogeneity of the controls (such as in the case of fertility and female labor force participation).

As suggested above and explained further below, the specific links between income inequality and education have been only very partially explored and the few conclusions drawn to date remain controversial at best. Therefore, the paper's main objectives are to determine in the Indonesian context whether or not income inequality matters for investments in education and, if so, to identify the main channel(s) through which any such effects are realized.

3. THE INDONESIAN SETTING AND THE IFLS DATA

Our aforementioned decision to focus on Indonesia would only be practical if available data allowed us to compare educational investments at the household level across local communities characterized by varying degrees of income inequality. A very important requirement for this is that income inequality be measured in a consistent way across communities. At the same time, we would also need to be able to control for the various individual, household and community characteristics that have been alleged to give rise to an apparent relationship between income inequality and education. Yet, thanks to Indonesian Family Life Survey (IFLS), all these requirements can be satisfied.

In particular, the IFLS data has been collected for 7200 households selected from some 321 different communities (districts) and 13 provinces in two rounds, 1993-94 (IFLS1) and 1997-98 (IFLS-2). The 13 provinces account for 85 percent of Indonesia's population of 201.2 million in 1997. The almost 95 percent success rate of the interviewers in the second round of surveys interviewers in completing interviews with the households interviewed in the first round is exceptionally high by international standards, thereby greatly increasing the usefulness of combining information from the two surveys.

Several features of the data are especially useful for our analysis. First, the two waves of data allow us to examine the effect of "initial" inequality (for 1991) and other determinants such as the distribution of public goods, household and community income and parent characteristics on investments in education between 1993 and 1997. In particular, we use data from IFLS-2 to generate our education investments and use data from IFLS-1 to generate measures of inequality and the other covariates in the educational investment equations.

Second, the information on educational investments is quite detailed. Specifically, there is information both from the household as a whole and for randomly selected children with sample households on the investments in education during the survey year and on each of several specific components like books, tuition, and uniforms.

Third, from the community survey component of IFLS-2, information is reported by community leaders on the existence, distance to, quality and cost of education and also on labor market conditions in each of the sample communities.

Fourth, since the samples in each of the 321 communities (districts or "kecamatans") consisted of between 20 and 30 households, the information on household income for each of the 7200 sampled households was aggregated to the community level, averaged and used to generate various income inequality indexes by community. Because of their lesser sensitivity to outliers (which quite possibly could have been the result of measurement error), we focus on the community-level Gini coefficients as our measure of income inequality.[viii] Since there were a few cases of very large reported incomes in the sample, to see how much the results could be affected by outliers in income, we have also recomputed the Gini and other inequality indexes once the outliers were eliminated. The results were not substantially affected by the elimination of outliers and here we report here only the results excluding the outliers. In addition to data on income, there is also complete data on household food expenditures (though, due to the omission of a catchall residual category in the 1993 expenditures data, not for total expenditures).

Fifth, IFLS-1 also provides information on other individual, household and community level data that may be used to construct relevant control variables, some of which may also serve as links between income inequality and investment.

The samples consisted of all children attending school in sample households for whom data on educational expenses in the 1996/1997 school year was available (5340). For computing the rate of return on education, we used a sample of 20,430 worker-year observations reported from sample households. Descriptive statistics on relevant variables are given in Table 7-1 below. For household and individual variables used in multiple equations, the means were not significantly different across the samples and hence we report here only one set of summary statistics for those variables (using the education sample).

4. METHODS AND RESULTS

In this section we consider the association between income inequality and both education and earnings. We develop the empirical equations and discuss their relationship to each other and to the issues and hypotheses under investigation.

A. Investments in Education

Education is in part an investment that reaps returns throughout life, including higher earnings, better marriage prospects, and various other benefits. We base our model on the two-step Beckerian model of parental altruism with perfect capital markets (Becker, 1991, Becker and Tomes, 1976, 1979). In this model, to maximize the joint wealth of the whole family parents first make investments in each child's human capital up to the point at which the marginal return on each child's human capital is equal to the interest rate. Next, they redistribute the maximized joint wealth across family members in a way that maximizes the utility of the household head. In this model, parental investment varies only with the returns to education. Parental investment in education, therefore, may vary for reasons limited to the following:

- Variations across children due, e.g., to differences in endowments or abilities.
- Variations across households due to inter-household differences in ability or environmental factors, such as parental education and income, that are common to all children in that household.
- Variations across communities or societies in labor market conditions.

Table 7-1. [Descriptive Statistics and Model Specifications]

	Mean (std. Dev. in parenthesis)	Education	Earnings	Income Gini-Infra-structure Correlation	Food Exp. Gini-Infra-structure Correlation
Community Level Variables					
Gini in 1993 (Income)	0.46 (0.10)	X	X		0.1028*
Gini in 1993 (Food Expenditure)	0.33(0.07)	X	X	0.1028*	
# of schools in community in 1993	7.76 (2.39)	X		-0.07	-0.05
Ratio of public/private schools	0.70 (0.19)	X		0.10*	-0.09
Average ebtanas score				0.04	0.14*
Factory Exist	0.28		X	-0.12*	-0.02
# of Factories	0.62(1.62)		X	-0.2*	-0.07*
Cottage industry exists	0.72		X	-0.03	0.08
# cottage industries	1.71(1.51)		X	-0.02	0.13*
Average income in community (in 1000's of rupiah)	145 (106)	X		-0.08	
Household Level Variables					
Income (in 100,000 rupiah)	0.09 (0.79)	X			
Mothers Education	5.47 (3.73)	X			-0.09*
Fathers' Education	6.46 (3.84)	X			-0.16*
Urban	0.44	X	X		
# of children	3.29 (1.62)	X			-0.06*
Individual Level Variables					
Child Age	10.14(2.54)	X			
Male Child	0.50	X			
Age at Earnings (in years)	40.8 (11.3)		X		

Table 7-1 [Cont.]

	Mean (std. Dev. in parenthesis)	Education	Earnings	Income Gini-Infrastructure Correlation	Food Exp. Gini-Infrastructure Correlation
Education	5.72 (4.48)		X		
Male Adult	0.66		X		

Notes: 3000 rupiah to a dollar in 1993

* indicates correlation is significant at the 10 % level

- Variations across communities or regions in the costs or availability of schooling.

While our model does not include variables that capture child-specific differences in endowments, it does include household and community differences of the types identified. While the IFLS data provides direct measures of the other inter-household and inter-community variations, inter-community differences in current and expected future labor market conditions have to be estimated from earnings data as described below.

As indicated above, two different measures of investments in education are used, the total annual value of investments in education for a randomly selected youth in the family and also the total value of investments in books alone. In both cases, we use the following model:

$$\log(Educ\_\exp)_{imc} = \beta_1 Inq_{c,93} + \beta_2 X_i + \beta_3 HHINC_m + \beta_4 HHINC_m{}^2 + \beta_5 SchoolInfra_c + u_{imc} \quad (1)$$

where Educ_exp is the measure of educational investments, $Inq_{c,93}$ is the Gini coefficient in community c in the year 1993, X is a set of characteristics of the mother, father and household including her age and education and urban location, community income, $HHINC$ is household income and its square is $HHINC^2$ and *SchoolInfra* is a vector of educational infrastructure variables including the number of schools in the community, the ratio of public to private schools and average ebtanas test scores (a measure of school quality in the community) and u_{imc} is a random error term. Since there is clustering of observations within a community, the reported standard errors in the tables are robust to within-community clustering of observations (16).

B. Earnings

As pointed out earlier, inequality may be expected to alter earnings by affecting labor market conditions. We explicitly test the hypothesis that inequality may lower earnings and/or the returns to education by estimating a standard earnings equation but now allowing inequality to have a direct and indirect (via education) effect on earnings. We use two different measures of inequality, i.e., both inequality in lagged household income of a particular year and food expenditure inequality which, because of the household's desire to smooth such consumption, may be considered a better measure of permanent income than measured annual income. The hypothesis is that there is less of an incentive to spend a rupiah on education in those Indonesian communities with high income inequality than ones with low income inequality. This can happen due to either (1) that inequality causes labor markets to clear at lower levels of wage rates or (2) that inequality affects the returns to education. According to (1) the effect is independent of the level of education, those in higher inequality communities earn less, whereas according to (2), the effect would be stronger at higher levels of education.

Data on individual income is not available for the year 1997, and hence a similar method as the one for education expenditure cannot be used. Instead, we use two variants of the original model. First, find the effect of community inequality in 1993 on the logarithm of earnings in 1993. Second, we use data on community inequality in 1991 on the logarithm of earnings for the years 1991, 1992, and 1993. The results are rather similar for the two approaches. Yet, since the earnings data for the latter is more comprehensive, only the results for the second approach are presented here.

Specifically, we try to explain variations in the log of average monthly earnings equation in the years 1991, 1992 and 1993 on the basis of the following equation:

$$\log(earnings)_{it} = \gamma_1 Inq_{c,91} + \gamma_2 Educ + \gamma_3 Age + \gamma_4 Educ * Inq_{c,91} + v_{it} \qquad (2)$$

This equation allows for inequality to have both a direct and indirect effect on earnings.

C. Results

Descriptive statistics on the variables used in the analysis are given in Table 7-1. Several points should be noted. First, the mean value of the Gini is 0.46 that is slightly higher than the mean value reported in official statistics. Second, there is variation in education and labor market infrastructure across communities. Further, in general the higher the income inequality within a community, the weaker is the infrastructure within the community. In addition, higher the income inequality, lower is the education of parents, and the number of kids in the household. The average age of while the average age of the child in school is 10 and the average age of individuals in the earnings equation is 41.

The empirical results for investments in education are given in Table 7-2 and those for earnings in Tables 7-3a and 7-3b. Since the measures of the dependent variables are in both cases fully continuous with no zero observations, equations (1) and (2) are estimated by OLS but with standard errors corrected for clustering.

Table 7-2 presents the results for educational investments. The results reported in the first two columns are those for equation (1) with the dependent variable defined as the logarithm of the total annual expenditure on education for the randomly selected child. The only difference between these two columns is that the Gini coefficient used in the first column is with respect to measured income in 1993 whereas that used in the second column is (for experimental purposes) based on food expenditures alone. The results in the third and fourth columns use identical specifications to those in the

Table 7-2. [Determinants of Expenditure on Education in the School Year 1996-97 (p values in parenthesis)]

Name of Explanatory Variable	Log (Total Education Expense on child) †	Log (Total Education Expense on child) ††	Log (Expense on Books on child) †	Log (Expense on Books on child) ††
Gini in 1993	-0.44 (0.030)	0.01 (0.976)	-0.55(0.011)	0.06(0.869)
Income (in 100,000 rupiah)	0.14 (0.028)	0.13(0.024)	0.10(0.252)	0.10(0.250)
Income square	-0.009(0.121)	-0.009(0.111)	-0.006(0.435)	-0.006(0.435)
Average Income in Community (in 100,000 rupiah)	0.0006(0.009)	0.0005(0.015)	0.0006(0.01)	0.0005(0.022)
Fathers' Education	0.030(0.000)	0.03(0.000)	0.035(0.000)	0.035(0.000)
Mothers' Education	0.036(0.000)	0.036(0.000)	0.043(0.000)	0.044(0.000)
Urban	0.47(0.000)	0.51(0.000)	0.45(0.000)	0.50(0.000)
Child Age	0.12(0.000)	0.12(0.000)	0.09(0.000)	0.09(0.000)
Number of children 0-19 in household	-0.05(0.000)	-0.05(0.000)	-0.04(0.000)	-0.05(0.000)
Male Child	-0.05(0.025)	-0.05(0.025)	-0.07(0.000)	-0.07(0.000)
# schools in community	0.01(0.130)	0.02(0.081)	0.02(0.104)	0.02(0.048)
# of public schools/# private schools	-0.43(0.001)	-0.44(0.000)	-0.17(0.164)	-0.18(0.136)
Average ebtanas test score within community	0.05(0.144)	0.04(0.266)	0.02(0.467)	0.014(0.730)
R2	0.41	0.41	0.31	0.31
N	5157	5157	5157	5157

†Using income inequality

†† Using food expenditure inequality

Table 7-3a. [Determinants of 1991-1993 Log Monthly Earnings (p value in parenthesis) Gini of Income]

Variable	Model 1	Model 2	Model 3
Constant	1.44(0.000)	2.11(0.000)	2.11(0.000)
Education (in years)	0.13(0.000)	0.06(0.000)	0.06(0.000)
Age	0.08(0.000)	0.08(0.000)	0.08(0.000)
Age Squared	-0.0009 (0.000)	-0.0009 (0.000)	-0.0009 (0.000)
Gini of income		-1.27(0.000)	-1.24(0.000)
Urban	0.66(0.000)	0.61(0.000)	0.61(0.000)
Gini of income *Education	-0.05(0.016)	0.08(0.03)	0.09(0.03)
Male	0.42(0.000)	0.42(0.000)	0.42(0.000)
Factory exists in community			0.08(0.345)
Total # of factories			0.06(0.033)
Cottage industry in community			0.19(0.000)
Total # of cottage industries			-0.04(0.041)
Average ebtanas test score within community			0.09(0.028)
R2	0.32	0.32	0.34
N	7235	7235	7235

Table 7-3b. [Determinants of 1991-1993 Log Monthly Earnings (p value in parenthesis) Gini of Food Expenditure]

Variable	Model 1	Model 2	Model 3
Constant	1.44(0.000)	1.97(0.000)	2.11(0.000)
Education (in years)	0.12(0.000)	0.06(0.000)	0.06(0.000)
Age	0.08(0.000)	0.08(0.000)	0.08(0.000)
Age Squared	-0.0009 (0.000)	-0.0009 (0.000)	-0.0009 (0.000)
Gini of food expenditure		-1.53(0.000)	-1.48(0.007)
Urban	0.70(0.000)	0.61(0.000)	0.61(0.000)
Gini of food expenditure *Education	-0.03(0.2)	0.12(0.018)	0.13(0.004)
Male	0.42(0.000)	0.42(0.000)	0.42(0.000)
Factory exists in community			0.05(0.537)
Total # of factories			0.07(0.007)
Cottage industry in community			0.19(0.007)
Total # of cottage industries			-0.04(0.027)
Average ebtanas test score within community			0.07(0.08)
R2	0.32	0.32	0.34
N	7235	7235	7235

first two columns though in this case the dependent variable is the logarithm of expenses on books alone, a measure that although narrower may be considered more comparable across children and schools. We also experimented with specifications in which provincial dummy variables were added to control for region. Since their inclusion had little effect on the results, these results are not reported in the table.

While the R^2 is somewhat higher when the dependent variable is the broader measure of educational expenditures, in other respects the results are very similar between the first and third columns and the second and fourth. The coefficients for the household income variables reflect positive though declining effects within the realistic range of incomes, and community income, parental education, urban location, and child age all have positive and significant effects in each case. On the other hand, educational expenditures are negatively affected by the number of children in the household, the dummy variable for a male child, and the public to private school mix in the community. Our measure of student cohort quality, namely, the average ebtanas score of sample children in the community, has a positive but not significant effect on both measures of educational expenditure. Of primary relevance are the coefficients of the inequality measures. Note that the effect of the income Gini is negative and highly significant in both cases, i.e., columns (1) and (3). Indeed, after controlling for all the other variables included, a one standard deviation increase in inequality reduces the total education expenditure on children by approximately 4.4 percent while expenditure on books is reduced by about 5.5 percent.

Note, however, that when the food expenditure Gini is used as in columns (2) and (4) the effect is no longer significant and is even positive. Given that the correlation between the two inequality measures is quite low (about 0.1 from Table 7-1) and that the standard deviation of the food expenditures is much smaller, it is not surprising that two sets of inequality effects are quite different and that for food expenditures weaker.

An additional year of fathers' education increases total educational expenditure by 3.0 percent and expenditure on books by 3.5 percent while an additional year of mothers' education increase expenditures on schooling by 3.6 percent and expenditure on books by 4.3 percent. A one year increase in the age of the child increases expenditure on books by 9 percent and total educational expenditure by 12 percent. Further, expenditure on books and total educational expenditure is respectively 7 and 5 percent less for boys. Finally, controlling for total number of schools in the community, a 0.1 increase in the ratio of public/private schools reduces total educational expenditure by approximately 4.3 percent but the change does not significantly affect expenditure on books. This suggests that the major difference in costs of public and private school may arise from tuition and other (non-book-related) issues.

Tables 7-3a and 7-3b present results where the dependent variable is the log of adult monthly earnings. Referring first to Table 7-3a, the results reported in Model 1 are for equation (3) which represents a fairly standard earnings equation but allows the returns to education to differ by community inequality levels. An additional year of education increases monthly earnings by 10 percent in a community at a level of inequality equal to 0.46 (the mean). However, for every one standard deviation increase in inequality (from its mean), there is a 3 percentage point fall in the returns to education.[ix] Individuals living in urban areas have 66 percent higher earnings than those in rural areas. Monthly earnings of males are 42 percent higher than that for females.

Once we allow for a direct effect of inequality on earnings as in Model 2, the interpretation is different. Here, we find that the returns to education actually rise with inequality levels but the overall level of earnings is lower. Since it can easily be seen that the latter effect outweighs the former within the relevant range, both sets of findings suggest that there is less incentive to invest in education in high inequality communities than in low inequality ones. In particular, holding constant other variables at their mean levels, a one standard deviation increase in the Gini reduces earnings by approximately 8 percent in Model 2. This could be the link between income inequality and investment in human capital in the case of educational investments.

Finally, in Model 3 we include additional controls for varying community characteristics, factories and cottage industries in the community and average ebtanas test scores (as a measure of school or student cohort quality). Having a factory or a cottage industry in the community and higher average test scores have positive and significant effects on earnings but earnings actually decline with additional cottage industries. Once we control for differences in labor market conditions across communities, a one standard deviation increase in the Gini reduces earnings by about 8 percent but by progressively less as the level of education rises. Individuals in a community with at least one factory earn 8 percent higher monthly earnings than individuals living in communities without those labor market conditions.

Turning now to Table 7-3b for the case in which inequality is measured in terms of food expenditure, in this case the effects of inequality on earnings are very similar to those of income inequality shown in Table 7-3a. Similarly, the effects on earnings of all the other variables are very similar, though varying slightly in terms of magnitudes of the effects.

5. CONCLUDING REMARKS

Our results show that greater inequality within a community is likely to have negative effects on household investments in education. The results are robust with respect to specification but not with respect to the inequality measure. In particular, when inequality is measured in terms of food expenditures instead of income, the effects are not statistically significant.

More importantly, we have also made some progress in identifying specific links or channels through which income inequality exercises its effects on household investments in education. In particular, our results are consistent with the hypothesis that higher income inequality communities also have lower educational infrastructure that in turn discourage household investments in education. This hypothesis is supported by the fact that the number of schools in the community (and hence shorter distance) has a positive effect on total expenditure at the community level. But, our results are also consistent with the hypothesis that labor markets clear at relatively lower wage rates in communities with higher income inequality. As a result, the incentives for undertaking educational investments are lower in communities with greater income inequality.

However, our results do not rule out alternative hypotheses that might explain why there is an observed negative effect of inequality on educational investments. For example, the effect might also operate via other channels, such as in lowering school quality that in turn reduces investments in education. It could also arise through lesser trust that could require greater direct supervision to control labor shirking instead of greater use of above equilibrium efficiency wages. Alternatively, it could arise as a mere statistical artifact as a result of out-migration of the best of the workforce from low wage communities. These issues could be explored in future research.

Admittedly also, measurement problems could affect our results. One of these could arise from recall bias. Yet, the recall problems in educational expenditures should be small since the period for recall was short (one year). The income and income inequality measures could also be subject to measurement error and since our inequality measures are constructed from the individual income measures for those households included in the sample for a given community. Chesher and Schluter (2002) provide a method to correct the inequality measures for measurement error. To get a sense of the extent of the problem, we applied their correction factor (assuming the variance of log measurement error = 0.04) to the Generalized Entropy (GE(2)) measure of inequality since the correction is most easily applied in that case. We found that there was a high correlation (0.85) between the corrected and uncorrected GE(2) indices. While not a rigorous test of the effect of measurement error on other inequality indices, our test suggests that the results are unlikely to be substantially affected by measurement error.

We did not have access to information on other important explanatory variables such as the intellectual or other abilities of the children in school or members of the labor force used in generating the earnings functions. However, there is little reason to believe that these unobserved variables would have influenced our estimates very much. Experimentation with other specifications, including alternative measures of income inequality, additional and alternative control variables and income data for different years as the basis for our inequality measures, reveals that the results with respect to income inequality on human capital investments are quite robust, with the single exception of the use of food expenditure data instead of income data in the educational investment model.

Since the educational investments considered here were relatively small, the fact that these investments have been found to be negative and significant, one might expect that the similar effects on other (bulkier) investments where financial constraints would be relatively more important would be even greater. However, at the same time, especially if the decision-makers are risk adverse and the investments risky, greater income inequality could imply greater investments overall holding average income constant.

As far as future research and validity checks are concerned, when the still incomplete data from the next round of the IFLS (IFLS-3) should become available, it should be possible to control for more presently unobserved variables and to determine whether the determinants of education have changed. At the same time, in view of potential endogeneity among some of the explanatory variables in our analysis, such data might facilitate the identification of additional instruments that would allow us to relax the exogeneity assumptions as Binswanger, Khandker and Rosenzweig (1993) have done.

Irrespective of the specific estimates reported here, we believe that the methodology for identifying links between income inequality and investments in human capital and thereby economic growth and development should prove useful in this and other contexts. As noted above, even if it should prove difficult to identify policies that would reduce income inequality in the short run without causing unwanted side effects, it is quite possible that knowledge of the links between inequality and specific types of investment may help identify more efficient and more cost-effective policies for avoiding the harmful effects of income inequality.

Finally, based on the results, attempts to open up high income inequality communities to outside employers, for example by investments in labor market information services, better transportation services or encouraging industrial employers to locate there, might well help. Such means of improving investment and growth through equity are very much within the tradition of Erik Thorbecke.

NOTES

[i] According to the same report (United Nations Development Programme 2003), Indonesia's HDI score was 0.464 in 1975, the increase since then being second only to Tunisia's. However, 26 percent of children under 5 were still reported to be significantly under weight for age, the infant mortality rate was 3.3 percent, and maternal mortality rate 380 per 100,000, and the education index still only 0.8, relatively low for countries at medium or above levels of human development.

[ii] These expectations of negative effects of inequality on growth coming through political instability, threats of violence and insurrection are by no means always likely to be fulfilled. Indeed, relative to a situation where there is no other form of political competition, the threat of revolt may induce governments to invest in infrastructure, education and various types of reforms that may be investment- and growth-increasing (Campos and Nugent, 2002, 2003). More generally, there might well be nonlinearities in these relationships.

[iii] On these various points, see Alesina and Perotti (1996) and Alesina and Rodrik (1994).

[iv] For example, in his classic works on the subject, Olson (1965, 1983) argued that inequality in income within the group would increase the probability of success in collective action and hence in financing and producing local public goods.

[v] These authors point out that the average value of a public good to members of a given community or group declines with the heterogeneity of their preferences. While they stress racial and ethnic sources of such heterogeneity, the differences could well lie in income differences within the group.

[vi] See, e.g., Rodgers (1973), Waldman (1992).

[vii] Among these are Judge, Mulligan and Benzeval (1998), Mellor and Milyo (2001).

[viii] To provide one way of checking the accuracy of household income data to generate our measure of inequality, we also generated inequality measures at the provincial level and compared these with the official statistics. We found these two measures to be highly correlated.

[ix] We tried a model adding the square of education and the interaction with inequality to model 5a. The coefficient on the square of education was negative (-0.002) but not significant, as was its interaction with inequality. Similarly, in model 5b, neither of these terms was significant.

REFERENCES

Acemoglu, Daron. 1995. "Reward Structures and the Allocation of Talent," *European Economic Review*, 39, pp. 17-34.

Alesina, Alberto, Reza Baqir and William Easterly. 1998. "Redistributive Public Employment", *NBER Working Paper* 6746.

Alesina, Alberto, Reza Baqir and William Easterly. 1999. "Public Goods and Ethnic Divisions" *Quarterly Journal of Economics*, 114:4, pp. 1243-1284.

Alesina, Alberto and Eliana LaFerrara. 2000. "Participation in Heterogeneous Communities", *Quarterly Journal of Economics*, 115:3, pp. 847-904.

Alesina, Alberto and Roberto Perotti. 1996. "Income Distribution, Political Instability and Growth." *European Economic Review*, 40, pp. 1203-1228.

Alesina, Alberto and Dani Rodrik. 1994. "Distributive Politics and Economic Growth", *Quarterly Journal of Economics*, 109, pp. 465-490.

Becker, Gary S. 1991. *A Treatise on the Family*. Cambridge: Harvard University Press.

Becker, Gary S. and Nigel Tomes. 1976. "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy*, 84:4, pp. 143-162.

Becker, Gary S. and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, 87:6, pp. 1153-1189.

Benabou, Roland. 1996a. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance", *American Economic Review*, 86:3, pp. 584-603.

Benabou, Roland. 1996b. "Inequality and Growth" National Bureau of Economic Research Working Paper 5658.

Binswanger, Hans P., S.R. Khandker, and M.R. Rosenzweig. 1993. "How Infrastructure and Financial Institutions Affect Agricultural Output and Investment in India", *Journal of Development Economics*, 41:2, pp. 337-366.

Birdsall, Nancy, Pinckney and Richard Sabot. 1999. "Equity, Savings and Growth" CSED Working Paper 8, Washington, D.C.: Brookings Institution.

Campos, Nauro F. and Jeffrey B. Nugent. 2002. "Who Is Afraid of Political Instability?" *Journal of Development Economics*, 67, pp. 157-172.

Campos, Nauro F. and Jeffrey B. Nugent. 2003. "Aggregate Investment and Political Instability: An Econometric Investigation", *Economica*, 70:279, pp. 533-550.

Chesher, Andrew and Christian Schluter. 2002. "Welfare Measurement and Measurement Error", *Review of Economic Studies*, 69, pp. 357-378.

De la Croix, David and Mathhias Doepke. 2003. "Inequality and Growth: Why Differential Fertility Matters", *American Economic Review*, 93:4, pp. 1091-1113.

Deaton, Angus. 2002. "Relative Deprivation, Inequality and Mortality", National Bureau of Economic Research, Working Paper 8099.

Deaton, Angus. 2003. "Health, Inequality and Economic Development" *Journal of Economic Literature*, 56:1, pp. 113-158.

Deininger, and Lynn Squire. 1998. "New Ways of Looking at Old Issues: Inequality and Growth", *Journal of Development Economics*, 57, pp. 259-287.

Judge, Ken, Jo-Ann Mulligan and Michaela Benzeval. 1998. "Income Inequality and Population Health." *Social Science Medicine*, 46:4, pp. 567-579.

Khan, Haider A. and Erik Thorbecke 1988. *Macroeconomic Effects and Diffusion of Alternative Technologies within a Social Accounting Matrix Framework*. Hants: Gower

Kuznets, Simon 1955. "Economic Growth and Income Inequality", *American Economic Review*, 45:1, pp. 1-28.

Li, H., Squire, L. and Zou, H. 1998. "Explaining International and Intertemporal Variation in Income Inequality" *The Economic Journal*, 108, pp. 26–43.

Mellor, Jennifer and Jeffrey Milyo 2001. "Re-Examining the Evidence of an Ecological Association between Income Inequality and Health", *Journal of Health Policy, Politics and Law*, 26:3, pp. 487-522.

Olson, Mancur 1965. *The Logic of Collective Action.* Cambridge: Harvard University Press.

Olson, Mancur 1983. *The Rise and Decline of Nations: The Political Economy of Economic Growth, Stagflation and Social Rigidities.* New Haven: Yale University Press.

Psacharopoulos, George 1985. "Returns to Education: A Further International Update and Implications", *Journal of Human Resources,* 20:4, pp. 584-604.

Psacharopoulos, George and H.A. Patrinos 2002. Returns to Investment in Education: A Further Update, Washington, D.C.: World Bank Policy Research Working Paper 2881.

Rodgers, G.B. 1973. "Income and Inequality as Determinants of Mortality: An International Cross-Section Analysis" *Population Studies,* 32, pp. 343-351.

Thorbecke, Erik 1988. "The Impact of Stabilization and Structural Adjustment Measures and Reforms on Agriculture and Equity" in E. Berg, ed. *Policy Reform and Equity.* San Francisco: ICS Press. 63-94.

Thorbecke, Erik 1992. *Adjustment and Equity in Indonesia.* Paris: OECD Development Centre.

Thorbecke, Erik and Theodore van der Pluijm 1993. *Rural Indonesia: Socioeconomic Development in a Changing Environment.* New York: New York University Press for the International Fund for Agricultural Development.

Thorbecke, Erik and Chatatong Charumilind 2002. " Economic Inequality and Its Socioeconomic Impact" *World Development,* 30:9, pp. 1477-1495.

United Nations Development Programme 2003. *Human Development Report 2003.* New York: Oxford University Press.

Waldmann, Robert J. 1992. "Income Distribution and Infant Mortality" *The Quarterly Journal of Economics,* 107:4, pp. 1283-1302.

Wilkinson, R.G. 1996. *Unhealthy Societies. The Afflictions of Inequality.* London: Routledge.

Wilkinson R.G. 2000. *Mind the Gap: Hierarchies, Health and Human Evolution.* London: Weidenfeld and Nicholson.

Chapter 8

POVERTY TRAPS AND SAFETY NETS

Christopher B. Barrett[1] and John G. McPeak[2]
[1]*Cornell University, cbb2@cornell.edu and* [2]*Syracuse University, jomcpeak@maxwell.syr.edu*

1. INTRODUCTION

Erik Thorbecke has been a leader among development economists for decades and an inspiration to those of us who have had the privilege to work with him and to witness his creativity, commitment and industriousness first hand. The rigor of his research on the economics of poverty and nutrition and the relevance of his work to the practical concerns of development practitioners and policymakers has distinguished Erik's scholarship from most others'. It is a great honor to contribute some reflections on the economics of poverty traps, nutrition-related health risk and safety nets to this event in recognition of Erik's achievements and contributions to our profession and to questions such as those that follow.

This paper explores three interrelated questions that have been central to Erik's past research and his current concerns: What is the etiology of chronic poverty and vulnerability? How does nutrition-related health risk affect patterns of chronic poverty and vulnerability? What are the implications for the design of development policy, especially safety net interventions? In recent years, economists have spilled much ink over both risk management and poverty analysis. Yet integration of these topics has remained distressingly limited. Most of the recent empirical development microeconomics research on risk has focused on variability in incomes or expenditures, the extent to which some portion of that variability might be uninsured among poor subpopulations, and the means by which insurance emerges. One of the main findings of this line of research is the rejection of the neoclassical consumption smoothing hypothesis. Meanwhile, the lion's share of recent poverty analysis has focused on technical issues surrounding poverty measurement, on the relationship between poverty and economic growth, and on poverty dynamics. In this paper, we seek to integrate these two threads explicitly.

Recent research has begun to point toward an economic rationale – as distinct from but complementary to a purely humanitarian rationale – for safety net interventions intended to reduce exposure to significant downside risk.[i] Protection against risk – and perhaps especially against nutrition-related health risk – plays a crucial role in stimulating accumulation of productive assets and adoption of improved agricultural production technologies, and thus in sparking sustainable growth in incomes and reduction in chronic poverty. When designed and implemented well, social protection in the form of safety nets can play an important role in stimulating economic growth as well as poverty reduction.

2. CHRONIC POVERTY AND POVERTY TRAPS

Increased availability of longitudinal data at household and individual level is changing empirical poverty research in ways that affect our understanding of poverty. For nearly twenty years, the staple poverty metric has been the Foster-Greer-Thorbecke (1984, hereafter FGT) family of decomposable measures encompassing the headcount, poverty gap and more distributionally-sensitive measures of poverty within a population. FGT measures offer a powerful instrument for poverty analysis, albeit one restricted to only a cross-sectional view of poverty. With advances in data availability and methods, poverty researchers are increasingly moving from such static or "snapshot" views of poverty to dynamic or "video" perspectives, tracking the path followed by the poor over time. This has led to a variety of important refinements.

One important refinement arises from the crucial distinction between transitory and chronic poverty (Grootaert et al. 1996, Baulch and Hoddinott 2000). The basic concept of chronic poverty is clear – poverty that persists for years, if not lifetimes – even if there is some variation across authors in operationalization of the concept, sometimes as mean income or expenditures below the poverty line over a time series, sometimes as all or a super-majority of observations below the poverty line. Whatever the precise definition one applies to the data, transitory poverty is plainly shorter-lived than chronic poverty. All else equal, a poor person would far rather experience transitory poverty rather than chronic poverty.

This reveals another dimension – besides familiar FGT headcount and poverty gap measures – in which the poverty of developing countries of the South differs qualitatively from that of the wealthy countries of the North.

In contrast to the United States, where the median time in poverty is 4.5 months (Naifeh, 1998), the median time in poverty in rural Bangladesh, Congo, Ethiopia, Kenya or Madagascar is one or more lifetimes. Back-of-the-envelope calculations implied by published data suggest monthly exit rates from poverty of 6.9 percent for the United States, meaning nearly 7 percent of those who are poor at the beginning of a month will exit poverty by the start of the next month. By contrast, equivalent exit rate estimates based on published panel data studies are only 1.3 percent in Côte d'Ivoire and only 0.7 percent for KwaZulu Natal state in South Africa. Panel data sets we have been assembling in recent years indicate monthly exit rates from poverty that are merely 0.6 and 0.4 percent in rural Madagascar and rural Kenya, respectively.[ii] Although these comparisons are necessarily crude, they nonetheless underscore an important qualitative point. It is not just the *magnitude* of poverty but, perhaps even more importantly, the *duration* of poverty that differentiates much of the developing world from the United States and other wealthy countries. Poverty that persists for such long periods of time gives particular salience to the concept of a *poverty trap*, on which we reflect more momentarily.

The second important recent refinement one finds in the literature is a parallel distinction of *structural* poverty from *stochastic* poverty. According to Carter and May (1999), the structurally poor lack asset endowments sufficient to generate expected income or expenditures above the poverty line, although observed income may exceed the poverty line due to random shocks. The stochastically poor, by contrast, have observed income or expenditures below the poverty line even though their asset holdings suffice, in expectation, for them to be nonpoor. This structural-stochastic distinction introduces an all-important mapping from income or expenditure measures to asset measures which can be subject to less measurement error (Sahn and Stifel 2000) and which lend themselves to thinking about poverty dynamics as they relate to asset dynamics, which underpin the concept of poverty traps.

The structural-stochastic poverty distinction leads naturally to the third relevant recent refinement in the economics of poverty: measurement of household vulnerability to poverty. The crucial insight of this emerging literature is that substantial vulnerability to poverty does not imply that people necessarily suffer poverty, just that they face real and costly risk, and that they likely behave accordingly. Studies adopting this approach develop vulnerability measures of households' probability of being poor in one period based on observable household characteristics in one or more prior periods (Christiansen and Boisvert 2000, Pritchett et al. 2000, Chaudhuri 2001, Christiansen and Subbarao 2001, Chaudhuri et al. 2002), perhaps augmented by valuation of the risk to which the household is exposed (Ligon and Schechter 2002). Vulnerability measures thus incorporate conditional moments beyond the mean, establishing the conditional probability of falling below a poverty line, the cost of uncertain welfare, or both.

Chronic and structural poverty raises the prospect of *poverty traps*. The pivotal feature of poverty traps is the existence of one or more critical wealth thresholds that people have a difficult time crossing from below. Above the threshold, asset growth takes people toward a high-productivity steady state where they are non-poor and, at most, only moderately vulnerable to poverty, while below the threshold, people sink toward a low-productivity poverty trap characterized by frequent, if not constant, spells of poverty. The threshold is the point at which expected path dynamics bifurcate, i.e., where the time derivate of expected wealth changes sign. Poverty traps thus imply the existence of multiple dynamic equilibria. Chronic vulnerability to poverty describes the status of households who struggle to accumulate or protect assets that will allow them to stay out of chronic, structural poverty but who remain below a critical wealth threshold.

Longstanding hypotheses about multiple dynamic equilibria are receiving renewed attention in the economics literature.[iii] Highly suggestive empirical evidence is now emerging that indeed Myrdal, Nurkse, Rosenstein-Rodan and Young may have been correct about the existence of distinct accumulation trajectories, one or more of which lead to what we have termed poverty traps.[iv] Further theoretical and empirical findings on the causes and consequences of multiple dynamic equilibria and their relationship to poverty traps offer the potential for improved development policy to diminish the prevalence of such traps. In particular, empirical corroboration of the existence of poverty traps -- bifurcations in expected asset accumulation patterns as reflected in thresholds defined on current period asset holdings – would signal the necessity of renewed activism by donors and governments to address insufficiency of asset holdings and financial markets access among the chronic poor.

The idea of poverty traps and multiple dynamic equilibria ultimately hinges on the role of productive assets. Given initial conditions on productive assets, the productivity of those assets, and expected asset dynamics, we can define asset trajectories that characterize household poverty status over time. Asset poverty thus plays a central role in understanding poverty as it is more commonly measured using income or expenditure data.

If asset holdings underpin poverty status, then asset dynamics underpin poverty dynamics. We can, therefore, get a reasonably good first answer to the question of "do there exist multiple dynamic equilibria associated with poverty traps?" by studying asset dynamics and looking for multiple dynamic equilibria in asset space. The central role asset holdings play in generating income and expenditure patterns, and that asset dynamics play in welfare dynamics more broadly, underscores a fundamental weakness of the canonical consumption smoothing model. The canonical version of the permanent income hypothesis, dating back to Friedman (1957), from whence predictions of consumption smoothing originate, rests on the crucial assumption that stochastic income draws are independent and identically distributed (iid) across time. That might be true if risk were purely associated with the returns on an intertemporally fixed set of assets. As soon as one allows for asset risk, however, then the iid assumption regarding stochastic income necessarily falls. If asset shocks have any persistence from one period to the next – for example, anything beyond very short term illness with no lasting effects –and income is a function of assets, then income draws are not iid. Rather, they will be strongly, positively autocorrelated at a minimum. More likely, the conditional distribution of income will shift over periods in response to changes in the underlying stock of productive assets. As a consequence, consumption does not follow a martingale process; rather it depends on (at least) wealth (Deaton 1992, Bhargava and Ravallion 1993). People know that and behave accordingly.

We briefly explore the issue of asset dynamics among a poor population using data collected quarterly from March 2000-December 2001 among 177 pastoralist households in six sites in the arid and semi-arid lands of northern Kenya.[v] The primary nonhuman assets held by pastoralists are their herds of livestock. The relationship between herd size, measured in tropical livestock units (TLU),[vi] and daily per capita income is strong and monotonically increasing, as depicted in Figure 8-1. Bigger herds generate a greater flow of milk, the primary source of income (in kind) in the east African rangelands.[vii]

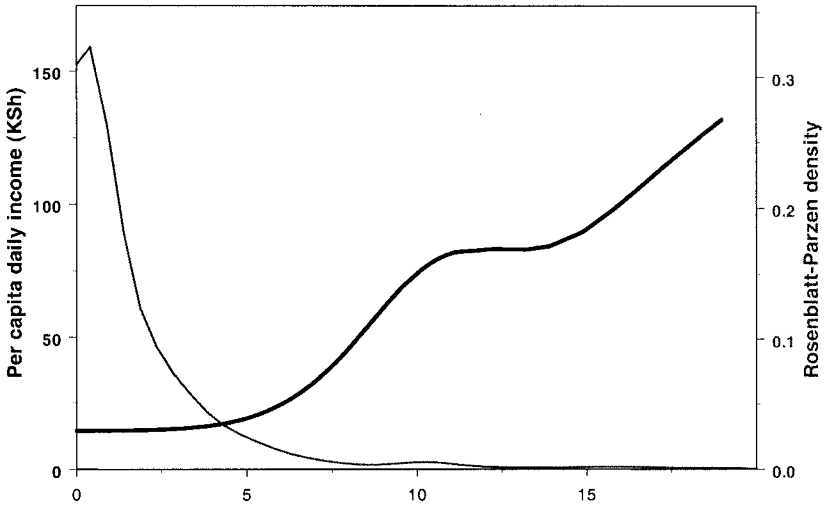

Figure 8-1. [Income - Herd Size Relationship]

The simple bivariate nonparametric kernel regression depicted in Figure 8-1 suggests that per capita daily income is convex in per capita TLU holdings over most of the data range (the density of household livestock holdings is also plotted, against the righthand axis), corroborating prior conjectures of endogenously increasing rates of expected return on assets (McPeak and Barrett 2001). While this merits further exploration controlling for additional covariates, this finding suggests that income may increase at more than a one-for-one rate as wealth increases, at least over the upper portion of the wealth distribution. Indeed, the points at which expected change in income seems to increase with additional livestock holdings correspond reasonably well with rules of thumb common in the multidisciplinary literature on pastoralism, which points to a mobility threshold of 2-5 TLU/person below which households are typically unable to engage in long-distance trekking to take advantage of spatiotemporal variability in forage and water availability, and an invulnerability-to-immobility threshold at 9-12 TLU per capita (see for example Upton, 1986; Assefa, 1990; Fratkin and Roth 1990). This interval corresponds precisely with the primary range of apparent convexity in the income-herd size relationship shown. Of course, relatively few households are able to take advantage of this.

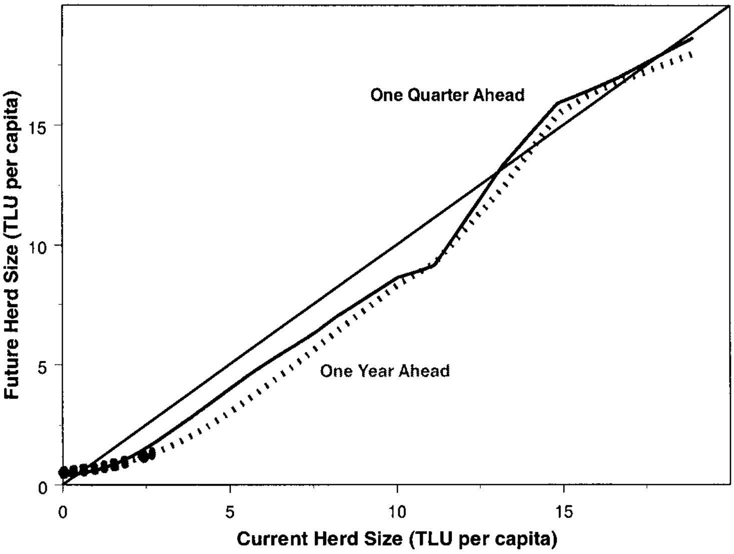

Figure 8-2. [Herd Dynamics in Northern Kenya]

The strong, seemingly convex relationship between assets and income raises a natural question about asset dynamics: who can expect to enjoy growth in herd sizes and thus even more rapid growth in incomes? Figure 8-2 offers an example of asset dynamics characterized by multiple equilibria in our Kenyan data. The black 45-degree line in Figure 8-2 represents dynamic equilibria, where expected future herd size equals current herd size. Observations that lie above the 45-degree line indicate growth in asset stock over time; observations below it reflect asset decumulation. Most households cluster around the 45-degree line, enjoying a stable herd size in both one quarter (blue solid line) and one year (red dashed line) transitions. Consequently, the conditional median (not shown) – as distinct from the conditional means depicted in the solid lines – tracks the 45-degree line through most of the data. But herd collapse is far more common than rapid herd growth, as reflected in the noticeably greater density of points below the 45-degree line than above it. The herd collapse recorded in the data largely reflects livestock losses that occurred during the first half of the study period due to a widespread drought in northern Kenya. Importantly, the results suggest an asymmetry in asset risk conditioned upon asset holdings. The asymmetry in asset risk is illustrated by the expected herd dynamics, which show a clear S-shaped pattern, based on the nonparametric kernel regression of future herd size on current period herd size at both 3-

month (quarter) and 12-month (year) leads. This pattern is very similar to that Lybbert et al. (2004) find using 17-year herd histories from a qualitatively similar system in neighboring southern Ethiopia.

These asset dynamics reveal multiple dynamic equilibria. Stable equilibria appear at approximately 1 and 17-18 TLU per capita, with an unstable equilibrium around 12 TLU per capita, a bit above – although not statistically significantly different from – the TLU/person invulnerability-to-immobility threshold mentioned earlier. The unstable equilibrium reflects a critical threshold. Herd sizes that reach or exceed the threshold will, on average, grow to the higher, stable equilibrium herd size, which yields expected per capita daily income of roughly US $1.50/day (see Figure 8-1). But only about one percent of our sample attains this high-level equilibrium. When herd sizes fall below the threshold level of roughly 12 head, future per capita herd sizes steadily decrease in expectation, to the point where per capita herd sizes and expected daily per capita income are only about 1 TLU and $0.25, respectively. The herd size distribution reflected in Figure 8-1 suggests that few northern Kenyan pastoralists are able to surmount that critical threshold to reach the high, stable dynamic equilibrium. Instead, most people find themselves trapped in extreme poverty.

People can be trapped either because they are born into extreme poverty and have a difficult time accumulating assets or because they suffer significant shocks that cast them below the critical threshold point before they are able to accumulate enough of a herd to cross the threshold, as in Dercon (1998). Because stochastic asset shocks play a central role in understanding why pastoralists routinely suffer accumulation failures, asset risk is central to a solid understanding of poverty dynamics in an environment such as northern Kenya, where frequent droughts, violent cattle raids and human disease epidemics confront pastoralists with extraordinarily great risk of asset loss (McPeak and Barrett 2001, Smith et al. 2001).

3. HUMAN CAPITAL, SUBSISTENCE CONSTRAINTS AND POVERTY TRAPS

The preceding discussion and evidence help to illustrate the concepts of poverty traps and of critical thresholds in asset space, with a specific application to livestock, the primary non-human asset of east African pastoralists. This same intuition regarding the monotone (and potentially convex) relation between assets and flow measures of welfare (e.g., expenditures or income) and the possible existence of S-shaped asset dynamics carries over to other key productive assets. One category of assets with great relevance to poverty analysis focuses on assets embodied in people: human capital. In this section we therefore reflect on how human capital as represented by health status can be viewed as another asset that can be used to conceptualize what distinguishes the poor from the non-poor. We focus especially on how health risk, especially nutrition-related health risk that threatens human capital, can affect accumulation and risk management patterns and can lead households into a poverty trap.

No asset risk more threatens human livelihoods than health risk for the simple reason that human capital is the most valuable asset among the poor, for two key reasons. First, the poor commonly own little other than their labor power, lacking land, livestock, and significant financial or physical assets. Their livelihoods depend almost entirely on wage earnings and transfers. Physical capacity to work underpins their livelihoods and is based on good health and nutrition.

The second reason for human capital's inordinate importance arises from its complementarity with other productive assets. Few assets yield returns without some complementary input of labor. Even the most fertile soils yield no crops without planting and harvest labor and the most productive livestock give no milk without labor. Hence the notion of Lockean property rights, that by mixing one's labor with land (or any other latent but previously unexploited resource), it transforms the resource into a productive asset. When human capital is diminished by health or nutrition shocks, the poor's livelihood systems are at least threatened, and commonly degraded.

Furthermore, human capital assets are potentially subject to important irreversibilities.[viii] A household can lose its herd or its land and yet remain able to reconstitute a herd or secure access to new land in time. But while many health shocks are mild and quickly overcome in time, similar recovery of human capital losses is often impossible in the wake of permanent physical disability or acute illness (e.g., blindness, accidental loss of limbs or brain damage, cretinism due to iodine deficiency), much less death. The irreversibility of some types of health shocks creates an especially salient critical threshold in asset space – akin to a subsistence constraint – that has a profound effect on welfare dynamics. Future asset stocks and income flows may depend on current consumption, at least for certain populations near critical morbidity or mortality thresholds.

Although most of the risk management literature has focused on income risk and the use of asset stocks to buffer consumption against stochastic income, an important sub-literature addresses the somewhat more complex problems associated with asset risk among poor populations (Dercon 1998, Carter and May 2001, Zimmerman and Carter 2003, McPeak 2004, Lybbert et al. 2004). Economists have long appreciated that asset risk influences consumption and accumulation patterns because future income is endogenous to current asset shocks and subsequent consumption choices

(Phelps 1962, Levhari and Srinivasan 1969, Sandmo 1969, 1970). When income shocks and asset shocks occur contemporaneously, then forward-looking agents will balance familiar consumption smoothing behaviors associated with agents' desire to equalize the discounted expected utility of consumption across periods – given income – with asset smoothing behaviors arising from an inextricable desire to smooth expected income across periods. As a consequence, household consumption over time will tend to be relatively more volatile (i.e., less smoothed) in communities where asset risk is greater.

As the coping strategies literature makes abundantly clear, very poor populations will liquidate virtually any asset – in extreme cases, even sell themselves or their children into slavery – when they might otherwise cross beneath a critical nutritional threshold, a point where they run intolerable risk of permanent impairment due to injury or illness (sometimes referred to in the dynamics literature as an "absorbing state"). Yet the poor will also vary consumption dramatically above that threshold in an effort to protect productive assets essential to minimizing nutrition-related health risk in future periods, typically reducing food consumption as a first line of defense against long-term asset loss (Maxwell 1995, Barrett 2002). In sum, in the face of asset risk, the very poor may destabilize consumption – while being careful not to cross the subsistence threshold in the current period – so as to defend their productive asset stock and thereby maximize the probability of future survival. This leads to great consumption volatility among the most vulnerable subpopulations, and forces them to make decisions that risk current health status to increase prospects for future survival.

Consumption instability is especially worrisome among the world's poor because severe (macro- or micro-nutrient) undernutrition is strongly and causally associated with acute health shocks and because the poor depend disproportionately on labor earnings for their livelihoods. Ill health is both cause and consequence of extreme poverty. As people become poor, they become more likely to suffer serious illnesses and injuries and such illnesses and injuries aggravate pre-existing poverty. This bidirectional causality lies at the heart of the literature on nutritional efficiency wages.[ix]

One source of low exit rates from poverty originates in this bidirectional causality. Over a long period of time, it can even lead to the intergenerational transmission of poverty, as parents' income is positively related to children's nutrient intake and likelihood of receiving immunizations and other preventive and curative health care, and early childhood episodes of illness and undernutrition have persistent effects on stature and cognitive development that significantly affect expected lifetime earnings (Dasgupta 1997, Strauss and Thomas 1997, Martorell 1999, Glewwe et al. 2001, Hoddinott and Kinsey 2001, Gertler and Gruber 2002, Dercon and Hoddinott 2004).

The fact that serious health shocks – many of which are causally related to undernutrition – cannot be fully insured and thus can have significant and persistent effects is fairly intuitive. But health shocks matter not only because of their ex post impact when they do occur but, of broader relevance, because people adapt their livelihood strategies in response to their assessment of the risks they face under alternative strategies. As we discussed above, vulnerability to future poverty is as important to understand as current and past poverty. People optimally manage assets and consumption with an eye toward the effects on future human capital stock, not just to equalize expected utility of current consumption over time. The resulting behaviors may sometimes appear as if agents are making intertemporal resource allocation decisions subject to a subsistence constraint, i.e., as if consumption is constrained not to fall below some threshold level in any period. The term "subsistence constraint" nonetheless lends itself to misinterpretation. In the context of poverty traps, it is less a strict physiological consumption constraint in a single period than an endogenous behavioral pattern caused by the persistence of shocks to human capital and dynamic patterns of asset trajectories. Income draws in this setting are not stochastic independent and identically draws across time, undercutting the canonical model of consumption smoothing. Decisions made in response to one's perceived vulnerability to falling into a poverty trap have important implications for well being over time. In particular, ex ante health and nutritional risk mitigation strategies can have a profound effect on accumulation patterns. This happens in at least two different ways.

First, peoples' risk preferences affect their choice of activity and consumption patterns. If decreasing absolute risk aversion best describes risk preferences, as the bulk of the relevant literature suggests, then the poor will tend to pay more to reduce risk exposure than will the rich. Such payments typically come in the form of foregone earnings. For example, Binswanger and Rosenzweig (1993) found that a one standard deviation increase in weather risk induces Indian households of median wealth to reduce expected farm profits by an estimated 15 percent, while household in the bottom quartile reduce expected farm profits by 35 percent. The wealthiest quartile households, on the other hand, have adequate independent risk coping mechanisms, so they adjust input use patterns hardly at all to increased exogenous risk. Carter (1997) estimated household willingness to pay for certain food availability in Burkina Faso at better than twenty percent of income. He found that the cost of incomplete self-insurance seemed to be on the order of ten percent or more of income. Several recent studies – notably Elbers and Gunning (2003) – similarly observe that households can trap themselves in chronic poverty through their rational approach to risk management.

Second, the effects of health and nutritional risk on behavior and welfare dynamics may come not through risk preferences per se, but through awareness of and behavioral response to the existence of critical thresholds in human capital space associated with irreversibilities in physical functioning. As a direct consequence of the absorbing states associated with irreversible health shocks, the poor may forego high-return investments that would demand greater short-term sacrifices in consumption than they dare undertake for fear of coming too near the threshold of permanent (or near-permanent) health shocks. For example, Zimmerman and Carter (2003) demonstrate how uninsurable asset risk leads poor people to hold highly unproductive asset portfolios. In a similar spirit, Moser and Barrett (2003) demonstrate the importance of subsistence constraints in explaining nonadoption of a high-yielding, low-input method of rice production in Madagascar.

The presence of de facto subsistence constraints associated with increased likelihood of irreversible health shocks causes important interhousehold variation in risk management behavior. In particular, while the poor will still tend to smooth income more than the rich – due to risk averse preferences and financial barriers to entry into high return/high risk activities – they will also tend to smooth consumption less, relative to their income, than do the rich, choosing instead to buffer productive assets on which future well-being depends as a strategy for staying away from the perilous threshold of permanent impairment. The threat of irreversible human capital loss tends to induce generalized asset smoothing – as distinct from consumption smoothing – among those near subsistence thresholds, as the poor aim to protect critical resources as a bulwark against future nutrition-related health risk.

Figure 8-3, drawn from Barrett et al. (2004), demonstrates the relation of wealth status to consumption variability by plotting the nonparametric kernel regression of the coefficient of variation (CV) for both income and expenditures – computed from the quarterly panel data observations from northern Kenya – on initial period household herd size. The positive correlation between wealth and income risk is apparent in the upward slope of the blue, dashed line depicting income CV. The gap between the income CV regression line and the (red, solid) expenditure CV regression line reflects consumption smoothing behavior. While richer households take on greater income risk than poorer households do, the rich nonetheless enjoy lower intertemporal variability in expenditures. Consumption smoothing is a normal good, increasing in wealth in spite of prospectively greater absolute risk aversion among the poor. We hypothesize that this occurs precisely because of the threat posed by nutrition-related health risk. Among the poorest households, intertemporal income variability is actually less than expenditure volatility, signaling that the most vulnerable households will destabilize consumption in order to protect crucial productive assets on which their future survival will depend.

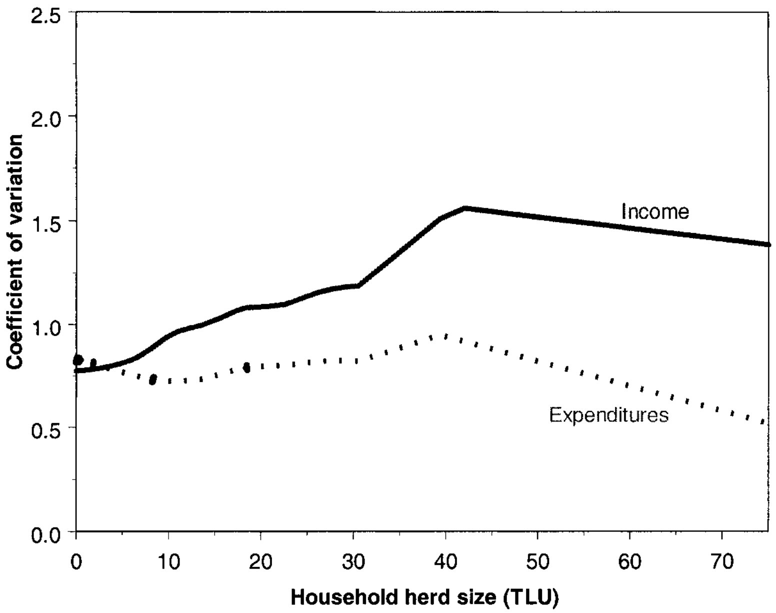

Figure 8-3. [Wealth-dependent risk management]

In summary, the existence of critical irreversibilities in human capital dynamics — generating multiple equilibria similar to the S-shaped asset dynamics shown previously for livestock in northern Kenyan pastoralist households – affects household risk management. Because households know that health and nutrition shocks occur with positive probability that is a function of current allocation decisions and that some asset shocks are irreversible, they adapt their behaviors accordingly, choosing activity and asset portfolios that limit income risk, foregoing high-return investments that would demand significant short-term sacrifice beyond what is prudent and safe, and willingly destabilizing consumption in order to protect assets so as to minimize the probability of falling into a poverty trap. However, in so doing, they increase the probability that they will suffer a negative shock to their health. Beyond some threshold level of asset holdings, they can begin to afford to undertake higher risk and higher return livelihood strategies, leading to locally increasing returns to assets such as those depicted in Figure 8-1. Thus poorer households must decide which is less threatening to their future prospects, parting with productive assets or risking their health.

4. THE ECONOMIC RATIONALE FOR SAFETY NETS

The existence of significant asset risk and the behavioral response these risks elicit lead directly to the strongest economic rationale for safety nets. The fact that many shocks – especially serious health shocks, many of which are causally related to undernutrition – cannot be fully insured and thus have significant and persistent effects is fairly intuitive. This provides the prevailing humanitarian rationale for interventions, based on what is commonly termed a "rights-based approach" enshrined in the 1948 Universal Declaration of Human Rights (Article 25) and the 1966 International Covenant on Economic, Social and Cultural Rights, (Article 11) and reaffirmed at the 1996 World Food Summit and the follow-on summit held in 2001. Unmet demand for ex post recovery assistance also provides one important economic rationale for publicly-provided safety nets in the form of emergency food assistance, employment guarantee schemes, and publicly funded health care for the indigent.

We want to pursue a slightly different, supplementary tack in arguing the economic rationale for safety nets. The poor tend to be much more exposed than the rich are to asset risk and thus face a higher probability of being cast below critical thresholds due to adverse shocks caused, for example, by drought, floods, hurricanes, epidemics or war.[x] In the absence of effective safety nets, people routinely fall not only into poverty, but beyond critical asset thresholds and into chronic poverty. Effective safety net programs can generate significant indirect benefits by reducing vulnerable peoples' need to mitigate downside risk through costly portfolio management and activity choice or to sell off scarce productive assets when current earned income shortfalls are so serious that they would otherwise suffer irreversible health effects.

Development policy has long focused on structural interventions intended to increase the poor's asset holdings through direct transfers, public health and education services, land reform or other such redistributive programs. Such policies may indeed be necessary to assist many of the chronically poor (Barrett forthcoming). There exists no good evidence of which we are aware, however, as to the relative poverty reduction effectiveness of direct efforts to build up the assets of the poor through redistributive transfers, as compared to indirect efforts to induce endogenous asset accumulation by the poor through reduced exposure to downside asset risk. This is an interesting and important question that rigorous, policy-oriented researchers in the Thorbecke tradition might usefully tackle.

The issue is not only the relative effectiveness of alternative types of interventions.[xi] Redistributive programs may only be effective in achieving sustainable, long-term reductions in chronic poverty when complemented by safety nets. Moreover, absent effective safety nets, redistributive policies to build up the poor's asset holdings may prove rather like opening the drain on a tub with the faucet still running. Many people may exit chronic poverty due to pro-poor transfer policies, but new ones will enter chronic poverty just as quickly in the absence of effective safety nets and households teetering on the edge of chronic poverty will choose risk management strategies that predictably fail to stimulate asset and income growth. At a minimum, effective safety nets should block pathways into poverty.

Moreover, knowledge that such safety nets exist may allow some households to move out of poverty due to behavioral changes that come about in response to reducing the risk of crossing some critical asset threshold. When financial markets fail and people cannot borrow to trade consumption between periods or contract for insurance against adverse shocks, the financial market failure gets displaced into exaggerated activity in some other area(s), causing costly allocative inefficiencies that retard asset accumulation and income growth. If safety nets can effectively reduce the risk faced by households such that they no longer need to worry about a brief shock having calamitous, permanent consequences, then we should observe poor people undertaking more high-return investments that require short-term sacrifice, a reallocation of portfolios away from safer, lower-return assets and towards higher-yielding assets and activities, and reduced asset smoothing behavior. These are important, open, empirical questions as to whether safety nets can provide a pathway out of poverty for those whose endogenous risk management strategies leave them chronically poor.[xii]

But safety nets can only generate desirable poverty reduction benefits if they are credible. Poor people must believe that a promised safety net will indeed be available when needed and that it will function as promised. Otherwise, they will not reduce their practice of costly risk mitigation strategies that contribute to chronic poverty.

A crucial issue here is targeting. Targeting concerns the who, the when, the what, and the how questions surrounding transfers: is aid reaching people who need it (and not flowing to people who do not need it), when they need it, in appropriate form, and through effective modalities? All real world transfer programs suffer targeting errors for the simple reasons that (i) information is costly to collect and to process, (ii) it is impossible to have perfect information about all people at all times (i.e., to know who is and is not needy) and (iii) actual allocations are made for multiple reasons, only one of which is objective need (Barrett and Maxwell forthcoming). Especially where means-based screening of prospective beneficiaries proves administratively infeasible – as is true in most low-income countries – then intra-community heterogeneity and factor market failures tend to generate significant errors of inclusion even in self-targeting program designs (Barrett and Clay 2003). Because a safety net program without targeting errors is practically infeasible, there exists a difficult tradeoff between wasteful and distortionary errors of inclusion and potentially damaging errors of exclusion. There's no clearly superior direction in which to err. The difficulty of this tradeoff makes minimization of targeting errors essential (Barrett and Maxwell forthcoming).

In the northern Kenyan communities we study, food aid provides the primary (in many cases, the only) safety net. In northern Kenya, food aid responds to rainfall shocks. When period-average rainfall across a large area drops, food aid shipments begin. So food aid responds to climate shocks. While rainfall indeed has a major effect on asset and welfare dynamics in the rangelands of the Horn of Africa, the effects are more subtle than is commonly acknowledged. First, there is considerable microclimatic variability. Some areas may get just enough rain at just the right time that they suffer no serious loss of water or forage during a "drought". Relief agencies commonly fail to take such microvariability into consideration in geographic targeting of food aid. Second, agents vary markedly in their capacity to manage common rainfall shocks – either ex ante, through mitigation efforts, or ex post through coping strategies – and thus asset risk appears to be overwhelming idiosyncratic (i.e., household-specific) rather than covariate (McPeak and Barrett 2001, Smith et al. 2001, Lybbert et al. 2004). Because asset dynamics are less closely correlated across households than is often assumed, safety net interventions such as food aid distribution that respond to aggregate shocks necessarily introduce considerable errors of inclusion through poor targeting. Third, when aggregate shocks do cause asset loss due, for example, to herd die-offs, the income shock is persistent, dampening only if and as herds recover. But for the nontrivial minority who lose their herds completely, or who suddenly fall below the critical mobility threshold, the shock may be nonstationary, in which case food aid as a short-term palliative fails to match the long-term needs for an alternative livelihood.

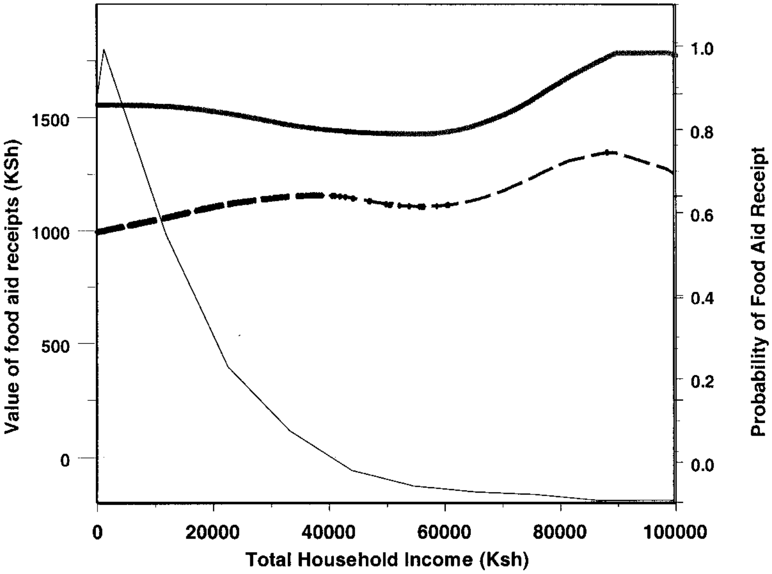

Figure 8-4. [Food Aid Targeting in Northern Kenya]

In our sample, every household received food aid in at least one quarter, 2000-2001, a period beginning in the midst of one of the worst droughts to hit the region in a generation.[xiii] Yet in spite of ubiquitous participation and a severe drought, the median share of total household income represented by food aid was merely 11%. Figure 8-4 shows the nonparametric kernel regressions of the probability of receiving food aid (blue thicker quasi-horizontal line) and of the monetary value of food aid receipts (red, thinner, dashed quasi-horizontal line) on household earned income[xiv] over a three-month period, as well as the density of household earned income (black, thinner curve). There was no appreciable difference across the income distribution in either the likelihood of receiving food aid in a given quarter nor of the value received. If anything, households with very high earned income were more likely than those with low earned incomes to receive food aid and their expected food aid receipts were somewhat higher than those received by the poorest households. These findings echo evidence reported recently from Ethiopia that similarly show food aid going to the rich with at least the same frequency and at nearly the same rate as to the poor (Clay et al. 1999, Jayne et al. 2001). In spite of massive shipments into the region and widespread concerns that pastoralists suffer food aid "dependency", the resource is distributed "a mile wide and an inch deep", as one NGO official described it to us recently, so broadly as to become ineffective in truncating downside risk faced by pastoralists. As a consequence, the safety net fails to provide necessary insurance against downside risk and pastoralists engage in costly self-insurance through herd accumulation (McPeak and Barrett 2001, Lybbert et al. 2004).

Conceptually, safety nets can play an extremely valuable role in mitigating asset risk, in keeping short-term shocks from leading to chronic poverty through endogenous asset decumulation or low-return production and portfolio strategies. There are examples of safety net schemes that seem to work, perhaps especially those based on public employment guarantees.[xv] But as presently designed and implemented, food aid based safety nets appear largely ineffective in either preventing people from falling into poverty traps or at lifting people out of poverty traps in places like northern Kenya. In addition, the amount of aid delivered does not appear to be particularly credible as a means to induce behavioral change that will help households grow themselves out of extreme poverty. Food aid may indeed contribute to better consumption outcomes and anthropometric status (Dercon and Krishnan 2003, Quisumbing 2003), but it largely fails to help move recipients out of chronic poverty.

5. CONCLUSION

In this chapter we have tried to outline the integration of risk management and poverty reduction issues, focusing especially on how the experience of and exposure to asset risk may trap certain households in chronic poverty and vulnerability. If there exist multiple dynamic equilibria of the sort conjectured by classical development theorists such as Myrdal, Nurkse, Rosenstein-Rodan and Young, then even minor perturbations to individual asset stocks and the prospect of significant, if transitory, income shortfalls can both have lasting welfare effects on subpopulations with limited initial endowments and scant or no access to credit and insurance. Drawing on original, high-frequency panel data from very poor pastoralist communities in northern Kenya, we find suggestive evidence both of multiple dynamic equilibria consistent with the hypothesis of poverty traps and of risk management patterns consistent with the notion that risk exposure and experience may play a significant role in trapping such populations in chronic poverty and vulnerability.

The normative prescriptions of minimalist neoclassical models generally fail in the presence of risk and uncertainty, especially where insurance market failures are compounded by credit market failures, so that people are unable to move consumption across periods in response to transitory shocks to assets or incomes, and where important irreversibilities break down intertemporal separability, thereby affecting risk management and asset and welfare dynamics. As a consequence, there exists a real public role in risk management for the poor (Ahmad et al. 1991, Dercon 2004).

Credible and ubiquitous safety nets can both (i) respect the human right to food and (ii) reduce downside risk, thereby inducing portfolio reallocation, raising the real prospect of income growth through capital accumulation and new technology adoption and market participation. In theory, effective safety nets can enlarge the basin of attraction toward higher dynamic equilibria and reduced vulnerability. For most of the poor, the key safety net lies in nutrition and health space because their most important irreversible asset is their human capital. The need for high frequency consumption to maintain health can necessitate asset divestiture in the face of sharp, transitory income shocks, causing short-lived shocks to have persistent, if not irreversible, effects. Moreover, awareness of crucial thresholds affects risk management, causing the poor to choose low-return production strategies and asset portfolios that merely reinforce their chronic poverty and vulnerability. Unfortunately, the record of safety net provision has been checkered, so we have limited empirical evidence as to how effective in practice safety nets can be in inducing the poor to choose higher-return livelihood strategies. This is an area ripe for intensive research.

New research in this area will need to take seriously, however, the centrality of asset risk – not just income risk – to chronic poverty and vulnerability. In particular, researchers need to start thinking about vulnerability relative to critical thresholds, not only relative to (inherently arbitrary) poverty lines. There's great inertia behind the use of poverty lines, but this doesn't seem like the most fruitful way to proceed. We need to begin establishing whether critical thresholds indeed exist and, if they do, to identify them with sufficient precision so that we can locate the appropriate level and trigger for safety nets. Because the key vulnerability is less with respect to a poverty line (although that is certainly important and informative) than with respect to the points at which welfare dynamics birfucate, research needs to focus on increasingly on these dynamics and on how to implement effective, credible safety nets to keep people from falling into chronic poverty.

ACKNOWLEDGEMENTS

We thank Alain de Janvry and Ravi Kanbur for their invitation to participate in this event, Erik Thorbecke for his ongoing inspiration to so many of us, Michael Carter, Stefan Dercon and Dan Maxwell for helpful discussions that have informed parts of this paper, Alok Bhargava, Jan

Willem Gunning, Martin Ravallion, an anonymous referee and conference participants for helpful comments, and Erin Lentz, Andrew Mude and Jacqueline Vanderpuye-Orgle for excellent research assistance. This work has been made possible, in part, by support from the United States Agency for International Development (USAID), through grant LAG-A-00-96-90016-00 through the BASIS CRSP, grants DAN-1328-G-00-0046-00 and PCE-G-98-00036-00 through the Pastoral Risk Management Project of the Global Livestock CRSP, and grant HFM-A-00-01-00132-00 to the Strategies and Analyses for Growth and Access (SAGA) cooperative agreement. The views expressed here and any remaining errors are mine and do not represent any official agency.

NOTES

[i] The papers in Dercon (forthcoming) offer a range of evidence on and vision of this rationale.

[ii] We can derive a back-of-the-envelope estimate of the monthly exit rate using the simple equation Poverty rate$_t$ = (1- exit rate)^t poverty rate$_0$, where t reflects the number of months since the initial observation. The reported back-of-the-envelope calculations are based on transition matrices reported in Grootaert and Kanbur (1993) for Côte d'Ivoire (with a poverty line equivalent to $1.12/day per capita), Naifeh (1998) for the United States (with a poverty line equivalent to $15.05/day per capita), and Carter and May (1999) for South Africa (with a poverty line equivalent to $2.23/day per capita). The Madagascar data reflect five-year panel data from sites in the central and southern highlands (Vakinankaratra and Fianarantsoa, respectively, with a poverty line equivalent to $0.50/day per capita). The northern Kenya data reflect quarterly data over two years from six sites (with a poverty line equivalent to $0.50/day per capita).

[iii] See Loury (1981), Romer (1986), Lucas (1988), Azariadis and Drazen (1990), Banerjee and Newman (1993), Galor and Zeira (1993), Durlauf (1996), Hoff and Sen (2002) and Mookherjee and Ray (2002).

[iv] See, for example, Dercon (1998), Barrett et al. (2001), Barrett et al. (2004), Carter and May (2001), Lybbert et al. (2002), Dercon and Hoddinott (2004).

[v] The survey methods and data are described in detail in Mude *et al.* (2003).

[vi] Tropical livestock units (TLU) standardize animals by species mean live weight, permitting aggregation across species and thus herd size comparisons across agroecologically dissimilar areas between which prevailing species herd composition vary. The standard weighting system, used here, is 1 TLU= 1 cattle = 0.7 camels = 10 goats = 11 sheep.

[vii] Expenditures follow a similarly shaped pattern, so this result is not specific to the welfare indicator used.

[viii] See Gersovitz (1983), Dasgupta (1993, 1997), Glomm and Palumbo (1993), Dasgupta (1997), Chavas (2000), Barrett (2002) and Zimmerman and Carter (2003) for richer discussions of the problems associated with undernutrition and irreversibilities (i.e., absorbing states and other forms of hysteresis) in human health dynamics and resulting intertemporally nonseparable preferences.

[ix] Liebenstein (1957) emphasized the economic importance of food biochemistry for labor productivity. See Dasgupta and Ray (1986, 1987), Dasgupta (1993), Ray and Streufert (1993), and Dasgupta (1997) on the details of the nutritional efficiency wage hypothesis and its relationship to poverty traps. Criticisms of this hypothesis include Bliss and Stern (1978), Rosenzweig (1988), and Strauss and Thomas (1995).

[x] The IFRCRCS (2002) reports that more than 98 percent of the people affected by different types of environmental (e.g., droughts, earthquakes, floods, avalanches) and technological (e.g., industrial or transport accidents) disasters worldwide, 1992-2001, lived in low and medium human development nations. Although airline crashes in the United States and deaths from severe heat waves in France capture the headlines, the overwhelming majority of shocks are experienced in the developing world.

[xi] Barrett (forthcoming) presents a simple formulation of income dynamics and draws a useful distinction between "cargo net" and "safety net" interventions intended to address different sources of intertemporal variability in welfare. The differences between alternative policies are developed in more detail there.

[xii] It is also critically important to identify and mitigate any adverse incentives that may result from the implementation of such safety nets.

[xiii] For example, in our three most arid sites (Kargi, Logologo and North Horr, respectively), cumulative rainfall between May 1999 and September 2000 was only 38-63 millimeters. The other three, semi-arid sites (Dirib Gombo, N'gambo and Suguta Marmar) had significantly higher total precipitation, but still well below annual or seasonal averages.

[xiv] The conditioning variable here is earned income (that is, excluding transfers). By way of reference, median quarterly income is just over KSh8700, the top quintile had quarterly income of KSh23500 or greater and the top five percent earned KSh50000 or more.

[xv] See Ravallion (1991, 1999), Ravallion et al. (1993), Besley and Coate (1992) and von Braun (1995) for detailed descriptions of the theory and evidence on public employment guarantee ("workfare") schemes.

REFERENCES

Azariadis, Costas and Allan Drazen (1990), "Threshold externalities and economic development," *Quarterly Journal of Economics* 105, 3: 501-526

Assefa, Mulugeta (1990). *Borana Cattle Herds: Productivity, constraints, and possible interventions.* Master's thesis Colorado State University, Fort Collins, CO.

Banerjee, Abhijit V. and Andrew F. Newman (1993), "Occupational Choice and the Process of Development," *Journal of Political Economy* 101, 2: 274-298.

Barrett, Christopher B. (2002), "Food Security and Food Assistance Programs," in B.L. Gardner and G.C. Rausser, eds., *Handbook of Agricultural Economics* (Amsterdam: Elsevier).

Barrett, Christopher B. (forthcoming), "Rural Poverty Dynamics: Development Policy Implications," *Agricultural Economics.*

Barrett, Christopher B.; Mesfin Bezuneh, and Abdillahi A. Aboud (2001), "Income Diversification, Poverty Traps and Policy Shocks in Côte d'Ivoire and Kenya" *Food Policy*, 26, 4: 367-384.

Barrett, Christopher B. and Daniel C. Clay (2003), "Self-Targeting Accuracy in the Presence of Imperfect Factor Markets: Evidence from Food-for-Work in Ethiopia," *Journal of Development Studies*, 39, 5: 152-180.

Barrett, Christopher B., Paswel P. Marenya, John G. McPeak, Bart Minten, Festus M. Murithi, Willis Oluoch-Kosura, Frank Place, Jean Claude Randrianarisoa, Jhon Rasambainarivo and Justine Wangila (2004), "Welfare Dynamics in Rural Kenya and Madagascar," Cornell University working paper.

Barrett, Christopher B. and Daniel G. Maxwell (forthcoming), *Food Aid After Fifty Years: Recasting Its Role.* London: Routledge.

Baulch, Bob and John Hoddinott, eds. (2000), *Economic Mobility and Poverty Dynamics in Developing Countries.* London: Frank Cass.

Besley, Timothy and Stephen Coate (1992), 'Workfare vs. Welfare: Incentive Arguments for Work Requirements in Poverty Alleviation Programs', *American Economic Review* 82, 2: 249–61.

Bhargava, Alok and Martin Ravallion (1993), "Does Household Consumption Behave As A Martingale? A Test For Rural South India," *Review of Economics and Statistics* 75, 3: 500-504.

Carter, Michael R. and Julian May (1999), "Poverty, Livelihood and Class in Rural South Africa," *World Development* 27, 1: 1-20.

Chaudhuri, Shubham (2001), "Empirical methods for assessing household vulnerability to poverty," Columbia University working paper.

Chaudhuri, Shubham; Jyotsna Jalan and Asep Suryahadi (2002), "Assessing household vulnerability to poverty from cross-sectional data: A methodology and estimates from Indonesia" Columbia University working paper.

Chavas, Jean Paul (2000), "The microeconomics of food security," *Australian Journal of Agricultural and Resource Economics* 44, 1: 1-30.

Christiansen, Luc and Richard Boisvert (2000), "On measuring household food vulnerability: Case evidence from northern Mali," Cornell University working paper.

Dasgupta, Partha (1993), *An Inquiry Into Well-Being and Destitution* (Oxford: Oxford University Press).

Dasgupta, Partha (1997), "Nutritional status, the capacity for work, and poverty traps," *Journal of Econometrics* 77 (1): 5-37.

Deaton, Angus (1992), "Household Saving in LDCs: Credit Markets, Insurance and Welfare," *Scandinavian Journal of Economics* 94, 2: 253-273.

Dercon, Stefan, ed., (2004), *Insurance Against Poverty* (Oxford University Press).

Dercon, Stefan (1998), "Wealth, risk and activity choice: cattle in Western Tanzania," *Journal of Development Economics*, 55, 1: 1–42.

Dercon, Stefan and John Hoddinott (2004), "Health, Shocks and Poverty Persistence," in S. Dercon, ed., *Insurance Against Poverty* (Oxford University Press).

Dercon, Stefan and Pramila Krishnan (2003), "Food Aid and Informal Insurance," University of Oxford Centre for the Study of African Economies working paper.

Durlauf, Steven (1996), "A Theory of Persistent Income Inequality," *Journal of Economic Growth* 1, 1: 75-93.

Elbers, Chris and Jan-Willem Gunning (2003), "Estimating Vulnerability," Free University of Amsterdam working paper.

Foster, James; Joel Greer and Erik Thorbecke (1984), "A class of decomposable poverty measures," *Econometrica* 52 (3): 761-766.

Fratkin, Elliot and Eric Roth. (1990). Drought and Economic Differentiation among Ariaal Pastoralists of Kenya. *Human Ecology.* 18(4): 385-402.

Friedman, Milton (1957), *A Theory of the Consumption Function.* Princeton: Princeton University Press.

Galor, Oded and Joseph Zeira (1993), "Income distribution and macroeconomics," *Review of Economic Studies* 60, 1: 35-52.

Glewwe, Paul; Hanan Jacoby, and Elizabeth King (2001), "Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis," *Journal of Public Economics* 81, 3: 345-368.

Glomm, Gerhard and Michael G. Palumbo (1993), "Optimal intertemporal consumption decisions under the threat of starvation," *Journal of Development Economics* 42, 2: 271-291.

Grootaert, Christian and Ravi Kanbur (1995), "The Lucky Few amidst Economic Decline: Distributional Change in Côte d'Ivoire as Seen through Panel Data Sets, 1985-88," *Journal of Development Studies* 31, 4: 603-19.

Hoddinott, John and Bill Kinsey (2001), "Child Health in the Time of Drought," *Oxford Bulletin of Economics and Statistics* 63, 3: 409-436.

Hoff, Karla and Amartya Sen (2002), "A Simple Theory of the Extended Family System and Market Barriers to the Poor" World Bank working paper.

International Federation of the Red Cross and Red Crescent Societies (2002), *World Disasters Report 2002*. Geneva: IFRCRCS.

Koopmans, Tjalling C. (1957), *Three Essays on the State of Economic Science*. New York: McGraw-Hill.

Ligon, Ethan and Laura Schechter (2002), "Measuring Vulnerability," University of California at Berkeley working paper.

Loury, Glenn C. (1981), "Intergenerational Transfers and the Distribution of Earnings," *Econometrica* 49, 4: 843-867.

Lucas, Robert (1988), "On the mechanics of economic development," *Journal of Monetary Economics*, 22, 1: 2-42.

Lybbert, Travis J.; Christopher B. Barrett, Solomon Desta and D.Layne Coppock (2004), "Stochastic Wealth Dynamics and Risk Management Among A Poor Population," *Economic Journal.*

Martorell, Reynoldo (1999), "The nature of child malnutrition and its long-term implications," *Food and Nutrition Bulletin* 20, 2: 288-292.

McPeak, John G. (2004), "Contrasting income shocks with asset shocks: Livestock sales in northern Kenya," *Oxford Economic Papers*, 56: 263-284.

McPeak, John G. and Christopher B. Barrett (2001), "Differential Risk Exposure and Stochastic Poverty Traps Among East African Pastoralists," *American Journal of Agricultural Economics*, 83: 674-679.

Mookherjee, Dilip and Debraj Ray (2002), "Contractual Structure and Wealth Accumulation," *American Economic Review* 92, 4: 818-849.

Mude, Andrew; John McPeak, Jacqueline Vanderpuye-Orgle, Christopher B. Barrett, Amare Yirbecho, Getachew Gebru, and Erin Lentz. (2003). Codebook for Data Collected Under the USAID Global Livestock Collaborative Research Support Program "Improving Pastoral Risk Management on East African Rangelands". Cornell University mimeo.

Naifeh, Mary (1998), "Dynamics of Well-Being, Poverty 1993-94: Trap Door? Revolving Door? Or Both?" Current Population Reports, Household Economic Studies. U.S. Census Bureau, Washington, D.C.

Pritchett, Lant; Asep Suryahadi and Sudarno Sumarto (2000), "Quantifying vulnerability to poverty: A proposed measure, with application to Indonesia," SMERU working paper.

Ravallion, Martin (1991), 'Reaching the Rural Poor through Public Employment: Arguments, Lessons, and Evidence from South Asia', *World Bank Research Observer* 6, 1: 153–76.

Ravallion, Martin (1999), 'Appraising Workfare' *World Bank Research Observer* 14, 1: 31–48.

Ravallion, Martin; Gaurav Datt and Subham Chaudhuri (1993), 'Does Maharashtra's Employment Guarantee Scheme Guarantee Employment? Effects of the 1988 Wage Increase', *Economic Development and Cultural Change* 41, 2: 251–75.

Ray, Debraj and Peter Streufert (1993), "Dynamic equilibria with unemployment due to undernourishment," *Economic Theory* 3, 1: 61-85.

Romer, Paul (1986), "Increasing returns and long run growth," *Journal of Political Economy* 94, 4: 1002-37.

Smith, Kevin; Christopher B. Barrett, and Paul W. Box (2001), "Not Necessarily in the Same Boat: Heterogeneous Risk Assessment Among East African Pastoralists," *Journal of Development Studies*, 37, 5: 1-30.

Solow, Robert (1956), "A contribution to the theory of economic growth," *Quarterly Journal of Economics*, 70, 1: 65-94.

Strauss, John and Duncan Thomas (1997), "Health, Nutrition and Economic Development," *Journal of Economic Literature* 36, 2: 766-817.

Upton, Martin (1986). Production Policies for Pastoralists: The Borana Case" *Agricultural Systems* 20:17-35

von Braun, Joachim (ed.) (1995), *Employment for Poverty Reduction and Food Security*, International Food Policy Research Institute, Washington DC.

Zimmerman, Frederick and Michael R. Carter (2003), "Asset smoothing, consumption smoothing and the reproduction of inequality under risk and subsistence constraints," *Journal of Development Economics* 71, 2: 233-260.

Chapter 9

PROGRESS IN THE MODELING OF RURAL HOUSEHOLDS' BEHAVIOR UNDER MARKET FAILURES

Alain de Janvry and Elisabeth Sadoulet[i]
University of California, Berkeley, alain@are.Berkeley.edu and sadoulet@are.Berkeley.edu

1. HOUSEHOLD BEHAVIOR UNDER MARKET FAILURES

It is well recognized that, in the developing country context, rural households are systematically exposed to market imperfections and constraints, referred to as "failures", and their behavior cannot be understood without reference to the specificity of these failures (Thorbecke, 1993). In some cases, markets do not even exist. In others, high transactions costs must be incurred in accessing markets. In yet others, there are constraints on the quantities that can be exchanged. Market failures are so pervasive for farm households that they have been used as a definitional characteristic of peasantries (Ellis, 1993). In this context, key to the analysis of peasant household behavior is to identify the resource allocation, consumption, investment, and exchange strategies that they devise to reduce the welfare costs of these failures. These countervailing strategies demonstrate peasants' considerable creativity in attempting to derive maximum benefit from the meager resources they control in a particularly adverse context. Indeed, there exists a booming academic industry consisting in marveling about this "creativity in the context of adversity".

Modeling rural household behavior in the context of market failures implies non-separability between production and consumption decisions. This class of non-separable models was first introduced in Singh, Squire, and Strauss's (1986) seminal book. It has been followed by an explosion of efforts in theory and empirical analysis to characterize the behavior of farm households. What did we learn from these models? How have they helped the specification of empirical strategies? How far have we progressed in measuring the transactions costs that are the causes of market failures? And what contributions have they made to the formulation of policies and programs for rural development and the struggle against rural poverty? It is to these questions that we try to answer in this paper on progress in rural household modeling. Interesting, in particular, is to try to assess in how much is peasant ingenuity in devising countervailing strategies able to shelter them from the welfare costs of market failures. We will see that many of these questions are still poorly addressed, leaving us with a rich research agenda, and some promising departures toward new answers.

In this review, we briefly present the conceptual framework used in formulating non-separable models. We then review many of the results obtained in using these models to understand farm household behavior, design policies and programs, define strategies for empirical analysis, and measure the transactions costs that are the causes of market failures. Finally, we identify new research initiatives that show promise in better explaining and measuring household behavior under market failures.

2. CONCEPTUAL FRAMEWORK

2.1 The concept of non-separability

In terms of model specification, presence of market failures leads to what has been called non-separability (Yotopoulos and Lau, 1974). A household model is said to be non-separable when the household's decisions regarding production (use of inputs, choice of activities, desired production levels) are affected by its consumer characteristics (consumption preferences, demographic composition, etc.). By contrast, in a separable model, the household behaves as a pure profit maximizing producer. The profit level achieved in turn affects consumption, but without feedback on production decisions.

Consider for example the case of variable transactions costs on the market for a food product. Take as a second market failure inexistence of a land market. In Figure 9-1, we represent the supplies of food, $S(p, z_{qi})$, coming from three households, $i = 1, 2, 3$, that own farms of different sizes z_{qi}. To facilitate comparison across households, we assume that they all have the same demand for food, $D(p, z_c)$, which depends on the characteristics z_c of the household as a consumer. Let p^v be the effective sale price (i.e., the market price p^m net of transactions costs t_p^v incurred in selling) and p^a the effective purchase price (that includes the transactions costs t_p^a incurred in buying) of the food product.

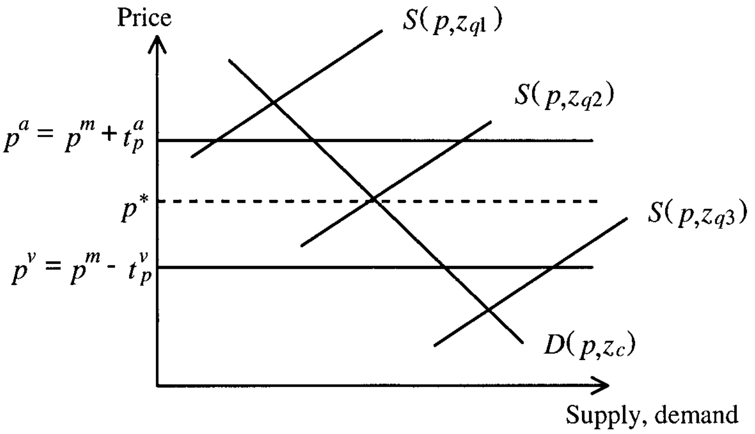

Figure 9-1. [Variable transactions costs and market participation]

This simple graph shows that the decision to participate in the market depends on the relative position of the household's supply and demand functions, and hence on its endowments in productive resources z_q and on its demand characteristics z_c. Because of transactions costs, there exists a non-zero price interval where households do not participate in the market. For these households (of type z_{q2} in the figure), it is optimum to remain in self-sufficiency and to adjust production and consumption decisions to each others. Their behavior is, consequently, of the non-separable type, and their internal equilibrium defines a shadow price $p^*(z_q, z_c)$ specific to each of them. Hence, to the heterogeneity in household resource endowments corresponds a heterogeneity in market participation decisions. A second source of heterogeneity can come from differences in transactions costs t_p^a and t_p^v across households.

In a similar fashion, variable transactions costs on the labor market induce a category of households to opt for an autarkic equilibrium with a shadow price of labor that breaks separability between their consumption and production decisions (Lopez, 1984).

As to fixed transactions costs, they enter in the household model as follows. Considering the case of a single market price (i.e., a case where there are no variable transactions costs), the relative positions of the household's supply and demand curves at that price determine the marketed surplus (which is negative in case of a purchase). When entering the market implies fixed costs, participation will only be preferred to autarky if the value of sales or purchases is sufficient to induce a gain for the household larger than the fixed costs incurred, which defines a minimum level of exchange MS_m to make the transaction worthwhile:

$$\left|q(p, z_q) - c(p, z_c)\right| \geq MS_m.$$

These minimum levels of sale or purchase determine again a category of households -- defined by their asset endowments, their preferences, and technology -- that choose autarky. They, too, will have a non-separable behavior. If both variable and fixed transactions costs exist, presence of fixed costs widens the price range over which household autarky is observed. While transactions costs may not be the only reason for food self-sufficiency, they are likely to be a major determinant of such behavior.

The third category of market failure, the case of a constraint on participation, is easy to understand. Once the maximum level of participation has been reached, the household must solve its internal equilibrium problem between residual supply and demand. This equilibrium defines a shadow price which is a function of the household's resource endowments, its characteristics in demand, and the level of the constraint. In this case as well, the choice problem consists in selecting an idiosyncratic non-separable equilibrium.

Lack of an insurance market, in combination with a market constraint or an imperfection on the credit market, induces households to manage their production decisions to reduce their consumption risk. Indeed, if households do not have access to complete insurance or credit mechanisms that allow them to smooth their consumption ex-post relative to a shock, they will adjust their income generation strategy to reduce income fluctuations (Alderman and Paxson, 1994). This adjustment generally implies a bias toward activities or technologies that are less risky, a greater diversification of income sources, in particular toward off-farm incomes that are less risky than farm incomes or with low covariation with those, storage behavior that accounts for food security considerations (Renkow, 1990; Saha and Stroud, 1994), and lower levels of investment, in particular in soil conservation (Holden, Shiferaw, and Wik, 1998; Shively, 2001).

2.2 Outline of a non-separable household model

Formalizing the behavior of a household under market failure requires specification of the particular types of failures to which the household is confronted. However, most of the commonly encountered market failures can be represented in the following generic model. The household maximizes the expected present value of a stream of utilities (1) under constraints (2) to (8) as follows:

$$\max_{c_t, q_t, x_t, m_t, s_t} E\sum_t \beta^t u(c_t; z_c) \qquad (1)$$

$$\sum_{i=1}^{N}\left[\left(\left(p_{it}^{m} - t_{pit}^{v}\right)\delta_{it}^{v} + \left(p_{it}^{m} + t_{pit}^{a}\right)\delta_{it}^{a}\right)m_{it} - t_{fit}^{v}\delta_{it}^{v} - t_{fit}^{a}\delta_{it}^{a}\right] + T_t + s_t = 0, \quad \forall t, \qquad (2)$$

$$q_{it} - x_{it} + E_{it} - m_{it} - c_{it} = 0, \quad i = 1 \ldots N, \quad \forall t, \qquad (3)$$

$$m_{kt} \leq \overline{M}_{kt}, \quad k \in K, \quad \forall t, \qquad (4)$$

$$G(q_t, x_t; z_{qt}) = 0, \quad \forall t, \qquad (5)$$

$$A_{t+1} = (1 + r_t)(A_t + s_t) \quad \forall t, \qquad (6)$$

$$A_t \geq A_{\min}, \quad \forall t, \qquad (7)$$

$$c_{it}, q_{it}, x_{it} \geq 0, \quad i = 1 \ldots N, \quad \forall t. \qquad (8)$$

In this model, the decision variables are the consumption vector c_t, the production vector q_t,[ii] the input vector x_t, the vector of marketed surpluses m_t (which is negative in case of a purchase), and the vector of savings s_t (which are negative in case of borrowing), for each year t. In equation (1), β is the rate of discount, and u the utility in period t, which is a function of consumption and of the characteristics z_c of the household's preferences. In the budgetary constraint for year t (equation 2), for each of the N goods or factors i, sales are characterized by $\delta^v = 1$ and $\delta^a = 0$ and purchases by $\delta^v = 0$ and $\delta^a = 1$. The sales price effectively received by a household is the market price p^m net of variable transactions costs t_p^v (which are proportional to the quantity exchanged), and the purchase price effectively paid is the market price net of variable transactions costs t_p^a. In addition, the household incurs fixed transactions costs t_f^v or t_f^a when he participates in the market. T represents exogenous sources of revenue, and s savings. Equation (3) specifies the equilibrium between availability and use of each of these products or factors, including the initial endowment E. Participation constraints to some markets can be written as inequalities (4), where K defines the set of goods that are constrained and M_{kt} the values of the constraints. Equation (5) represents the technology that links production, inputs, and fixed production factors z_q. Equation (6) specifies the law of motion of the (unproductive) asset A, as a function of the yield r_t and annual savings s_t. Inequality (7) specifies the constraint on the level of indebtedness.

This formulation thus captures both variable and fixed transactions costs (in equation (2)), risk aversion behavior (in the objective function (1)), participation constraints in some markets (4), and the credit constraint (7). Use of such a general model is, however, difficult and inefficient. For this reason, we will see that the analysis of specific combinations of market failures is better done in specialized models.

3. IMPLICATIONS OF NON-SEPARABILITY: BEHAVIOR AND POLICIES

3.1 Implications of market failures on transactions in non-failing markets

In the context of non-separability, the household's ability to respond to production incentives in a market without failures is affected by failures in other markets. This allows to explain aspects of household behavior that would otherwise appear irrational from an economic perspective. This is what led anthropologists from the substantivist school to reject the possibility of analyzing peasant behavior on the basis of economic rationality, calling instead on the role of rituals in exchanges or on an objective of simple reproduction of traditional needs (Polanyi et al., 1958). Construction of household models with failures on food, labor, or manufactured goods markets allows to explain these same behavioral patterns in the context of economic calculus. It is the incomplete performance of markets that induces patterns of behavior apparently contrary to economic logic, not a logic specific to peasant households that would remain to be uncovered.

Take for example the case of a farm household that produces food and cash crops and that faces two market failures, one in the food market and the other in the labor market (de Janvry, Fafchamps, and Sadoulet, 1991). An anthropologist from the substantivist school would observe with reason that a rise in the price of cash crops does not induce any notable response in the production of the latter. However, this lack of response does not come from absence of economic rationality. It derives first from lack of a food market which prevents the household from reallocating its land toward cash crops, in a context where income rose and also the desire to consume more food; and second from lack of a labor market that would allow to employ workers from outside in a context where the desire for additional leisure refrains increasing time worked by the household. Supply response in cash crops cannot come from a decline in food production nor from an increase in family labor. It is consequently confined to technological change, as observed in Guatemala by von Braun et al. (1989), or to an increase in the use of purchased inputs such as fertilizers that are partial substitutes to land and labor. It is consequently not surprising that peasant households' responses to price incentives in cash crops are low under these failures in other markets, to the despair of governments. Taylor and Adelman (2002) use a similar specification to analyze the impact of Mexican trade (NAFTA) and transfer (Procampo) policies on household behavior when labor and food markets may be missing. They find, not surprisingly, that these policy shocks under markets failures have, contrary to expectations of policy makers, remarkably small impacts on production and rural incomes. Embedding household models in a Computable General Equilibrium model, thus providing a rigorous micro-macro linkage, brings these results to the aggregate level. Löfgren and Robinson (2002) show that large transactions costs and the resulting regime switches in market participation by households in response to price and productivity changes create low aggregate response and discontinuities that differ markedly from the smooth responses with separable household models. Their results stress the importance of transactions costs in determining the aggregate gains from increases in the price of cash crops on international markets.

The same phenomenon occurs when it is the market for manufactured consumption goods that is failing. This was the case not only in the Eastern European economies, but also under the policy of industrialization by import substitution that endeared the price of industrial consumption goods in much of the developing world. In this case, forced savings create a disincentive to production of cash crops, resulting in an inelastic supply of these crops (Berthélémy and Morrisson, 1987; Azam and Besley, 1991).

Policy implications from these cross-market effects can be counter-intuitive. Technological change in the production of food crops helps increase the elasticity of supply response in cash crops. In the Sahelian zone of Burkina Faso, we showed that technological change in water harvesting for food crops helps increase the supply response in livestock which is the cash activity for these peasant households (Dutilly-Diane, Sadoulet, and de Janvry, 2003). A decline in the price of manufactured consumption goods, for instance through trade liberalization when there were import tariffs, creates incentives to the production of cash crops. Conceptualizing household behavior in non-separable models that take into account cross effects between failing markets (food, labor, manufactured consumption goods) and non-failing markets (cash crops, fertilizers) opens perspectives for the interpretation of household behavior and for the choice of policy instruments to use in raising their responsiveness to market incentives.

3.2 Nutrition, health, and productivity

Nutrition and health, like other dimensions of human capital, influence time worked and the productivity of labor, and hence also households' incomes and poverty. However, it is only when a household faces certain types of market failures that production decisions are affected by its own human capital endowment. This will be the case, for example, when health affects management capacity, a direct input in production that is not accessible through the market, or when family and hired labor are imperfect substitutes in production and health affects the wage received by efficiency unit (piece rates). This source of non-separability has been tested and rejected by Pitt and Rosenzweig (1986) and by Deolalikar (1988), while Strauss (1986) shows strong evidence of the positive effect of caloric intake on farm labor productivity. In a dynamic multi-stage household model, Behrman, Foster and Rosenzweig (1997) show evidence of differential productivity effects of calorie consumption at the harvest and planting stage in a sample of households from Pakistan. This suggests that farmers face large costs of transferring resources across stages, implying that improving the operation of the credit market would increase the productivity of small farmers.

3.3 Intensity of factor use and the inverse relation between yield and farm size

If there exists one relation that has captivated the imagination of development economists, it is existence of an inverse relation between yield and farm size. While empirical evidence is far from universal, this relation has been observed in a multiplicity of contexts of traditional agriculture (Berry and Cline, 1979; Carter, 1984; Benjamin, 1995; Barrett, 1996; Lamb, 2003). From an economic policy perspective, implications of this relation are enormous in as much as it helps justify redistributive land reform in terms of efficiency gains, in addition to the obvious equity gains.

From a theoretical point of view, this relation has been attributed to presumption that the opportunity cost of family labor working on the farm is less than the prevailing wage (Barrett, 1996, points to an alternative explanation due to the presence of price risk). Smaller farms thus rationally use a production process that is more labor intensive and, in traditional agriculture where labor is the main variable input, obtain higher yields than larger farms that use hired labor. What household models under market failures contribute is a rigorous conceptualization of this hypothesis based on the differential opportunity cost of labor across farm sizes, and a broadening of the causes that can result in this inverse relation.

To focus on the fundamental role of labor allocation in a farm household, the consumption side of the household is reduced to a trade off between income and leisure. This is done by replacing the objective function in equation (1) by an additive utility in income and the utility for leisure. Transactions costs on labor come from the need to supervise hired labor. In Eswaran and Kotwal's (1986) model, the second market failure is an access to working capital that is proportional to productive assets needed as collateral. The combination of these two market failures leads to a differentiation across farm households, starting from the poorest who do not have enough resources to even make productive use of their own land, through small part time farmers that participate in the labor market, family farms that are self-sufficient in labor, up to large farms that hire labor that they must supervise. Along this gradient of asset endowments, that define farm sizes, one observes a rising labor cost, and hence an inverse relation between yield and farm size. It is important to recall that the mere existence of supervision costs would not be sufficient to produce an inverse relation (Feder, 1985). Indeed, if there were no other market failure, one should see farm sizes adjusting to labor availability.

It is obvious that empirical verification of this relation requires controlling for diseconomies of scale due to production technology and land quality. While there exists some empirical evidence both on existence of this relation and of supervision costs and other imperfections on the labor market (Frisvold, 1994), the link between these two phenomena is hard to establish. Yet, it would be important to know what part of this inverse relation is due to labor market imperfections, as opposed to technological choice, land quality, or measurement errors for example (Carter, 1984; Benjamin, 1995; Lamb, 2003).

3.4 Management of the intra-annual credit constraint

How a credit constraint is managed should be analyzed in a model that endogenizes savings and credit. However, if one wants to concentrate on the problem of intra-annual credit mainly linked to the seasonality of agricultural costs and incomes, leaving aside considerations of risk and investment, a static annual model reveals a number of interesting phenomena.

The following model is simplified in only considering one period *t*, and in replacing the savings determination equations (6) and (7) by a static borrowing constraint, for example by setting savings s_t equal to zero. A simple way of formalizing the liquidity constraint for the lean season is to consider the inputs and the goods produced and consumed during the two seasons as different commodities. Let I_k be the set of factors and production or consumption goods during the lean season. Monetary transactions on these products and inputs are constrained by availability of financial liquidity:

$$(4) \qquad \sum_{i \in I_k} \left[\left(p_i^m - t_{pi}^v \right) \delta_i^v + \left(p_i^m + t_{pi}^a \right) \delta_i^a \right] n_i \leq \bar{L},$$

where $\bar{L}$ represents the liquidity available at the beginning of the period. Constrained optimization leads to the definition of a shadow price for liquidity, λ_c, and to what can be called decision prices, equal to transactions prices increased by the implicit value of liquidity:

$$\left(p_i^m - t_{pi}^v \right)\left(1 + \lambda_c \right) \text{ for sales, and } \left(p_i^m + t_{pi}^a \right)\left(1 + \lambda_c \right) \text{ for purchase.}$$

Hence, even though transactions are done at market prices net of transactions costs, decisions are taken using higher prices. In these conditions, the household will bias its decisions in favor of activities that generate (for sales) or save (for purchases) liquidity. Using a simulation model calibrated for Moroccan households, we show that the credit constraint reduces the capacity of households to take advantage of opportunities offered by a rise in the price of cereals as a consequence of trade liberalization policies, and increases family labor participation in the labor market to generate a flow of liquidity necessary for the agricultural activities (de Janvry, Fafchamps, Raki, and Sadoulet, 1992). These results stress the importance of accompanying the incentive policies for agricultural production with a complementary policy of access to credit for working capital in order to derive full benefit from the new incentives.

Feder, Lau, Lin, and Luo (1990) verified empirically the relevance of this model for a sample of Chinese households. They show that the liquidity available at the beginning of the season affects the level of production of households that are constrained in accessing credit, but not that of other households. Vakis (2002) shows that while credit constraints dramatically decrease adoption rates of potato high yielding varieties in Peru, farmers alleviate these constraints by diversifying their source of through inclusion of cash generating activities such as milk production. Note that identification of the regime to which households belong cannot be done on the basis of the mere observation that households receive credit or not: households that receive credit may be constrained or not on the amount received, and households that do not receive any credit may be excluded by available supply or be on their demand functions. Feder et al. are in a favorable situation where existence of a constraint was made explicit in the survey. In other situations, estimation of the regime to which households belong has to be done on the basis of observed behavior (Carter and Olinto, 2003).

The intra-seasonal specificity of this theory is better captured in a model that explicitly considers the asymmetry between the two seasons, as done by Key (2000). He shows that, when households are constrained on the credit market, the opportunity cost of liquidity is given by the conditions under which savings are made. In terms of economic policy for rural households, this result points to the importance of improving the institutions and instruments used for savings, and not only access to credit.

3.5 Response to food price risk through self-sufficiency

Under insurance market failures, a quasi universal circumstance for small holders, a household subject to a credit market constraint is unable to perfectly smooth its consumption. It is consequently led to use additional instruments, in particular in adjusting its income generation strategy and production decisions. Analysis of the household's adjustment to this situation, including optimal management of savings-credit decisions and production decisions, requires an inter-temporal model. Nevertheless, a number of intuitions about adjustments in production behavior can be derived from a simplified model that does not include optimization on behavior toward savings-credit. One can use a static model without inter-period transfers, like the one in the previous section. The household allocates its resources in production to optimally manage a trade-off between income level and variability. This static approach likely exaggerates the need for adjustment in production since it forces all risk management on a single instrument.

A particularly interesting case in the context of a household model is that of response to food price risk. Under food price risk, portfolio theory predicts that a risk averse producer will reduce production. Consider, however, a farm household that consumes all or part of its food production. And only consider those households that participate in the food market, as sellers or buyers, since they are the only ones subject to price risk on this market. Finkelshtain and Chalfant (1991) show that, in response to price risk, these farm households reduce their food production, but less than would a producer who is not also a consumer. And this all the more that their share of home consumption in production is high and that they are risk averse. Intuition for this result is as follows: the fact of producing food with a price that fluctuates creates a positive correlation between income (imputed for home consumption) and the price of the consumption good. This correlation "protects" the consumer from price fluctuations in as much as income rises and falls along with the price of the consumption good. Food production thus acquires an insurance value, additional to its normal contribution to income. This additional marginal value induces a bias in resource allocation toward food production, thus partially correcting the negative impact of risk on production. This reasoning rationalizes a frequent observation like what farm households perceive food self-sufficiency as a source of protection against price risks in food markets. Fafchamps (1992) develops this model when there are two competing crops, a cash crop and a food crop, with risks on both prices and yields. He shows that the rise in food production at the cost of cash crop production is all the more important that price risk is high and that correlation between price and yield is high, a phenomenon that is accentuated by market segmentation. The same reasoning on self-insurance applies to production of feed crops that serve as inputs into other activities. Note, however, that this self-insurance via the home production of consumption goods has an efficiency cost in as much as it creates a distortion in resource allocation toward food production. Models of storage motivated by risk aversion and considerations of food security have been developed by Renkow (2000) and Saha and Stroud (1994).

These analyses highlight the importance of policies to reduce transactions costs and to promote market integration in order to lower price fluctuations and the correlation between prices and local production. They also show the role of a "consumption" credit system that would free households from the need to self-insure their subsistence needs.

In spite of its conceptual importance, this approach has not been accompanied by systematic empirical analyses. Is the phenomenon quantitatively important? Is insurance through food self-sufficiency only of marginal importance relative to other determinants of food production? One would like to have answers to these questions. Difficulty with empirical analysis consists in the need to jointly estimate production and consumption behavior. In a recent analysis on a sample of villages in Pakistan, Kurosaki and Fafchamps (2002) estimated this type of structural model and showed that price risk on feed in itself reduces by 20% the area planted in basmati rice, resulting in a 9% welfare loss. If these results are confirmed in other studies, it will become clear that the good performance of markets as politically sensitive as those for food and feed should have high priority in economic policy.

3.6 Risk management in a context of insurance and credit market failures

As indicated above, inter-temporal modeling is needed to analyze responses to risk using credit and savings as instruments. Under insurance market failure, responding to risk calls upon two types of mechanisms. Mechanisms to protect (or smooth) consumption at a given level of income, and mechanisms to reduce exposure to risk through adjustments in income strategies. Consumption smoothing is done through insurance (mutual insurance in particular), and through credit and savings. Income strategies to reduce risk include contracts (e.g., sharecropping), the choice of activities, and production decisions. These two dimensions of risk-reducing strategies are not independent. Indeed, in as much as adjustments in production imply efficiency costs, optimal use of these risk management instruments depends importantly on possibilities and costs of ex-post smoothing.

We do not consider in this paper the literature on consumption smoothing (and the credit and savings instruments used) where implications on production decisions are not taken into account (see Deaton, 1992). This excludes, in particular, the very important literature on the consequences of market failures in credit and insurance on nutrition and health. We do not consider either the literature on contracts, even though risk management through contracts such as sharecropping influences production decisions (see Hayami and Otsuka, 1993). It calls on a conceptual framework that does not directly belong to household modeling. We consequently limit ourselves to analyses that explicitly link risk management to production decisions in the context of a household model.

In these inter-temporal analyses, the basic model is generally simplified by only considering one consumption good, one production good, and neither transactions costs nor constraints on markets for these goods. Thus, equations (2) to (4) in the general model are replaced by:

$$(2') \qquad p_{ct}c_t + s_t = \pi_t + w_t L_t, \quad \forall t,$$

where π_t represents profit in period t, and $w_t L_t$ potential income to family labor L_t paid at the wage w_t. Production behavior (5) is directly specified in a profit function:

$$(5') \qquad \pi_t = \pi\left(p_t, A_t, \varepsilon_t, d; z_q\right) \quad \forall t,$$

where p_t represents the price of the production good, A_t productive assets, and ε_t risk. The decision variable d is added in an ad-hoc fashion to indicate the possibility of choice of an income strategy (choice of technology, level of diversification in crops, etc.).

In this model (composed of equations (1), (2'), (5'), (6), and (7)), one can distinguish two processes that link risk management to the production and income strategy. One comes from the use of productive assets (A) as savings instruments, and the other from the choice of productive resources and labor allocation in order to reduce income risk, i.e., the choice of d in the profit function.

An important feature of rural households' saving behavior is that nearly all the assets they own contribute directly to production. However, only assets that can easily be transformed into cash can additionally provide insurance services. This insurance function gives them an additional value and induces a distortion in the optimum assets portfolio relative to a portfolio for expected profit maximization. Use of a productive asset for insurance purposes has a double implication for income generation. On the one hand, its insurance function induces the household to over-invest in this liquid asset; on the other hand, in as much as accumulation and decapitalization follow the needs for insurance, households that have suffered a series of negative shocks are under-capitalized relative to the optimum for production. These shocks can push the household into a vicious circle of decapitalization, higher risk, and poverty traps.

This differential insurance function of assets is illustrated in a pioneering article by Rosenzweig and Wolpin (1993) where they contrast investments in oxen and irrigation pumps in India's ICRISAT villages. Using the same information, Fafchamps and Pender (1997) show that a potentially highly profitable investment in irrigation pumps is made both less attractive due to its irreversibility and more difficult because of its indivisibility. In this context, a policy of investment subsidies would have little impact, when access to credit (for insurance) would induce a strong increase in investment. Dercon (1998) uses a similar model to show that poor households specialize in low return, low risk activities as imperfect credit market limit their entry into the lumpy and profitable investment in cattle.

In the context of insurance market failures, households are induced to follow less risky income strategies. Morduch (1990) thus showed that credit constrained households adopt a more diversified crops portfolio and use less high yielding varieties that are more profitable but also more risky. Rosenzweig and Binswanger (1993) show that the degree of adjustment that households make in their investment decisions in response to risk depends on their ability to smooth consumption. The contribution of their analysis consists in the identification of two different aspects of the relation between wealth and portfolio risk. On the one hand, wealth can have a direct impact on portfolio choice, in as much as richer households are less risk averse. They can consequently choose a more risky assets portfolio. On the other hand, and this is the phenomenon we are interested in here, wealth that also indicates access to ex-post consumption smoothing instruments, enables households to engage in less ex-ante risk management. This relation is established by showing that the link between climate risk and portfolio choice declines with household wealth. Recognizing that the agricultural cycle extends over the year, Fafchamps (1993) estimates a production model with progressive revelation of uncertainties and sequential decision-making. He shows that poor farm households in Burkina Faso respond to climate risk by choosing flexible production techniques that enable them to adjust their decisions in response to random events.

These arguments are, in essence, present in Eswaran and Kotwal's (1989) paper where they use an extreme form of liquidity contrast in considering that savings can only take the form of non-productive monetary reserves or of irreversible productive investments. They use a two-periods model to show that development of consumption credit as an instrument for insurance would facilitate adoption of new technologies for two reasons: first by releasing for investment resources otherwise held to respond to shocks; and second by allowing households to assume higher risks. Thus, risk aversion and the need for self-insurance affect the structure of capital accumulation, and they force rural households into long term income generation strategies.

4. GUIDE FOR EMPIRICAL ANALYSIS STRATEGIES

4.1 The first tests of non-separability

The concepts of a household's subjective equilibrium and of the shadow price of family labor derive from Chayanov's (1925) analyses of peasant households' time allocation between production and leisure. Through the

1950's and 1960's, this analysis of household behavior under market failures has been revisited, formalized, and expanded, notably by Nakajima (1970), to culminate in Singh, Squire, and Strauss' edited book (1986). This collection of articles, considered as the fundamental reference in the field of household behavior, offers a formal analysis of the household model under market failure, and considers different types of market failures that lead to non-separability. The sources of non-separability discussed in that book are in particular: the combination of an imperfect substitution between off-farm labor and farm labor for household members, with an imperfect substitution between hired labor and family labor on the farm; presence of a margin between the prices of labor sold and hired (that can come from transportation costs) resulting in a price band and in a self-sufficiency zone; credit rationing combined with fixed factors in production; and insurance market failures combined with a credit constraint.

The first tests of non-separability used the reduced form of a household model. One of the consequences of non-separability is that the household's characteristics in consumption z_c affect production decisions. These tests yielded mixed results. For example, Lopez (1984) rejected separability with Canadian data, Benjamin (1992) could not reject it with a sample of Javanese rural households as well as Bowlus and Sicular (2003) with a sample from China, and Grimard (2000) rejected it for Côte d'Ivoire. One could likely explain these differences by calling upon differences in context, even though in this particular contrast one would have expected more market imperfections in Java than in Canada. However, the more fundamental weakness of these studies is that they do not recognize heterogeneity in the household population, even though it is so clearly emphasized in the theory. As we have seen above in the case of a price band, non-separability is an idiosyncratic characteristic of a household and not of the market. Even when the context is not specific to each household, as would be the case if transactions costs were the same for all, households' responses, especially in terms of their decision to participate or not to the market, are specific to each. In this context, the result one can expect from a global estimation is unclear at best.

Recognizing household heterogeneity in the analysis creates difficulties. It requires a more structural approach that considers more explicitly the prevailing types of market imperfections. We will see in the following sections what progress has been made in this direction.

4.2 Direct measurement of the shadow price of labor

Rural labor markets are ridden with imperfections. It is important to first recognize the heterogeneity of what is called "labor": heterogeneity of workers' qualifications and tasks, heterogeneity due to seasonality. For some of these labor categories, a market simply does not exist, particularly for child labor and, in many cultural contexts, for female labor outside the family enterprise. Unemployment on the labor market places a constraint on time worked. Difficulty to find workers in peak seasons creates a constraint for employers. In addition, some supervision is necessary to counteract moral hazard behavior among hired labor (Frisvold, 1994). All these market imperfections create a situation whereby the marginal productivity of labor is not necessarily equal to the wage observed on the market. We will see in the following sections that credit market imperfections can also lead to a separation between marginal productivity and labor cost. Whatever the cause, comparisons between labor productivity and the observable market wage reveal market imperfections and situations of non-separability.

Many analyses have thus attempted to estimate the marginal productivity of labor, by estimating either a production function (Jacoby, 1993; Skoufias, 1994; Lambert and Magnac, 1992; Bhattacharyya and Kumbhakar, 1997) or a cost function. From a production function, Jacoby and Skoufias derive a measure of the marginal productivity of labor for every household. For the subset of household members that also work outside the farm, and for whom we have an observed wage (mainly men), the comparison can be made by regressing marginal productivity on wage. All these analyses reject equality between wage and marginal productivity of labor, and hence reject separability of production decisions. The approach followed by Lambert and Magnac allows for a better characterization of heterogeneity. This is because they can calculate for every household the average value and the standard deviation of marginal productivity. This allows them to test individually for each household the equality between marginal productivity of labor and wage offered. They derive from this a classification of households depending on whether separability is rejected or not. Applied to the LSMS for Côte d'Ivoire, they find that one can reject equality between labor productivity and wage for 90% of men, but only for 50% of women, the others having a productivity significantly lower than the wage. Bhattacharyya and Kumbhakar use a parametrization of shadow prices that they make a function of fixed factors and exogenous prices. Estimation of an indirect production function (function of input prices and fixed factors) allows them to identify the shadow prices for all factors. In their data for Bengal, they find large distortions on labor and draft animals relative to other inputs, with the surprising result that large farms under-value these two inputs more than small farms.

Even though the approach is undeniably interesting, the quality and credibility of these results crucially depends on specification of a production function. One has to admit that to represent an agricultural production process that unfolds over several months, with shocks, sequential decisions, irreversible choices, and complementarities and substitutabilities among inputs, by a production, cost, or profit function as simple as a Cobb-Douglas, or even a Translog, can make every production economist skeptical. The consequence is that the quality of estimations is weak, and inaccuracies in the measurement of marginal productivities very large.

4.3 Transactions costs, market participation, and supply and demand functions

The analysis of transactions costs and of the magnitude of transactions made on a particular market is illustrated in Figure 9-1. Note first an important result, though frequently neglected in empirical analyses: there is a single reference price for the production and the consumption of a particular household. This means that, for a net buyer, the opportunity cost of production is the purchase price of the product; and that for a net seller; the opportunity cost of consumption is the sales price. As a result, analyses of supply that consider the "producer price" as the reference price for all producers, and analyses of demand that take the "consumer price" for all households, are incorrect.

Consider for instance the case of a food item that is produced and consumed (without initial endowments and not used as an input). Assume that all other goods are exchanged on markets at exogenous and observable prices (that can be sales or purchase prices). Solution of the model leads to the definition of a shadow price p^* for the self-sufficient household equilibrium:

$$q\left(p^*, p_q; z_q\right) = c\left(p^*, p_c, y^*; z_q\right),$$

where p_q and p_c represent price vectors for the other goods produced and consumed, and y^* is the household's income in which food production q is valued at the shadow price. Following the notations in equation (2) above, this income can be written as:

$$y^* = p^* q + p_q' q_q + T,$$

where q_q represents the vector of other production goods (factors and products). This shadow price is thus function of the prices of all consumption and production goods, of fixed factors z_q and z_c, and of exogenous transfers T. With no fixed transactions costs, one can show that participation in the food market depends on the value of this shadow price relative to the two boundaries of the price band, $p^m - t_p^v$ and $p^m + t_p^a$. With fixed costs, market participation will only occur if the gain in utility is sufficient to compensate for the fixed costs. Denoting indirect utility by V, which is function of prices and income, the rule for market participation as a seller is thus:

$$V\left(p^m - t_p^v, \left(p^m - t_p^v\right)q + p_q' q_q + T - t_f^v; z_c\right) \geq V\left(p^*, y^*; z_c\right). \qquad (9)$$

The left hand side term measures the utility level for the seller household for whom the price of the food product would be $p^m - t_p^v$ and income would be reduced by the fixed cost t_f^v. The right hand side term measures utility under autarky. We can thus show that entry into the market as seller occurs when p^* reaches not $p^m - t_p^v$ but an even lower level, and that, upon entering, the seller sells a finite quantity, not an infinitesimal one. A similar reasoning on the side of purchasing shows that the household will only enter the market as a buyer when its shadow price will be sufficiently above the upper boundary $p^m + t_p^a$ of the price band.

The supply function, that includes these market participation decisions, takes the shape ABB'C'CD as shown in Figure 9-2.

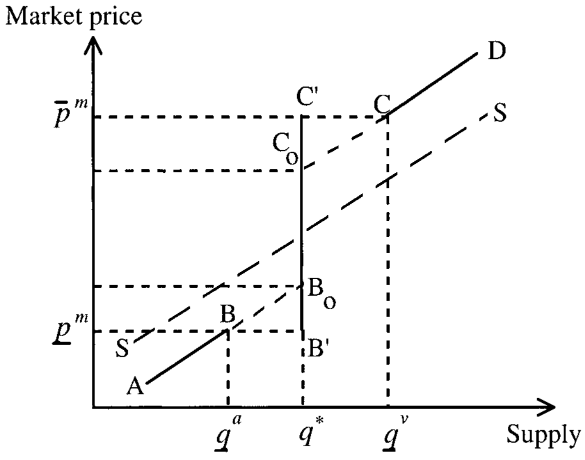

Figure 9-2. [Supply response when there are fixed and variable transactions costs]

When the market price is below the $\underline{p}^m$ threshold, the household is a net buyer and produces along the line segment AB. When the market price exceeds this threshold, the household becomes self-sufficient and produces q^*. Note the discontinuity in production from $\underline{q}^a$ to, which indicates that, the moment the household enters the market as a buyer, he buys the quantity

$q^* - \underline{q}^a$. The household remains in self-sufficiency between the two thresholds $\underline{p}^m$ and $\bar{p}^m$, beyond which he enters the market as a seller with production $\underline{q}^v$, and a minimum sales threshold $\underline{q}^v - q^*$. We are far from the simple supply function SS that would hold if there were no transactions costs.

Different empirical strategies have been used to estimate this complex supply function (and hence also the underlying behavioral response SS without transactions costs). Note first that a simple estimation that does not account for transactions costs would consist in fitting a line to the observations. It would estimate an elasticity much lower than that which motivates the underlying producer behavior.

A first empirical problem is about the measurement of transactions costs. Few transactions costs are directly observed and measured in household surveys. More fundamentally, while some of the transactions costs are observable (such as transportation costs, labor supervision costs, travel time, etc.), others such as the time and effort necessary to collect information and to carry negotiations are hardly so, and they are likely to be large as shown in a study of Peruvian transactions on potato markets done by Vakis, Sadoulet, and de Janvry (2003).

One can, however, observe the factors that determine the transactions costs. Using this approach, Goetz (1992) estimated the participation decisions to the grain market in Senegal as seller or buyer, using a bivariate probit corresponding to the reduced form in equation (9). Conditional on this decision, quantities sold or purchased can then be estimated. An interesting result from this analysis is to be able to decompose the impact of a rise in the price of grains between entry of new sellers and increase in the sale of producers already engaged in the market.

Making more explicit use of the existence of production thresholds $\underline{q}^v$ and $\underline{q}^a$ identified above, we were able to identify separately the determinants of fixed and variable transactions costs though a joint estimation of these thresholds and of the supply function for the corn market in Mexico (Key, Sadoulet, and de Janvry, 2000). While we did not get to the point of actually measuring the size of these transactions costs, those are clearly identified on Figure 9-2 (the distance B_0C_0 is exclusively determined by proportional transaction costs, while distances BB' and CC' are exclusively determined by fixed transactions costs in entering the market as buyer and seller, respectively). Fixed and variable transactions costs can thus be measured through the estimation of such a semi-structural model. Note also that if one draws the demand curve, both transactions costs can also be recovered from the observation of purchases by buyers and sales by sellers.

In a study of the rural labor market, we followed rigorously the concept of price band and estimated the participation decision as an ordered probit (for the shadow price p^*) with idiosyncratic thresholds for each household, representing the opportunity cost of selling and buying labor (Sadoulet, de Janvry, and Benjamin, 1998). Using the same theoretical framework, but observing the purchasing and selling prices for all individual households, and hence implicitly the proportional transaction costs they would face, Renkow, Hallstrom and Karanja (2004) are able to quantify the fixed transactions costs facing households from Kenya. They find that on average fixed transactions costs are equivalent to a 15% ad-valorem tax (but as high as 70% for some households), while purchasing prices are on average 35% above selling prices. Finally, Skoufias (1995), using a simpler specification of transactions costs, estimated directly net land rental, jointly for tenants and landlords, as a friction model.

All these approaches differ in their characterization of transactions costs and especially in their econometric specification of the distribution of error terms. None is yet fully satisfactory in either of these dimensions. None has yet succeeded in rigorously estimating transactions costs. One can expect, however, that this approach could allow to identify non-observable transactions costs as revealed through the behavior they induce on markets.

An even more fundamental difficulty which remains to be overcome is the correct definition of regime. Indeed, in the price bands approach, one makes the assumption that there is no constraint on the level of market participation. This assumption is likely to be valid for product markets, but it is generally too restrictive for the labor market. When there is unemployment, many households may not be able to work the desired amount, even though they participate in the labor market. Market participation in this case does not allow to identify the constrained regime. What is necessary, then, is to identify the regime on the basis of the determinants of behavior (separable or non-separable) and not observed participation (see Carter and Olinto, 2003, for the credit market; Vakis, Sadoulet, de Janvry and Cafiero, 2002, for the labor market). A more satisfactory approach would be to make sure that household surveys collect information not only on quantities purchased or sold, but also on the possible existence of constraints, thus avoiding having to rely on econometric techniques to reveal these constraints.

5. WHAT HAVE WE LEARNED FROM MODELING HOUSEHOLD BEHAVIOR UNDER MARKET FAILURES?

We have seen that farm households, in the context of developing economies, are systematically embedded in complex configurations characterized by incomplete and failing markets (Thorbecke, 1993). This opens the door to the analysis of their behavior through the formulation of non-separable models where consumption and production decisions are jointly determined to accommodate these failures. We have reviewed the theoretical and empirical contributions that followed this approach, taking as a starting point the seminal advances made by contributors to the book by Singh, Squire, and Strauss (1986).

A first conclusion that derives from this review is that many observed behavioral patterns would be incomprehensible without taking into account the specificity of the context of failing markets where farm households make their decisions. In the papers reviewed, we have seen that this is in particular the case for: (1) The way health, nutrition, and education decisions affect production choices. (2) How market participation decisions are made and the conditions under which participation occurs. (3) Supply response to price incentives when there are fixed and proportional transactions costs. (4) Reasons for the existence of an inverse relation between productivity and farm size. (5) Implications of failures on some markets for behavior on complete markets. (6) Management of the credit constraint when insurance markets are not accessible. (7) Responses to price risks on food markets when risk coping instruments are missing.

There are of course many aspects of household behavior that still require interpretation, leaving us with an array of unresolved Schultzian puzzles. This includes, for instance, imperfect risk sharing within households (Goldstein, 2000), and observations of low levels of adoption of profitable activities, such as pineapple production in Ghana (Goldstein and Udry, 1999) and fertilizer use in Kenya (Duflo, 2003), that cannot be explained by insurance or credit market failures. Further theoretical and empirical work on household behavior under market failures is thus needed to address these puzzles.

The possibilities for farm households to adjust their behavior in order to optimize the use of their resources in a context of imperfect markets, however creative and ingenious their strategies may be, are sharply limited by their assets positions and by the contexts where they operate. Their limited access to productive assets and the unfavorable contexts where they operate result in high levels of risk aversion and credit market failures. In this fashion, poverty changes the set of options available to households, making poverty hard to escape (Duflo, 2003). One can wonder how close to a first best situation do these countervailing strategies allow households to get. Ten percent or ninety percent? And what is the cost of these responses? Here is where we still lack empirical evidence. This is consequently the second conclusion that we derive from this review. The important theoretical advances on household behavior under market failures that have been made starting in the early 1980's, and that were subsequently pursued in a multiplicity of directions that we reviewed here, leave us with a large gap in empirical validation and confirmation of the importance of these phenomena. Many theoretical propositions that derive from non-separability are all too often blindly accepted as truths when they remain to be empirically verified, their order of magnitude in explaining observed behavior remains to be ascertained, and their usefulness in the design of policies to be shown. Heterogeneity of rural household behavior, given their control over assets and the particular contexts where they make decisions, remains to be characterized. Gains to be derived from differentiated policies will not materialize until this heterogeneity has been sufficiently quantified.

If transactions costs are large, they need to be measured and explained. We have argued that attempting to observe them directly will always underestimate their importance, quite likely by large amounts. We showed, however, that they can be derived from observed behavior. This, however, requires the construction of structural models where behavior is specified. Lack of direct data is thus replaced by structural models, and the quality of the models will determine the quality of the measurements made. These household models need to capture the relevant dimensions of behavior, given assets and a particular context. Our third conclusion is thus that making advances in the measurement of transactions costs will require improved specification of structural models, which in turn requires better understanding of what processes make transactions costs so pervasively large for poor farm households. This may require imaginative field work and interdisciplinary efforts to better specify the dimensions of behavior under market failures. Structural household models have been estimated by Fafchamps (1993) and Kurosaki and Fafchamps (2002). In spite of progress in computational capacity, these models remain cumbersome to estimate. Specification and estimation of semi-structural models as in Key, Sadoulet, and de Janvry (2000) are more likely to provide a compromise between precision in the specification of behavior and empirical expediency.

It is thus important to conclude cautiously when observing rural households' response strategies to market failures, however clever they may be. These countervailing strategies likely only compensate for a small fraction of the market failures under which households operate, in part because the effectiveness of these strategies is itself limited by poverty, and they are generally implemented at very high costs in foregone expected incomes. Cunningness under low assets endowments and extensive market failures is unlikely to be sufficient to enable rural households to overcome poverty.

NOTES

[i] We thank John Strauss for useful comments and references.

[ii] Time is treated as a commodity used as input under the form of labor and as a consumption good under the form of leisure. Labor market participation is considered as an activity that uses labor as the only input.

REFERENCES

Alderman, Harold and Cristina Paxson. 1994. “Do the Poor Insure? A Synthesis of the Literature on Risk and Consumption in Developing Countries.” In *International Economics Association,* Moscow meeting, Proceedings Vol. 4.

Azam, Jean-Paul, and Besley, Timothy. 1991. "Peasant Supply Response under Rationing: The Role of the Food Market." *European Journal of Political Economy,* 7: 331-343.

Barrett, Christopher. 1996. “On Price Risk and the Inverse farm-Size-Productivity Relationship.” *Journal of Development Economics*, 51(2): 193-215.

Behrman Jere, Andrew Foster, and Mark Rosenzweig. 1997. “The Dynamics of Agricultural Production and the Calorie-Income Relationship - Evidence From Pakistan.” *Journal of Econometrics.* 77(1): 187-207.

Benjamin, Dwayne. 1992. "Household Composition, Labor Markets, and Labor Demand: Testing for Separation in Agricultural Household Models." *Econometrica,* 60(2): 287-322.

Benjamin, Dwayne. 1995. "Can Unobserved Land Quality Explain the Inverse Productivity Relationship?" *Journal of Development Economics,* 46(1): 51-84.

Berry, Albert, and William Cline. 1979. *Agrarian Structure and Productivity in Developing Countries.* Baltimore: The Johns Hopkins University Press.

Berthélemy, Jean-Claude, and Christian Morrisson. 1989. *Agricultural Development in Africa and the Supply of Manufactured Goods.* Washington D.C.: OECD Publications and Information Center.

Bhattacharyya, Anjana, and Subal Kumbhakar. 1997. “Market Imperfections and Output Loss in the Presence of Expenditure Constraint: A Generalized Shadow Price Approach.” *American Journal of Agricultural Economics,* 79(3): 860-71.

Bowlus, Audra, and Terry Sicular. 2003. “Moving Toward Markets? Labor Allocation in Rural China,” *Journal of Development Economics*, 71(2): 561-583.

Carter, Michael, and Pedro Olinto. 2003. “Getting institutions "right" for whom credit constraints and the impact of property rights on the quantity and composition of investment.” *American Journal of Agricultural Economics*, 85(1): 173-186.

Carter, Michael. 1984. “Identification of the Inverse Relationship Between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production.” *Oxford Economic Papers* 36: 131–146.

Chayanov, A.V. 1925. *The Theory of Peasant Economy.* Translated by D. Thorner *et al.* (1966). D. Irwin: Homewood, IL.

de Janvry, Alain, Marcel Fafchamps, and Elisabeth Sadoulet. 1991. "Peasant Household Behavior with Missing Markets: Some Paradoxes Explained." *Economic Journal,* 101(409): 1400-17.

de Janvry, Alain, Marcel Fafchamps, Mohammed Raki, and Elisabeth Sadoulet. 1992. "Structural Adjustment and the Peasantry in Morocco: A Computable General Equilibrium Model." *European Review of Agricultural Economics,* 19: 427-453.

Deaton, Angus. 1992. *Understanding Consumption.* Oxford: Clarendon Press.

Deolalikar, A. B. 1988. "Nutrition and Labor Productivity in Agriculture: Estimates for Rural South India." *Review of Economics and Statistics,* 70(3): 406-413.

Dercon, Stefan. 1998. Wealth, Risk and Activity Choice: Cattle in Western Tanzania." *Journal of Development Economics,* 55(1): 1-42

Duflo, Esther. 2003. "Poor but Rational?" Department of Economics, MIT.

Dutilly-Diane, Céline, Elisabeth Sadoulet, and Alain de Janvry. 2003. "Household Behavior under Market Failures: How Improved Natural Resource Management in Agriculture Promotes the Livestock Economy in the Sahel". *Journal of African Economies,* 12: 343-370

Ellis, Frank. 1993. *Peasant Economics: Farm Households and Agrarian Development.* Cambridge University Press.

Eswaran, M., and A. Kotwal. 1986. "Access to Capital and Agrarian Production Organization." *Economic Journal,* 96: 482-498.

Eswaran, M., and A. Kotwal. 1989. "Credit as Insurance in Agrarian Economies." *Journal of Development Economics* 31: 37-53.

Fafchamps Marcel. 1992. "Cash Crop Production, Food Price Volatility and Rural Market Integration in the Third-World." *American Journal of Agricultural Economics,* 74(1): 90-99.

Fafchamps, Marcel. 1993. "Sequential Labor Decisions under Uncertainty - An Estimable Household Model of West-African Farmers." *Econometrica,* 61(5): 1173-1197.

Fafchamps, Marcel, and John Pender. 1997. "Precautionary Saving, Credit Constraints, and Irreversible Investment: Theory and Evidence from Semi-arid India." *Journal of Business and Economic Statistics,* 15(2): 180-194.

Feder, Gershon, L.J. Lau, J.Y. Lin, and X. Luo. 1990. "The Relationship between Credit and Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium." *American Journal of Agricultural Economics,* 72(4): 1151-1157.

Feder, Gershon. 1985. "The Relationship between Farm Size and Farm Productivity: The Role of Family Labor, Supervision, and Credit Constraint." *Journal of Development Economics,* 18: 297-313.

Finkelshtain, Israel, and Chalfant James. 1991. "Marketed Surplus Under Risk: Do Peasants Agree with Sandmo?" *American Journal of Agricultural Economics,* 73(3): 557-567.

Frisvold, George. 1994. "Does supervision matter? Some hypothesis tests using Indian farm-level data." *Journal of Development Economics,* 43(2): 217-238

Goetz, Stephan. 1992. "A Selectivity Model of Household Food Marketing Behavior in Sub-Sahara Africa." *American Journal of Agricultural Economics,* 74(2): 444-452.

Goldstein, Markus, and Christopher Udry. 1999. "Agricultural Innovation and Resource Management in Ghana". Development Studies Institute, London School of Economics.

Goldstein, Markus. 2000. *Intra-household Allocation and Farming in Southern Ghana.* Ph.D. dissertation, University of California at Berkeley.

Grimard, Franque. 2000. "Rural Labor Markets, Household Composition, and Rainfall in Côte d'Ivoire." *Review of Development Economics,* 4(1): 70-86.

Hayami, Yujiro, and Keiji Otsuka. 1993. *The Economics of Contract Choice: An Agrarian Perspective.* Oxford: Clarendon Press.

Holden, Stein, Bekele Shiferaw, and Mette Wik. 1998. "Poverty, market imperfections and time preferences: of relevance for environmental policy?" *Environment and Development Economics,* 3(1): 105-130

Jacoby, Hanan. 1993. "Shadow Wages and Peasant Family Labour Supply: An Econometric Application to the Peruvian Sierra." *Review of Economic Studies,* 60: 903-921.

Key, Nigel, Elisabeth Sadoulet, and Alain de Janvry. 2000. "Transactions Costs and Agricultural Household Supply Response". *American Journal of Agricultural Economics,* 82(2): 245-259.

Key, Nigel. 2000. "Savings, Credit, and the Self-Finance of Crop Production in Mexico." *Savings and Development.*

Kurosaki, Takashi, and Marcel Fafchamps. 2002. "Insurance Market Efficiency and Crop Choices in Pakistan." *Journal of Development Economics,* 67(2): 419-453.

Lamb R.L. 2003. "Inverse Productivity. Land Quality, Labor Markets, and Measurement Errors." *Journal of Development Economics,* 71(1): 71-96.

Lambert, Sylvie, and Thierry Magnac. 1994. "Measurement of Implicit Prices of Family Labour in Agriculture: An Application to Côte d'Ivoire." In Caillavet, Gyomard, and Lifran (eds.) *Agricultural Household Modelling and Family Economics,* Amsterdam: Elsevier.

Löfgren, Hans, and Sherman Robinson. 1999. "To Trade or Not to Trade: Non-Separable Household Models in Partial and General Equilibrium". Washington D.C.: International Food Policy Research Institute.

Lopez, Ramón. 1984. "Estimating Labour Supply and Production Decisions of Self-Employed Farm Producers." *European Economic Review,* 24: 61-82.

Morduch, Jonathan. 1990. "Risk, Production, and Saving: Theory and Evidence from Indian Households." Mimeo, Harvard University.

Nakajima, C. 1970. "Subsistence and Commercial Family Farms: Some Theoretical Models of Subjective Equilibrium." In Wharton, ed. *Subsistence Agriculture and Economic Development.* Chicago: Aldine.

Pitt, M.M., and Mark Rosenzweig. 1986. "Agricultural prices, food consumption, and the health and productivity of Indonesian farmers." In I. Singh, L. Squire, and J. Strauss, eds., *Agricultural Household Models,* Baltimore: Johns Hopkins University Press.

Polanyi, Karl, Conrad Arensberg, and Harry Pearson. 1958. *Trade and Markets in the Early Empires.* New York: The Free Press.

Renkow, Mitch. 1990. "Household Inventories and Marketed Surplus in Semisubsistence Agriculture." *American Journal of Agricultural Economics,* 72(3): 664-675.

Renkow, Mitch, Daniel Hallstrom and Daniel Karanja. 2004 "Rural Infrastructure, Transactions Costs and Market Participation in Kenya," *Journal of Development Economics,* 73(1): 349-367.

Rosenzweig, Mark, and Hans Binswanger. 1993. "Wealth, Weather Risk and the Composition and Profitability of Agricultural Investments." *Economic Journal,* 103(416): 56–78.

Rosenzweig, Mark, and Wolpin Kenneth. 1993. "Credit Market Constraints, Consumption Smoothing and the Accumulation of Durable Production Assets in Low Income Countries: Investment in India." *Journal of Political Economy,* 101(2): 223-244.

Sadoulet, Elisabeth, Alain de Janvry, and Catherine Benjamin. 1998. "Household Behavior with Imperfect Labor Market". *Industrial Relations,* 37(1): 85-108.

Saha, A., and J. Stroud. 1994. “A Household Model of On-Farm Storage Under Price Risk.” *American Journal of Agricultural Economics*, 76(3): 522-534.

Shively, Gerald. 2001. “Poverty, consumption risk, and soil conservation.” *Journal of Development Economics*, 65(2): 267-290

Singh, I., Lynn Squire, and John Strauss, (eds.). 1986. *Agricultural Household Models.* Baltimore. MD: The Johns Hopkins University Press.

Skoufias, Emmanuel. 1994. “Using Shadow Wages to Estimate Labor Supply of Agricultural Households.” *American Journal of Agricultural Economics,* 76(2): 215–227.

Skoufias, Emmanuel. 1995. “Household Resources, Transactions Costs, and Adjustment through Land Tenancy.” *Land Economics,* 71 (February): 42-65.

Strauss, John. 1986. “Does better nutrition raise farm productivity?” *Journal of Political Economy,* 94(2): 297-320.

Taylor, J. Edward, and Irma Adelman. 2003. "Agricultural Household Models: Genesis, Evolution and Extensions" *Review of Economics of the Household,* 1(1).

Thorbecke, Erik. 1993. “Impact of State and Civil Institutions on the Operation of Rural Markets and Nonmarket Configurations.” *World Development,* 21(4): 591-605.

Vakis, Renos, Elisabeth Sadoulet, Alain de Janvry, and C. Cafiero. 2002. “Searching for failures in the Peruvian labor market via mixture models”. http://are.berkeley.edu/~rvakis/mixtures.pdf

Vakis, Renos, Elisabeth Sadoulet, and Alain de Janvry. 2003. “Transactions Costs and the Role of Information: Evidence from Peru.” http://are.berkeley.edu/~rvakis/tc.pdf

Vakis, Renos. 2002. "Overcoming Credit Market Failures: A Paradigm of Diversification for Technology Adoption in Peru". http://are.berkeley.edu/~rvakis/adoption.pdf.

von Braun, J., D. Hotchkiss, and M. Immink. 1989. *Nontraditional Export Crops in Guatemala: Effects on Production, Income, and Nutrition.* Washington D.C.: IFPRI Research Report No. 73.

Yotopoulos, Pan, and Lawrence J. Lau. 1974. “On Modeling the Agricultural Sector in Developing Economies: An Integrated Approach of Micro and Macroeconomics, *Journal of Development Economics*, 1(2): 105-127.

Chapter 10

LABOR LAWS AND LABOR WELFARE IN THE CONTEXT OF THE INDIAN EXPERIENCE

Kaushik Basu
Cornell University, kb40@cornell.edu

1. INTRODUCTION

Markets often work in peculiar ways. A policy that seems obviously good for some groups of people may turn out, when the dust settles and an equilibrium is established, to be detrimental to their welfare. Economics would not have been an interesting subject if this were never the case. One market where such pathologies often occur is the labor market. So it is not surprising that labor market legislation is one area where well-meaning but erroneous policies abound.

In many parts of the world, workers do not seem to have done particularly well and have in fact often lost out in relative terms. This can be because the policy makers did not care but it can also be because they cared but misunderstood the way the labor market works and so their interventions did not work the way they expected they would.

Observers often argue that trade unions use their muscle power to get benefits for organized labor to the detriment of other workers. But in reality it is not evident that organized labor has done that well either. One reason for this is that the relation between the legal and contractual environment of a nation and the well-being of workers is sufficiently complex that trade unions do not always understand what is good for them and so do not demand what is in their interest.

The aim of this paper is to construct plausible theoretical models, using India as the backdrop of stylized facts, to show that this may indeed be the case. It will be argued that India's myriad labor laws, meant to protect laborers, may have actually hurt them. The argument will be presented in terms of a theoretical model and, as such, should be of interest to other developing and transition economies as well. I take the view in this paper, in keeping with the normative position that Erik Thorbecke has espoused time and again (Thorbecke, 2003; 2003a) that a society or government has a special responsibility towards the disadvantaged sections of a nation. Hence, in general, laborers should be the target of welfare-enhancing government intervention.[i]

Several pieces of labor legislation in India were drafted expressly to make the laying off of laborers difficult. If an employer found that a worker (1) was shirking from putting in enough effort or (2) did not have adequate skill for the job in question, in many situations he would not be allowed to dismiss the worker (or be allowed to do so only at a considerable cost), no matter what the initial contract with the worker. I shall show that the eventual labor market equilibrium that emerges in an economy with such legislation may actually cause workers to have a lower welfare than in an economy with less protective legislation; and that *between* legislating to prevent layoffs and legislating to maintain minimum wages, the latter may be the more desirable policy from the point of view of worker welfare.

Section 2 presents some institutional details of the Indian labor market. In the model in Section 3 it will be assumed that labor effort is fixed and so only (2), above, is the relevant issue. It will be shown that in a labor market model, which *prima facie* captures the broad realities of the Indian economy, an employer's inability to dismiss workers who turn out not to possess the required skill could, in equilibrium, hurt all workers, including the unskilled. The essential argument goes as follows. If worker dismissal is disallowed or very costly, firms which need specialized skills and talents may operate on a smaller scale or, worse, close down. This would of course hurt the skilled workers and, by turning them out to the unskilled labor market, could also lower wages in the latter, thereby hurting all workers. There are other possible routes to a similar conclusion. If a firm faces a fluctuating-demand environment and is prevented by law from laying off workers, it may once again close down or function only on a small scale, thereby causing a contraction in the demand for labor and depressing wages (Basu, Fields and Debgupta, 2001). Likewise, consider the case where workers can shirk effort, that is, (1), above, is relevant. In such a situation it is possible that workers may want to be forced to work hard. That would make labor a more coveted input and could increase the demand for labor and wages so much that it would more than compensate the workers for the higher effort.[ii]

Before proceeding further I want to emphasize that the modeling here is based on realistic assumptions, but assumptions all the same. Hence, by altering these we can get different results. Nevertheless, it is interesting to see that legislation which is seemingly pro-worker *may* end up hurting the same workers it is supposed to help. Even if, in reality, this is not always so, the fact that this can happen under realistic assumptions should alert us to the fact that labor laws need more careful scrutiny, theoretical and empirical, to sort out which ones *actually* help workers and which ones hurt.

2. THE INSTITUTIONAL BACKGROUND

Beginning with policy adjustments to stave off a foreign exchange and fiscal crisis in mid-1991, the Indian government has gone on to attempt major economic reforms during the last decade. While significant changes have been effected in several sectors (see Basu, 2004), notably those dealing with international trade and investment, one area which has resisted reform is that of labor markets and labor legislation. This is a matter of some concern since it is arguable that, in the long run, reforms in this area will matter more than those in many other sectors.

Up to now, and in sharp contrast to many other Asian countries, India has failed to deploy her large labor resources to compete better on the domestic and international markets. As a consequence Indian workers remain poor, underemployed and often unemployed.[iii] On the other hand, in India a large number of labor laws have been enacted with the express purpose of protecting labor.[iv] It is the aim of this paper to suggest that these two facts may not be unrelated. To the extent that these laws are well-meaning and their intent is widely supported, it will be argued here that the condition of the Indian worker represents a major *intellectual* failure -- the failure to appreciate that overt protection can ultimately do harm when the market has fully responded to the policies and laws and settled down to an equilibrium. In particular, it will be shown that enabling retrenchment and layoffs may result in larger employment and higher wages in the resulting equilibrium. Casual empiricism certainly does not contradict such a hypothesis. The East Asian and South East Asian countries where employment has grown and wages risen are also the countries which, in contrast to India, have fewer protective laws. As Edgren (1989, p. 1) notes at the start of a major I.L.O. study: "legislation governing hiring and firing, minimum wages and the scope for collective bargaining differs between different parts of Asia, with the countries of the Indian subcontinent having stricter regulations of employers' rights to hire and fire, and granting unions wider scope for bargaining."

The kind of legislation that is of central interest to me here is *India's Industrial Disputes Act*, 1947. This act, along with the amendments of 1976 and 1982, places restrictions on layoffs and dismissal by large firms. What is important is that these restrictions are exogenous in the sense that they override any contract between the employer and workers. For firms employing 50 or more workers there are predetermined compensations which the employer has to give to workers when they are laid off. For firms employing more than 100 workers the Act requires the employer to take prior permission from government for layoffs and retrenchment of labor and for closing down the firm. And, as Datta Chaudhuri (1994) points out, "government permission is seldom given", and in most employment-related disputes, government gets involved and treats the handling of labor as a child custody problem in a divorce suit. In India, government intervention in labor markets goes much deeper than would appear through studying the labor laws. The *Indusrial Disputes Act* allows the Labor Departments of the Centre and State governments to intervene not only in labor disputes but also in anticipated labor disputes. In addition there is the problem of political and ministerial intervention (Ramaswamy, 1984). Finally, as mentioned above, the judiciary often takes a custodial attitude to labor. In 1992 in a case involving a bankrupt private firm, a judge of the Calcutta High Court, Mr. Justice Hazari, argued that, if another private firm took over the firm, there would be no guaranteeing that that firm would not, in turn, go bankrupt and cause workers to be laid off. He, therefore, directed the Government of West Bengal to take over the firm and "run it with the existing workers" (Datta Chaudhuri, 1994).

It will be shown that it is likely that such laws and practices hurt not only workers who are not protected by the law (because, for instance, they work in small firms), but also the workers who are allegedly protected by law. The fact that protective labor legislation may have hurt India's overall growth and efficiency has been pointed out by many observers (see, e.g., Lucas, 1988; Ahluwalia, 1991; Papola, 1994). They are probably right but my argument here is distinct because I am claiming that such legislation may have hurt the very constituency that it was meant to protect, to wit, labor.[v] Hence, Kannan's (1994) observation that wages in the eighties have not in general kept pace with labor productivity, put forward as a critique of the view that increasing protective legislation has hurt growth and efficiency, and Ghose's (1994) finding that employment per unit of gross value added in manufacturing fell, monotonically, throughout the eighties sit very comfortably with the theoretical findings of this paper. I must stress that what I am arguing is not for firms to be given the freedom to hire and fire as they wish, but for firms and workers to have greater freedom to sign contracts concerning layoffs, retrenchment and closure, without these being overruled by exogenously determined conditions as wantonly as they currently are.

Suppose some workers in a large firm ask their employer to pay them a higher wage and, in turn, they promise to go away without compensation whenever the employer wishes to sack them. Even if both the workers and the employer benefit by such a contract, it is unlikely in contemporary India that an employer will agree to such a contract. This is because if, after paying a higher wage for some time, the employer actually gives notice to his workers (perhaps because demand for the product has fallen off), the workers can appeal to government and government is very likely to cite the *Industrial Disputes Act* and declare such a dismissal illegal *no matter what the prior agreement between workers and employer.* Indeed there are not too many credible ways for workers to give up their *right* not to be dismissed. Over here we must distinguish between "not resisting dismissal" and "giving up the right not to be dismissed". A worker can of course choose not to resist dismissal; but what is interesting is that he may not (and, in the case of India, he *is* not) able to waive the right not to be dismissed.

One kind of reform that my model in Section 3 prompts is to allow employers and workers to sign any contract concerning dismissal conditions and have the state or judiciary uphold such a contract. A less radical but, nevertheless desirable reform would be to leave much of the law, for example, the *Industrial Disputes Act,* as it is but to add on a clause which gives workers the right to waive the right not to be dismissed as conferred on them currently by the Act.[vi] Such provisos are not unheard of. In the U.S. a student has the right to see the recommendation a professor writes for her, but she also has the right to waive this right.

Before proceeding to construct a formal model, it is worth asking ourselves what is the correct market structure to assume in describing the interaction between firms and workers. Since most of the layoff and dismissal laws apply to large firms, as discussed above, a model with only atomistic firms will not be the right one. At the same time we know that there are lots of firms which do not come under the purview of laws such as the *Industrial Disputes Act* by virtue of being too small. Hence, the right model seems to be one with some large or dominant firms that are capable of affecting the market wage, along with a wage-taking fringe of smaller firms. In addition, it will be assumed that it is the large firms which need specialized or skilled labor.

3. MODEL: TURNOVER AND QUALITY

3.1 Basic Concepts

There are certain kinds of skills which are not captured by usual indices like university degrees or I.Q. test scores, but which nevertheless matter to the employer. The skills needed for dealing with people, for remembering little tasks to be performed, and for punctuality often fall in this category. People better endowed with these skills may not necessarily find it easier to acquire education, so even in equilibrium such skills may not be strongly correlated with the degrees and diplomas held by a worker, thus making standard job-market signaling models irrelevant for our present purpose. The only way for an employer to know exactly how much skill of the above kinds a worker possesses is to employ the worker. For reasons of modeling simplicity I shall assume that skill can be of only two levels. Hence workers are either skilled or unskilled. Let us assume that a fraction t of the labor force is skilled. Let the aggregate supply curve of labor be given by

$$s = s(w), \qquad s'(w) > 0 \tag{1}$$

where w is wage.

By assuming that both skilled and unskilled workers are otherwise homogenous or that their labor-leisure choices are identical, we know that if a wage of w is fixed, the supply of skilled labor is $ts(w)$ and the supply of unskilled labor is $(1 - t)s(w)$.

Next, suppose that there are two kinds of firms -- ones where the worker's skill matters (these are also the large or `dominant' firms) and ones where they do not (these are the small or 'fringe' firms). From each skilled worker the *dominant* firms can get an output of $r > 0$. They have no use for unskilled labor, who produce 0 in such firms. On the other hand a *fringe firm* has no special use for skilled labor. Each worker, skilled or unskilled, produces an output less than r.

Let me sketch the intuitive argument first. If the dominant firms are not allowed to lay off workers, they will be forced to make do with a labor force in which a fraction t will be skilled and fraction $1 - t$ unskilled (since the only way to ensure that you have all skilled workers is to employ people, check them out, lay off the unskilled, employ new people in their place, check them out, and so on). This will typically result in a smaller demand for labor from the dominant firms. Hence the supply of labor in the fringe job market will be greater. This will tend to push down wages. Hence, a law preventing layoffs can actually push down *all* wages. It is interesting here to note that the detailed empirical study of Fallon and Lucas (1993) reveals that demand for labor in large firms fell as legislation preventing labor dismissal was made stronger. And Fallon and Lucas (p. 269) go on to conjecture: "This decline may be understood in terms of [...] reluctance to hire in case employees prove to be poor matches with their job demands, a mismatch which cannot readily be reversed".[vii]

In the model that follows I shall make special assumptions to keep the algebra simple but the aim is to formalize the above general argument.

As explained in the previous section, the fringe firms will be treated as wage-takers, whereas the dominant firms are like oligopsonists in the labor market. Hence, we are considering a labor market similar to the product-model of Nichol (1930) and Stigler (1950) and subsequently extended and discussed by Encaoua and Jacquemin (1980), Dixit and Stern (1982), and Basu (1993).

Let the fringe firms' aggregate demand for labor be given by:

$$d = d(w), \qquad d'(w) < 0 \tag{2}$$

Hence, if w is the wage, the supply of labor in excess of what is needed by the fringe firms is given by

$$\psi(w) = s(w) - d(w) \tag{3}$$

Let us suppose that there are k dominant firms. If these firms choose workers randomly, the expected output from a worker is $rt(= rt + 0(1 - t))$. If these firms are not allowed to lay off workers then the expected output from each employed worker is rt. If, on the other hand, the firms can freely dismiss workers then, we shall assume, worker productivity to be r. This is because through successive laying off of unproductive workers, the dominant firms' labor force will converge to a purely skilled group.

Each dominant firm's cost of production consists of an entry fee of $K(> 0)$ and the wage bill.

Our model consists of two periods. In period 1, each of the k dominant firms have to decide whether to enter the industry or not. Then in period 2 the ones that enter decide how much labor to employ, keeping in mind that their decision will affect the labor demand of the fringe firms. Our aim is to characterize the *subgame perfect* equilibrium of this two-period game. Actually, we have two games. One in which layoffs are allowed and one in which the law prevents layoffs. Our aim is to compare the equilibria in these two games.

There are, in reality, many intermediate cases. For instance, a law could make layoffs costly, instead of disallowing layoffs. As discussed above, *India's Industrial Disputes Act* does exactly that for firms employing between 50 and 100 workers. Fortunately, such cases are easy to discuss once we have worked out the two polar models.

I shall characterize the subgame perfect equilibria by first analyzing the Nash equilibria in the second period game when no layoffs are permitted (Section 3.2) and when layoffs are permitted (Section 3.3) and then turn to the outcomes in the full two-period model (Section 3.4). If m dominant firms enter the industry and layoffs are not permitted, the second period game is described as *G(m,N)*. The second period game with *m* dominant firms and layoffs permitted is called *G(m,L)*.

3.2 The Game *G(m,N)*

There are *m* dominant firms that have entered the industry, and confront the labor supply function $\psi(w)$, described in (3). Let the inverse of this function be given by $\varphi(\cdot\ )$. If the firms hire $n_1,\ldots,n_m$ units of labor, recalling that no layoffs are allowed in this model the profit function of firm *i* is given by

$$\pi_i^N(n_1,\ldots,n_m) = rtn_i - \phi(n_1 + \ldots + n_m)n_i \tag{4}$$

This is the model used by Encaoua and Jacquemin (1980), though in their model the dominant firms and the fringe compete in the *product* market. I shall here focus on the symmetric Nash equilibrium of this m-player game. In the examples that I consider below such an equilibrium always exists and, in fact, is the only Nash equilibrium. Hence, each firm employing n^N labor is an *equilibrium* if and only if, for all $i \varepsilon \{1,\ldots,m\}$.

$$\pi_i^N(n^N,\ldots,n^N) \geq \pi_i^N(n^N,\ldots,n^N,\ n_i, n^N,\ldots,n^N), \quad \textit{for all } n_i.$$

If n^N is the equilibrium in *G(m,N)*, define[viii]

$$\pi^N(m) \equiv \pi_i^N(n^N,\ldots,n^N).$$

Hence, $\pi^N(m)$ is the profit earned in the period 2 equilibrium by each firm when there are *m* firms and no layoffs (hence the superscript *N*) are allowed. In equilibrium the fringe firms hire $d(\phi(mn^N))$ laborers and pay wage equal to $\phi(mn^N)$.

While this completes the description of the outcome of *G(m,N)*, let me introduce a geometric description of the problem which could aid intuition later in the more complicated case.

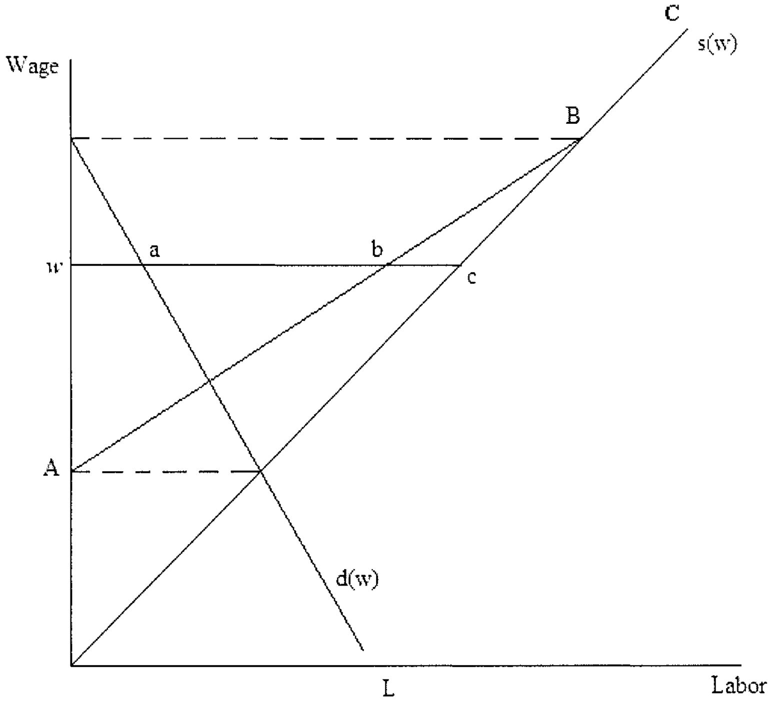

Figure 10-1.

If the aggregate supply curve, *s(w)*, and the demand curve of the fringe firms, *d(w)*, are as shown in Figure 10-1, then the supply curve faced by the dominant firms is given by ABC where AB is drawn such that for any wage *w*, the line *wa* is equal to *bc*. The final equilibrium is then the usual oligopsony equilibrium for *m* firms facing the supply curve ABC.

3.3 The Game *G(m,L)*

There are *m* dominant firms that have entered the industry and they are allowed to layoff workers. This means that (in the end) these firms will only employ skilled workers and get an output of *r* from each employed worker. Hence, the supply curve of the labor that is of relevance to these firms is given by *ts(w)*. What we have to be careful about is that if the wage that the dominant firms pay drops too low then some of the skilled laborers may prefer to go to the fringe firms. Keeping this in mind, let us now work out the aggregate demand function that these dominant firms face.

Define the wages, $\overline{w}$ and $\underline{w}$, as, respectively,

$$d(\overline{w}) = (1-t)s(\overline{w})$$

and

$$d(\underline{w}) = s(\underline{w})$$

In this model with layoffs the skilled workers can get a different wage from unskilled workers. Let w^1 be the wage earned by each skilled worker and w^0 the wage of an unskilled worker.

If $w^1 > \overline{w}$, the supply of skilled workers to the dominant firms is $ts(w^1)$ and $w^0 = \overline{w}$. If $w^1 > \overline{w}$, the supply of workers to the dominant firms is $s(w^1) - d(w^1)$ and $w^0 = w^1$.[ix] The dominant firms will ensure that the workers they employ are skilled workers. This is feasible since

$$s(w^1) - d(w^1) < ts(w^1),$$

which is an implication of $w^1 \geq \overline{w}$.

The information in the above paragraph is summed up, by writing the supply function of skilled workers faced by the dominant firms as:

$$S(w^1) = \begin{cases} ts(w^1), & if\ w^1 \geq \overline{w} \\ s(w^1) - d(w^1), & if\ w^1 < \overline{w} \end{cases} . \tag{5}$$

If $\theta(\cdot\ )$ is the inverse of $S(w^1)$, the profit function of firm i is given by

$$\pi_i^L(n_1,...,n_m) = m_i - \theta(n_1 + ... + n_m)n_i, \tag{6}$$

the superscript L being a reminder that this is a model with layoffs.

Figure 10-2 illustrates (5). The right-hand panel of Figure 10-2, shows *d(w)* and *s(w)*, as in Figure 10-1, and in addition illustrates *ts(w)*. The left-panel shows *d(w)* and *(1 - t)s(w)*.

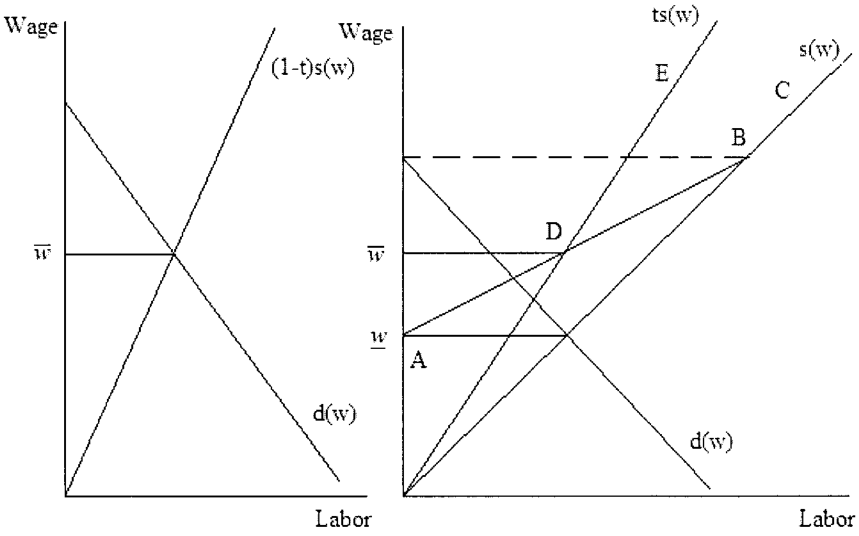

Figure 10-2.

The supply curve (5) is given by the line ADE. Note that in the case without layoffs the supply curve was given by ABC. Hence, the supply now faced by the dominant firms is smaller. But, of course, all laborers employed by the dominant firms now are skilled laborers.

Following the exercise in Section 3.2, we define n^L to be *equilibrium* employment by each firm in game *G(m,L)* if, for all *i*,

$$\pi_i^L(n^L,...,n^L) \geq \pi_i^L(n^L,...,n^L,\ n_i, n^L,...,n^L),$$

for all n_i.

Define,

$$\pi^L(m) = \pi_i^L(n^L,...,n^L).$$

Hence, $\pi^L(m)$ is the profit earned in the period 2 equilibrium by each firm when there are *m* firms and layoffs are permitted.

If $\theta(mn^L) > \overline{w}$, the fringe firms pay a wage of $\overline{w}$ in equilibrium. Otherwise they pay a wage of $\theta(mn^L)$.

3.4 Subgame Perfect Equilibria With or Without Layoffs

It is now easy to work out the subgame perfect equilibria of the two two-period games that we have, one with the second period game $G(m,N)$ and the other with $G(m,L)$. Call the two two-period games G^N and G^L, respectively.

Consider first the case where no layoffs are allowed, that is, game G^N. Let the ordered pair, (t,n), denote the number of firms, t, that decide to enter the industry in period 1 and the number of units of labor, n, that each entrant firm employs in period 2. Recall that if a firm enters and earns π in the post-entry game, the firm's net profit is π - K, and also that k is the total number of potential entrants.

(t^*,n^*) is described as a *subgame perfect* equilibrium of G^N if

i. $t^* = k$ and $\pi^N(k) - K \geq 0$, or

ii $t^* = 0$ and $\pi^N(1) - K \leq 0$, or

iii. $0 < t^* < k$ and $\pi^N(t^*) - K \geq 0$ and $\pi^N(t^* + 1) - K \leq 0$, and

iv. n^* is an equilibrium in $G(t^*, \mathrm{N})$.

The subgame perfect equilibrium of G^L is defined in the same way with π^L replacing π^N and $G(t^*,L)$ replacing $G(t^*,N)$.

If legislation like the *Industrial Disputes Act* is repealed, the labor market game will switch from G^N to G^L. Our aim is to compare the resulting new equilibrium with the equilibrium in G^N. While all kinds of welfare changes are possible, I focus here on the relatively counter-intuitive one and demonstrate that laborers -- all laborers -- can be better off under G^L.

I shall demonstrate this by confining attention to a class of linear models. In particular, I shall assume

$$s(w) = bw \tag{7}$$

$$d(w) = A - Bw \tag{8}$$

Hence (4) now becomes

$$\pi_i^N(n_1,\ldots,n_m) = \left[rt - \frac{\sum n_j}{b+B} - \frac{A}{b+B} \right] n_i.$$

It must be recalled that this is valid, assuming that equilibrium wage does not exceed A/B, because otherwise fringe demand is zero. That is,

$$\frac{m}{m+1}[(b+B)rt - A] \leq \frac{bA}{B}$$

It is easy to compute the amount produced and the profit earned by each firm under symmetric Nash equilibrium or what is simply called the equilibrium in Section 3.2. These are given by

$$n^N = \frac{r(b+B)-A}{m+1} \tag{9}$$

$$\pi^N(m) = \left[\frac{rt(b+B)-A}{m+1} \right]^2 \cdot \frac{1}{b+B} \tag{10}$$

Let us now turn to the case where layoffs are permitted. From the definition of w and using (7) and (8) we get:

$$\overline{w} = \frac{1}{(1-t)b+B}$$

From (5) and (6), it follows that

$$\pi_i^L(n_1,\ldots,n_m) = \begin{cases} \left[r - \dfrac{\sum n_j}{tb} \right] n_i, & if \ \dfrac{\sum n_j}{tb} \geq \dfrac{A}{(1-t)b+B} \\[2ex] \left[r - \dfrac{\sum n_j + A}{b+B} \right] n_i, & if \ \dfrac{\sum n_j}{tb} < \dfrac{A}{(1-t)b+B} \end{cases}$$

Let us suppose that the top line of the above function is relevant. Then

$$n^L = \frac{rtb}{m+1}$$

and

$$\pi^L(m) = \frac{r^2tb}{(m+1)^2}$$

and

$$\frac{mr}{m+1} \geq \frac{A}{(1-t)b+B} \tag{11}$$

The last inequality merely ensures that the top line of the π_i^L function is relevant. If (11) is untrue, n^L and $\pi^L(m)$ will be given by (9) and (10).

Suppose we have a situation where (11) holds and also π^L(m) > π^N (m). That is,

$$r^2tb > \frac{[rt(b+B)-A]^2}{b+B}$$

In that case, we can find a $K > 0$ such that

$$\pi^L(1) > K > \pi^N(1) \tag{12}$$

or

$$\frac{r^2tb}{4} > K > r^2\left[\frac{rt(b+B)-A}{2}\right]^2 \cdot \frac{1}{b+B} \tag{13}$$

If, for instance, $r = 2,\ t = \frac{1}{2},\ b = 2, B = 8,\ A = 8$

and

$$K = \frac{3}{4},$$

then (11) and (12) are true (with $m = 1$ in (11)).

If (11) and (12) are true, there will be no dominant firms in the industry if layoffs are banned, since $\pi^N(1) - K < 0$. But if layoffs are allowed, entry is bound to occur since $\pi^L(1) - K > 0$. So if layoffs are banned all workers will be getting a wage of $\underline{w}$, where, as before, $d(\underline{w}) = s(\underline{w})$. If such a ban is revoked, some dominant firms, which need skilled labor, will enter the industry. They will of course pay their workers a higher wage. And, by virtue of their employing workers, the supply of labor to the fringe sector will fall. Hence, the fringe sector wage will rise.

By confining attention to situations where $\pi^L(1) > K > \pi^N(1)$, we here get the result that there will be no entry of firms using skilled labor if layoffs are not permitted. In that case we can assert that allowing for layoffs will cause an expansion in the sector using skilled labor thereby pulling up wages for all workers. The scenario is not at all unlike what has been observed in many developing countries, such as India. There is skilled labor and a lot of scope for using the labor; but the labor laws prevent adequate use of this labor.

In the present model we focused on one kind of causation. Firms need to try out workers, dismiss some, employ others and so on in order to improve the skill-level of their labor force. Laws which make layoffs impossible discourage the emergence and growth of such firms. There are other kinds of causation, for instance, that involving the workers' choice of effort which could push us towards the same conclusion. The latter is studied in Section 4.

3.5 Costly Layoff

Up to now we focused on the polar cases of "costless layoffs" and "no layoff". Now suppose there is a law that requires that firms pay a compensation of c units to every worker that is dismissed. The firm will then face a choice of either employing randomly picked workers and not laying them off in which case a fraction $(1 - t)$ of its work force will be unproductive as in game $G(m,N)$ or incurring the compensation cost but moving towards a skilled labor force. If the firm wishes to employ n_i workers who are skilled, it could, for instance, employ n_i dismiss the $(1 - t)n_i$ who turn out to be unproductive, employ $(1 - t)n_i$ new workers, dismiss $(1 - t)^2 n_i$ of them who turn out unproductive, and so on. The total number of dismissals will be:

$$\left(\frac{1-t}{t}\right)n_i = (1-t)n_i + (1-t)^2 n_i + (1-t)^3 n_i + \ldots$$

Ignoring the process by which this is achieved and treating it as instantaneous we would write firm *i's* profit function as

$$\pi^i = \left[r - \theta(n_1 + \ldots + n_m) - \frac{c(1-t)}{t} \right] n_i$$

where $\theta(\cdot\ )$ is the same as in (6). Whether a firm finds it worthwhile to replace its unskilled worker with a skilled worker depends on whether

$$r - rt \geq c(1-t)/t.$$

It is quite obvious that if c is large enough we could have the same kind of result as demonstrated above. The dominant firms will either not enter at all or not expand, thereby causing a glut of labor on the fringe market and depressing wages.

3.6 Minimum Wages

The implementation of legal minimum wages is quite ubiquitous across nations. In India, the *Minimum Wages Act*, 1948, empowers the government to announce legally minimum wages and periodically revise these.[x] Failure on the part of an employer to adhere to this is "punishable with imprisonment [up to] six months, or a fine [up to] five hundred rupees, or with both". Despite this, in practice, the minimum wage law is frequently contravened in India.[xi]

In the above model it can be shown that, unlike anti-layoff laws or custom, minimum wage laws have a desirable effect on labor welfare *if they are applied only to the dominant firms and in small measure.* There could be other, more ubiquitous reasons for having minimum wage laws, as argued in Basu, Genicot and Stiglitz (2003), but the analysis here is confined within the structure of the above model.

This is easy to see. The usual reason why some economists argue against the use of a minimum wage is transparent by an appeal to the conventional demand and supply curve analysis of labor-market equilibrium. In such a model, a minimum wage law can indeed raise wages but this also causes unemployment. Hence it benefits some workers but only by hurting other workers.

The above reasoning however does not carry over to the model built in this paper, when the minimum wage law is applied to the oligopsonists. To demonstrate this with the simpler model, suppose that layoffs are not permitted. Consider the game *G(m,N)*, that is, a game with no layoffs and m large firms. Figure 10-3 reproduces *s(w)* and *d(w)* from Figure 10-1.

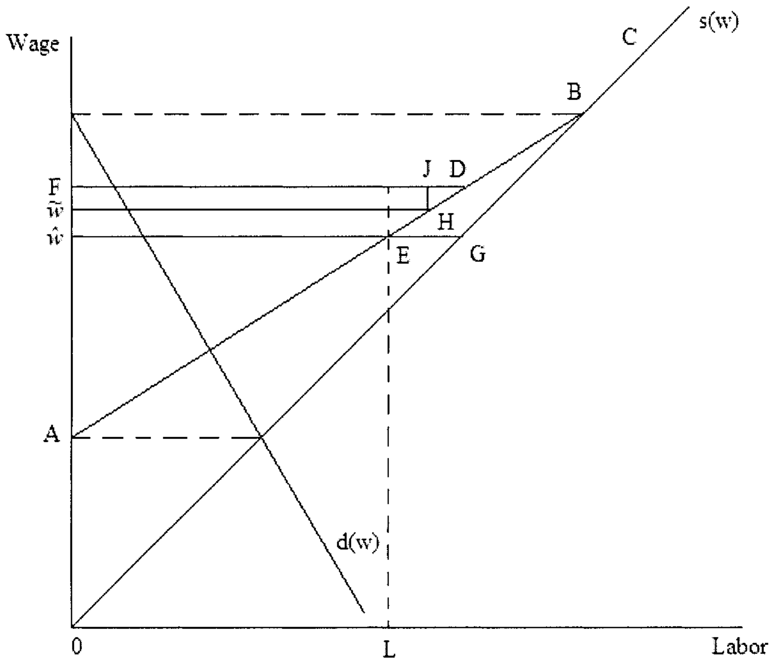

Figure 10-3.

In addition, the figure marks *F* on the wage axis, which shows the expected amount earned by a firm from a randomly chosen laborer (i.e. *OF* = *rt*). Since the supply curve of labor faced by the oligopsonists is *ABC*, a standard analysis suggests that the oligopsonists will employ *OL* laborers (*OL* being *m/(m + 1)* of the length of the line *FD*) with each oligopsonist employing *OL/m* laborers, and equilibrium wage being $\hat{w}$. Even the fringe sector will pay $\hat{w}$ (and employ *EG* laborers) in equilibrium.

Now consider a minimum wage, $\tilde{w}$, a little above $\hat{w}$. That may result in some of the (large) firms to not enter the industry. If $\tilde{w}$ is so high that no firm enters the industry then wage will drop to *OA* in Figure 10-3 and laborers will be worse off. But as long as that does not happen, that is, some large firms continue to produce, the laborers benefit because of the minimum wage legislation. If $\tilde{w}$ is such that $\tilde{w} > \hat{w}$ and $\tilde{w}HJF = K$ (the entry cost), then any minimum wage which is below w, benefits the laborers.[xii]

The above analysis suggests that if the welfare of workers is of concern to the government, it may be best to repeal anti-layoff legislation (that is, as explained earlier, the terms and conditions of layoffs should be decided by the workers and their employer through free contracting) but to have some minimum wage requirements for large firms (not for the small fringe firms).

4. CONCLUSION

Modern economists are aware -- or at any rate ought to be -- that markets often fail, and even when they do not fail they may result in intolerable inequities. From this truism to jump to the conclusion that government should be brought in wherever the market is expected to fail is however a fallacy. Just as markets can fail, so can governments. It is entirely possible that government will act as a handmaiden of vested interests or will simply be ineffective and falter. In the context of labor, markets do often fail. This means that we may need certain kinds of non-market interventions to ensure minimal labor standards and basic worker rights. Yet, the steps from this realization to its execution can be fraught with pitfalls, unless we do this on the basis of very careful theorizing and documenting of facts. The aim of this paper was to sound this warning bell.

This paper evaluated labor laws concerning layoffs and minimum wage legislation in India and, on the basis of a theoretical model, reached the following policy conclusions. It argued that legislation or even *customary practice* which makes the laying off of labor illegal or (exogenously) costly may be harmful for the workers. The same may happen if employers are *a priori* given the freedom to fire workers at will. Instead, workers and employers should have the freedom to develop their own contract concerning the conditions for the dismissal of labor. There will be reason to set limits on the range of contracts allowable (without this, contracts run the risk of becoming so complex that workers may not fully comprehend what they are committing themselves to), but there is reason to believe that a larger freedom than what is currently permitted would benefit workers. I would expect that many firms needing specially skilled labor would, in the first place, come into existence and, secondly, agree to give higher wages and expand their labor force if such an amendment to existing laws were to be made.

This paper argued that there may be a need to impose some legislative restrictions on minimum wages. In particular, workers will benefit by having some minimum wage restrictions, as long as these are not set too high and as long as these are confined to the large firms. Of course, if the layoff legislation is amended as suggested here, the market wage paid by the large firms may well outstrip the legal minimum wage naturally. This is consistent with the experience of Taiwan and Korea (Ranis, 1994).

From these broad do's and don'ts, to answer specific questions, such as how high should minimum wages be, or, when allowing workers to sign contracts to give up the existing right not to be dismissed, what restrictions should be placed on the allowable range of contracts, we need to move to empirical and descriptive research. The purpose of this paper was to provide an analytical basis for such work.

ACKNOWLEDGEMENTS

This is a topic on which I have, over the years, had discussions with Gary Fields, Mrinal Datta Chaudhuri, Shub Debgupta, and Martin Rama. I would like to thank them and also Jan Svejnar for helpful comments.

NOTES

[i] This is not to deny that there are people who are even more disadvantaged (such as the unemployed, the aged, and the infirm), but as a class, workers are, in most societies, poor enough to deserve attention.

[ii] Though my analysis is in terms of pre-existing legislation, this is closely related to the question of trade union empowerment and its effect on labor welfare: see Rama (2000) for discussion.

[iii] There is, in addition, some evidence of employment growth not having kept pace with growth *per se* in the country. One reason for this is a growing tendency in firms to replace labor with capital (Papola, 1989; Nagaraj, 1993). For a detailed micro study of evolving labor market conditions, see Mathur (1991). The problem of layoffs in connection with the closure of firms is documented in Anant et al (1993). For an excellent study of the whole range of labor legislation in India and its effect of worker welfare, see Singh (2000).

[iv] Legislation which is explicitly meant to protect labor interests include *The Trade Unions Act*, 1926; *Industrial Employment (Standing Orders) Act*, 1946; *Industrial Disputes Act*, 1976.

[v] This is also the line taken by Besley and Burgess (2003), who, taking advantage of the fact that labor legislation in India is on the 'concurrent list' (that is, the law can, in principle, be modified by the states) do an inter-state comparison. It is, however, worth keeping in mind that the variability possible across states in reality is less than what is permitted in principle. This is the reason why the erstwhile Chief Minister of Andhra Pradesh, Chandrababu Naidu, had been arguing that labor legislation should be made entirely a

state matter. I believe this is a good suggestion though with the proviso that there should be laws to ensure that the competition across states does not become ruinous.

[vi] The distinction between not giving workers certain rights, *R*, and giving them the rights, *R*, and giving them the further right to waive *R* is discussed in Basu (1997).

[vii] If we assume some wage rigidity, these arguments can easily translate to the effect of anti-dismissal laws on unemployment. Such a law (see, e.g., Layard, Nickell and Jackman, 1994, p. 108) clearly has opposing *ex ante* and *ex post* effects, so that the net impact on employment can be ambiguous. Hence, a law making labor dismissal harder may result in a larger unemployment. Though Layard, Nickell and Jackman reach a similar conclusion, they point out that empirical studies do not provide any clear resolution of this ambiguity.

[viii] We are proceeding under the assumption of there being a unique equilibrium – an assumption that holds true in the class of examples considered below.

[ix] If $w^1 < \underline{w}$, the supply to the dominate firms is, of course, zero.

[x] The Act also specifies other details, like hours of work and, over time rates.

[xi] Compliance with this law turns out to be far from perfect even in developed countries. Ashenfelter and Smith's (1979) study of the U.S. labor market of the early seventies shows that there was 65 percent compliance. They are careful to define compliance not as percentage of laborers who get a wage above the statutory minimum, but as the percentage of workers who earn the statutory minimum wage or lose their jobs among workers who would earn less than the statutory minimum wage in the absence of a minimum wage law. Though I cannot cite evidence, it seems natural to me to expect that minimum wage regulation would be violated more widely than anti-dismissal legislation. This is because when striking a deal for a new job, it may be in the worker's interest to comply with a less than minimum wage (because otherwise he may not get the job) but when served a dismissal notice a worker has no interest in complying with the employer's demand.

[xii] A recent study by Card and Krueger (1994) finds that an upward revision of the minimum wage in New Jersey increased employment in the fast-food industry (see also Card, 1992). As the authors point out, this is not surprising if the labor market is oligopsonistic (though in their study there are other changes which cast some doubt on the market being an oligopsony). See Drazen (1986) for an argument why minimum wage legislation may be efficient.

REFERENCES

Ahluwalia, I.J. (1991), **Productivity and Growth in Indian Manufacturing**, Oxford University Press, New Delhi.

Anant, T.C.A., Gangopadhyay, S. and Goswami, O. (1993), `Industrial Sickness in India: Characteristics, Determinants and History, 1970-1990', mimeo: Indian Statistical Institute, New Delhi.

Ashenfelter, O. and Smith, R.S. (1974), `Compliance with the Minimum Wage Law', **Journal of Political Economy**, Vol. 87.

Basu, K. (1993), **Lectures in Industrial Organization Theory**, Blackwell: Oxford, U.K. and Cambridge, U.S.A.

Basu, K. (1997), `Some Institutional and Legal Prerequisites of Economic Reform in India', in H.E. Bakker and N.G. Schulte Nordholt (eds.), **Corruption and Legitimacy**, SISWO Publications, Amsterdam.

Basu, K. (2004), `Indian Economic Reforms: Up to 1991 and Since', in K. Basu (ed.), **India's Emerging Economy: Performance and Prospects in the 1990s and Beyond**, MIT Press, Cambridge, MA, and Oxford University Press, New Delhi.

Basu, K., Fields, G. and Debgupta, S. (2001), `Alternative Labor Retrenchment Laws and their Effects on Wage and Employment', mimeo: Cornell University.

Basu, K., Genicot, G. and Stiglitz, J. (2003), `Minimum Wage Laws and Unemployment Benefets When Labor Supply is a Household Decision' in K. Basu, P.B. Nayak and R. Ray (eds.) **Markets and Governments**, Oxford University Press, New Delhi.

Besley, T. and Burgess, R. (2003), `Can Labor Regulation Hinder Economic Performance? Evidence from India', BREAD Working Paper No. 44.

Card, D. (1992), `Do Minimum Wages Reduce Employment? A Case of California, 1987-89', **Industrial and Labor Relations Review**, Vol. 46.

Card, D. and Krueger, A.B. (1994), `Minimum Wages and Employment: A Case Study of the Fast-Foot Industry in New Jersey and Pennsylvania', **American Economic Review**, Vol. 84.

Datta Chaudhuri, M. (1990), `Market Failure and Government Failure', **Journal of Economic Perspectives**, Vol. 4.

Datta Chaudhuri, M. (1994), `Labor Markets as Social Institutions in India', CDE Working Paper No. 16, Delhi School of Economics.

Dixit, A. and Stern, N. (1982), `Oligopoly and Welfare: A Unified Presentation with Applications to Trade and Development', **European Economic Review**, Vol. 19.

Drazen, A. (1986), `Optimal Minimum Wage Legislation', **Economic Journal**, Vol. 96.

Edgren, G. (1989), `Structural Adjustment, the Enterprise and the Workers' in G. Edgren (ed.), **Restructuring Employment and Industrial Relations**, I.L.O., Geneva.

Ehrenberg, R. (1994), **Labor Markets and Integrating National Economies**, Brookings Institution, Washington, D.C.

Encaoua, D. and Jacquemin, A. (1980), `Degree of Monopoly, Indices of Concentration and Threat of Entry', **International Economic Review**, Vol. 21.

Fallon, P.R. and Lucas, R.E.B. (1993), `Job Security Regulations and the Dynamic Demand for Industrial Labor in India and Zimbabwe', **Journal of Development Economics**, Vol. 40.

Ghose, A.K. (1994), `Employment in Organized Manufacturing in India', **Indian Journal of Labor Economics**, Vol. 37.

Kannan, K.P. (1994), `Levelling Up or Levelling Down? Labor Institutions and Economic Development in India', **Economic and Political Weekly**, 23 July.

Layard, R., Nickell, S. and Jackman, R. (1994), **The Unemployment Crisis**, Oxford University Press.

Lucas, R.E.B. (1988), `India's Industrial Policy', in R.E.B. Lucas and G.F. Papanek (eds.), **The Indian Economy: Recent Developments and Future Prospects**, Oxford University Press.

Mathur, A. (1991), **Industrial Restructuring and Union Power: Micro-economic Dimensions of Economic Restructuring and Industrial Relations in India**, I.L.O., Geneva.

Nagaraj, R. (1993), `Employment and Wages in Manufacturing Industries in India', Discussion Paper 98, Indira Gandhi Institute of Development Research, Bombay.

Nichol, A.J. (1930), **Partial Monopoly and Price Leadership**, Press of Smith-Edwards Co., Philadelphia.

Papola, T.S. (1989), `Restructuring in Indian Industry: Implications for Employment and Industrial Relations' in G. Edgren (ed.), **Restructuring Employment and Industrial Relations**, I.L.O., Geneva.

Papola, T.S. (1994), `Structural Adjustment, Labor Market Flexibility and Employment', **Indian Journal of Labor Economics**, Vol. 37.

Rama, M. (2000), `Downsizing in the Presence of Monopoly Rights: The Road to Riches', mimeo: Development Research Group, World Bank.

Ramaswamy, E.A. (1984), **Power and Justice**, Oxford University Press.

Ranis, G. (1994), `Labor Markets, Human Capital and Development Performance in East Asia', mimeo, Yale University.

Singh, J. (2000), Some Aspects of Industrial and Labor Markets in India: Perspectives from Law and Economics, Ph.D. Dissertation, Delhi School of Economics, Delhi University.

Shapiro, C. and Stiglitz, J.E. (1984), `Equilibrium Unemployment as a Worker Discipline Device', **American Economic Review**, Vol. 74.

Stigler, G. (1950), `Monopoly and Oligopoly by Merger', **American Economic Review**, Vol. 23.

Thorbecke, E. (2003), `Poverty Analysis and Measurement within a General Equilibrium Framework', in C.M. Edmonds (ed.), **Reducing Poverty in Asia**, Edward Elgar, Cheltenham, U.K.

Thorbecke, E. (2003a), `Conceptual and Measurement Issues in Poverty Analysis', mimeo: Cornell University.

Chapter 11

MACRO MODELS AND MULTIPLIERS: LEONTIEF, STONE, KEYNES, AND CGE MODELS

Sherman Robinson
University of Sussex, sherman.robinson@sussex.ac.uk

1. INTRODUCTION

A major theme in Erik Thorbecke's career has been extensive work on social accounting matrices (SAMs), both from the perspective of data generation and as providing the underpinnings for a wide range of empirical models. His work is in a long tradition of work with SAMS, including seminal contributions by Richard Stone and Graham Pyatt (a coauthor of Thorbecke's), and a great deal of work by the rest of us. The SAM can be seen both as a data base and as a logical framework for economywide economic models, and Thorbecke has contributed to the development of both.

A SAM can be seen as an extension of Leontief's input-output accounts, filling in the links in the circular flow from factor payments to household income and back to demand for products. The SAM delineates flows across product and factor markets, and provides the statistical underpinnings for multi-sector, multi-factor, computable general equilibrium (CGE) models, much as the national accounts provide the data framework for macro-econometric models. Since a SAM includes all economic flows—indeed, provides the organizing framework for the system of national accounts—and is organized around the accounts of all economic "actors" in the economy, the aggregate national accounts and macro models based on them are necessarily embodied in the SAM framework. The converse is also true:

any SAM-based model must, of necessity, incorporate macro aggregates and, implicitly or explicitly, incorporate relationships among them. In particular, CGE models, which are always based on a SAM framework, are theoretically grounded in Walrasian general equilibrium theory, but nonetheless incorporate macro aggregates and macro issues.

There is a longstanding theoretical tension between Walrasian models, with their usual assumptions of market-clearing price-adjustment mechanisms in product and factor markets, and macro models where there is unemployment and very different notions of equilibrium. The literature on CGE models is replete with debates about the macro properties of these models, and a number of different schools of thought have emerged concerning how, or indeed whether, one should incorporate macro features into these SAM-based models. No clear consensus has emerged, which is hardly surprising since the debate really concerns the theoretical divide between Walras and Keynes, and the micro foundations of macro models—or the lack thereof.

In this paper, I discuss a range of models in the SAM framework, from macro-multipliers to CGE models, focusing on the difficulties, pitfalls, and potential insights from different attempts to incorporate macro relationships into SAM-based general equilibrium models. I start with simple multiplier models, and argue that these models capture the essence of the problem of capturing macro phenomena in more elaborate models. I review the various schools of thought, relating the different approaches to the general problem, evident in multiplier models, of specifying macro "closure" in general equilibrium models. Finally, I discuss the theoretical strains that must be accommodated as macro phenomena are incorporated into SAM-based, flow-equilibrium, models.

2. SAMS AND MULTIPLIER MODELS

Table 11-1 presents a simple macro SAM—"macro" because it includes only macro aggregates and excludes intermediate inputs (the input-output table). The SAM is square, entries represent payments from column accounts to row accounts, and the corresponding row and column sums must balance since they represent the double-entry, receipt-expenditure accounts of the various economic actors. For the macro SAM, the row and column balances represent the various macro balances in the national income and product accounts. Table 11-2 spells them out.

Table 11-1. [Macro SAM]

	Activities	Commodity	Factors	Hshld	Govt	S-I	World
Activities		D					E
Commodity				C	G	I	
Factors	X						
Household			Y				
Government	T^X			T^H			
S-I				S^H	S^G		S^F
World		M					

Definitions:	
D: production sold domestically	C: consumption
E: exports	G: government demand
X: production (GDP at factor cost)	I: investment demand
T^X: indirect taxes	S^H: household savings
T^H: direct taxes on households	S^G: government savings
M: imports	S^F: foreign savings
Y: factor payments to households	S-I: savings-investment account

Table 11-2. [Balances in the Macro SAM]

1. $GDP = X + T^X = D + E$
2. $D + M = C + G + I$
3. $GDP + (M - E) = C + G + I$
4. $Y = X = GDP$ (factor cost)
5. $Y = C + T^H + S^H$
6. $T^X + T^H = G + S^G$
7. $I = S^H + S^G + S^F$
8. $M = E + S^F$

The SAM is a compact way to present the national accounts, and nicely traces out the circular flow from production activities to factor payments to incomes of "institutions"—households, government, savings-investment ("S-I")—and back to demand for commodities. GDP at market prices (equation 1, Table 11-2) equals the value of production (*X*, or GDP at factor cost) plus indirect taxes. The "commodity" account represents total "absorption"—the total supply of commodities for use in the economy—and its sum (equations

2 and 3 in Table 11-2) provides a measure of aggregate welfare. As an "actor" in the economy, the commodity account can be seen as a department store which buys domestically produced goods and imports (down the column) and then sells them to other actors in commodity markets (across the row). The "rest of the world" is included as a separate actor, providing imports, buying exports, and financing the difference through foreign savings.

The SAM incorporates the three macro balances: government deficit, trade deficit, and savings-investment balance. The macro balances are expressed as flows—the SAM does not include asset accounts—and any macro relationship in this framework will be in flow terms. In particular, the savings-investment (S-I) account should be seen as representing the "loanable funds" market. The account collects savings from various sources (government, private, and foreign) and spends the accumulated savings on capital goods (I). The SAM provides no information about who "owns" the capital goods or in which sectors they are installed. Investment demand in the SAM is by sector of origin, not sector of destination, so the SAM cannot provide information about changes in sectoral capital stocks, or their valuation. While much of macro theory is concerned with the operation of asset markets, SAM-based models need only incorporate the implications for flows across goods and factor markets (commodities and factor services). There is a potential division of labor between modelers working with flows and those working with assets that has been exploited in much of the SAM-based modeling literature, but the division is somewhat artificial—flow and asset market equilibria are obviously linked—and there are theoretical tensions between the two approaches.

All models in the SAM framework must "explain" how balance is achieved in the three macro accounts. Given that the SAM is always balanced, determining two of the macro balances necessarily determines the third. The SAM represents a closed system—all economic transactions are included—and models in this framework will incorporate Walras' Law in some form. They need (indeed, only can) explain one less than the total number of accounts in the SAM. The task facing modelers, then, does not seem very difficult. A model that may include many sectors and factors, and is based on a truly elegant body of general equilibrium theory, must incorporate and determine two out of the three macro balances. Yet how to achieve this apparently simple task has engendered a large and contentious literature.

A very simple approach to modeling in the SAM framework is to expand the SAM to include demands for intermediate inputs and many activities, commodities, factors, and perhaps households; and then to assume that each actor whose account is given in the SAM behaves according to fixed (column) coefficients. For example, producers ("activities") demand inputs in fixed input-output coefficients, commodities are made up of domestically produced and imported goods in fixed proportions, factor income is distributed to households in fixed proportions; households consume and save in fixed shares of total income, and so forth across the SAM. One can interpret these column coefficients as fixed expenditure shares—all behavior is essentially Cobb-Douglas—or assume that prices are fixed exogenously, and so the coefficients also represent real input-output coefficients for the accounts where there is a real counterpart to the expenditures (activities, factors, commodities).

The SAM presented as coefficients describes behavior of the different actors, and provides a description of their behavior across markets. Equilibrium is defined as row-column balance in all SAM accounts. Given fixed prices, row-column balance in commodity and factor markets also represents supply-demand balance, and in the other accounts it represents just receipt-expenditure balance. But not all the accounts can be included in the behavioral model—something has to be made exogenous or the model is over determined. Traditionally, in multiplier models, some of the accounts in the SAM are assumed to be determined exogenously. The SAM is partitioned, separating endogenous and exogenous accounts. The result is a linear model in which the endogenous accounts are a function of the values of the exogenous accounts. Which accounts are specified as endogenous and exogenous determines the macro "closure" of the model.

In the "open" Leontief input-output model, only the activity and commodity accounts are endogenous and all final demand accounts, including exports, are specified as exogenous. In SAM-multiplier models, pioneered by Stone, some combination of the government, S-I, or world accounts is assumed exogenous. In the open input-output model, the column for factor accounts is not needed, since the household account is exogenous. In SAM-multiplier models, it is included since the household account is usually left endogenous.[xcii] In these models, a change in demand (column entries) by any exogenous account yields equilibrium changes in all endogenous accounts, with changes in the commodity accounts that are generally larger than the exogenous shock—hence the name "multiplier" models. "Equilibrium" in a SAM-multiplier model is defined as row-column (or supply-demand) balance in all endogenous accounts. Given the SAM structure, it must also be true that the solution balances the sum of all the exogenous accounts. The imbalances of these exogenous accounts can be computed from the various SAM coefficients, but their separate balance must be achieved through some unspecified mechanism outside the endogenous part of the multiplier model.

SAM-multiplier models are driven by changes in exogenous demand (column entries) and solve for a resulting change in supply and demand that balances all endogenous accounts. There are no constraints on supply in the model. At fixed prices, all activities are assumed to be able to produce as much as needed to meet the changed demand. Factor demands can be computed from the input-output factor coefficients, but output supply is not constrained by any limits on factor availability.

The richness of SAM multipliers comes from their tracing out chains of linkages from changes in demand to changes in production, factor incomes, household's incomes, and final demands. There is a literature on decomposing these multiplier chains—a literature to which Thorbecke has contributed—and these methods have been widely used to analyze growth linkages in developing countries.[xciii]

While the motivation for this work was not to analyze macro equilibrium, the SAM-multiplier models look very much like the simple Keynesian model where unemployment is assumed and output is determined by demand. It is easy to show in simple models that the SAM-multiplier model will generate a demand multiplier equal to one over one minus the marginal propensity to consume—the Keynesian demand multiplier. A change in exogenous demand by the investment account will yield a multiplied increase in income necessary to generate the increased savings to "finance" the additional investment.

The SAM-multiplier models handle achieving macro equilibrium through induced changes in incomes and demand. Prices are fixed, supply is unconstrained, and savings-investment flow equilibrium is achieved through changes in injections into the loanable funds market given fixed savings rates. No supply side, no price adjustments, no assets, no treatment of time or dynamics. Next, we add the supply side in a model with production and prices, but that still focuses on flows—the CGE model.

3. THE CLASSIC CGE MODEL AND MACRO BALANCES

A CGE model is Walrasian in spirit, incorporating all the flows in the SAM, including production, distribution, and demand; and determining equilibrium wages and prices by simulating the operation of factor and product markets—the classic neoclassical general equilibrium model of production and exchange. The model can only determine relative prices, and some price or price index is chosen to define the numeraire. The absolute price level is indeterminate and must be specified exogenously. In the classic model, the supply and demand equations are all homogeneous of degree zero in prices—double all prices and equilibrium production and demand does not change—so the absolute price level does not matter to the real side. In macro terminology, the model displays strong neutrality of money. Introducing some mechanism to determine the absolute level of prices such as a simple transactions demand for money plus a fixed money supply would determine the absolute price level, but would not affect relative prices or any real magnitudes.

Typically, classic CGE models specify fixed supplies of primary factors of production (e.g., labor and capital) and assume that all markets "clear" in that prices and wages (defined broadly to include rental rates for all factors) adjust to achieve supply-demand equilibrium in all product and factor markets. In macro terms, the model will always generate full employment of all factors and hence the economy is always operating on the production possibility frontier.[xciv] Many applications of CGE models focus on introducing various distortions to the price system and calculating the resulting inefficiencies and loss of welfare. Assuming full employment, however, "inefficiency" is always in terms of being at the wrong place on the production possibility frontier, not from ending up at some point inside the frontier.

3.1 Imports, Exports, and the Balance of Trade

Extending the classic Walrasian CGE model to incorporate foreign trade was a major part of the work program in the development of CGE models. Various approaches were tried, but there is now a broad consensus on the general outlines of a "trade focused" CGE model. Such a model incorporates imperfect substitutability between domestically produced and imported goods, citing early work on specifying import demand functions by Paul Armington.[xcv] The Armington insight was extended to the treatment of exports, and most trade-focused CGE models specify import demand based on sectoral CES (constant elasticity of substitution) "import aggregation" functions and export supply based on sectoral CET (constant elasticity of transformation) "export transformation" functions. There was some debate about whether this treatment was consistent with neoclassical trade theory, but the model is now widely recognized as being an extension of the Salter-Swan model. It is a theoretically consistent generalization of the "standard" trade model with nontraded goods, introducing degrees of substitutability and transformability rather than assuming a rigid dichotomy between tradable and nontradable goods. The theoretical properties of this model have been worked out in detail.[xcvi]

While exports and imports can be accommodated in the classic CGE model, the introduction of a new actor, the "world", does raise a new problem: What do we do about the balance of trade? Trade theory usually ducks the problem by assuming that it is always zero, but applied models must deal with it since the trade balance is rarely zero in actual data. The simplest solution, widely practiced, is to assume that the trade balance is exogenous and the resulting flow of funds is given to (or taken from) some other actors, typically the savings-investment account. This treatment is shown in the SAM in Table 11-1, where the trade balance is identified as "foreign savings" and added to the S-I account.

The introduction of the trade balance raises a macro issue. Treating it as exogenous means that, for some reason outside the theoretical framework of the CGE model, the budget constraint of one actor includes an exogenous transfer. While transfers among actors are easily incorporated into the classic CGE model, the trade balance is special and is the subject of much study in macro theory. Treating it as an exogenous transfer finesses a lot of questions about why foreigners are willing to hold claims against future exports in a model that does not include assets or time. While the treatment in the CGE model is theoretically coherent, the introduction of the trade balance does raise legitimate concerns among macro economists about the limitations of using only a flow-equilibrium specification.

Adding exports, imports, and the trade balance also raises the issue of how the receipt-expenditure account of the new actor, the world, is brought into balance, or equilibrated. As with the Salter-Swan model, trade-focused CGE models include a new equilibrating variable, the real exchange rate, which is the relative price of aggregates of traded and nontraded goods. There is an implicit functional relationship between the real exchange rate and the trade balance. Increasing foreign savings always yields an appreciation of the real exchange rate—the price of nontraded goods rises relative to the price of traded goods (exports and imports).[xcvii] Exports fall as producers shift production toward domestic markets and imports rise as consumers shift demand in favor of imports, bringing the trade balance into equilibrium with the new exogenous higher level of foreign savings.

Most, trade-focused CGE models introduce the exchange rate as an explicit variable, with units of domestic currency per unit of foreign currency. However, the "currency" is not money but simply defines the units of domestic and world prices—domestic prices in local currency units and world prices in foreign currency units (e.g. dollars). The model still contains no assets or money, and the exchange rate is not a "financial" variable in any sense. Changes in the exchange rate work only by changing the relative prices of traded to nontraded goods on domestic markets, affecting export supply and import demand.

The trade-focused CGE model is still very much Walrasian in spirit, determining only relative prices that achieve flow equilibria, and the trade balance can be seen as simply the income-expenditure constraint of the new actor (the "world"). However, that interpretation is strained—the trade balance is also a macro phenomenon, reflecting the operation of asset markets that are outside the CGE theoretical structure. The new actor does not appear to be an optimizing entity in any sense, but simply demands and supplies traded goods (usually at fixed world prices) and finances a specified gap.[xcviii] The trade balance is a "macro" phenomenon unexplained in the CGE model, and the exchange rate is a "macro" variable in the sense that it is an equilibrating variable indirectly affecting domestic producers and consumers to achieve a specified macro balance—the balance of trade. The exchange rate is not a "signal" in some specific market, especially not any kind of financial market, but operates indirectly though its effect on the relative prices of traded and nontraded commodities.

3.2 Savings, Investment, and Government

In addition to the trade balance, CGE models applied to actual economies incorporate savings and the demand for investment goods. The introduction of the S-I account, which collects savings and purchases investment goods, is standard. A new flow equilibrium condition is added to the model—the flow of savings must be made to equal the flow demand for investment goods—and some mechanism is introduced to achieve savings-investment balance. Typically, CGE models specify fixed savings rates by households and assume that whatever is saved is then spent on investment goods. The result is a "savings driven" model of aggregate investment demand.[xcix]

As with the introduction of the trade balance, this treatment is theoretically coherent within the flow-equilibrium specification of the CGE model, but raises a host of questions. Why should actors in a static model save? Why should their savings rate be fixed as shares of income? Why should they purchase investment goods rather than increase consumption? Given the common specification of utility functions as depending on consumption, any diversion of funds to purchase investment goods must decrease welfare.[c] Furthermore, the flow-of-funds specification of the savings-investment account finesses a host of questions. Who owns the new capital stock? Do actors care about asset portfolios?

Savings does not depend on the interest rate or profit rate. Is that a reasonable specification? It is easy to introduce a "loanable funds" market into the CGE model, making saving and investment flows a function of some clearing variable which is interpreted as an interest rate.[ci] The model is still expressed in flows, and the clearing "interest rate" variable equilibrates savings-investment flows, not asset portfolios. Incorporating asset markets requires the introduction of time and dynamics, and also financial variables such as money, inflation, and interest rates. Do we need explicit consideration of financial markets? These issues are central to macroeconomics, and it is difficult conceptually to keep them out of a model that includes savings and investment.

The government in a classic CGE model collects taxes, makes and receives transfer payments, and purchases goods and services. It is hard to see the government as being a utility maximizing actor, so most CGE models treat government as following specified rules of behavior.[cii] For example, a common specification is that government expenditure is fixed in real terms, including transfers; government revenue is determined by fixed tax rates; and government savings is determined residually as the gap between revenue and expenditure. Again, this rule-based treatment raises a number of macro questions. The model treats the government deficit or surplus as coming from the loanable funds market. Thus any government deficit "crowds out" private investment. Is this a reasonable specification from a macro perspective? In this flow-of-funds specification, there is no consideration of how the government finances its deficit. No money, no money creation, no financial markets, and no interest rates.

The discussion above has described a typical CGE model that achieves macro balances (or macro "closure") in a particular way. The model assumes full employment, with wages and prices adjusting to achieve equilibrium in factor and product markets. The balance of trade is fixed exogenously, which determines foreign savings. The real exchange adjusts to achieve the specified trade balance through its affect on aggregate imports and exports. The government has a simple rule-based specification: fixed real expenditure, fixed tax rates, and government savings determined residually. Households have fixed savings rates, which determine private savings. Finally, given that all the components of savings are determined by various rules and behavioral parameters, aggregate investment is specified as "savings driven" and equal to the sum of private, government, and foreign savings.

To summarize the discussion so far, the introduction of macro balances and the three deficits into the CGE model is feasible, remaining within the classic general equilibrium paradigm, with its focus on relative prices and flow equilibria. However, theoretical strains are apparent and one can quickly sense the unease of even sympathetic macro economists, and hostility from others, when this model is presented. The research question is: As you introduce macro balances into the classic general equilibrium framework, at what point do you have to open up the model to include "macro" phenomena such as assets, asset markets, money, interest rates, inflation, expectations, and dynamics? And, given the long historical tension between micro and macro theory—Walras versus Keynes—how should one proceed in formulating empirical models?

4. RECONCILING CGE AND MACRO MODELS

Classic CGE models do not incorporate the sorts of financial and macro variables that would support analysis of macro stabilization. A CGE model, however, provides a good framework for analyzing issues of structural adjustment: the impact of shocks and policies that work through changing prices and market incentives to affect resource allocation and the structures of demand, production, and trade. Short-run stabilization problems involve links between the financial and real sides of the economy: for example, nominal rigidities, non-neutrality of money, financial credit constraints on production, and so forth. The existence of such real-financial links serious strains the Walrasian paradigm underlying the classic CGE modeling framework.

The debate on how best to use empirical models to analyze both stabilization and adjustment issues is active and far from settled. A number of schools of thought have emerged on how best to adapt and use CGE models for these purposes, and how to reconcile or integrate the CGE and macro modeling frameworks.

The orthodox school. This school of thought can be summed up in the maxim: "There is only one model, and its prophet is Walras." In this view, the Walrasian CGE model is theoretically elegant and complete, and any attempt to add macro features and financial variables simply corrupts the model. CGE models should only be used to analyze issues of allocative efficiency, relative prices, and the structure of employment, production, and demand in an environment of well-functioning markets. Macro issues should be left to macro economists.[ciii]

The eclectic school. In this view, one should build integrated models that incorporate the best elements from Walrasian CGE models and a variety of macro and financial models. The multisector CGE model can be used to provide the supply side of a much richer integrated macro-CGE model that includes assets, including money, asset markets, interest rates, inflation, expectations, and any other features drawn from modern macro theory that seem appropriate to the issues being analyzed.[civ] These models all embed a CGE model in a dynamic macro model that includes financial variables and asset markets. This literature is very active, and many macro-econometric models now include some kind of CGE model to provide the supply side.

The ecumenical school. The philosophy of this school is to use separate CGE and macro-financial models and keep them separate, but specify ways through which the models can talk to one another and cooperate. For example, the models can be linked through variables that are endogenous in one but exogenous in the other.[cv] A macro model might determine the price level and various macro aggregates as endogenous, and these variables are then specified as exogenous variables in a CGE model. The CGE model, in turn, might determine endogenously variables such as the wage rate and various prices which are taken as exogenous in the macro model. The two models can communicate, and perhaps be solved simultaneously. The advantage of this approach is that the two modeling systems are kept separate, with no need to mix paradigms in a single model.

In the past decade, for good reasons, the influence of the orthodox school has declined, while the eclectic school has clearly grown. First, it has been widely recognized that, as argued above, any applied economywide model must incorporate macro balances and notions of macro equilibrium. While it is possible to include such macro flow equilibria in the Walrasian CGE model, the theoretical strains are serious. If one has to extend the general equilibrium model to incorporate macro phenomena, one might as well go all the way, and incorporate elements of modern macro theory as required. Second, advances in modeling software and solution algorithms have made the specification and solution of forward-looking dynamic CGE models incorporating asset markets and inter-temporal optimization feasible. Third, increased data availability and advances in econometric methods have made it feasible to estimate the parameters of such models. The research frontier in this area is advancing rapidly.

Any attempt to incorporate assets and asset markets into CGE models requires extension of the underlying SAM framework to incorporated asset accounts. Earlier work in this tradition used a "financial SAM" of "FSAM" which extends the flow accounts to include the capital/asset accounts of all major economic actors.[cvi] Such an extended data framework is often difficult to implement in practice, since it requires reconciliation of national income and product accounts with financial flow accounts.

While there is active research developing integrated macro-CGE models, the ecumenical approach still has much to recommend it. In particular, extending the classic CGE model to include dynamics and macro features requires opening up the model to incorporate the possibility of unemployment and feedbacks from financial variables to the real side of the economy. The problem is that the classic CGE model, with its assumptions of market clearing and focus on relative prices, is an uneasy host for such macro phenomena. One approach to this problem is to use a separate macro model to determine variables such as aggregate employment, investment, government expenditure, and the three macro balances, and then impose these results onto the CGE model structure. While this "top down" approach is consistent with the ecumenical school, figuring out how to adapt a CGE model to incorporate these effects is also a crucial element in constructing any integrated macro-CGE model.

The eclectic school research program thus needs to build on the ecumenical approach, which can provide examples of ways of adapting the classic CGE model to be an adequate host for including macro phenomena. One can usefully start from the notion of macro closure of the CGE model: "What kinds of macro closures can we specify for the CGE model that will make it behave in ways that reflect the outcomes of macro models." Starting with a simple macro problem, we have come full circle: "How can we make a CGE model behave like a Keynesian multiplier model, where demand elements are an important determinant of aggregate employment and supply?"

5. MACRO CLOSURE, MULTIPLIERS, AND CGE MODELS

The issue of macro closure in CGE models brings together the different strands of analysis: SAM multiplier models, CGE models, and macro models.[cvii] At one extreme, it is possible to make a CGE model behave like a fixed-price multiplier model by assuming that primary input prices are all fixed and that supplies of primary inputs are unconstrained—a specification that is reasonable for models of regions within a country.[cviii] Multiplier analysis in such models requires specification of how demand shocks feed through the economy. Similarly, to use a CGE model either with or within a macro model requires extension of the core model to incorporate macro features. Specifying the macro closure of the CGE model is an essential part of that process.

There is a large literature on issues of macro closure of CGE models.[cix] The essential problem that has to be tackled is that the classic CGE model, in which all markets clear, yields a full-employment equilibrium and market-clearing prices, while short-run macro models typically involve wage and price rigidities, partial adjustment mechanisms, and equilibrium without market clearing, including unemployment. The two paradigms embody very different notions of equilibrium.[cx]

There is a literature on "structuralist" CGE models which embody elements of short-run macro models, including "demand driven" Keynesian equilibria with unemployment.[cxi] These models do not explicitly incorporate financial variables and asset markets, but manage to work within the flow-equilibrium structure of CGE models. They effectively impose a macro story onto the CGE model structure—the ecumenical modeling philosophy.

To see how these models work, we will consider two simple CGE models, a closed-economy model and an open economy model. For each model, we will consider four alternative macro closures that illustrate how the results from different macro models can be grafted onto the CGE model.

5.1 Closed Economy Model

Table 11-3 presents the equations of a one-sector supply-side macro model, Table 11-4 lists the variables, and Table 11-5 presents the SAM associated with the model. The model is "short run" in that capital is assumed to be fixed and labor is the only variable input into production. There are eleven endogenous variables and eleven equations. The model satisfies Walras' Law and thus one equation is redundant. Instead of dropping an equation, we add an additional endogenous variable, *WALRAS,* that must have the value zero at the solution. The top part of Table 11-4 lists the variables that are potentially involved in specifying the macro closure of the model. To define a macro closure, two variables from this list must be specified as endogenous, and all the rest are exogenous.

The first two macro closures in Table 11-4, Neoclassical and Johansen, both assume that both product and labor markets clear, with wages adjusting to yield full employment. In these variants, macro issues are only "compositional" in that they determine the composition of final demand (among C, I, and G), but have no effect on aggregate employment and hence GDP. In the Neoclassical Closure, aggregate investment is determined by aggregate savings, which in turn are determined endogenously through the fixed savings rate out of after-tax income and the government deficit. In Johansen Closure, aggregate investment is assumed to be fixed exogenously, and the savings rate is assumed to adjust to generate the required savings.[cxii] Johansen (1974) explicitly argues that macroeconomic fiscal and monetary polices, presumably outside of the CGE framework, will ensure that savings are generated to "balance" investment.

The second pair of macro closures are more interesting in that they involve Keynesian demand multipliers that affect employment and GDP. In the first Keynesian closure, Keynes 1, the wage is chosen as numeraire, which is common in structuralist macro models by Taylor (1990, 2004). Aggregate investment is fixed exogenously, but instead of adjusting the savings rate, the labor supply is assumed to be endogenous. Adjustment in the real wage is the macro equilibrating mechanism, while the aggregate price level is the equilibrating variable. Assume, starting from equilibrium, that real investment is increased. Savings must be increased to finance the increased investment. The only way that savings can increase is through an increase in income, which requires an increase in employment and output, which requires a fall in the real wage (given that firms are on their demand curve for labor), which in turn requires an increase in the aggregate price level. Assuming no government, this model will behave exactly like the simple Keynesian multiplier model. An increase in investment will yield a new equilibrium level of output which equals 1/(1-mpc) times the increase in investment. Note that the Keynesian multiplier will be exactly the same if the price level is chosen as the numeraire and the wage becomes the macro equilibrating variable. The equilibrating mechanism is the same—the real wage adjusts to generate the employment necessary to generate the income necessary to generate savings equal to aggregate investment.

One problem with the Keynesian macro (multiplier) closure is that any increase in employment is associated with a fall in the real wage—firms are assumed to be on their demand curves for labor, and the wage must fall to induce an increase in labor demand. It has been argued in the macro literature that this result is empirically unrealistic. The Keynes 2 closure specifies a different mechanism. In this case, the price level is chosen as numeraire and the real wage is assumed to be fixed. The aggregate labor supply is assumed to be free, so employment is not fixed. In contrast to Keynes 1 closure, however, firms are assumed not to be on their demand curves for labor. A "labor distortion" parameter, gamma (γ), is introduced which measures the degree to which the wage deviates from the marginal product of labor. Gamma adjusts so that firms are "induced" (given the first order conditions in equation 2) to hire the labor at the fixed wage necessary to generate the income (and output) necessary to generate the savings necessary to finance investment.

The interpretation of gamma depends on the macro story that is used to justify this mechanism. Following Barro and Grossman (1976), one can assume that product and labor markets are out of equilibrium, and that firms are forced off their labor demand curves. Gamma simply measures how far off they are, but is not a "signal" in any sense in the CGE model. Another interpretation, following Malinvaud (1977), is that firms are demand constrained and rationed in the product market.[cxiii] In this interpretation, gamma measures the degree of rationing in the product market. In both cases, the equilibrating mechanism is the same—employment is demand determined, the Keynesian multiplier operates, and the real wage is fixed.

Table 11-3. [One Sector Macro Model]

$$X = F(L, \bar{K}) \tag{1}$$

$$W_L = \gamma \cdot P^X \cdot \frac{\partial F(L, \bar{K})}{\partial L} \tag{2}$$

$$W_K \cdot \bar{K} = P^X \cdot X - W_L \cdot L \tag{3}$$

$$\tilde{Y}^H = W_K \cdot \bar{K} + W_L \cdot L \tag{4}$$

$$\tilde{T} = t^H \cdot Y^H \tag{5}$$

$$\tilde{S}^G = \tilde{T} - P^X \cdot G \tag{6}$$

$$\tilde{S}^H = MPS^H \cdot \left(\tilde{Y}^H - \tilde{T}\right) \tag{7}$$

$$P^X \cdot C = \tilde{Y}^H - \tilde{T} - \tilde{S}^H \tag{8}$$

$$P^X \cdot I = \tilde{S}^G + \tilde{S}^H + WALRAS \tag{9}$$

$$L = L^S \tag{10}$$

$$X = C + I + G \tag{11}$$

Table 11-4. [Macro Closure Rules for One-Sector Model]

		Full employment		Unemployment	
Variable	**Description**	**Neoclassical**	**Johansen**	**Keynes 1**	**Keynes 2**
Potential macro closure variables					
P^X	Price of output	Numeraire	Numeraire		Numeraire
W_L	Wage of labor, L			Numeraire	Fixed
L^S	Supply of labor, L	Fixed	Fixed		
γ	Wage distortion	Fixed	Fixed	Fixed	
I	Investment demand		Fixed	Fixed	Fixed
MPS^H	Savings rate	Fixed		Fixed	Fixed
$\overline{K}$	Supply of capital, K	Fixed	Fixed	Fixed	Fixed
G	Government demand	Fixed	Fixed	Fixed	Fixed
t^H	Tax rate on income	Fixed	Fixed	Fixed	Fixed
Other endogenous variables					
X	Output				
L	Demand for labor, L				
C	Consumption demand				
$\tilde{Y}^H$	Household income				
$\tilde{S}^G$	Government saving				
$\tilde{S}^H$	Household saving				
$\tilde{T}$	Tax revenue				
WALRAS	Walras' Law variable				
W_K	Wage of capital, K				

Table 11-5. [Social Accounting Matrix (SAM) for One-Sector Model]

	Expenditures				
Receipts	Activities	Factors	Household	Savings	Government
Activities			$P^X \cdot C$	$P^X \cdot I$	$P^X \cdot G$
Factors	$W_K \cdot \overline{K} + W_L \cdot L$				
Household		$\tilde{Y}^H$			
Savings			$\tilde{S}^H$		$\tilde{S}^G$
Government			$\tilde{T}$		
Total	$P^X \cdot X$	$\tilde{Y}^H$	$\tilde{Y}^H$	$P^X \cdot I$	$\tilde{T}^G$

5.2 Open Economy Model

Table 11-6 presents the second model, which adds exports, imports, and foreign savings. Table 11-7 lists the variables, and the model SAM is presented in Table 11-8. The model is an extension of the 1-2-3 model of Devarajan, Lewis, and Robinson (1987, 1990).[cxiv] In the 1-2-3 model (1 country, 2 products, and 3 commodities), the economy produces aggregate output, X (equation 1), which is "transformed" (equation 2) into two products: D which is sold domestically and E which is exported. A third commodity, M, is imported but not produced domestically. The commodities D and M are imperfect substitutes in demand (equation 3), while D and E are imperfectly "transformable" in production delivered to domestic and foreign markets (equation 2). This model is a good simplified representation of most trade-focused CGE models. The 1-2-3-2 version adds a production function and two factors, K and L.

The model has 20 equations and 20 endogenous variables (world prices of exports and imports are assumed fixed). A new macro balance is included, the trade balance or foreign savings, and a new equilibrating variable, the exchange rate. As discussed above, in this class of trade-focused CGE model, there is a relationship between the real exchange rate and the trade balance. Any increase in foreign savings is associated with an appreciation (decrease) in the real exchange rate. Given the new macro balance, to define macro closure, three of the potential closure variables must be specified as endogenous.

As with the closed-economy model, there are two full-employment macro closures and two involving unemployment. In the Neoclassical Closure, the trade balance (foreign savings) is assumed fixed, and the real exchange rate adjusts to achieve equilibrium. In this case, with fixed foreign savings, investment is savings driven, and the model will behave very much like the closed-economy model.

The second closure, "Foreign", is more interesting. In this case, investment is fixed and foreign savings is endogenous. Instead of the domestic savings rate adjusting to achieve savings-investment equilibrium as in Johansen Closure, the equilibrating mechanism is through changes in foreign savings. The macro equilibrating variable is the real exchange rate. A change in investment will induce a change in the real exchange rate to generate a change in the trade balance (foreign savings) to balance the change in investment. The model assumes full employment—the wage varies to clear the labor market, given the exogenous labor supply.[cxv]

Table 11-6. [1-2-3-2 Model with Factor Markets and Macro Closures]

$$X = F_X(L,\bar{K}) \tag{1}$$

$$X = G_X(D,E) \tag{2}$$

$$Q = F_Q(D,M) \tag{3}$$

$$P^X \cdot X = P^D \cdot D + P^E \cdot E \tag{4}$$

$$P^Q \cdot Q = P^D \cdot D + P^M \cdot M \tag{5}$$

$$P^M = R \cdot \bar{\pi}^M \tag{6}$$

$$P^E = R \cdot \bar{\pi}^E \tag{7}$$

$$W_L = P^X \cdot \frac{\partial X}{\partial L} \tag{8}$$

$$W_K = P^X \cdot \frac{\partial X}{\partial K} \tag{9}$$

Table 11-6. [Cont.]

$$\frac{E}{D} = g_X(P^D, P^E) \tag{10}$$

$$\frac{M}{D} = f_Q(P^D, P^M) \tag{11}$$

$$\tilde{Y}^H = W_L \cdot L + W_K \cdot \bar{K} \tag{12}$$

$$\tilde{T} = t^H \cdot \tilde{Y}^H \tag{13}$$

$$\tilde{S}^G = \tilde{T} - P^Q \cdot G \tag{14}$$

$$\tilde{S}^H = MPS^H \cdot (\tilde{Y}^H - \tilde{T}) \tag{15}$$

$$P^Q \cdot C = \tilde{Y}^H - \tilde{T} - \tilde{S}^H \tag{16}$$

$$P^Q \cdot I = \tilde{S}^H + \tilde{S}^G + R \cdot S^F + WALRAS \tag{17}$$

$$\bar{\pi}^M \cdot M - \bar{\pi}^E \cdot E = S^F \tag{18}$$

$$L = L^S \tag{19}$$

$$Q = C + I + G \tag{20}$$

Table 11-7 [Macro Closure Rules for 1-2-3-2 Model]

		Full employment		**Unemployment**	
Variable	**Description**	**Neoclassical**	**Foreign**	**Keynes 1**	**Keynes 3**
Potential macro closure variables					
P_X	Price of X	Numeraire	Numeraire		
W_L	Wage of L			Numeraire	Numeraire
R	Exchange rate				Fixed
I	Investment demand		Fixed	Fixed	Fixed
S^F	Trade balance	Fixed		Fixed	

Table 11-7.[Cont.]

		Full employment		**Unemployment**	
Variable	**Description**	**Neoclassical**	**Foreign**	**Keynes 1**	**Keynes 3**
L^S	Supply of labor	Fixed	Fixed		
G	Government demand	Fixed	Fixed	Fixed	Fixed
$\bar{K}$	Capital in X	Fixed	Fixed	Fixed	Fixed
MPS^H	Savings rate	Fixed	Fixed	Fixed	Fixed
t^H	Tax rate on income	Fixed	Fixed	Fixed	Fixed
Other variables					
X	Aggregate output				
D	Domestic good				
E	Export good				
M	Import good				
Q	Composite good				
L	Labor in X				
P_Q	Price of Q				
P_D	Price of D				
P_M	Price of M				
P_E	Price of E				
$\bar{\pi}_M$	World price of M	Fixed	Fixed	Fixed	Fixed
$\bar{\pi}_E$	World price of E	Fixed	Fixed	Fixed	Fixed
W_K	Wage of K				
C	Consumption demand				
$\tilde{Y}^H$	Household income				
$\tilde{S}^G$	Government saving				

Table 11-7. [Cont.]

		Full employment		Unemployment	
Variable	**Description**	**Neoclassical**	**Foreign**	**Keynes 1**	**Keynes 3**
$\tilde{S}^H$	Household saving				
$\tilde{T}$	Tax revenue				
WALRAS	Walras' Law variable				

Table 11-8 [SAM for 1-2-3-2 Model]

	Expenditures						
Receipts	Activities	Com	Fctrs	Hshld	S-I	Govt	World
Activities		$P^D \cdot D$					$P^E \cdot E$
Commodity				$P^Q \cdot C$	$P^Q \cdot I$	$P^Q \cdot G$	
Factors	$W_K \cdot \overline{K} + W_l \cdot L$						
Household			$\tilde{Y}^H$				
Saving (S-I)				$\tilde{S}^H$		$\tilde{S}^G$	$R \cdot S^F$
Government				$\tilde{T}$			
World		$P^M \cdot M$					
Total	$P^X \cdot X$	$P^Q \cdot Q$	$\tilde{Y}^H$	$\tilde{Y}^H$	$P^Q \cdot I$	$\tilde{T}$	

The first Keynesian Closure, Keynes 1, operates in the same manner as in the closed-economy model. The trade balance is assumed to be fixed exogenously, with the real exchange rate adjusting to achieve exports and imports consistent with external equilibrium. Changes in the price level change the real wage, generating employment, output, income, and savings consistent with the specified aggregate investment. Since foreign savings are fixed, they play no role in this adjustment.[cxvi] The Keynesian multiplier will operate pretty much the same as in the closed-economy model. In this variant, as in the closed economy, any increase in employment will be associated with a fall in the real wage.

The second Keynesian Closure, Keynes 3, adds a new twist. In this closure, which follows the treatment in a structuralist macro-CGE model of Brazil, the wage is chosen as numeraire, the exchange rate is set exogenously, and foreign savings is endogenous.[cxvii] In this model, as the price level adjusts to change the real wage, as in the Keynesian equilibrating mechanism, the real exchange rate also changes. An increase in the price level lowers the real wage, increasing employment, income, and savings. However, the price rise also appreciates the real exchange rate, increasing imports, reducing exports, and thus inducing an increase in foreign savings. Achieving macro equilibrium after an increase in aggregate investment is now achieved through two sources of savings: increased income generating higher savings and increased foreign savings. The result is a much smaller Keynesian income-employment multiplier.[cxviii]

6. CONCLUSION

There are a few lessons from the literature on SAM multipliers and macro closure in CGE models.

Any realistic economywide model must include macro aggregates and some mechanism for achieving macro equilibrium. The SAM structure is a powerful framework for delineating the links between micro and macro "actors" and provides the underlying data for any micro-macro modeling, including multiplier models.

There are strong links between structuralist macro models and SAM multiplier models. SAM multiplier models represent one end of a continuum, a fix-price model that is completely driven by demand. At the other end is a classic, full-employment, Walrasian CGE model. The different macro closure models range along the continuum.

Any CGE model in which factor markets are assumed to clear will generate full employment. Any macro model linked to such a CGE model can only have compositional effects—macro forces will affect the composition of final demand (C, I, and G), but not employment or GDP.

To extend the CGE model to incorporate aggregate employment effects requires some specification of neoclassical disequilibrium in the factor markets, which certainly stretches the Walrasian paradigm underlying the CGE model.

It is feasible to impose a variety of structuralist macro features on CGE models. While the justification for these features comes from macro theory outside the flow equilibrium structure of the CGE model, it is not necessary to incorporate features such as financial flows, asset markets, money, inflation, and interest rates directly into the CGE model. Many features of

Keynesian demand-driven multiplier models can be accommodated within the flow-equilibrium structure of a CGE model.

The literature on macro closure indicates the robustness of an ecumenical modeling strategy, keeping macro models and CGE models separate, but linking them by imposing the results from the macro models on the CGE models.

The work program on integrating CGE and macro-financial models, while active, has proceeded very slowly, with fits and starts. Part of the problem is that the literature on macroeconomic theory and practice has not always been very helpful. Macro theorists came to recognize the deficiencies of demand-driven models with very weak supply sides, but then swung to the opposite extreme, building long-run, dynamic rational expectations models with strong roots in neoclassical growth theory, but with weak empirical foundations. The pendulum is again in motion, and there are now a number of examples of empirical dynamic macro models which incorporate simple CGE models to provide the supply side. From the CGE side, there is an active literature on dynamic CGE models which incorporate many features from the theoretical literature on dynamic models, including forward-looking expectations.[cxix] Recent advances in modeling software have made the implementation of large, dynamic, empirical multisector models feasible. It seems to be a propitious time to link the two strands of work.

NOTES

[i] In "type 2" input-output multiplier models, the factor accounts and the household account are left endogenous.

[ii] See Defourny and Thorbecke (1984) on path decomposition of SAM multipliers. See also Pyatt and Roe (1977), Chapter 4, and Robinson and Roland-Holst (1988), who decompose total multipliers into direct and indirect components.

[iii] Adding a labor supply function does not change the essential properties of this model. It is still a full-employment model, solving for a market wage that clears the labor market.

[iv] Armington (1969).

[v] See, for example, Dervis, de Melo, and Robinson (1982); de Melo and Robinson (1989); Devarajan, Lewis, and Robinson (1990, 1993); de Melo and Tarr (1992); and Thierfelder and Robinson (2002).

[vi] The theoretical properties of the real exchange rate in this model are worked out in Devarajan, Lewis, and Robinson (1993).

[vii] Expanding the "world" by adding more countries and solving for world prices endogenously in a multi-country CGE model does not make the interpretation of trade balances any easier. Such global models typically do not include explicit treatment of asset markets, and so specify and solve for only flow equilibria in world commodity markets. Most world trade models typically specify fixed national trade balances,

recognizing that that they largely depend on the operation of asset markets that are outside the CGE structure.

[viii] This is an example of a macro "closure" of the CGE model. Other examples will be discussed below.

[ix] It is feasible, and sometimes done, to specify utility functions such as the extended linear expenditure system (ELES) that explicitly include savings, but there is still no tracking of the uses of the savings.

[x] See, for example, Lewis (1985) who introduced loanable funds markets into a CGE model to examine issues of financial repression and credit constraints.

[xi] There are exceptions in the public finance literature where government is treated as analogous to a household, with its own utility function. See Shoven, and Whalley (1992).

[xii] Examples of work expressing this view include Whalley and Yeung (1984), Bell and Srinivasan (1984), and Srinivasan (1982).

[xiii] Examples of work in this tradition include McKibbin and Sachs (1991); McKibbin and Wilcoxen (1998); Agénor and Montiel (1996); Borguignon, de Melo, and Suwa (1991); Thorbecke (1991). Robinson (1991) surveys models which incorporate asset markets in a dynamic CGE framework.

[xiv] Robinson and Tyson (1984) formalized the approach, and Powell (1981) describes applications using the Orani model of Australia.

[xv] See Bourguignon, de Melo, and Suwa (1991), Thorbecke (1991), and Robinson (1991) for a discussion of FSAMs and their use in integrated CGE/macro models. Taylor (2004), Chapter 2, uses an elegant integrated flow/asset SAM to provide an analytic framework for a discussion of structuralist macro models.

[xvi] For other discussions of the links, see Thorbecke (1985) and Robinson and Roland-Holst (1988).

[xvii] See, for example, Thorbecke (1998) and Hoffmann, Robinson, and Subramanian (1996).

[xviii] See, for example, Sen (1963), Taylor (1990), Rattso (1982), Robinson (1989, 1991), and Dewatripont and Michel (1987).

[ivx] Malinvaud (1977) discusses the different notions of "equilibrium" in macro and general equilibrium models.

[xx] See Taylor (1983, 1990).

[xxi] This closure is named after Leif Johansen, who used it in the first CGE model (originally published in 1960). See Johansen (1974).

[xxii] A standard practice in structuralist macro models is to assume that the output price is set by some markup rule, rather than through profit maximization, and then output is demand driven. Again, see Taylor (1983, 1990).

[xxiii] See also de Melo and Robinson (1989).

[xxiv] Examples of applied models using this macro closure include Ahluwalia and Lysy (1981) and Devarajan and de Melo (1987). Devarajan and de Melo specify the exchange rate as numeraire and assume that the real exchange rate adjusts through changes in the aggregate price level.

[xxv] This is not quite true. Foreign savings are defined as fixed in "foreign" currency units. Changes in the exchange rate will change their value in domestic currency units.

[xxvi] See Lysy and Taylor (1980).

[xxvii] Adelman and Robinson (1988) compare the impact of this closure to the standard Keynesian closure in models of Brazil and Korea.

[xxviii] See, for example, Diao, Roe, and Yeldan (1999).

REFERENCES

Adelman, Irma, and Sherman Robinson. 1988. "Macroeconomic Adjustment and Income Distribution: Alternative Models Applied to Two Economies." *Journal of Development Economics*, Vol. 29, pp. 1-22.

Armington, Paul S. (1969). "A Theory of Demand for Products Distinguished by Place of Production." *IMF Staff Papers*, Vol. 16, pp. 159-176.

Ahluwalia, Montek and Frank Lysy (1981). "Employment, Income Distribution, and Programs to Remedy Balance of Payments Difficulties." In W. R. Cline and S. Weintraub, eds., *Economic Stabilization in Developing Countries*. Washington: Brookings Institution.

Agénor, Pierre-Richard and Peter J. Montiel (1996). *Development Macroeconomics*. Princeton: Princeton University Press.

Barro, Robert and Herschel Grossman 1976). *Money, Employment and Inflation*. Cambridge: Cambridge University Press.

Bell, Clive and T. N. Srinivasan (1984). "On the Uses and Abuses of Economywide Models in Development Planning Analysis." In M. Syrquin, L. Taylor, and L.E. Westphal (eds.), *Economic Structure and Performance*. New York, NY: Academic Press.

Bourguignon, Francois, Jaime de Melo, and A. Suwa (1991). "Modelling the Effects of Adjustment Programs on Income Distribution." *World Development* Vol. 19, No. 11, pp. 1527-1544.

Defourny, Jacques and Erik Thorbecke (1984). "Structural Path Analysis and Multiplier Decomposition within a Social Accounting Matrix Framework." *The Economic Journal*, Vol. 94 (March), pp. 111-136.

Dervis, Kemal, Jaime de Melo, and Sherman Robinson (1982). *General Equilibrium Models for Development Policy*. Cambridge: Cambridge University Press. Reprinted by the World Bank, 1989.

Devarajan, Shantayanan and Jaime de Melo (1987). "Adjustment with a Fixed Exchange Rate: Cameroon, Côte D'Ivoire, and Senegal." *World Bank Economic Review*, Vol. 1, pp. 447-487.

Devarajan, Shantayanan, Jeffrey D. Lewis, and Sherman Robinson (1990). "Policy Lessons from Trade-Focused Two-Sector Models." *Journal of Policy Modeling*, Vol. 12, No. 4, pp. 625-657.

Devarajan, Shantayanan, Jeffrey D. Lewis, and Sherman Robinson (1993). "External Shocks, Purchasing Power Parity, and the Equilibrium Real Exchange Rate." *World Bank Economic Review*. Vol. 7, No. 1, pp. 45-63.

Dewatripont, Mathias and Gilles Michel (1987). "On Closure Rules, Homogeneity, and Dynamics in Applied General Equilibrium Models." *Journal of Development Economics*, Vol. 26, No. 1 (June), pp. 65-76.

Diao, Xinshen, Terry Roe, and Erinc Yeldan (1999). "Strategic Policies and Growth: An Applied Model of R&D-Driven Endogenous Growth." *Journal of Development Economics*, Vol. 60, pp. 343-380.

Hoffman, Sandra, Sherman Robinson, and Shankar Subramanian (1996). "The Role of Defense Cuts in the California Recession: Computable General Equilibrium Models and Interstate Factor Mobility." *Journal of Regional Science*, Vol. 36, No. 4, pp. 571-595.

Johansen, Leif (1974). *A Multi-Sectoral Study of Economic Growth*. Second enlarged edition. Amsterdam: North-Holland.

Lewis, Jeffrey D. (1985). "Financial Liberalization and Price Rigidities in a General Equilibrium Model with Financial Markets." Development Discussion Paper No. 211, Harvard Institute for International Development, Harvard University, Cambridge, MA.

Lysy, Frank and Lance Taylor (1980). "A Computable General Equilibrium Model for the Functional Distribution of Income: Experiments for Brazil, 1959-71." In L. Taylor et al., eds., *Models of Growth and Distribution for Brazil.* Oxford: Oxford University Press.

Malinvaud, Edmond (1977). *The Theory of Unemployment Reconsidered.* Oxford: Basil Blackwell.

McKibbin, Warwick J. and J. Sachs (1991). *Global Linkages: Macroeconomic Interdependence and Cooperation in the World economy.* Washington: Brookings Institution.

McKibbin, Warwick J. and Peter Wilcoxen (1998). "The Theoretical and Empirical Structure of the G-Cubed Model." *Economic Modelling,* Vol. 16, No. 1, pp. 123-148.

de Melo, Jaime, and Sherman Robinson (1989). "Product Differentiation and the Treatment of Foreign Trade in Computable General Equilibrium Models of Small Economies." *Journal of International Economics.* Vol. 27, Nos. 1-2, pp. 47-67.

de Melo, Jaime and David Tarr (1992). *A General Equilibrium Analysis of U.S. Trade Policy.* Cambridge: MIT Press.

Powell, Alan (1981). "The Major Streams of Economy-Wide Modelling: Is Rapprochement Possible?" In J. Kmenta and J. B. Ramsey, eds. *Large Scale Econometric Models: Theory and Practice.* Amsterdam: North-Holland.

Pyatt, Graham (1988). "A SAM Approach to Modeling." *Journal of Policy Modeling,* Vol. 10, pp. 327-352.

Pyatt, Graham and Alan Roe (1977). *Social Accounting for Development Planning.* Cambridge: Cambridge University Press.

Pyatt, Graham and Jeffery Round (1988). *Social Accounting Matrices: A Basis for Planning.* Washington, DC: World Bank.

Pyatt, Graham and Erik Thorbecke (1976). *Planning Techniques for a Better Future.* Geneva: ILO.

Rattso, Jorn (1982). "Different Macroclosures of the Original Johansen Model and Their Importance in Policy Evaluation." *Journal of Policy Modelling,* Vol. 4, pp. 85-98.

Robinson, Sherman. 1989. "Multisectoral Models." Chapter 18, in *Handbook of Development Economics.* H. Chenery and T.N. Srinivasan (eds.). Amsterdam: North-Holland Publishing Co.

Robinson, Sherman (1991). "Macroeconomics, Financial Variables, and Computable General Equilibrium Models." *World Development.* Vol. 19, No. 11, pp. 1509-1525.

Robinson, Sherman and David W. Roland-Holst. (1988). "Macroeconomic Structure and Computable General Equilibrium Models." *Journal of Policy Modeling,* Vol. 10, No. 3, pp. 353-375.

Sen, Amartya K. (1963). "Neo-Classical and Neo-Keynesian Theories of Distribution." *Economic Record,* Vol. 39, pp. 54-64.

Shoven, John B. and John Whalley (1992). *Applying General Equilibrium.* Cambridge: Cambridge University Press.

Srinivasan, T. N. (1982). "General Equilibrium Theory, Project Evaluation and Economic Development." In M. Gersowitz et al. (eds.), *The Theory and Experience of Economic Development: Essays in Honour of Sir W. Arthur Lewis.* Allen and Unwin.

Taylor, Lance (2004) *Reconstructing Macroeconomics: Structuralist Proposals and Critiques of the Mainstream.* Cambridge, MA: Harvard University Press.

Taylor, Lance (1990). "Structuralist CGE Models." In L. Taylor, ed., *Socially Relevant Policy Analysis.* Cambridge, MA: MIT Press.

Thierfelder, Karen and Sherman Robinson. 2003. "Trade and Tradability: Exports, Imports, and Factor Markets in the Salter-Swan Model." *The Economic Record*, Vol. 79, No. 244, March, pp. 103-111.

Thorbecke, Erik (1991). "Adjustment, Growth, and Income Distribution in Indonesia." *World Development.* Vol. 19, No. 11.

Thorbecke, Erik (1998). "Social Accounting Matrices and Social Accounting Analysis." In W. Isard, et al. (eds.), *Methods of Interregional and Regional Analysis.* USA: Ashgate, Aldershot, Brookfield.

Thorbecke, Erik (1991). "Adjustment, Growth and Income Distribution in Indonesia." *World Development*, Vol. 19, No. 11 (November), pp. 1595-1614.

Thorbecke, Erik (1985). "The Social Accounting Matrix and Consistency-Type Development Planning Models." In Graham Pyatt and Jeffery I. Round, eds., *Social Accounting Matrices: A Basis for Planning.* Washington: World Bank.

Thorbecke, Erik and D. Berrian (1990). "The Impact of Structural Adjustment Policies on Poverty and Nutrition Analyzed within a General Equilibrium Framework." In P. Pinstrup-Andersen, ed., *Macroeconomic Policy Reforms, Poverty, and Nutrition: Analytical Methodologies.* Cornell Food and Nutrition Policy Program, Monograph 3 (February), pp. 49-71.

Thorbecke, Erik and Haider A. Khan (1989). "Macroeconomic Effects of Technology Choice: Multiplier Analysis and Structural Path Analysis within a SAM Framework." *Journal of Policy Modelling*, Vol. 11, No. 1, pp. 131-156.

Tyson, Laura D., and Sherman Robinson (1983). "Modeling Structural Adjustment: Micro and Macro Elements in a General Equilibrium Framework." In H. Scarf and J. Shoven (eds.), *Applied General Equilibrium Analysis.* Cambridge: Cambridge University Press.

Whalley, John and Bernard Yeung (1984). "External Sector 'Closing' Rules in Applied General Equilibrium Models." *Journal of International Economics*, Vol. 16. pp. 123-138.

Chapter 12

MULTIPLIER EFFECTS AND THE REDUCTION OF POVERTY

Graham Pyatt and Jeffery I. Round
University of Warwick, grahampyatt@ukonline.co.uk and J.I.Round@warwick.ac.uk

1. INTRODUCTION

Two extensions of the fixed-price multiplier analysis set out in Pyatt & Round [1979] are introduced in this paper. The first is an interpretation of each element $m_{i,j}$ of a multiplier matrix **M** as the sum of all elements of an $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$-type transformation of the core of **M** where details of the various mappings that generate the circular flow on income are to be found. The interpretation of each $m_{i,j}$ is facilitated accordingly. The second contribution is to show how fixed-price multipliers can be used to explore changes in poverty that result from the stimulation of alternative production activities. Combining these two themes leads us to an explanation of how and why different stimuli may have different implications for poverty.

These themes are important in their own right. But they gain particular relevance from the contemporary concern to reduce poverty in poor countries *via* the development of *Poverty Reduction Strategies*. In this context, fixed-price multiplier analysis has been highlighted in the World Bank's 'toolkit' as a useful way of evaluating the poverty impact of alternative policies (see Round [2003]). This context is one reason for choosing to address these themes in our contribution to this volume. A second consideration is that both themes have been of particular concern to Erik Thorbecke over the years as evidenced by his joint work with Jacques Defourny on structural path analysis (see especially Defourny & Thorbecke [1984]); with Hong-Sang Jung on 'multiplier decomposition methods to analyse poverty alleviation' (Thorbecke & Jung [1996]); and more recent arguments set out in Thorbecke [2003] which favour a general equilibrium modelling approach to the explanation of changes in poverty, as an alternative to the fixed-price multipliers previously adopted (and adapted) by Thorbecke & Jung.[cxx]

The paper has two main sections following this Introduction. In the first of these (which is Section 2) we set out a derivation of fixed-price multipliers as a feature of most if not all general equilibrium models and show how the multiplier matrix **M** can usefully be expressed as the product of three matrices $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_3$. Hence we are able to show that each element of **M** is the sum of all elements of an $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$-type transformation of $\mathbf{M}_2$.

The analysis developed in Section 2 is deployed in Section 3 to explain changes in poverty. Here we focus on the head-count measure of the incidence of poverty since this has been adopted by the Development Committee in formulating its *Millennium Development Goals.* However, the exposition admits any measure that is additively separable across household groups. This generality is important because, while it is useful to monitor the head-count measure of poverty *ex post,* the minimisation of the head-count ratio is problematic as a suitable policy objective.[cxxi]

The paper concludes with a brief discussion of some pros and cons of using a complete general equilibrium model instead of the partial framework that is offered by fixed-price multiplier analysis.

2. MULTIPLIERS AND THEIR DECOMPOSITION

2.1 Fixed price multipliers

Table 12-1 sets out in schematic form a social accounting matrix framework within which our discussion of multipliers can be organised. The matrix is partitioned to distinguish five categories of accounts *viz* product accounts, which record purchases and sales for all goods and non-factor services; factor accounts, which do the same for factor services; a set of current accounts for institutions which receive all forms of income and generate outlays; a corresponding set of capital accounts for institutions; and one or more accounts to record all international transactions. There are therefore twenty-five sub-matrices within the social accounting matrix and these can be grouped into four categories as shown in the table.[cxxii] Sub-matrices located in the north-west quadrant of the table record all transactions within and between accounts that belong to one or other of the first three of the five categories noted above. An implication is that the matrix **T** in Table 12-1 contains all details of the circular flow on income within the economy. In contrast, the matrix **F** in the south-east quadrant records all transactions within and between the last two of the five categories of accounts. Collectively, the latter constitute the flow of funds.

Transactions between the first three and last two categories occupy the remaining quadrants. The north-east quadrant sets out details of those outlays by domestic capital accounts and the rest of the world that imply either a demand for goods and services or income for domestic institutions. These outlays are referred to collectively as injections and denoted here by **X**. They are complemented in the south-west quadrant by those outlays of accounts in the first three categories that result in revenue for accounts in the last two categories. These are referred to as leakages and denoted here by the matrix **L**. Because Table 12-1 is a social accounting matrix, corresponding row and column totals must be equal. Hence, in aggregate, leakages and injections must be equal.

Within this framework, the construction of a general equilibrium model requires an algebraic specification of the way in which each cell of **T**, **X**, **L** and **F** is determined. So, if $t_{i,j}$ is a particular element of **T**, expressions of the form

$$t_{i,j} = t_{i,j}(\mathbf{p},\mathbf{y};\boldsymbol{\theta}) \qquad [1]$$

will be part of the model, where **p** is a vector of prices for goods and services, **y** is the vector of marginal totals identified in Table 12-1, and $\boldsymbol{\theta}$ is a set of parameters. It now follows that, by summing along rows of Table 12-1, we can obtain a set of equations

$$y_i = \sum_k t_{i,k}(\mathbf{p},\mathbf{y};\boldsymbol{\theta}) + x_i \qquad [2]$$

where y_i is the ith element of **y** and x_i is the sum of all elements of the ith row of **X**. It then follows from total differentiation of [2] that, for any general equilibrium model of an economy, if the parameters $\boldsymbol{\theta}$ are fixed, then

$$d\mathbf{y} = \mathbf{C}d\mathbf{y} + \mathbf{E}d\mathbf{p} + d\mathbf{x} \qquad [3]$$

where **x** is the vector formed by summing **X** along rows, the (i,j)th element of **C** is

$$c_{i,j} = \frac{\partial}{\partial y_j}\sum_k t_{i,k}(\mathbf{p},\mathbf{y};\boldsymbol{\theta}) \qquad [4]$$

and the (i,l) th element of **E** is similarly defined by

$$e_{i,l} = \frac{\partial}{\partial p_l} \sum_k t_{i,k}(\mathbf{p},\mathbf{y};\boldsymbol{\theta}) \qquad [5]$$

It therefore follows from [3] that if the matrix

$$\mathbf{M} = (\mathbf{I} - \mathbf{C})^{-1} \qquad [6]$$

exists, then

$$d\mathbf{y} = \mathbf{M}(\mathbf{E}d\mathbf{p} + d\mathbf{x}) \qquad [7]$$

The matrix **M** is referred to as a fixed price multiplier matrix for two reasons: firstly, because **M** defines the impact of any change in the injections **x** on each element of the vector **y** when prices are held constant; and, secondly, when **M** exists, an equivalent expression to [6] is

$$\mathbf{M} = \mathbf{I} + \mathbf{C} + \mathbf{C}^2 + \ldots + \ \textit{ad inf} \qquad [8]$$

so that, if all elements of **C** are non-negative, then each element of **M** must be at least as large as the corresponding element of an identity matrix. This second consideration defines the sense in which **M** is a multiplier matrix.

2.2 The decomposition of a multiplier matrix

The matrices **C** and **T** are both square, the same size, and conform to the same accounting taxonomies. Both can be partitioned, therefore, according to the distinctions noted above between product accounts, factor accounts and the current accounts of institutions. Moreover, for a wide class of economic model, each of the partial derivatives $c_{i,j}$ will be zero if the corresponding element of **T** is zero. It follows that for all models which satisfy this condition **C** can be written as a 3×3 matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{PP} & \mathbf{0} & \mathbf{C}_{PI} \\ \mathbf{C}_{FP} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{IF} & \mathbf{C}_{II} \end{bmatrix} \qquad [9]$$

in which four of the sub-matrices are zero. The remaining five are of two types. Two of them *viz* the sub-matrices $\mathbf{C}_{PP}$ and $\mathbf{C}_{II}$ represent transfers within a given class of accounts: the matrix $\mathbf{C}_{PP}$ captures inter-industry (raw material) transactions while $\mathbf{C}_{II}$ represents current transfers between institutions. The three remaining sub-matrices are $\mathbf{C}_{PI}$, $\mathbf{C}_{FP}$ and $\mathbf{C}_{IF}$. Together they characterise the circular flow of incomes and outlays whereby any increase in expenditure by institutions generates extra demand for products *via* the mapping $\mathbf{C}_{PI}$. Next, the supply response to this increase in demand generates an increase in demand for factor services *via* the mapping $\mathbf{C}_{FP}$ and hence generates some extra income for institutions *via* $\mathbf{C}_{IF}$. This extra income will, in turn, generate a further increase in the consumption expenditures of institutions, which acts as a trigger for further rounds of the multiplier process, the size of which will depend on whether the elements of **C** remain constant as effective demand increases within the system.

This characterisation of the structure of **C** can be formalised by writing

$$\mathbf{C} = \left(\mathbf{C} - \hat{\mathbf{C}}\right) + \hat{\mathbf{C}} \qquad [10]$$

where

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_{PP} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C}_{II} \end{bmatrix} \quad \text{and} \quad \mathbf{C} - \hat{\mathbf{C}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{C}_{PI} \\ \mathbf{C}_{FP} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{IF} & \mathbf{0} \end{bmatrix} \qquad [11]$$

which implies that $\hat{\mathbf{C}}$ captures the transfer elements of **C** while the elements of $\mathbf{C} - \hat{\mathbf{C}}$ generate the circular flow.

It follows directly from this decomposition that

$$\mathbf{M} = (\mathbf{I} - \tilde{\mathbf{C}})^{-1}(\mathbf{I} - \hat{\mathbf{C}})^{-1} \qquad [12]$$

where

$$\tilde{\mathbf{C}} = (\mathbf{I} - \hat{\mathbf{C}})^{-1}(\mathbf{C} - \hat{\mathbf{C}}) = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \tilde{\mathbf{C}}_{PI} \\ \tilde{\mathbf{C}}_{FP} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{IF} & \mathbf{0} \end{bmatrix} \qquad [13]$$

provided that

$$\tilde{\mathbf{C}}_{PI} = (\mathbf{I}-\mathbf{C}_{PP})^{-1}\mathbf{C}_{PI}; \quad \tilde{\mathbf{C}}_{FP} = \mathbf{C}_{FP}; \quad \text{and} \quad \tilde{\mathbf{C}}_{IF} = (\mathbf{I}-\mathbf{C}_{II})^{-1}\mathbf{C}_{IF} \qquad [14]$$

The matrix $\mathbf{C}$ must therefore have the same structure of zero and non-zero elements as the matrix $\mathbf{C}-\hat{\mathbf{C}}$.

Equation [12] provides an initial decomposition of the matrix $\mathbf{M}$ into a transfer effects matrix $(\mathbf{I}-\hat{\mathbf{C}})^{-1}$ and a complementary matrix $(\mathbf{I}-\tilde{\mathbf{C}})^{-1}$that can be further developed by noting that because $\tilde{\mathbf{C}}$ has the same partitioned structure as $\mathbf{C}-\hat{\mathbf{C}}$, the matrix $\tilde{\mathbf{C}}^3$ must be block diagonal: it is given by

$$\tilde{\mathbf{C}}^3 = \begin{bmatrix} \tilde{\mathbf{C}}_{PI}\tilde{\mathbf{C}}_{IF}\tilde{\mathbf{C}}_{FP} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{C}}_{FP}\tilde{\mathbf{C}}_{PI}\tilde{\mathbf{C}}_{IF} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{C}}_{IF}\tilde{\mathbf{C}}_{FP}\tilde{\mathbf{C}}_{PI} \end{bmatrix} \qquad [15]$$

The inverse $(\mathbf{I}-\tilde{\mathbf{C}})^{-1}$ can therefore be rewritten as

$$(\mathbf{I}-\tilde{\mathbf{C}})^{-1} = \left(\mathbf{I}-\tilde{\mathbf{C}}^3\right)^{-1}\left(\mathbf{I}+\tilde{\mathbf{C}}+\tilde{\mathbf{C}}^2\right) \qquad [16]$$

and it now follows from [12] and [16] that

$$\mathbf{M} = \mathbf{M}_3\mathbf{M}_2\mathbf{M}_1 \qquad [17]$$

where

$$\mathbf{M}_3 = \left(\mathbf{I}-\tilde{\mathbf{C}}^3\right)^{-1}; \mathbf{M}_2 = \left(\mathbf{I}+\tilde{\mathbf{C}}+\tilde{\mathbf{C}}^2\right); \text{and } \mathbf{M}_1 = (\mathbf{I}-\hat{\mathbf{C}})^{-1} \qquad [18]$$

Equation [17] is the three-part decomposition of $\mathbf{M}$ established in Pyatt & Round [1979]. One of its properties is that, if all elements of $\mathbf{C}$ are non-negative, then all elements of $\mathbf{M}_1$, $\mathbf{M}_2$ and $\mathbf{M}_3$ will be at least as large as the corresponding elements of an identity matrix. All three matrices qualify as multiplier matrices, therefore. And each has a distinct role in generating $\mathbf{M}$. The matrix $\mathbf{M}_1$ has already been identified as a transfer effects multiplier which is given by

$$\mathbf{M}_1 = (\mathbf{I}-\hat{\mathbf{C}})^{-1} = \begin{bmatrix} (\mathbf{I}-\mathbf{C}_{PP})^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & (\mathbf{I}-\mathbf{C}_{II})^{-1} \end{bmatrix} \qquad [19]$$

while $\mathbf{M}_2$ is a matrix of cross-effects that explains why and how the stimulation of one part of the system has repercussions for all others. From [13] and [18] it follows that $\mathbf{M}_2$ is given by

$$\mathbf{M}_2 = \left(\mathbf{I} + \tilde{\mathbf{C}} + \tilde{\mathbf{C}}^2\right) = \begin{bmatrix} \mathbf{I} & \tilde{\mathbf{C}}_{PI}\tilde{\mathbf{C}}_{IF} & \tilde{\mathbf{C}}_{PI} \\ \tilde{\mathbf{C}}_{FP} & \mathbf{I} & \tilde{\mathbf{C}}_{FP}\tilde{\mathbf{C}}_{PI} \\ \tilde{\mathbf{C}}_{IF}\tilde{\mathbf{C}}_{FP} & \tilde{\mathbf{C}}_{IF} & \mathbf{I} \end{bmatrix} \qquad [20]$$

Finally, from [15] and [18], the matrix $\mathbf{M}_3$ must be block-diagonal, as is $\mathbf{M}_1$. The former is given by

$$\mathbf{M}_3 = \left(\mathbf{I} - \tilde{\mathbf{C}}^3\right)^{-1}$$

$$= \begin{bmatrix} \left(\mathbf{I} - \tilde{\mathbf{C}}_{PI}\tilde{\mathbf{C}}_{IF}\tilde{\mathbf{C}}_{FP}\right)^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{I} - \tilde{\mathbf{C}}_{FP}\tilde{\mathbf{C}}_{PI}\tilde{\mathbf{C}}_{IF}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \left(\mathbf{I} - \tilde{\mathbf{C}}_{IF}\tilde{\mathbf{C}}_{FP}\tilde{\mathbf{C}}_{PI}\right)^{-1} \end{bmatrix} \qquad [21]$$

Its role is to represent the consequences of a change in **x** travelling around the entire system to reinforce the initial injection.

2.3 Alternative developments

Some years' ago Richard Stone suggested that a useful variant of [17] could be obtained by rewriting our result as

$$\mathbf{M} = (\mathbf{M}_3 - \mathbf{I})\mathbf{M}_2\mathbf{M}_1 + (\mathbf{M}_2 - \mathbf{I})\mathbf{M}_1 + (\mathbf{M}_1 - \mathbf{I}) + \mathbf{I} \qquad [22]$$

which provides an additive decomposition of **M** with four elements (Stone [1985]). The last of these corresponds to the initial injection; the second-to-last to transfer effects; the third-to-last to cross effects; and the fourth-to-last (which is the first term of the expansion) to circular effects. So, if $m_{i,j}$ is the (i, j) th element of **M** then

$$\begin{aligned} m_{i,j} &= \mathbf{d}_i'\mathbf{M}\mathbf{d}_j \\ &= \mathbf{d}_i'(\mathbf{M}_3 - \mathbf{I})\mathbf{M}_2\mathbf{M}_1\mathbf{d}_j + \mathbf{d}_i'(\mathbf{M}_2 - \mathbf{I})\mathbf{M}_1\mathbf{d}_j + \mathbf{d}_i'(\mathbf{M}_1 - \mathbf{I})\mathbf{d}_j + \mathbf{d}_i'\mathbf{d}_j \end{aligned} \qquad [23]$$

where $\mathbf{d}_k$ is a vector the kth element of which is one and all other elements of $\mathbf{d}_k$ are zero. The result [23] implies, therefore, that an injection into account j has no initial impact on account i unless $i = j$. Otherwise $\mathbf{d}_i'\mathbf{d}_j = 0$ which implies that the initial injection will also have no transfer effect on *i* unless *i* and *j* belong to the same partition of **T**.

Stone's linearization of [17] has proved to be a useful way of presenting results. But, for some purposes, the original multiplicative formulation has advantages that follow from the implication of [17] that

$$m_{i,j} = \mathbf{d}_i'\mathbf{M}\mathbf{d}_j = \mathbf{d}_i'\mathbf{M}_3\mathbf{M}_2\mathbf{M}_1\mathbf{d}_j = \mathbf{i}'\left(\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}\right)\mathbf{i} \qquad [24]$$

where $\mathbf{i}$ is a vector all elements of which are one, while $\mathbf{r}$, $\mathbf{A}$ and $\mathbf{s}$ are given by

$$\mathbf{r}' = \mathbf{d}_i'\mathbf{M}_3; \quad \mathbf{A} = \mathbf{M}_2 \quad \text{and} \quad \mathbf{s} = \mathbf{M}_1\mathbf{d}_j \qquad [25]$$

Each $m_{i,j}$ must therefore be equal to the sum of all elements of an $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$-type transformation of the matrix $\mathbf{M}_2$ when the vector $\mathbf{r}'$ is formed from the ith row of $\mathbf{M}_3$ and the vector **s** is formed from the j th column of $\mathbf{M}_1$. In other words, a complete accounting for $m_{i,j}$ can be constructed for any *i* and *j* from three elements: (i) the ith row of the matrix $\mathbf{M}_3 = (\mathbf{I} - \tilde{\mathbf{C}}^3)^{-1}$; (ii) the entire matrix $\mathbf{M}_2 = (\mathbf{I} + \tilde{\mathbf{C}} + \tilde{\mathbf{C}}^2)$; and (iii) the *j*th column of the matrix $\mathbf{M}_1 = (\mathbf{I} - \hat{\mathbf{C}})^{-1}$. The last of these shows how the consequences of a particular injection will be amplified as a result of transfer effects within the category of accounts in which the initial stimulus arises. The second explains how these initial effects will spread to accounts in other categories. And the leading term $\hat{\mathbf{r}}$ quantifies the consequences for account *i* of the circulation around the entire system of the *stimuli* generated *via* the first two mechanisms.

All three mechanisms are important for diagnostic reasons since they allow us to account for $m_{i,j}$ in microscopic detail. However, it is evident from [21] that most elements of the ith row of $\mathbf{M}_3$ will be zero, while [19] has similar implications for the *j*th column of $\mathbf{M}_1$. It is not difficult, therefore, to devise a more compact version of [24] without any loss of information, depending on the specific categories to which accounts *i* and *j* belong. To illustrate this point we can assume that *i* is a particular institution ($i \in$ I) and *j* is a particular production activity ($j \in$ P). Recalling that both $\mathbf{M}_1$ and $\mathbf{M}_3$ are block diagonal matrices it follows from [17] that the element $m_{i,j}$ of **M** will now be an element of the sub-matrix $\mathbf{M}_{I,P}$ of **M** where

$$\mathbf{M}_{I,P} = \mathbf{M}_{3(I,I)}\mathbf{M}_{2(I,P)}\mathbf{M}_{1(P,P)} \qquad [26]$$

Hence

$$m_{i,j} = \left(\mathbf{d}_i'\mathbf{M}_{3(I,I)}\right)\mathbf{M}_{2(I,P)}\left(\mathbf{M}_{1(P,P)}\mathbf{d}_j\right) \qquad [27]$$

which can also be written in the form $\mathbf{i}'\left(\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}\right)\mathbf{i}$ where now

$$\mathbf{r}' = \mathbf{d}_i'\mathbf{M}_{3(I,I)}; \quad \mathbf{A} = \mathbf{M}_{2(I,P)} \quad \text{and} \quad \mathbf{s} = \mathbf{M}_{1(P,P)}\mathbf{d}_j \qquad [28]$$

The cell $m_{i,j}$ is therefore equal to the sum of all elements of a new $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$-type transform in which $\mathbf{r}$ is a row of $\mathbf{M}_{3(I,I)}$, $\mathbf{A}$ is equal to $\mathbf{M}_{2(I,P)}$, and $\mathbf{s}$ is the relevant column of $\mathbf{M}_{1(P,P)}$. Since $\mathbf{M}_{2(I,P)}$ is a 9×24 matrix the result [27] provides a disaggregation of $m_{i,j}$ into $9 \times 24 = 216$ components for each *i* and *j* that falls within the relevant accounting categories.

Tables 12-2(a), 12-2(b) and 12-2(c) provide empirical estimates of the three matrices on the right-hand side of [26] and, therefore, all the details needed to generate the 9×24 disaggregation [27] of $m_{i,j}$ for all *i* and *j* when *i* is an institution and *j* is a production activity. For example, if *i* is the institution 'small scale farming households' (institution 25) and *j* is the activity 'food processing' (production activity 53) the 9×24 disaggregation of $m_{25,53} = 0.258$ is provided by the details set out in Table 12-3. Table 12-3(a) shows the matrix $\mathbf{M}_{2(I,P)}$ bordered by the two vectors $\mathbf{r}_{25}$ and $\mathbf{s}_{53}'$, the former being the relevant row of Table 12-2(a) and the latter the relevant column of Table 12-2(c). Accordingly, a unit injection into 'food processing' generates multiplier effects $\mathbf{s}_{53}$ on the various production activities, the magnitude of which can be read-off from the relevant column of the input-output inverse $(\mathbf{I} - \mathbf{C}_{PP})^{-1} = \mathbf{M}_{1(P,P)}$ as reported in Table 12-2(c). These consequences of the original injection are then translated by the $\mathbf{A}$ part of the $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$ transform *i.e.* by the matrix $\mathbf{M}_{2(I,P)}$ into increments of income for the various institutions. And, finally, the transmission of these increments right around the system - the complete circular flow - generates the implications for 'small scale farm households' that are captured by the multiplier $\hat{\mathbf{r}}$ *i.e.* by the row of $\mathbf{M}_{3(I,I)}$ (Table 12-2(a)) that corresponds to account 26.

Table 12-3(b) sets out the 9×24 detail of the $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$ transformation generated by the components assembled in Table 12-3(a). Evidently only 31 of the 216 elements of this table are sufficiently large to be recorded as contributions to the aggregate multiplier effect $m_{25,53} = 0.258$ given that all entries in the table have been rounded to three digits. Moreover, it is evident from inspection that the linkage from 'food processing' to 'small scale farm households' *via* all forms of factor income *i.e.* the row 25, column 53 element of **A** (which is $\mathbf{M}_{2(I,P)}$) is not the most important. More powerful linkages are generated by the increased intermediate demand for 'food crops' (sector 47) and, to a lesser extent, for 'other crops' (sector 48) as a result of the stimulation of the food processing sector. This derived demand evidently creates significant extra income for all of the household groups (institutions 24 to 31) which, in turn, generate extra income for 'small scale farm households' as shown in Table 12-3(b).

3. SOME IMPLICATIONS FOR POVERTY

3.1 Determinants of poverty

One of the many reasons for constructing a general equilibrium model and/or for undertaking a fixed-price multiplier analysis is to explore the impact of change on some social evaluation criterion. In particular, we can explore the impact of changes $d\mathbf{x}$ on an overall measure of poverty Q.

To give explicit form to this possibility we can assume that the measure of poverty Q is additively decomposable across groups of households so that

$$Q = \sum_i Q_i \qquad [29]$$

where Q_i is a measure of poverty among households within a particular socio-economic group i. For simplicity, we assume here that

$$Q_i = n_i P_i \qquad [30]$$

where n_i is the number of people in socio-economic category i and P_i is the proportion who are poor. It follows at once that Q_i is the number of poor people in socio-economic category i and that

$$dQ_i = n_i dP_i + P_i dn_i \qquad [31]$$

There are, therefore, two aspects to the change dQ_i *viz* changes in P_i and changes n_i.

With respect to the first of these it is implicit that dP_i will depend both on changes in average income for households in category i and also on changes in prices, if only to the extent that such changes will shift the poverty line and hence change the proportion of those in category i who are poor. However, the importance of price changes is qualified by the fact that within most general equilibrium models one is free to choose a particular price (index) as *numeraire*. So, if the level of income necessary to avoid poverty (the poverty line) is chosen as the *numeraire* price changes will matter only in so far as the poverty lines for different socio-economic groups move differentially. Such effects are necessarily ignored in a fixed-price model.

If the implications of relative price changes are set aside, the most important influence on P_i is likely to be any change in the scale of incomes within category i. The consequences of such a change can be estimated for any specific P_i under the assumption that the dispersion of income within each category i is unaffected by the change in scale. Given this assumption it follows that

$$dP_i = \frac{\partial P_i}{\partial \lambda_i} d\lambda_i \text{ and } \frac{d\lambda_i}{\lambda_i} = \frac{d(y_i/n_i)}{y_i/n_i} \qquad [32]$$

where λ_i is the scale of incomes among households in category i and y_i is their total income. Hence, by elimination of $d\lambda_i$ from the expressions [32]

$$\frac{dP_i}{P_i} = \varepsilon_i \left(\frac{dy_i}{y_i} - \frac{dn_i}{n_i} \right) \qquad [33]$$

where ε_i is the partial elasticity of P_i with respect to λ_i (which is usually referred to as the poverty elasticity for households in category i).

By substituting for dP_i in [33] according to the result [31] it now follows that

$$\frac{dQ_i}{Q_i} = \left[\varepsilon_i \frac{dy_i}{y_i} + (1 - \varepsilon_i) \frac{dn_i}{n_i} \right] \qquad [34]$$

which can usefully be rewritten as

$$\frac{dQ_i}{Q_i} = \left[(1 + |\varepsilon_i|) \frac{dn_i}{n_i} - |\varepsilon_i| \frac{dy_i}{y_i} \right] \qquad [35]$$

since ε_i will invariably be negative.

Accordingly, if the total income y_i of group i remains constant, the number of poor in group i will grow faster than n_i because population growth will tend to reduce y_i/n_i relative to what it would otherwise have been. Conversely, if population growth is zero, Q_i will be reduced if y_i increases. More generally, it follows from [35] that Q_i will be reduced if and only if

$$\frac{dy_i}{y_i} > \left(1 + \frac{1}{|\varepsilon_i|}\right)\frac{dn_i}{n_i} \tag{36}$$

so that total income y_i must grow faster than the population n_i if poverty is to fall within socio-economic category *i*. This is a stringent condition that might be softened by a possibility we have yet to consider, which is that individuals and families can migrate and may, therefore, be able to improve their circumstances by moving into a different socio-economic category (depending on how these categories have been defined). Equally, however, if the inequality [36] is not satisfied for any household type it is unlikely that migration will ameliorate the situation significantly.

It is evident from [35] that the alternative to reducing poverty by reducing the growth of n_i is to increase the growth of y_i one way or another. In particular, we know from [7], [17] and [26] that for given $\mathbf{x}_P$ and $\mathbf{x}_I$

$$d\mathbf{y}_I = \mathbf{M}_{I,P} d\mathbf{x}_P = \mathbf{M}_{3(I,I)} \mathbf{M}_{2(I,P)} \mathbf{M}_{1(P,P)} d\mathbf{x}_P \tag{37}$$

Changes $d\mathbf{x}_P$ in the demand for products will therefore generate increases in the incomes of institutions *via* the mappings $\mathbf{M}_{1(P,P)}$, $\mathbf{M}_{2(I,P)}$ and $\mathbf{M}_{3(I,I)}$ so that, from [35]

$$\frac{dQ_i}{Q_i} = (1 + |\varepsilon_i|)\frac{dn_i}{n_i} - \frac{|\varepsilon_i|}{y_i}\mathbf{d}_i'\mathbf{M}_{I,P} d\mathbf{x}_P \tag{38}$$

which can be written in the alternative form

$$g(\mathbf{q}) = (\mathbf{I} + |\hat{\boldsymbol{\varepsilon}}|)g(\mathbf{n}) - (\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P)g(\mathbf{x}_P) \tag{39}$$

where the *j*th element of the vector $g(\mathbf{a})$ is the proportionate rate of change of the *j*th element of the vector **a**, and the ith element of the vector $\boldsymbol{\mu}$ is $|\varepsilon_i|/y_i$. The result [39] therefore implies that for given $\mathbf{x}_P$ and $\mathbf{x}_I$ the proportional rate of change of the number of poor people within socio-economic group *i* will depend on the proportionate rate of change of population within that group and on the stimulation of the economy *via* changes in $\mathbf{x}_P$. More precisely, the second term on the right-hand side of [39] shows how the stimulation of any one or more production activities would reduce poverty in each socio-economic group. Specifically, the formulation implies that the proportionate rate of decline of poverty as measured by $g(\mathbf{q})$ will depend on the proportional rates of increase of the elements of $\mathbf{x}_P$ *i.e.* on $g(\mathbf{x}_P)$ and on the expression $\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$ which, it can be noted, is an $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$-type transform of $\mathbf{M}_{I,P}$.

Two variants of the result [39] can also be noted at this point. Firstly, if the ith element of $g(\mathbf{p})$ is the proportionate rate of growth of P_i, we can replace $g(\mathbf{q})$ on the left-hand side of [39] by $g(\mathbf{p})$ if we simultaneously subtract $g(\mathbf{n})$ from the right-hand side. And, secondly, if we pre-multiply both sides of equation [39] by a matrix $\hat{\boldsymbol{\sigma}}$ such that the *i*th element of $\boldsymbol{\sigma}$ is

$$\sigma_i = Q_i / Q \qquad [40]$$

then σ_i is the proportion of total poverty that is attributable to households in socio-economic group *i* and equation [39] now becomes

$$\hat{\boldsymbol{\sigma}} g(\mathbf{q}) = \hat{\boldsymbol{\sigma}}\left(\mathbf{I} + |\hat{\boldsymbol{\varepsilon}}|\right) g(\mathbf{n}) - \left(\hat{\boldsymbol{\sigma}}\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P\right) g(\mathbf{x}_P) \qquad [41]$$

But

$$\mathbf{i}'\hat{\boldsymbol{\sigma}} g(\mathbf{q}) = \boldsymbol{\sigma}' g(\mathbf{q}) = \sum_i \frac{Q_i}{Q}\frac{dQ_i}{Q_i} = \frac{dQ}{Q} \qquad [42]$$

so that the proportionate rate of change of the aggregate number of people who are poor is

$$\frac{dQ}{Q} = \boldsymbol{\sigma}'\left(\mathbf{I} + |\hat{\boldsymbol{\varepsilon}}|\right) g(\mathbf{n}) - \mathbf{i}'\left(\hat{\boldsymbol{\sigma}}\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P\right) g(\mathbf{x}_P) \qquad [43]$$

Accordingly, for given n, the impact of changes $g(\mathbf{x}_\mathbf{P})$ on the aggregate number of poor will depend on the size of the changes $g(\mathbf{x}_\mathbf{P})$ and the elements of an $\hat{\mathbf{r}}\mathbf{A}\hat{\mathbf{s}}$-type transformation in which

$$\hat{\mathbf{r}} = \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\sigma}}; \quad \mathbf{A} = \mathbf{M}_{I,P} \text{ and } \mathbf{s} = \mathbf{x}_P \qquad [44]$$

The *i*th element of **r** is therefore the product

$$r_i = \sigma_i \mu_i \qquad [45]$$

where σ_i is the proportion of the poor who belong to socio-economic group *i* and μ_i can be shown to be given by

$$\mu_i = \frac{dP_i}{d\lambda_i} \qquad [46]$$

i.e. by the sensitivity of the proportion of households in group *i* who are poor to a change in their average income that is the result of a scale change of all incomes within that group.

It follows from these results that the reciprocal of μ_i has the dimension 'rupiah *per* year' while each element of $\boldsymbol{\sigma}$ is a pure number. Hence all elements of the expressions $\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$ and $\hat{\boldsymbol{\sigma}}\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$ that appear in equations [39] and [43] must also be pure numbers.

3.2 Empirical results

As noted in the Introduction to this paper, Thorbecke & Jung [1996] have previously explored some implications for poverty of sectoral growth in Indonesia, basing their results on the social accounting matrix compiled by Keuning and Thorbecke. Here we present some new results based on our own analysis of the same data.

Table 12-4 provides estimates of $|\varepsilon_i|$, y_i and σ_i (and hence implicitly for μ_i) for each socio-economic category of households. The information required to compile these estimates is taken directly from Thorbecke & Jung [1996; Table 12-2] where it is noted that the estimates of σ_i have been constructed from secondary-source material. These particular estimates are problematic accordingly and our results based on them can only be illustrative. They suggest that most poor people are to be found in the socio-economic categories 'small scale farmers' and, to a lesser extent, among 'agricultural employees' and 'rural, non-agricultural, low skilled' households. There is accordingly a *prima facia* case for focusing a poverty reduction strategy on helping these groups. But the extent of poverty in a particular group is not necessarily the only criterion. The poverty elasticity for each group also plays a role and these elasticities tend to be inversely related to the incidence of poverty because *ceteris paribus* they are directly related to the mean income within each group.

To the extent that the poverty elasticities might suggest a shift in the focus of a poverty reduction strategy away from those socio-economic groups in which incomes are generally low there is a case for ignoring their influence and emphasising instead the importance of raising living standards within those groups in which average incomes are lowest. This case rests on the limitations of the incidence of poverty as a social evaluation criterion. And because it is a strong case, the impact of changes $d\mathbf{x}$ on each Q_i and on Q as set out in the remainder of this section are arguably less important than their impact on the income vector $\mathbf{y}$ as discussed in the previous section of this paper.

Table 12-5 presents an estimate of the (transpose of) matrix $\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$ together with details of the vector $\mathbf{x}_P$ (which are included here as an *aide memoir*) and the row vector $\mathbf{i}'\left(\hat{\boldsymbol{\sigma}}\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P\right)$ which, from [43], is needed to calculate the impact of changes $g(\mathbf{x}_P)$ on the proportionate rate of change of the number of poor *i.e.* to calculate dQ/Q.

To illustrate the way in which this table should be read we can consider the implications of a 1% increase of x_{53} implying a 1% increase in exogenous demand for the products of the 'food processing' sector. Reading across the row of the table corresponding to sector 53 we see that the level of exogenous demand in 1980, as recorded in the Indonesia social accounting matrix, was 335.85 Rp billion. A 1% increase of this amount would therefore be small relative to a 1% increase in x_{52} (mining) or x_{58} (building & construction) since exogenous demand for the products of these sectors was much larger in 1980. The next eight columns of the table set out our estimates of the percentage change in poverty for each socio-economic group of households that would result from a 1% increase in final demand for the products of each productive sector *i.e.* the (transpose of) the matrix $\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$.

The final column of Table 12-5 shows the aggregate, economy-wide effects on poverty of a 1% change in the final demand for the products of each sector *i.e.* the (transpose of) the vector $\boldsymbol{\sigma}'\hat{\boldsymbol{\mu}}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$. These aggregate results imply that a 1% increase in final demand for the products of *all* sectors *i.e.* a scale increase, economy-wide, of the exogenous demands $\mathbf{x}$ of 1% would reduce the number of poor by 0.97 %.

This last figure is, of course, a (weighted) average of the figures for individual household groups and productive sectors. However, significant variation around this average is evident in the table. The detail suggest that for most household types the largest percentage reductions in poverty would be achieved by a 1% increase in exogenous demand for 'building and construction'. But there are exceptions. Poverty among 'agricultural employees' is most sensitive to the proportionate rate of change of exogenous demand for 'other crops'. And, for 'high skilled' households poverty is more sensitive to 'education & health" or 'mining' depending on whether the household is rural or urban. It is, therefore, not surprising that, as shown by the final column of the table, the reduction of poverty nation-wide is most sensitive to changes in the exogenous demand for the products of sector 48 (other crops), sector 52 (mining) and sector 58 (building & construction).

These are interesting characteristics of the relationship between the production structure and household poverty. But, from a policy perspective, it would be wrong to infer from such results that other sectors might be ignored. This would be wrong for two reasons, the first being that any increase in any element of $\mathbf{x}_P$ will potentially help to raise incomes and reduce poverty. The second consideration is that every increase in $\mathbf{x}_P$ can imply an increase in exports and/or gross fixed capital formation, depending on the source of the increase.[cxxiii] If exports increase, this should help improve the balance of payments situation. An increase in investment may be desirable or even necessary to create jobs, remove bottlenecks and stimulate development. So, other things equal, any opportunity for increasing $g(\mathbf{x}_P)$ is potentially a development opportunity.

However, within the context of a fixed-price multiplier analysis, reductions in poverty depend on $\hat{\sigma}\hat{\mu}\mathbf{M}_{I,P}\hat{\mathbf{x}}_P$ as well as on $g(\mathbf{x}_P)$. There are, therefore, several other determinants of poverty to consider. Of these, the vector $\mathbf{x}_P$ is datum for a particular time and place. And changing our social evaluation criterion would change the $\hat{\sigma}\hat{\mu}$ component of the above expression but, otherwise, would have no effect. It particular, it would not detract from the central importance of

$$\mathbf{M}_{I,P} = \mathbf{M}_{3(I,I)}\mathbf{M}_{2(I,P)}\mathbf{M}_{1(P,P)} \qquad [26]=[47]$$

i.e. of the mapping from productive sectors *via* factor markets to households. Input-output linkages contribute to this mapping *via* $\mathbf{M}_{1(P,P)}$ while the multiplier effects $\mathbf{M}_{3(I,I)}$ which are generated by the complete circular flow of income will augment the income increases that are generated by the preceding terms. But the core of this mapping is the transformation

$$\mathbf{M}_{2(I,I)} = (\mathbf{I} - \mathbf{C}_{II})^{-1}\mathbf{C}_{IF}\mathbf{C}_{FP} \qquad [48]$$

which has three elements. Their collective implication is that if any increases in $\mathbf{x}_P$ are to raise incomes they must, in the first instance, generate an effective demand for factor services (the mapping $\mathbf{C}_{FP}$) and this will be of direct benefit to households in particular socio economic categories (*via* the mapping $\mathbf{C}_{IF}$) only in so far as they are the ones who provide the factor services that are in demand. More specifically, labour markets are crucial because unskilled labour is the only asset that the poor possess. So, one way or another, a poverty reduction strategy that does not rely on fiscal transfers and safety nets (*i.e.* the redistribution of income *via* $(\mathbf{I}-\mathbf{C}_{II})^{-1}$) must increase the effective demand for unskilled labour. In the longer term, other 'redistribution with growth' strategies can come into play, endowing the poor with an enhanced asset base (changing $\mathbf{C}_{IF}$) through education and/or land reform. But, whatever the mechanism, if the incomes of the poor are to be raised, then the effective demand for the factor services they are able to supply must increase if any increases $g(\mathbf{x}_P)$ are to be of any benefit to them. And that will not happen if bottlenecks of any kind restrict the mapping $\mathbf{C}_{IF}\mathbf{C}_{FP}$.

4. A REMAINING ISSUE

Erik Thorbecke's most recent paper on the subject of this essay promotes a very different position to that of Thorbecke & Jung [1996], emphasising the need to move beyond fixed-price formulations to investigate the determinants of poverty *via* a general equilibrium model.[cxxiv] Since our own inclination is to suggest caution in this regard some comment on the issue provides an appropriate note on which to conclude this essay. While prices may be sticky, they are rarely fixed. Indeed, in so far as a key policy concern within a *poverty reduction strategy* is to change incentives *via* prices, there is an important sense in which changing prices may be of the essence.

Thorbecke [2003] emphasises two aspects of changing prices *viz* the potential importance of supply (capacity) constraints and the implications of changing prices for the simulation of changes in poverty. With respect to the first of these issues we can note here that it is entirely possible to include capacity constraints in a fixed-price model.[cxxv] The simplest formulation is to assume that if (or when) a sector reaches capacity any increase in demand for its products will have no effect on the derived demand for raw materials or for factor services. Instead, the excess demand will leak into imports, for example. By implication, all elements of those columns of $\mathbf{C}_{PP}$ and $\mathbf{C}_{FP}$ that refer to productive sectors that have reached capacity should be set equal to zero, with obvious implications for $\mathbf{M}_{1(P,P)}$ and $\mathbf{M}_{2(I,P)}$: there can be no 'trickle-down' from sectors that are capacity-constrained to institutions *via* employment, for example. This formulation suggests a stark contrast between sectors that are capacity-constrained and those that are not. However, it does catch the essence of a problem which, in a fully price-endogenous model, would be mediated by the price mechanism but not, otherwise, altered fundamentally.

Turning now to the problem of deciding how to simulate changes in poverty, it is clearly naive to assume that scale changes in the distribution of income within each socio-economic group are the only important consideration. In a general equilibrium context in which wage rates and levels of unemployment are endogenous one way forward is to make use of this extra information *via* micro simulations or otherwise to improve our estimates of poverty effects. This is one direction in which progress can be and is being made. But there is less attention being paid to the fact that each socio-economic group of households probably justifies having its own poverty line, while markets for goods and services are typically not as complete as general equilibrium modellers are tempted to assume. Indeed, the separability of each household's production and consumption activities that is characteristic of most general equilibrium models is not persuasive as a description of the subsistence sector in poor countries (Pyatt [2003]). And here we can note that these issues would remain essentially unchanged if we were to abandon the incidence of poverty as our social evaluation criterion and focus instead on raising the real incomes of households in the poorest groups.

We can also note in conclusion that the above issues arise in a context of ubiquitous concerns over the sensitivity of relative prices to 'model closure'. Moreover, the ambiguities that arise from this are exacerbated by the limitations of known techniques for unravelling 'what is driving what' in any particular general equilibrium model. In these circumstances there is something to be said for using simpler and/or partial models that are well understood and, therefore, amenable to detailed diagnostics of the type illustrated in this paper. Beyond that, we might venture to suggest that improving primary data and greater sophistication in the construction of social accounting matrices are probably more important avenues towards a better understanding of basic issues and mechanisms.

Table 12-1. [A Schematic Social Accounting Matrix]

			(i)	(ii)	(iii)	(iv)	(v)	Totals
			Production activities	Factors of production	Domestic institutions		Rest of the world	
					Current accounts	Capital accounts		
(i)	Production activities		**T**			**X**		**y**
(ii)	Factors of production							
(iii)	Domestic Institutions	Current accounts						
(iv)		Capital accounts	**L**			**F**		**z**
(v)	Rest of the world							
	Totals		**y'**			**z'**		

Table 12-2. [Estimates of the multiplier matrices $M_{3(I,I)}$, $M_{2(I,P)}$ and $M_{1(P,P)}$ for Indonesia 1980]

(a) $\mathbf{M}_{3(I,I)}$

Institutions

Institutions	24	25	26	27	28	29	30	31	32
24	1.09	0.08	0.06	0.05	0.06	0.05	0.04	0.03	0
25	0.28	1.24	0.19	0.15	0.17	0.14	0.13	0.09	0
26	0.15	0.13	1.11	0.08	0.10	0.08	0.08	0.06	0
27	0.30	0.26	0.20	1.16	0.19	0.15	0.14	0.10	0
28	0.21	0.21	0.17	0.15	1.18	0.16	0.15	0.13	0
29	0.05	0.06	0.05	0.04	0.05	1.05	0.04	0.03	0
30	0.23	0.23	0.20	0.18	0.22	0.20	1.21	0.18	0
31	0.12	0.13	0.12	0.11	0.12	0.11	0.11	1.09	0
32	0.32	0.31	0.27	0.24	0.29	0.26	0.27	0.23	1

(b) $\mathbf{M}_{2(I,P)}$

Products

Institutions	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
24	0	0.07	0.07	0.03	0.02	0.06	0	0	0	0	0	0	0.01	0.16	0.01	0.01	0.01	0.01	0	0.01	0	0.02	0.01	0.01
25	0	0.25	0.09	0.10	0.06	0.11	0	0.01	0.01	0	0	0	0.01	0.03	0.01	0.01	0.01	0.04	0.01	0.03	0	0.03	0.02	0.03
26	0	1.10	0.06	0.08	0.04	0.07	0	0.02	0.01	0	0	0	0.01	0.01	0	0	0.01	0.03	0.03	0.03	0	0.02	0.01	0.01
27	0	0.21	0.18	0.25	0.12	0.19	0	0.01	0.01	0	0	0	0.01	0.01	0	0	0	0.02	0.02	0.02	0	0.03	0.01	0.01
28	0	0.04	0.03	0.03	0.07	0.03	0.02	0.05	0.06	0.01	0.01	0.03	0.09	0.04	0.08	0.07	0.10	0.19	0.06	0.14	0.03	0.06	0.07	0.14
29	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0	0	0.02	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.02	0.01	0.02	0.21	0.03
30	0	0.01	0.01	0.01	0.03	0.01	0.01	0.03	0.08	0.03	0.02	0.07	0.06	0.00	0.09	0.08	0.11	0.24	0.15	0.22	0.08	0.17	0.09	0.16
31	0	0	0.01	0	0.02	0.01	0.04	0.01	0.02	0.02	0.01	0.06	0.02	0.01	0.03	0.02	0.03	0.09	0.07	0.06	0.07	0.13	0.27	0.06
32	0	0	0.08	0.02	0.17	0.05	0.68	0.05	0.07	0.06	0.12	0.26	0.09	0.14	0.12	0.14	0.12	0.24	0.03	0.08	0.25	0.23	0.05	0

Table 12-2 [Cont.]

(c) $\mathbf{M}_{1(P,P)}$

Products	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
46	1.02	0.15	0.24	0.21	0.32	0.38	0.03	**0.22**	0.26	0.22	0.23	0.12	0.16	0.10	0.11	0.13	0.12	0.01	0.11	0.05	0.07	0.02	0.05	0.10
47	0	1.17	0.00	0.04	0	0	0	**.041**	0	0	0	0	0.01	0	0	0	0	0	0.13	0	0.01	0	0.02	0.00
48	0	0	1.19	0.02	0	0	0	**.017**	0.11	0	0.01	0	0	0	0	0	0	0	0.07	0	0	0	0.01	0.01
49	0	0	0	1.34	0	0	0	**0.01**	0.03	0	0	0	0	0	0	0	0	0	0.16	0	0	0	0.02	0
50	0	0	0.01	0	1.13	0.02	0	**0.00**	0.01	0.01	0	0	0.21	0.04	0.03	0.03	0.08	0	0.01	0	0.01	0.01	0.01	0.01
51	0	0	0	0	0.00	1.08	0.00	**0.01**	0	0	0	0	0	0	0	0	0	0	0.04	0	0.00	0	0.00	0
52	0.01	0.01	0.01	0	0.01	0.01	1.03	**0.01**	0.04	0.02	0.21	0.08	0.12	0.19	0.23	0.14	0.15	0.01	0.01	0.03	0.04	0.01	0.02	0.03
53	0	0	0	0.06	0	0.01	0	**1.05**	0	0	0	0	0	0	0	0	0	0	0.20	0	0.01	0	0.02	0
54	0.01	0	0	0	0	0.01	0	**0.01**	1.48	0	0.01	0.01	0	0	0.00	0	0	0	0.01	0.01	0.01	0.00	0.02	0.04
55	0.03	0.01	0.02	0.01	0.03	0.03	0.03	**0.03**	0.04	1.24	0.03	0.09	0.14	0.09	0.07	0.33	0.18	0.02	0.03	0.05	0.13	0.03	0.12	0.21
56	0.05	0.05	0.06	0.02	0.05	0.04	0.02	**0.05**	0.19	0.09	1.13	0.40	0.38	0.37	0.44	0.33	0.35	0.03	0.07	0.16	0.21	0.04	0.10	0.16
57	0.01	0	0	0	0	0	0	**0**	0.01	0	0.01	1.16	0.01	0	0	0	0	0.01	0.03	0.01	0.01	0.01	0.01	0.03
58	0.01	0	0	0	0	0	0	**0**	0	0	0	0	1.00	0	0.01	0	0	0.01	0.01	0	0.02	0.05	0.02	0
59	0	0	0	0	0	0	0	**0**	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	**0**	0	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	**0**	0	0	0	0.03	0	0	0	1.00	0	0	0	0	0.01	0	0	0
62	0	0	0	0	0	0	0	**0**	0	0	0	0.01	0	0	0	0	1.00	0	0	0	0	0	0.01	0
63	0.87	0.12	0.20	0.18	0.28	0.32	0.03	**0.19**	0.23	0.19	0.20	0.10	0.14	0.09	0.10	0.11	.011	1.01	0.10	0.08	0.15	0.02	0.05	0.09
64	0.01	0	0	0	0	0	0	**0.00**	0	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	1.00	0.01	0.02	0.01	0.00	0.01
65	0.13	0.02	0.03	0.03	0.04	0.05	0	**0.03**	0.04	0.03	0.03	0.02	0.02	0.01	0.02	0.02	0.02	0.01	0.02	1.02	0.01	0.01	0.01	0.01
66	0.05	0.01	0.01	0.01	0.02	0.02	0	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	1.07	0.01	0.01	0.01
67	0.04	0.01	0.02	0.02	0.03	0.02	0.03	**0.02**	0.02	0.01	0.02	0.03	0.03	0.04	0.04	0.03	0.03	0.04	0.03	0.04	0.07	1.06	0.02	0.03
68	0	0	0	0	0	0	0	**0.00**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	0
69	0.03	0.01	0.02	0.01	0.03	0.01	0.01	**0.01**	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.19	0.02	0.01	0.01	1.01

Note: figures in bold type are used in the derivation of $m_{25,53}$ in conjunction with the matrix $\mathbf{M}_{2(I,P)}$ (see section 2.4)

Table 12-3. [Detailed decomposition of the element m $_{25,53}$]

(a) $\mathbf{r}'_{25} = \mathbf{d}'_{25}\mathbf{M}_{3(I,I)}$; $\mathbf{A} = \mathbf{M}_{2(I,P)}$; $\mathbf{s}_{53} = \mathbf{M}_{1(P,P)}\mathbf{d}_{53}$

	r_{25}	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69
s'_{53}		0.223	0.411	0.174	0.008	0.004	0.006	0.011	1.054	0.005	0.026	0.051	0.004	0.003	0.001	0	0	0	0.190	0.004	0.030	0.012	0.020	0.001	0.010
24	0.151	0	0.066	0.073	0.031	0.019	0.060	0.001	0.002	0.003	0.001	0.001	0.004	0.005	0.162	0.006	0.006	0.007	0.009	0.003	0.007	0.004	0.016	0.013	0.008
25	0.133	0	0.254	0.092	0.104	0.058	0.111	0.002	0.008	0.009	0.002	0.002	0.004	0.012	0.034	0.009	0.008	0.011	0.038	0.014	0.025	0.005	0.035	0.017	0.026
26	1.106	0	0.102	0.062	0.080	0.041	0.067	0.003	0.016	0.014	0.003	0.002	0.001	0.011	0.013	0.005	0.003	0.006	0.029	0.025	0.029	0.003	0.024	0.010	0.007
27	0.085	0	0.210	0.180	0.245	0.117	0.187	0.002	0.010	0.009	0.002	0.001	0.001	0.007	0.010	0.003	0.002	0.004	0.019	0.016	0.018	0.002	0.030	0.009	0.006
28	0.100	0	0.038	0.032	0.034	0.069	0.029	0.015	0.047	0.055	0.013	0.014	0.030	0.093	0.043	0.082	0.073	0.096	0.186	0.063	0.144	0.035	0.057	0.065	0.137
29	0.082	0	0.011	0.012	0.013	0.014	0.010	0.008	0.006	0.006	0.002	0.003	0.019	0.008	0.010	0.006	0.005	0.007	0.025	0.012	0.018	0.011	0.022	0.211	0.027
30	0.076	0	0.006	0.008	0.007	0.030	0.008	0.014	0.031	0.075	0.026	0.016	0.070	0.062	0.003	0.093	0.083	0.108	0.235	0.149	0.220	0.076	0.171	0.091	0.156
31	0.056	0	0.002	0.008	0.004	0.017	0.005	0.041	0.012	0.024	0.015	0.013	0.059	0.019	0.007	0.026	0.024	0.030	0.087	0.070	0.060	0.071	0.133	0.270	0.061
32	0	0	0	0.077	0.015	0.173	0.046	0.682	0.050	0.068	0.056	0.122	0.262	0.087	0.138	0.120	0.137	0.124	0.242	0.032	0.080	0.246	0.229	0.051	0

Table 12-3 [Cont.]

(b) $\hat{\mathbf{r}}_{25}\mathbf{A}\hat{\mathbf{s}}_{53}$

	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	Total
24	0	0.007	0.004	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.012
25	0	0.129	0.020	0.001	0	0.001	0	0.010	0	0	0	0	0	0	0	0	0	0.009	0	0.001	0	0.001	0	0	0.173
26	0	0.008	0.002	0	0	0	0	0.003	0	0	0	0	0	0	0	0	0	0.001	0	0	0	0	0	0	0.015
27	0	0.013	0.005	0	0	0	0	0.002	0	0	0	0	0	0	0	0	0	0.001	0	0	0	0	0	0	0.020
28	0	0.003	0.001	0	0	0	0	0.008	0	0	0	0	0	0	0	0	0	0.006	0	0.001	0	0	0	0	0.020
29	0	0.001	0	0	0	0	0	0.001	0	0	0	0	0	0	0	0	0	0.001	0	0	0	0	0	0	0.003
30	0	0	0	0	0	0	0	0.004	0	0	0	0	0	0	0	0	0	0.006	0	0.001	0	0	0	0	0.012
31	0	0	0	0	0	0	0	0.001	0	0	0	0	0	0	0	0	0	0.001	0	0	0	0	0	0	0.003
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	0	0.161	0.032	0.002	0	0.001	0	0.030	0	0	0	0	0	0	0	0	0	0.025	0	0.003	0	0.002	0	0.001	0.258

Table 12-4. [Indonesia: Relevant statistics for household socioeconomic groups]

Account Number	Household group	Income (Rp millions)	Income Per capita (Rp thousands)	Headcount elasticity w.r.t. mean income	Poverty Shares
		y_i	$\bar{y}_i$	$\lvert\varepsilon_i\rvert$	σ_i
24	Agricultural employees	1575.98	101.2	0.384	0.167
25	Small-scale farmers	4192.47	103.01	0.431	0.433
26	Medium-scale farmers	2430.05	218.92	3.169	0.056
27	Large-scale farmers	4484.69	320.34	5.553	0.038
28	Rural, nonagricultural: low skill level	5277.06	189.14	2.461	0.183
29	Rural, nonagricultural: high skill level	1945.62	240.2	3.665	0.012
30	Urban: low skill level	6099.31	306.5	1.712	0.099
31	Urban: high skill level	4453.76	468.82	6.533	0.012
	Total	30458.94	207.49		1.000

Sources: Columns (1) and (2): Keuning & Thorbecke (1992); Table 3.9
Columns (3), (4) and (5): Thorbecke & Jung (1996); Table 2

Table 12-5. [Percentage change in numbers of poor people (*dQ/Q*) arising from a one per cent change in exogenous sectoral demand (dx_p/x_p)]

	Production activity	Initial Exogenous Sectoral demand (Rp billions) x_P	Household groups								
			Agricultural employees 24	Small-scale Farmers 25	Medium-scale Farmers 26	Large-scale Farmers 27	Rural, non-agricultural low skill 28	Rural, non-agricultural high skill 29	Urban: low skill 30	Urban: high skill 31	All Indonesia households
47	Food crops	154.83	-0.005	-0.008	-0.045	-0.085	-0.017	-0.019	-0.010	-0.029	-0.01
48	Other crops	1447.61	-0.048	-0.038	-0.297	-0.656	-0.144	-0.156	-0.092	-0.256	-0.11
49	Livestock	116.08	-0.003	-0.004	-0.031	-0.074	-0.013	-0.015	-0.008	-0.022	-0.01
50	Forest & wood	1298.72	-0.020	-0.025	-0.200	-0.412	-0.152	-0.135	-0.093	-0.245	-0.08
51	Fishery	132.94	-0.004	-0.004	-0.029	-0.060	-0.015	-0.015	-0.010	-0.026	-0.01
52	Mining	12236.31	-0.024	-0.029	-0.258	-0.366	-0.270	-0.360	-0.205	-1.191	-0.13
53	Food processing	335.85	-0.007	-0.009	-0.064	-0.111	-0.036	-0.033	-0.022	-0.058	-0.02
54	Textiles	168.20	-0.002	-0.002	-0.020	-0.029	-0.018	-0.014	-0.014	-0.033	-0.01
55	Paper & metal products	4082.79	-0.015	-0.019	-0.168	-0.230	-0.187	-0.0148	-0.159	-0.412	-0.09
56	Chemicals & minerals	2646.81	-0.011	-0.014	-0.115	-0.163	-0.131	-0.113	-0.101	-0.301	-0.06
57	Electricity, gas & water supply	25.16	0.000	0.000	-0.001	-0.002	-0.002	-0.002	-0.001	-0.005	0.00
58	Building & construction	4509.53	-0.041	-0.050	-0.410	-0.656	-0.476	-0.336	-0.281	-0.710	-0.20
59	Public works: agriculture	522.49	-0.025	-0.007	-0.049	-0.076	-0.039	-0.038	-0.021	-0.068	-0.02
60	Public works: transporation	848.61	-0.006	-0.007	-0.052	-0.079	-0.073	-0.049	-0.053	-0.130	-0.03
61	Public works: utilities & commerce	487.39	-0.003	-0.004	-0.028	-0.043	-0.040	-0.027	-0.030	-0.071	-0.02
62	Public works: other	654.84	-0.005	-0.006	-0.047	-0.075	-0.064	-0.043	-0.047	-0.108	-0.03
63	Trade & transport services	49.41	-0.001	-0.001	-0.006	-0.008	-0.007	-0.005	-0.005	-0.013	-0.00
64	Restaurant & hotel services	386.45	-0.006	-0.008	-0.067	-0.111	-0.042	-0.040	-0.038	-0.100	-0.03
65	Land transport	45.10	-0.001	-0.001	-0.006	-0.008	-0.007	-0.005	-0.005	-0.011	0.00
66	Other transport and communication	403.94	-0.003	-0.003	-0.025	-0.037	-0.028	-0.028	-0.026	-0.093	-0.01
67	Finance, real estate & business services	159.56	-0.002	-0.002	-0.016	-0.024	-0.012	-0.015	-0.014	-0.050	-0.01
68	Education & health	917.54	-0.012	-0.013	-0.097	-0.159	-0.089	-0.428	-0.071	-0.499	-0.05
69	Personal & household services	498.14	-0.005	-0.006	-0.041	-0.063	-0.060	-0.054	-0.045	-0.109	-0.03
	Total	32128.30	-0.25	-0.26	-2.07	-3.53	-1.92	-2.08	-1.35	-4.54	-0.97

Source: Column (2): Keuning & Thorbecke (1992) Table 3.4; other columns authors' calculations

ANNEX

The identification of accounts within the Indonesia social accounting matrix.

The Indonesia social accounting matrix compiled by Keuning & Thorbecke has seventy-five distinct accounts. Those which cover domestic production activities, factors of production and domestic private sector institutions, together with the account numbers that identify each of these within our text tables, are as follows:

(i) Production activities

46 Trade & transport margins; 47 Food crops; 48 Other crops; 49 Livestock; 50 Forestry & wood; 51 Fishing; 52 Mining; 53 Food processing; 54 Textiles; 55 Paper & metal products;

56 Chemicals & minerals; 57 Electricity, gas & water supply; 58 Building & construction;

59 Public works: agriculture; 60 Public works: transportation; 61 Public works: utilities & commerce; 62 Public works: other; 63 Trade & transport services; 64 Restaurant & hotel services; 65 Land transport; 66 Other transport & communication; 67 Finance, real estate & business services; 68 Education & health; 69 Personal & household services.

(ii) Factors of production

1 Agricultural workers, paid, rural; 2 Agricultural, paid, urban; 3 Agricultural, unpaid, rural;

4 Agricultural, unpaid, urban; 5 Manufacturing, paid, rural; 6 Manufacturing, paid, urban;

7 Manufacturing, unpaid, rural; 8 Manufacturing, unpaid, urban; 9 Clerical, paid, rural;

10 Clerical, paid, urban; 11 Clerical, unpaid, rural; 12 Clerical, unpaid, urban; 13 Professional, paid, rural; 14 Professional, paid, urban; 15 Professional, unpaid, rural; 16 Professional, unpaid, urban; 17 Unincorporated capital, land; 18 Unincorporated capital, housing; 19 Other unincorporated capital, rural; 20 Other unincorporated capital, urban; 21 Incorporated capital, private; 22 Incorporated capital, public; 23 Incorporated capital, foreign.

(iii) Domestic Institutions

24 Agricultural employees; 25 Farmers, small land-holding; 26 Farmers, medium land-holding; 27 Farmers, large land-holding; 28 Rural, non-agricultural low-skilled; 29 Rural, non-agricultural high-skilled; 30 Urban, low-skilled; 31 Urban, high-skilled; 32 Companies.

Source: Keuning & Thorbecke (1992); Tables 3.1 and 3.2.

NOTES

[i] It had been our intention to review all of this literature in this paper. However, constraints on length in particular have mitigated against doing so. We have, therefore, restricted ourselves here to a statement of how we suggest that others might set about exploring the links between the stimulation of production and the reduction of poverty.

[ii] The essential difficulty is that the head-count ratio does not satisfy the principle of transfers. A wide range of alternative social evaluation criteria is not open to this objection.

[iii] Several of these twenty-five sub-matrices will be empty if the construction of the matrix adheres to standard national accounting conventions.

[iv] A third possibility is that an increase in $\mathbf{x}_p$ might be the result of an increase in public expenditure which may or may not also be beneficial.

[v] See Thorbecke [2003].

[vi] The work of Pyatt *et al* [1973] on Iran provides an early illustration.

REFERENCES

Defourny, J. and E. Thorbecke [1984] "Structural path analysis and multiplier decomposition within a social accounting matrix framework", *Economic Journal* Vol. 94, No. 373 (March, 1984) pp. 111-36.

Keuning, S. and E. Thorbecke [1992] "The social accounting matrix and adjustment policies: the impact of budget retrenchment on income distribution", Chapter 3 of E. Thorbecke (ed) *Adjustment and Equity in Indonesia* Paris, O.E.C.D.

Pyatt, G. *et al* [1973] "Methodology for macro-economic projections", International Labour Office *Employment and Incomes Policy for Iran* Technical Paper #12, Geneva, International Labour Office.

Pyatt, G. [2003] "An Alternative Approach to Poverty Analysis", *Economic Systems Research*, Vol. 15, No. 4 (June) pp. 113-133.

_______ and J.I. Round [1979] "Accounting and fixed price multipliers in a social accounting matrix framework", *Economic Journal* Vol. 89, pp. 850-73. reproduced in extended form as Chapter 9 of Pyatt, G. and J.I. Round (eds.) (1985) *Social Accounting Matrices: A Basis for Planning* Washington, D.C., the World Bank

_________________ [2003] "Multiplier analysis and the design of social accounting matrices" (mimeograph) University of Warwick.

Round, J.I. [2003] "Social Accounting Matrices and SAM-based Multiplier Analysis", Chapter 14 in F Bourguignon, and L A Pereira da Silva (editors) *Techniques and Tools for Evaluating the Poverty Impact of Economic Policies*, World Bank and Oxford University Press.

Stone, R. [1985] "The disaggregation of the household sector in the national accounts", Chapter 8 of Pyatt, G. and J.I. Round (eds.) *Social Accounting Matrices: A Basis for Planning* Washington, D.C., the World Bank

Thorbecke, E. [2003] "Poverty analysis and measurement within a general equilibrium framework", Chapter 3 of Edmonds, C.E. (ed.) *Reducing Poverty in Asia* Cheltenham, Edward Elgar.

___________ and H-S Jung [1996] "A multiplier decomposition method to analyses poverty alleviation", *Journal of Development Economics* Vol. 48, No.2, (March,) pp. 279-300.

Chapter 13

DEVELOPING AN ACCOUNTING MATRIX FOR THE EURO AREA: ISSUES AND APPLICATIONS

Tjeerd Jellema, Steven Keuning, Peter McAdam and Reimund Mink[i]
European Central Bank, Tjeerd.Jellema@ecb.int, steven.keuning@ecb.int, peter.mcadam @ecb.int, Reimund.Mink@ecb.int

1. INTRODUCTION

An important part of external or policy shocks is transmitted throughout the economy via various channels of transactions. In order to analyse such channels and to predict the impact of shocks, it is expedient to know who recently exchanged what with whom and for what purpose. At the macroeconomic level, such an analysis obviously requires considerable aggregation. For that reason, the national accounts are often taken as a starting point. However, whereas conventional national accounts contain much information on production, income, expenditure and financial transactions, they are less well developed when it comes to the inter-sectoral linkages, or, in other words, they lack "from-whom-to-whom" accounts. In view of the intricate relationship between financial and non-financial transactions, the inter-sectoral linkages should ideally be shown in both the financial and the non-financial accounts.

The most expedient format for presenting inter-sectoral linkages is in a matrix. The European System of Accounts (ESA) 1995 (Eurostat, 1996) puts this as follows: "A matrix presentation permits each transaction to be represented by a single entry and the nature of the transaction to be inferred from its position. Each account is represented by a row and column pair and the convention is followed that resources are shown in the rows and uses are shown in the columns." In each account, the row total (total incomings) is then by definition equal to the column total (total outgoings) and each cell describes how much the 'sector' in the corresponding row received from the 'sector' in the corresponding column on the account concerned.

National Accounting Matrices (NAM) are defined as the presentation of a sequence of national accounts and balancing items in a matrix, which elaborates the linkages between a supply and use table and institutional sector accounts. NAMs are described in detail in Chapter XX of the 1993 System of National Accounts (SNA93) (United Nations *et al.*, 1993). A very useful option of such a presentation of the accounts in a matrix is that different types of actors and groupings thereof can be selected in each account, without giving up the coherence and integration of the complete accounting system. In other words, in each account a unit (e.g. an institutional unit, a kind-of-activity unit, a product, or a financial asset) and an aggregation of units (e.g. a sector, an industry, a product group, or a financial asset category) can be selected that are most relevant to the kind of transactions that are depicted in that account (income distribution, production, capital, financial transactions, and supply and use of goods and services). Thus it is not necessary to distinguish between certain subsectors in all accounts just because one would like to distinguish these subsectors in some accounts. A case in point is the financial corporation or the general government sector that should be broken down by subsector in the primary distribution of income and the financial transactions, but can be consolidated in e.g. the secondary distribution of income accounts. Another advantage of a matrix format is its suitability for mathematical treatment using matrix algebra, which is in turn quite expedient for its use in all kinds of analyses and when balancing the accounts. Steven Keuning (1996, 1997) presents a more elaborate discussion of the advantages of a matrix presentation and of the estimation of such a matrix in both current and constant prices.

This paper considers the usefulness of such a NAM in the framework of the ECB's monetary policy analysis. Obviously, that requires a euro area instead of a national perspective, which poses particular statistical challenges (e.g. on the split of national balance of payments data into intra- and extra-euro area transactions so that a euro area rest-of-the-world account can be compiled, and on the elimination of asymmetries in the bilateral external trade statistics, but also in the financial accounts) – henceforth, we will speak about a euro area accounting matrix (EAAM). The EAAM has been designed in such a way that it complies with the requirements centred on what is called the monetary transmission mechanism (MTM): how does the ECB's monetary policy affect the euro area economy; through what channels does policy operate? Monetary policy is transmitted from the "financial" part of the economy to the "real" part. This view leads to concentrate on the relative importance of the various links between these two parts. This follows earlier work in this area by Eric Thorbecke *et al.* (1992). The rest of this paper is organised as follows. Section 2 reviews the potential uses and the specific methodological issues raised by the construction of the multi-country EAAM. Section 3 presents a pilot version of an annual EAAM and also a detailed description of the transactions in order to guide the reader through the matrix presentation of the accounts. Section 4 concludes.[ii] The detailed EAAM, the classifications selected, the data sources used and more details on the compilation method of the EAAM are available at http://www.ecb.int/pub/wp/ecbwp356.pdf (four annexes to the paper Tjeerd Jellema et al. (2004). "Developing a euro area accounting matrix: issues and applications", ECB Working Paper Series No. 356).

2. THE EAAM: USES AND METHODOLOGICAL ASPECTS

Eric Thorbecke (1992) describes the traditional rationale for the compilation of National and Social Accounting Matrices as long-run structural issues such as income distribution, tax reform, welfare analysis, etc. Nevertheless, the widespread use of integrated systems of *financial or flow-of-funds accounts* in central banks as noted by Susan McIntosh *et al.* (1999) suggests that the approach has an added value for monetary policy analysis. In this section, we discuss such issues: first their potential uses (Section 2.1) and then methodological and data-related issues which arise from such usage (Section 2.2).

2.1 Potential uses of Euro Area Accounting Matrices

The main merit of an EAAM is that it can help to understand the structure of the economy, including the financial transactions, and its development over time (provided that a time series of such matrices is available). This may provide useful, for instance, to gain more insight into (changes of) the monetary transmission mechanism at work in the euro area. Second, an EAAM provides a consistent accounting framework that can be used in the calibration of general equilibrium models. Third, time series of EAAMs can be used for more elaborate empirical studies on relationships between sectors of the euro area economy.[iii]

In a complex and data-demanding environment, EAAMs provide a consistent and coherent statistical framework for both the real and the financial side of the economy.[iv] As such, they offer the user a single, macroeconomic accounting framework with harmonised statistical concepts and the most "appropriate" level of classification of economic sectors, economic events, assets and liabilities, and the like. Among other uses, this is well suited to the ECB's monetary policy strategy, as described in ECB (2003), which puts emphasis on monitoring a wide range of economic indicators. Providing that EAAMs become available at a sufficiently high, i.e. quarterly, frequency, and with an acceptable delay of one quarter, say, it may allow a crosscheck of these indicators.

To give an example, central banks are clearly interested in the process by which monetary policy affects prices and output in the economy – i.e., the "Monetary Transmission Mechanism" (MTM). Peter McAdam and Julian Morgan (2003) provide an extensive discussion of a number of (not necessarily competing) views on the transmission process. A more in-depth empirical analysis of these views has been hampered by lack of sufficient data in three respects. First, how do households and corporations reallocate their portfolio asset or leverage ratios in response to monetary policy changes (i.e. wealth composition effects); second, how sensitive are different parts of the economy to monetary policy changes (i.e. distributional effects); and finally, how can the validity of the various alternative MTM theories be tested. Without claiming that accounting matrices are the panacea to resolving such uncertainties or uniquely delineate the channels of importance for monetary policy, the "from-whom-to-whom" features that are an essential ingredient of accounting matrices clearly facilitate an analysis of the interrelations in the economy, including the various channels of the monetary transmission.

Furthermore, all economic models require an explicit accounting framework. That is to say, they must fashion the data around a structure that has economic meaning. Obviously, this also requires categorising and classifying the data according to various types: e.g. factors of production (labour and capital), institutional units (households, corporations and government), and types of transactions (financial or non-financial and between residents or between residents and non-residents). Providing such an accounting framework also has implications for the type of data that must be collected and their degree of disaggregation. The accounting framework then provides the basis for the subsequent modelling assumptions, including the choice which variables are considered as exogenous (e.g. interest rates, tax rates, certain cross-border capital transactions) and which as endogenous.

This, in turn, illustrates one of the key benefits of the accounting matrix framework for modelling purposes. The EAAM, notably in view of its from-whom-to-whom framework of transactions, can provide a consistent statistical skeleton for macro-economic forecasting or computable general equilibrium (CGE) models that incorporate both the financial and the non-financial side of the economy, and their interrelationships. By contrast, it is not uncommon for macro-econometric models to imperfectly model flows between different agents – this can be expressed in trade volumes not adding up, incomplete circular flows of incomes between private sectors and the government sector, etc (see the discussion in e.g. John Whitely, 1994).

Another aspect is that accounting matrices are often used to calibrate both the baseline and certain parameter values of CGE models. Examples of such calibrated parameters include marginal propensities to consume various goods, tax elasticities, and share and technology parameters of the production process. Notably, this presupposes that the baseline values of the accounting matrix represent a "normal" year. Whereas CGE- or macro-models additionally embody a number of behavioural and other model specifications (e.g., uncovered interest parity, households and firms' optimisation, adjustment costs and other frictions, policy rules, etc.) as well as more dynamic features, the underlying accounting matrix fixes the various channels of interest in the economy and the related taxonomies. Such factors – e.g., accounting structure and calibration – are discussed in Thorbecke (1985, 2000).

Of course, all tools involve trade-offs. Possible disadvantages of EAAMs are that they may require additional data collection on the counterparts of transactions, some (limited) additional data compilation efforts and some calibration to achieve consistency.[v] For the time being, euro area countries do not compile NAMs on a regular basis, let alone every quarter. The Netherlands is the main exception, with an annual compilation frequency. National pilot studies have been conducted by Belgium, Greece, Italy, the Netherlands, Portugal, Finland, and the United Kingdom in the context of the Leadership Group SAM (2002). A direct compilation at the euro area level may be more efficient, but would still require sufficient basic data, particularly for the most important Member States. Moreover, compiling a multi-country accounting matrix entails some particular methodological complexities. These are spelled out in the next section.

2.2 Specific Methodological Issues when Compiling a Multi-country Accounting Matrix

In principle a NAM incorporates several semi-integrated accounting systems at the national level. These are: the Supply and Use Table (SUT), the non-financial sector accounts, the financial accounts and the balance sheets. The latter three parts together are also labelled the Integrated Economic Accounts (IEA). These components are extensively documented in the respective manuals published by United Nations *et al.* (1993) and European Commission (1996). A SUT provides a detailed presentation of the supply and use of goods and services **by production activities**. The supply of goods and services consists of the output by domestic production activities and imports, and the use of goods and services consists of the intermediate consumption by production activities and the final use categories: final consumption, capital formation and exports. The IEA on the other hand provide an overview of all economic transactions **by institutional sector,** as well as other flows and opening and closing balance sheets for non-financial and financial assets and liabilities.[vi] In the NAM groups of transactions are shown by institutional sector of origin (outgoings) and by institutional sector of destination (incomings). The challenge in the construction of a NAM for a single country is the development of a series of transaction matrices that are consistent with the data contained in the IEA. Conceptually, this adds a full dimension (that of **the counterpart institutional sector**) to the IEA integration framework leading to from-whom-to-whom accounts. At present, only a few countries already compile such transaction matrices as an integral part of the IEA.

The construction of an accounting matrix for a multi-country area such as the euro area poses additional methodological challenges. A multi-country area is treated as a single economy, with a single economic boundary distinguishing transactions between resident sectors and between residents and the rest of the world. As a consequence, when building up an EAAM from NAMs for the Member States, transactions between residents of a Member States and residents or the other Member States should no longer be treated as cross-border transactions, but as transactions between residents within the multi-country area. A multi-country accounting matrix is therefore not equal to the sum of the NAMs of the constituent countries.

The first step when deriving an EAAM from a set of national accounts per Member State is that the transactions as reflected in the national ROW accounts must be subdivided into transactions between residents and euro area residents (the so-called intra transactions) and transactions between residents and residents outside the euro area (the so-called extra transactions). As said before, the intra euro area transactions should then be reflected as transactions between residents in the various transaction matrices of the EAAM.

To date, the compilation of the euro area ROW is a challenge because national SUT and IEA statistics do not always provide a geographical breakdown between intra- and extra-euro area transactions. Additional data (e.g. the balance of payments) need to be integrated to obtain the desired split by area, sector and transaction category. Unfortunately, the ROW account and the balance of payments do not always match at the national level.

However, even if the required intra- and extra-euro area breakdowns were readily available in all national data sets, their summation at the level of the euro area as a whole would reveal that total uses or changes in assets of certain intra-euro area transactions do not equal the corresponding total resources or changes in liabilities and net worth of certain intra-euro area transaction. Such 'asymmetries' on the intra-euro area ROW account arise because of the different recording of transactions by Member States.[vii] The elimination of these asymmetries implies that adjustments are made at the level of the euro area, either to the resident transaction or to the extra euro area transaction. In any case, these asymmetries must be eliminated before the intra transactions can be removed from the ROW account and before they can be reflected as transactions between resident institutional units in the relevant transaction matrices. For that reason, it would be ideal to dispose of intra euro area transaction data classified by **country of origin** and **country of destination.** Data at this level of detail conceptually allows for a matrix presentation that distinguishes between Member Countries and institutional sectors in relevant parts of the EAAM. It would enable a distinction of transactions between the different institutional sectors located in different Member States.

In practice this level of detail is not available in source data and estimation methods must be used to compute transaction matrices that describe both the transactions between resident sectors of a Member State as well as the intra–euro area transactions between institutional sectors. The compilation of an EAAM is further complicated because integrated SUT and IEA are not yet available in all euro area Member States. At present, ten (out of twelve) Member States produce annual non-financial sector accounts, while nine countries produce annual financial accounts. Likewise, only eight Member States produce SUT tables. Only one Member State currently produces transaction matrices as part of its regular statistical output. The timeliness of the production of these accounts is another concern. For instance the transmission deadline for annual SUT tables is three years, whereas the annual non-financial institutional accounts are made available after one year, and the annual financial accounts are available after nine months.

Another issue is that international organisations that are located in a given country are not considered to be resident in that country. However, if the multi-country area includes all Member States of the international organisation concerned, a corresponding international organisation should arguably be treated as resident in the multi-country area. In the case of the euro area, this concerns the ECB, whereas in the case of the European Union (EU) this extends to all other EU institutions. Therefore a complete set of accounts must be compiled for such international organisations, to be aggregated with the sets of accounts describing the Member States.

The production of more timely and frequent EAAMs will depend on the availability of quarterly non-financial and financial sector accounts as well as on the 'quarterisation' of euro area annual SUT frameworks using available quarterly indicators. Currently much work is ongoing in the development of quarterly institutional sector accounts for the euro area, in accordance with the priorities for EMU statistics as set by the Ecofin Council. A sub-set of financial accounts for the euro area, the Table on Financing and Investment (TFI), is already published by the ECB in its Monthly Bulletin.[viii] Work is in progress to extend the coverage of the TFI to all sectors and financial instruments.

3. AN EAAM FOR 1999

This section presents a rather aggregated EAAM for 1999. It incorporates annual data for transactions as shown in the SUT and in the production, income and accumulation (capital and financial) account by institutional sector. As a pilot exercise it brings together different statistical data sources available at the ECB. They are described in Annex 2 of Tjeerd Jellema *et al.* (2004). Due to an as yet limited data availability it was decided to opt for a rather straightforward EAAM layout, corresponding to Table 20.4 of the 1993 SNA. To specifically accommodate the potential use of the EAAM in the MTM analysis, the 'allocation of primary income' account has been split into two accounts: the 'allocation of interest income' account and the 'allocation of other primary income' account.[ix]

This section is divided in two parts. Part one describes the overall structure of the EAAM in terms of the sequence of accounts and of some selected balancing items.[x] Part two provides more details on the individual accounts. It also deals with specific features that are relevant to the monetary policy framework.

3.1 Overall structure of the EAAM

The aggregated EAAM is presented in Table 13-1 below.[xi] It contains 13 accounts describing the processes of production, income generation and use, and accumulation of assets and liabilities, for resident sectors and for the rest of the world. This aggregate EAAM can be seen as a roadmap for the more detailed tables shown in the following section. The amounts in each submatrix of the detailed EAAM add up to a single number in a cell of Table 13-1.

The EAAM presents incoming transactions in the rows and outgoing transactions in the columns. For instance the cell in row 1 and column 8 represents final consumption expenditure, which is an incoming transaction or a resource in the goods and services account, and an outgoing transaction or a use with respect to the institutional sectors in the use of income account.

Row 1 and column 1 contain the *Goods and Services Account*, and row 2 and column 2 contain the *Production Account*. Together they show (in the more detailed EAAM) the production structure of the euro area. Column 1 reveals that the total supply of goods and services is composed of euro area production (EUR 11,151 billions; cf. row 2) and euro area imports (996 billions; cf. row 11), both recorded at basic prices. In order to adjust for the difference between the basic prices valuation of supply and the purchasers' prices (i.e. 'market prices') valuation of demand, column 1 also contains two sets of adjustments: the adjustment for trade and transport margins (0 in the aggregate matrix, but not in the more detailed tables; see below) as recorded in row 1, and the adjustment for taxes on products less subsidies on products (EUR 674 billions) in row 3. Row 1 presents the uses of goods and services: intermediate consumption (5,567) as a cost to production activities in column 2, final consumption expenditure by households and governments (4,846) in column 7, changes in stocks and net acquisition of valuables (19) in column 8, gross fixed capital formation (1,318) in column 9 and exports of goods and services (1,071) to the ROW in column 11. Of course, total demand equals total supply (EUR 12,821 billions). Column 2 shows that the production costs of all industries equal intermediate consumption (EUR 5,567 billions), net value added (4,714; cf. row 3) and consumption of fixed capital (870; cf. row 9). The euro area GDP (EUR 6,258 billions) can be calculated by adding taxes on products less subsidies on products (674 billions; cf. cell [3,1]) and consumption of fixed capital (870 billions; cf. cell [10,2]) to net value added.

The *Generation of Income Account* in row 3 and column 3 describes how production factors (e.g. employees) generate income and hand it over to their institutional sectors (e.g. households). First, in row 3 net value added generated by domestic activities is augmented by income earned outside the euro area (EUR 8 billions; cf. column 11). Then, in column 3 these incomes are paid out to euro area institutional sectors (5,387; cf. row 4) and to the ROW (EUR 10 billions; cf. row 11).

Accounts 4 and 5 present the allocation of primary incomes to the institutional sectors. In the EAAM, the *Allocation of Interest Income Account* (#4) and the *Allocation of Other Primary Income Account* (#5) are separated to emphasise the special role of interest in the transmission mechanism of monetary policy. The allocation of interest income account precedes the allocation of other primary income account. Because interest payments are typically contractual, accrue continuously and are often established before the other types of property income are known, it may indeed be assumed that institutions first assess the income after interest transactions before deciding/establishing the allocation of major other property incomes in the euro area, such as dividends. Concerning rents, it may be noted that rents of dwellings, other buildings, machinery, etc. are already settled in the production account, so that the rents as a part of 'other property income' only cover land rent. Assuming that land rent is settled after interest payments may not always be appropriate, but this will thus not have a large impact in macroeconomic terms.

In row 4, the euro area institutional sectors receive the income generated by euro area production factors (EUR 5,387 billions; cf. column 3) and the interest (EUR 1,365 billions; cf. column 4) from other euro area sectors and from the ROW (EUR 144 billions; cf. column 11). In column 4, these sectors hand over the interest payable to other euro area sectors (in total: EUR 1,365 billions; cf. row 4) and to the ROW (EUR 154 billions; cf. row 11). The balance of net generated income and interest of the euro area (EUR 5,376 billions) is then put on the next account, that is, on row 5.

In row 5, the allocation of primary income is completed with the allocation of the other property incomes. The income transferred from column 4 is augmented with other property income receivable from other euro area sectors (EUR 751 billions; cf. column 5) and from the ROW (39 billions; cf. column 11). Likewise, in column 5 other property incomes payable are handed over to other euro area sectors (EUR 751 billions; cf. row 5) and to abroad (EUR 61 billions; cf. row 11). The balance equals net income of the euro area (EUR 5,355 billions), which is put on the secondary distribution of income account in row 6.

The *Secondary Distribution of Income Account* (row and column 6) contains all transfer payments between institutional sectors. These transfer payments include taxes on income and wealth, social security contributions and benefits, and miscellaneous current transfers. To a large extent, these transactions occur within the euro area (EUR 4,125 billions; cf. cell [6,6]), although some transfers are obtained from the ROW (EUR 41 billions; cf. column 11) and some transfers are made to the ROW (EUR 74 billions; cf. row 11). As usual, the balance of the secondary distribution of income account, net disposable income of the euro area (EUR 5,321 billions), is allocated to the next, *Use of Income Account* (cf. row 7).

From disposable income, final consumption expenditure (EUR 4,846 billions) by institutional sectors is financed in row 1. In order to allow for changes in the net equity of private pension funds (EUR 43 billions) that are shown as a transfer between the corporate sector and the household sector in the more detailed EAAM, they are also included here, in row 7. The final balance of the current accounts, net euro area saving (EUR 475 billions), is allocated to the capital account (cf. row 8).

On the *Capital Account* (#8), net saving is augmented with capital transfers receivable from other euro area sectors (EUR 198 billions; cf. row 8) and from the ROW (EUR 18 billions; cf. column 12) and with liabilities incurred net, (EUR 3,885 billions; cf. column 10). These funds are used, in the corresponding column, to finance changes in inventories (EUR 19 billions; cf. row 1), net fixed capital formation[xii] (EUR 448 billions; cf. row 9), inter-sectoral capital transfers payable to euro area sectors (EUR 198 billions; cf. row 8) and to the ROW (EUR 8 billions; cf. row 12) and the net acquisition of financial assets ('lending') (EUR 3,870 billions; cf. row 10). However, this still leaves a discrepancy of EUR 32 billions between financing and investment, which is shown, for the time being, in an additional, so-called *Discrepancy Row* at the bottom of the table.

The next, *Fixed Capital Formation Account* (#9) serves to show in the detailed EAAM which institutional sector allocates net investment to what industry (here aggregated to cell [8,9]), which together with the consumption of fixed capital (cell [9,2]) is then used to purchase capital goods (here aggregated to cell [1,9]).

The *Financial Account* (# 10) shows in the row the net acquisition of financial assets by euro area sectors (EUR 3,870 billions; cf. column 8) and by the ROW (EUR 662 billions; cf. column 12). Column 10 then records the net incurrence of liabilities by euro area sectors (EUR 3,885 billions; cf. row 8) and by the ROW (EUR 643 billions; cf. row 12). At the aggregate level, the statistical discrepancy in this account rounds to EUR 3 billions; cf. the bottom row of the table.

The accounts 11 and 12 are the *Current and Capital Accounts for the Rest of the World.* The transactions on these accounts are presented from the viewpoint of the ROW. Thus the euro area receivables are now shown in the columns and the payables in the rows. All entries on these accounts have been described above, except for the euro area current account balance (EUR 7 billions), which is shown as a negative balancing item of the ROW current account (cell [12,11]) that is put on its capital account. Similarly, net lending of the euro area can be computed as borrowing by the ROW (EUR 643 billions, cell [12,10]) minus lending by the ROW (662 billions, cf. cell [10,12]), that is, EUR –19 billions. However, it should be borne in mind that in this pilot EAAM the statistical discrepancy on the ROW capital account (analogous to the so-called errors and omissions of the balance of payments) still equals EUR –37 billions (cf. the bottom row of the table).

The bottom row shows that the statistical discrepancies in the capital account (EUR 33 billions; cf. column 8), the financial account (EUR 3 billions, cf. column 10) and the ROW capital account (EUR –37 billions, cf.

column 12) cancel out (except for rounding errors), as must be the case by definition. Therefore, there is no need for an additional column to show statistical discrepancies.

3.2 Detailed EAAM

This section describes some key components of the detailed EAAM and also intends to follow some transactions related to a specific sector throughout the system of accounts. The EAAM in its full detail is shown in Table A1 annexed to the paper of Tjeerd Jellema *et al.* (2004).

3.2.1 Supply and use of goods and services

The detailed submatrices in the first two accounts, the Goods and Services Account and the Production Account, correspond to the SUT matrix for the euro area. Table 13-2 contains the supply table, shown in the traditional way, that is, by transposing the columns of the EAAM Goods and Services Account. It describes in rows 1 to 6 the euro area supply (at basic prices) of six product groups by six euro area industries (columns 2 through 8) and from imports (column 10). In addition, column 1 presents the trade and transport margins, with a compensating negative entry in the row for the trade, hotels, restaurants and transport industry, and taxes less subsidies are added in column 9. A salient fact of the euro area, in comparison to most of its Member States, is the relatively small share of imports (less than 15% for all product groups and less than 8% overall) in total supply.

Table 13-3 below contains the use table, again presented in its traditional format. For example, reading along row 2 it is possible to identify the various uses of mining and manufacturing products. From columns 1 through 7, it appears that a large share of these products (EUR 1,823 billions) is used as intermediate inputs in the mining and manufacturing industry itself. General government consumes EUR 59 billions and households (including non-profit institutions serving households) consume EUR 1,761 billions of these products. An amount equal to EUR 526 billions is used for gross fixed capital formation (cf. column 11). Finally, column 12 shows that exports of manufactured goods equal EUR 827 billions.

The Use Table also describes the cost structure of the various activities, for example in column 3 that of manufacturing. Rows 1 through 9 of this column demonstrate that the main intermediate inputs into manufacturing are manufactured products. Value added is shown by category in rows 10 through 12 of this table and consists of compensation of employees (row 10), other taxes less subsidies on production (row 11) and (in this case: gross) operating surplus and mixed income. In the euro area mining and manufacturing industry, the latter category amounted to 25% of gross value added and 7.5% of total production costs in 1999.

3.2.2 Allocation of primary income accounts

In the EAAM, the allocation of primary income account is split into two sub-accounts: the allocation of interest income account and the allocation of other primary income account. The details of these accounts are presented in Table 13-4.

As a first step, the income generated in production is allocated to the institutional sectors (incl. an unspecified sector in row 1) and to the ROW in columns 1 through 3. Almost all compensation of employees (EUR 3,106 billions) is earned by euro area households, while the taxes on production predominantly accrue to its governments (EUR 768 billions). Corporations only receive operating surplus (EUR 879 billions). The financial intermediation services indirectly measured (FISIM) are the main component of the negative entry (EUR –205 billions) in cell [1,3] of this table.

In rows 1 through 5 and columns 4 through 7 the interest matrix is shown. It is clear that financial corporations are the major recipients of interest income (EUR 866 billions), followed by households (EUR 200 billions). Financial corporations also pay the major part of interest (EUR 524 billions), followed by government (EUR 274 billions), households (EUR 195 billions) and non-financial corporations (EUR 164 billions). More than a quarter of households' interest income (EUR 54 billions) is obtained from the governments. It is also remarkable that households are net receivers of interest income.

In rows and columns 6 through 9, the allocation of other primary income is dealt with. This largely concerns the distribution of dividends from corporations to their (mainly household) owners and the distribution of property income earned by life insurance corporations and pension funds to their policyholders. Rows 10 through 14 then show the primary income balances, which add up to net euro area income. Clearly, the primary distribution of income leads to an increase of household income as generated directly in production, mainly through the households' receipts of non-interest income.

3.2.3 Secondary distribution of income and use of income

The secondary distribution of income, as shown in the first part of Table 13-5, mostly affects governments and households, through taxation, social contributions and benefits and other current transfers. In row 12, government receives EUR 1,677 billions from households, EUR 529 billions from other (layers of) government and EUR 180 billions from corporations. Most of these revenues are also paid out as transfers, to households (EUR 1,111 billions) and other (layers of) government (EUR 529 billions). Part of the secondary distribution of income is carried out through financial corporations, notably through insurance corporations and pension funds. This is shown in row 11, where financial corporations receive EUR 208 billions from households, and in column 3, where households receive EUR 185 billions from financial corporations. The diagonal matrix composed of rows 14 through 17 and columns 2 through 5 then shows net disposable income by institutional sector. The secondary income is lower than the primary income for all sectors except for government.

Subsequently, consumption expenditure of the euro area governments and households is shown as part of their use of income in rows 1 through 9 and columns 6 through 9 of Table 13-5. In row 17 and columns 6 and 7, an adjustment is made to reflect the increase in the value of the pension fund reserves of households with corporations. This adjustment is necessary in order to match an increase in the claims of households on pension funds.

The use of income account is closed by means of net saving in row 18 (for ease of presentation not shown as a diagonal matrix here). Euro area households saved EUR 384 billions in 1999, corporations around EUR 109 billions, and governments only EUR 19 billions. This amounts to a household savings ratio of 8.9% and a government savings ratio of 1.5%.

3.2.4 The accumulation account due to transactions

Table 13-6 contains the accumulation account due to transactions in the traditional T-accounts presentation. Saving, net capital transfers and the net incurrence of liabilities (financing) are used to acquire non-financial and financial assets (investment). This means that the EAAM capital and financial accounts have been combined and that changes in assets are shown on the left-hand side and changes in liabilities on the right-hand side. This facilitates an extension with the other flow accounts (revaluation and other changes in volume of asset account) and the balance sheets in a future extension of the EAAM. For the rest, the numbers add up to the aggregates presented in Table 13-1 above. Net lending / net borrowing is also shown as the balancing item of the capital account and of the financial account.

Financial transactions are presented with a breakdown by financial instrument and original maturity. Financial corporations in the euro area acquired more than half of the financial assets and incurred a similar amount of liabilities (EUR 2,631 billions of EUR 4,530 billions), leading to a rather balanced financial account for this sector.

Non-financial corporations incurred liabilities amounting to EUR 853 billions. This mainly consisted of loans (EUR 414 billions) and shares and other equity (EUR 241 billions). The issuance of debt securities was a less important source of financing for euro area non-financial corporations (EUR 51 billions). Non-financial corporations invested considerably in debt securities and in shares and other equity (EUR 96 billions and EUR 296 billions, respectively). They granted EUR 169 billions of loans to other institutional units (incl. inter-company loans).

Net borrowing of general government was of a similar magnitude as that of the non-financial corporations (EUR 83 billions), and was mainly catered for by a large issuance of debt securities (EUR 110 billions), while general government paid back loans worth EUR 21 billions. Financial investment of euro area general government was rather small (EUR 37 billions).

Households acquired EUR 551 billions of financial assets, which mainly consisted of insurance technical reserves (EUR 246 billions), shares and other equity (EUR 193 billions) and currency and deposits (EUR 116 billions), compensating a net sale of debt securities (EUR 22 billions). In parallel, they incurred loans, trade credits and other advances for a total amount of EUR 280 billions. As a result, their net lending equaled EUR 271 billions.

When translating the euro area balance of payments into an EAAM financial account for the rest of the world, broken down by the instruments listed in Table 13-6, it appears that most of the investment of the rest of the world was in shares and other equity (EUR 237 billions), debt securities (EUR 175 billions) and currency and deposits (EUR 167 billions). The main financing instruments were shares and other equity (EUR 396 billions), loans (EUR 164 billions) and debt securities (EUR 139 billions), while the amount of currency and deposits held as liabilities was reduced by EUR 86 billions. All in all, net lending of the rest of the world, and thus net borrowing of the euro area, amounted to EUR 20 billions.

3.2.5 Discrepancies

When compiling the pilot EAAM, one had to face the existing discrepancies between the net lending/net borrowing compiled in the capital account and the same balancing item in the financial transaction account. Other discrepancies were caused by the asymmetries in the intra-euro area trade estimates, errors and omissions on the balance of payments and approximations that had to be made because of missing countries and an incomplete coverage of transactions. As the pilot EAAM is directly or indirectly based on national data, existing national discrepancies between the various data sets translate into euro area discrepancies. Fortunately, sometimes these national discrepancies also cancel out.

All in all, the 1999 discrepancies are still rather substantial for households and for non-financial corporations. While the discrepancies for financial corporations and for the rest of the world are also quite (too) high, this item is relatively small for the government sector.

It is obvious that this pilot EAAM required many assumptions that can hopefully be replaced by more comprehensive actual data in the future. For a presentation of the compilation methodology used the reader is referred to Tjeerd Jellema *et al.* (2004). The next section draws some preliminary conclusions.

4. CONCLUSION

This paper has set out a euro area accounting matrix (EAAM) and considered possible applications and issues arising. The main advantage of an EAAM, in comparison to the traditional national accounting framework, is its presentation of inter-sectoral linkages, not taking into account national borders with monetary union and focusing on the matrix presentation that shows the counterpart sectors to all types of transactions (from-whom-to-whom accounts). In turn, revealing these linkages assists in detecting the interrelationships between portfolio shifts and restructuring of their liabilities of various sectors as a consequence of, or in anticipation of, monetary policy decisions. In addition, this feature potentially enables also tracing the impacts of a monetary policy decision from the financial to the non-financial side of the economy and back.[xiii] Finally, the EAAM can provide a consistent statistical skeleton for macro-economic forecasting models that incorporate both the financial and the non-financial side of the economy.

As to date neither annual euro area supply and use tables nor institutional sector accounts were completely available, the process of compiling the non-financial account components for the EAAM has been lengthy, with many intermediate steps. In itself, these intermediate steps, such as the first ever compilation of non-financial institutional sector accounts for the euro area and the first ever attempt to construct a euro area Supply and Use framework, were already a valuable experience. In addition, the integration of this work with the ongoing and much more advanced compilation process of quarterly euro area financial accounts (ECB, 2000, 2001, 2002; Reimund Mink, 1999, 2002) has provided useful new insights.

The three guiding principles behind the compilation of the sector accounts in the EAAM have been the following:

- *Completion*, that is, supplementary estimates have been made for Member States that do not provide the full national accounts as outlined in the ESA95 transmission programme.
- *Transformation*, that is, the euro area has been transformed into a 'national' economy. Specifically, the ROW account in the EAAM only reflects transactions of the euro area with residents outside the euro area. For this purpose, cross-border transactions within the euro area have been transformed into 'resident' transactions between specific institutional sectors.
- *Consistency*, that is, accounting relationships have been used together with additional information and assumptions to yield internally consistent EAAM accounts, apart from relatively small, remaining discrepancies that could not yet be eliminated at this stage.

Yet, the results shown in this paper should be seen as a pilot version that serves as a basis for a further elaboration, both concerning its compilation assumptions and the plausibility of the outcomes. Besides, eliminating the statistical differences between the non-financial and the financial accounts, especially for the non-financial sectors and the ROW, remains high on the agenda.

The assembly of the EAAM would have been much easier if the different components such as euro area supply and use tables and institutional sector accounts, had been readily and regularly (at least annually) made available. These components are anyhow in high demand by the ECB and other users. This should then be step-by-step supplemented by sufficient and timely quarterly data. In that regard, existing EC Regulations on quarterly government data, the forthcoming EC Regulation[xiv] on quarterly institutional sector accounts and the ECB Guideline on quarterly financial accounts for the euro area are of great importance.

As regards the longer-term future, the ultimate objective is to take the framework to a quarterly frequency. Indeed, in that regard, the present work on an annual EAAM should be seen as a first step and as a diagnostic tool. A time-series of quarterly EAAMs would of course be more suited to timely and relevant policy analyses. The key to reaching this stage is data availability. For example, whilst data for financial transactions and balance sheets are mostly available at quarterly frequency, supply and use tables are less well developed at that frequency. Investigating the options for a quarterly representation of the EAAM is therefore an ongoing concern.

Table13-1. [Aggregate EAAM for 1999]

EAAM 1999 **EUR Billions**	0 Goods and Services (Products) 1	I Production (Industries) 2	II.1.1 Generation of Income Account (Transactions) 3	II.2.1.a Allocation Interest Income Account (Sectors) 4	II.2.1.b Allocation of Other Primary Income Account (Sectors) 5	II.2.2. Secondary Distribution of Income Account (Sectors) 6	II.3 Use of Income Account (Sectors) 7	III.1.a Capital Account (Sectors) 8	III.1.c Fixed Capital Formation (Industries) 9	III.2. Financial Account (Instruments) 10	V Rest of the World Account (Cur/Acct) 11	(Cap./Acct) 12	Totals
0 Goods and Services 1 Product Groups	Trade and Transport margins	Intermediate Consumption 5,567					Final Consumption expenditure 4,846	Changes in Inventories 19	Gross fixed capital formation 1,318		Exports of goods and services 1,071		12,821
I Production 2 Industries	Output 11,151												11,151
II.1.1 Generation of Income 3 Value Added Categories	Taxes less subsidies on products 674	*Net value added in basic prices* 4,714									Comp. of empl. and net taxes 8		5,396
II.2.1.a Allocation Interest Income 4 Institutional Sectors			*Net generated income in basic prices* 5,387	Interest 1,365							Interest 144		6,896
II.2.1.b Allocation Other Primary Inc 5 Institutional Sectors				*Net income after interest* 5,376	Other Property Income 751						Other Property Income 39		6,166
II.2.2 Sec. Distr Income 6 Institutional Sectors					*Net National Income* 5,355	Transfers 4,125					Transfers 41		9,520
II.3 Use of Income 7 Institutional Sectors						*Net disposable income* 5,321	Adj. in net equity on pension funds 43				Adj. in net equity on pension funds -		5,365
III.1.a Capital Account 8 Institutional Sectors							*Net Savings* 475	Capital Tranfers 198		Borrowing 3,885		Capital Transfers 18	4,560
III.1.c Fixed Capital Formation 9 Industries		Consumption of Fixed Capital 870						Net Fixed Capital Formation 448					1,318
III.2 Financial 10 Instrument Categories								Lending 3,870				Lending 662	4,529
S2.a ROW current Account 11	Imports of goods and services 996		Comp. of empl. and net taxes 10	Interest 154	Other Property Income 61	Transfers 74	Adj. in net equity on pension funds -						1,295
S2.b ROW capital account 12								Capital Transfers 8		Borrowing 643	Current Account - 7		643
Statistical Discrepancy	- 0	0	1	- 0	0	- 0	- 0	18	0	- 3	0	-	
Totals	12,821	11,151	5,396	6,896	6,166	9,520	5,365	4,561	1,318	4,529	1,295	643	

Table 13-2. [Aggregate EAAM for 1999 Supply Table]

Supply Table			Act 0	Act 1								Act II.1.1	Act V.a	Total
			Goods and Services	Production Account								Generation of Income	Current ROW	
			G4	AN	A1	A2	A3	A4	A5	A6				
			Trade, Hotels and Restaurants, Transport 1	Not specified activity, nominal sector 2	Agriculture Agriculture, hunting, forestry and fishery 3	Industry, including energy 4	Construction 5	Trade, Hotels and Restaurants, Transport 6	Financial, real-estate, renting and business activities 7	Other service activities 8	Taxes less Subsidies on Products 9	Imports 10	Total Supply at Purchasers Prices 11
G1	Agriculture, forestry, fisheries and aquacultur	1	85	-	252	0	-	1	0	0	-15	41	365
G2	Mining and quarrying, manufactured products	2	974	-	15	3,755	4	14	1	5	440	771	5,979
G3	Construction work	3	-	-	1	7	758	4	1	4	64	2	842
G4	Trade, Hotels, Restautrants, Transport	4	-1,068	-	1	48	3	2,127	2	5	44	54	1,215
G5	Finance, Rent, Business Services	5	2	-	0	39	2	21	2,303	32	107	70	2,576
G6	Other services	6	6	-	1	3	0	3	1	1,737	34	16	1,803
G7	Cif-FOB Adjustment	7	-	-	-	-	-	-	-	-	-	-1	-1
G8	Expenditures abroad	8	-	-	-	-	-	-	-	-	-	-	-
G9	Expenditures by non-residents	9	-	-	-	-	-	-	-	-	-	42	42
Total	Total Cost of Production = Output	10	-0	-	270	3,852	767	2,170	2,309	1,785	674	996	12,821

Euro Billion

Table 13-3. [Use Table]

Use Table			Act. I Production Account							Act II.3 Use of Income		Act III.1 Capital		Acc V.a ROW	Total
			AN	A1	A2	A3	A4	A5	A6	S13	S14, S1	P52+P5	P.51	P.6	
			Not specified activity, nominal sector 1	Agriculture , hunting, forestry and fishery 2	Industry, including energy 3	Construction 4	Trade, Hotels and Restaurants, Transport 5	Financial, real-estate, renting and business activities 6	Other service activities 7	Final Consumption of General Government 8	Final Consumption of Households and NPISH 9	Changes in inventories and net acq. Of Valuables 10	Gross Fixed Capital Formation 11	Exports of Goods and Services 12	Total Use in purchasers prices 13
G1	Agriculture, forestry, fisheries and aquaculture	1	- 0	30	171	1	11	1	5	0	116	3	8	19	365
G2	Mining and quarrying, manufactured products and	2	- 0	73	1,823	269	330	95	205	59	1,761	11	526	827	5,979
G3	Construction work	3	- 0	2	25	71	18	60	28	2	19	0	618	0	842
G4	Trade, Hotels, Restautrants, Transport	4	- 0	4	114	11	279	63	53	7	596	2	3	82	1,215
G5	Finance, Rent, Business Services	5	205	12	374	94	274	499	150	40	719	2	154	54	2,576
G6	Other services	6	- 0	3	39	4	27	44	102	1,138	427	0	9	9	1,803
G7	Cif-FOB Adjustment	7	-	-	-	-	-	-	-	-	-	1	-	- 2	- 1
G8	Expenditures abroad	8	-	-	-	-	-	-	-	-	- 81	-	-	81	-
G9	Expenditures by non-residents	9									42	-			42
D1	Compenstion of Employees	10	-	34	794	192	653	488	944						
D29-D3	Taxes less Subsidies on Production	11	-	- 5	22	4	24	44	7						
B2+B3	Oeprating Surplus and Mixed Income	12	- 205	77	274	99	394	726	149						
Total	Total Cost of Production = Output	13	-	229	3,634	745	2,012	2,020	1,642	1,247	3,600	19	1,318	1,071	12,821

Euro, Billion

Table 13-4. [Allocation of Primary Income]

Allocation of Primary Income Accounts			Act II.1.1 Generation of Income			Act II.2.1.a Allocation of Interest Income				Act II.2.1.b Allocation of other primary income				Act V.a Cur. ROW	
			D1	D2-D3	B2+B3	S11	S12	S13	S14_15	S11	S12	S13	S14_15	S2	Total
			Compenstion of Employees	Taxes less Subsidies on Production	Oeprating Surplus and Mixed Income	Non Financial Corp.	Financial Corp.	Government	Households and NPISH	Non Financial Corp.	Financial Corp.	Government	Households and NPISH		
			1	2	3	4	5	6	7	8	9	10	11	12	13
S1N	S1N	1			- 207	-	207	-	-						-
S11	Non Financial Corp.	2			779	1	41	24	-					8	853
S12	Financial Corp.	3			100	160	323	187	195					115	1,081
S13	Government	4		768	- 3	0	16	9	-					3	794
S14_15	Households and NPISH	5	3,106		843	3	143	54	-					18	4,167
S11	Non Financial Corp.	6				633	-	-	-	75	4	0	1	26	739
S12	Financial Corp.	7				-	252	-	-	51	46	-	0	6	355
S13	Government	8				-	-	519	-	26	6	0	2	0	553
S14_15	Households and NPISH	9				-	-	-	3,972	347	188	0	4	7	4,519
S11	Non Financial Corp.	10								197	-	-	-		
S12	Financial Corp.	11								-	93	-	-		
S13	Government	12								-	-	553	-		
S14_15	Households and NPISH	13								-	-	-	4,511		
S2	Rest of the World	14	6	4		56	98	-	-	43	18	-	0		225
Total	Total	15	3,112	771.853	1512.43	853	1,081	794	4,167	739	355	553	4,519	183	

Euro, Billion

Table 13-5. [Secondary Distribution of Income and Use of Income]

EAAM 1999			Act II.2.a	Act II.2.1.a				Act II.2.1.a				Act V.a Cur. ROW	Total
Secondary Distribution and Use of Income				Secondary distibution of Income				Use of Income					
				S11	S12	S13	S14_15	S11	S12	S13	S14_15	S2	
			Net National Income 1	Non Financial Corp. 2	Financial Corp. 3	Government 4	Households and NPISH 5	Non Financial Corp. 6	Financial Corp. 7	Government 8	Households and NPISH 9	10	11
G1	Agriculture, forestry, fisheries and aquaculture	1								0	116		
G2	Mining and quarrying, manufactured products	2								59	1,761		
G3	Construction work	3								2	19		
G4	Trade, Hotels, Restautrants, Transport	4								7	596		
G5	Finance, Rent, Business Services	5								40	719		
G6	Other services	6								1,138	427		
G7	Cif-FOB Adjustment	7								-	-		
G8	Expenditures abroad	8								-	- 81		
G9	Expenditures by non-residents	9								-	42		
S11	Non Financial Corp.	10	197	-	25	7	62					1	291
S12	Financial Corp.	11	93	30	4	6.25109	208					4	346
S13	Government	12	553	138	42	529	1,677					22	2,962
S14_15	Households and NPISH	13	4,511	55	185	1,111	47					13	5,921
S11	Non Financial Corp.	14		68	-	-	-						68
S12	Financial Corp.	15		-	84	-	-						84
S13	Government	16		-	-	1,266	-						1,266
S14_15	Households and NPISH	17		-	-	-	3,904	12	31				3,946
B8N	Net Savings	18						56	53	19	347		
S2	Rest of the World	19		1	6	43	24						
Total	Total	20		291	346	2,962	5,921	68	84	1,266	3,946		

Table 13-6 [Accumulation account due to transactions by sector]
EUR billions, 1999

Investment
Changes in assets

Net saving, capital transfers and financing
Changes in liabilities and net worth

Total	S.1	S.11	S.12	S.13	S.13/ S.14	S.2	ESA95 code	Transaction	Discre-pancy	Total	S.1	S.11	S.12	S.13	S.13/ S.14	S.2
							B8.n	**Saving, net**		475	475	56	53	19	347	
							B12	**External balance**								-7
223	206	6	23	152	25	18	**D.9**	**Capital transfers**		223	216	61	4	103	49	8
1337	1337	718	38	160	421		**P.5**	**Gross capital formation**								
1318	1318	704	38	159	417		P.51	Gross fixed capital formation								
16	16	14	0	1	1		P.52	Changes in inventories								
4	4	0	0	0	3		P.53	Valuables								
-870	-870	-497	-30	-107	-236		**K.1**	**Consumption of fixed capital**								
0	0	4	0	-1	-3		**K.2**	**Net acquisition of non-produced non-financial assets**								
0	*17*	*-115*	*26*	*-82*	*189*	*-17*	***B.9***	***Net lending (+) / net borrowing (-)***								
691	*673*	*232*	*31*	*203*	*206*	*18*		***Total***	*-7*	*698*	*690*	*117*	*57*	*122*	*395*	*1*
							B.9F	**Net lending (+) / net borrowing (-)**		2	-14	-126	-76	-83	271	20
0	-1		-1			1	**F.1**	**Monetary gold and SDRs**								
775	608	22	441	29	116	167	**F.2**	**Currency and deposits**	1	774	860	5	836	18		-86
1,138	963	96	878	10	-22	175	**F.3**	**Securities other than shares**	6	1,131	993	51	832	110	0	139
676	501	87	426	10	-22	175	F.33	Securities other than shares, excluding FD	0	676	534	45	379	109	0	142
106	48	33	40	1	-26	58	F.331	Short-term	7	99	103	22	125	-44	0	-4
570	453	54	385	9	5	117	F.332	Long-term	-7	576	430	23	254	153	0	146
462	462	9	453	0	0		F.34	Financial derivatives (FD)	7	455	459	6	452	0	0	-3
875	810	169	627	11	2	65	**F.4**	**Loans**	-8	883	719	414	76	-21	250	164
264	234	76	155	4	0	30	F.41	Short-term		279	209	142	48	9	10	71
611	576	94	472	7	3	35	F.42	Long-term		604	511	272	28	-30	241	93
1,230	993	296	539	-36	193	237	**F.5**	**Shares and other equity**	9	1,220	825	241	583	0	0	396
898	663	270	401	-42	35	235	F.51	Shares and other equity, excluding MFS		777	409	241	168	0	0	368
332	329	26	138	7	159	2	F.52	Mutual funds shares (MFS)		443	415		415			28
260	260	6	8	0	246	0	**F.6**	**Insurance technical reserves**	-3	263	263	9	252	0	1	0
229	229				229	0	F.61	Net equity of households in life insurance and pension fund reserves	1	227	228	9	217	0	1	0
31	31	6	8	0	18	0	F.62	Prepayments of insurance premiums	-4	35	35		35			0
255	238	138	64	22	15	17	**F.7**	**Other accounts receivable/payable**	-4	259	226	132	52	13	28	30
4,532	3,870	727	2,555	37	551	662	**F**	**Total**	2	4,530	3,885	853	2,631	120	280	643

NOTES

[i] Without implicating, we thank Dieter Gerdesmeier, Celestino Gíron, Linda Kezbere, Kalle Siljander and, especially, two anonymous referees and the Editorial Board of the ECB working paper series, one anonymous referee for this conference volume as well as seminar participants at the conference in honour of Erik Thorbecke at Cornell University, October 2003 for their valuable contributions. The views expressed in this paper are not necessarily those of the ECB. A longer version of this paper appeared in the ECB's working paper series, Tjeerd Jellema *et al.* (2004).

[ii] A further development of this NAM into a Social Accounting Matrix (SAM) may further enhance the monetary policy analysis, particularly through its additional data on employment and wage rates by type of labour.

[iii] Please note that a country breakdown can be inserted in (some accounts of) the EAAM, if and when relevant for the analysis. This then follows the general rule in EAAMs, namely that in each account a breakdown is adopted that is most relevant for the economic processes that are described in this account. For example, a country breakdown may be more relevant in some non-financial accounts for institutional sectors than in the financial accounts or in the classification of product groups.

[iv] Obviously, this advantage applies to an integrated set of financial and non-financial national accounts more generally, albeit that a matrix integrates the so-called supply and use tables and sector accounts in a single presentation format, a matrix allows for a more flexible selection of the most suitable classification in each account and allows for a more in-depth analysis of inter-sectoral linkages, e.g. spill-over effects of external or policy shocks.

[v] More actual data are available on the counterparts of transactions than is commonly assumed. For instance, the counterpart sector of government transactions can often be inferred from the nature of the transaction (e.g. the type of tax) and the same applies to many financial and concomitant property income transactions (e.g. conducting mortgage loans and paying interest on them). On the other hand, some degree of estimation based on assumptions may be indispensable for smaller items, like 'other transfers'.

[vi] Conceptually the non-financial (flow) accounts and the financial flow accounts yield identical balances by institutional sector (net lending). In practice, though, this is not achieved in all (euro area) countries.

[vii] Important asymmetries occur in the intra euro area trade statistics (Intrastat). Intra-EU exports ('dispatches') have consistently grown faster than intra-EU imports ('arrivals').

[viii] See ECB (2001) for a description of the TFI.

[ix] See ECB (2000) for an overview of the data that are collected by the ECB.

[x] See Keuning and De Ruijter (1988) and Leadership Group SAM (2002) for a more general discussion on the structure and classifications to be used in such accounting matrices.

[xi] Table 13-1 is also available to download from our working paper at the address: http://www.ecb.int/pub/wp/ecbwp356.pdf.

[xii] In this EAAM, this item includes the acquisition of net non-produced non-financial assets.

[xiii] See e.g. Jacques Defourny and Erik Thorbecke (1984) for a method to trace 'policy impact paths' on the basis of a framework as laid out in this paper.

[xiv] Proposal for Parliament and Council Regulation 2003/0296 (COD)

REFERENCES

Bacharach, M., (1970). *Bi-proportional matrices and input-output change*, Cambridge University Press.

Defourny, J. and E. Thorbecke. (1984). "Structural Path Analysis and Multiplier Decomposition within a Social Accounting Matrix Framework", *Economic Journal*, 94, 373, 111-136.

European Central Bank (2000). "Statistical information collected and compiled by the ESCB," *ECB Monthly Bulletin*, May.

European Central Bank (2001). "Financing and financial investment of the non-financial sectors in the euro area", *ECB Monthly Bulletin*, May, 75-82.

European Central Bank (2002). "Saving, financing and investment in the euro area", *ECB Monthly Bulletin*, August, 65-76.

European Central Bank (2003). "The outcome of the ECB's evaluation of its monetary policy strategy", *ECB Monthly Bulletin*, June, 79-92.

European Commission (1996). Council Regulation (EC) No 2223/96 on the European system of national and regional accounts in the Community (ESA 1995)

Jellema, T., S. J. Keuning, P. McAdam and R. Mink (2004). "Developing a euro area accounting matrix: issues and applications", ECB Working Paper Series No. 356. Available at http://www.ecb.int/pub/wp/ecbwp356.pdf

Keuning, S. J. and W. de Ruijter (1988). "Guidelines to the Construction of a Social Accounting Matrix", *The Review of Income and Wealth*, 34, 1, March.

Keuning, S. (1991) "Proposal for a Social Accounting Matrix which fits into the next system of national accounts", *Economic Systems Research*, 3, 3.

Keuning, S. J. (1996). *Accounting for Economic Development and Social Change*, IOS Press, Amsterdam.

Keuning, S. J. (1997). "SESAME: an Integrated Economic and Social Accounting System", *International Statistical Review*, 65, 1, 111-121.

Leadership Group SAM (2002). *Handbook on Social Accounting Matrices and Labour Accounts*, Populations and social conditions 3/2003/E/N 23, Eurostat, Luxembourg.

McAdam, P. and J. Morgan (2004). "The Effects of Euro Area Interest Rate Changes: Evidence from Macroeconomic Models", *National Institute Review*, 187, January, 70-80.

McIntosh, S. H., J. M. Scherschel, and A. M. Teplin (1999). "Use of Flow of Funds Accounts For Policy Making At The Federal Reserve", presented at *Central Bank Uses of Financial Accounts*, European Central Bank, 22nd November.

Mink, R. (1999). "Monetary Union Financial Accounts for ECB Monetary Policy Analysis", presented at *Central Bank Uses of Financial Accounts*, European Central Bank, 22nd November.

Mink, R. (2002). "Quarterly Monetary Union Financial Accounts for ECB Monetary Policy Analysis", *IFC Bulletin*, 12, October, 98-115.

Stone, R., J. E. Meade and D.G. Champernowne (1942). "The precision of National Accounts Estimates", *Review of Economic Studies*, 9 (2), 11-125.

Stone, R. (1976). "The development of economic data systems". In *Social Accounting for Development Planning with special reference to Sri Lanka* (G. Pyatt et. al, eds.). Cambridge, Cambridge University Press.

Thorbecke, E., (1985). "The Social Accounting Matrix and Consistency-Type Planning Models", in G. Pyatt and J. I. Round (Eds.), *Social Accounting Matrix: A Basis for Planning*, World Bank Publications.

Thorbecke, E. and associates (1992). *Adjustment and Equity in Indonesia*, Organisation for Economic Cooperation and Development, Paris.

Thorbecke, E. (2000). "The Use of Social Accounting Matrices in Modelling", paper presented at *the 26th General Conference of The International Association for Research in Income and Wealth*, Cracow, Poland, 27 August to 2 September.

Thorbecke, E., B. Kim, D. Roland-Holst and D. Berrian (1992). "A Computable General Equilibrium Model Integrating Real and Financial Transactions" in Thorbecke, E. and associates, *Adjustment and Equity in Indonesia*, Organisation for Economic Cooperation and Development, Paris.

United Nations, Eurostat, International Monetary Fund, Organisation for Economic Cooperation and Development and World Bank (1993). *System of National Accounts 1993*, Series F, No. 2, Rev. 4, United Nations, New York.

Whitely, J. (1994). *A Course in Macro Economic Modelling and Forecasting*, Harvester.

Chapter 14

GLOBALIZATION, ECONOMIC REFORM, AND STRUCTURAL PRICE TRANSMISSION: SAM DECOMPOSITION TECHNIQUES WITH AN EMPIRICAL APPLICATION TO VIETNAM

David Roland-Holst[1*] and Finn Tarp[2]
[1]Mills College and University of California, dwrh@mills.edu and [2]University of Copenhagen, Finn.Tarp@econ.ku.dk

1. INTRODUCTION

The advent of globalization and market reforms in many countries has significantly raised awareness of the mechanism of price transmission. Informal pressures arise from border prices, and more direct challenges are posed by conformity requirements in WTO negotiation and a myriad of regional trade arrangements. Individuals as well as public and private institutions have started to think about prices and markets in ways that only theoretical economists did a century ago. As economic linkages continue to proliferate, especially across international boundaries, we all become more acquainted with how markets interact with one another and how livelihoods are connected through the price system. Nevertheless, the consequences of market reform are difficult to anticipate, especially in former centrally planned economies. They used to rely on administered prices, which served a variety of policy objectives. Thus, the political economy of transition to a market economy is complex, and the lack of historical experience with price dynamics is a stumbling block in reform efforts. Better visibility for policymakers of the incidence of price transmission is clearly an important need.

In a series of seminal contributions, Graham Pyatt and Jeffery I. Round (1979), Jacques Defourny and Erik Thorbecke (1984), and Round (1985) each elaborated methods of economy-wide empirical analysis based on the social accounting matrix (SAM). They have been intensively used to examine the income generating process in scores of countries, and justifiably so.[i] It can also be noted that the unifying aspect of this work has been a focus on quantity-driven income determination, the primal side of the underlying accounting system. Despite their flexibility and tractability, SAMs have generally not been used to examine price transmission. This is somewhat surprising. Price formation is an essential issue in economic policy. Thus, in this paper, we propose to start filling this gap and demonstrate how a dual perspective on SAM multiplier methods can help shed light on direct and indirect price and cost linkages across an economy. We use Vietnam as case, including a new and detailed SAM estimated for the year 2000. The result is a practical framework for incidence analysis that should help improve visibility for policy makers seeking to facilitate economic reform and structural adjustment.

The structural perspective advocated here is of special relevance to economies with extensive price administration, and even more so when, as is the case in Vietnam, prices represent only one aspect of an extensive set of institutional rigidities. For an economy like this, which is at the early stages of both domestic and external market reform, price transmission is likely to be significant in the transition process. First, in the past factor prices did not conform to the neoclassical paradigm. They were implicitly indexed to commodity prices or cost-of-living effects, so the initial misalignment between the administered prices and their real or hypothetical market counterparts is likely to be large. Second, reforms geared at addressing institutional rigidities and market failures can be expected to facilitate transmission of price shocks rather than moderate them. These issues are of great policy relevance in economies opening up to global market forces, and even more so when they are simultaneously undertaking extensive domestic market reforms as in Vietnam. Of course, no real-world economy is purely neoclassical or purely structural, but by analyzing the structural components in isolation, we gain additional knowledge about the transmission mechanism for prices and its implications for policy and welfare.

Section 2 presents the dual formulation of the SAM multiplier models, whereas Section 3 recasts two leading decomposition techniques in this framework. They are applied in Section 4, which reveals patterns of transmission from external price shocks to domestic incomes and a variety of producer, factor, and consumer price/cost indices. Section 5 concludes.

2. METHODOLOGY

A standard SAM offers a disaggregate view of value flows in a given base period, detailing the direct linkages among its component sectors and institutions and pointing out the scope of the underlying indirect interactions. Inflows from exogenous sectors that stimulate the level of activity of a production sector, for instance, will also induce additional factor incomes that, once distributed among households, will be used to finance new final demand for producer goods and services. Table 14-A1 in the Annex depicts a partitioned simplified MacroSAM with four classes or groups of accounts, namely production, factors, households, and a consolidated account of the remaining sectors (government, capital and foreign accounts). Columns of the SAM indicate payments and rows tally receipts.[ii] For each group total spending is necessarily equal to total receipts, i.e., column and row totals of the matrix are equal.

A SAM-based quantity model is derived from the SAM-table by distinguishing endogenous and exogenous groups and assuming activity levels may vary while prices are fixed. This assumption is justified in the presence of excess capacity and unused resources in production activities. Suppose group 1 is chosen as endogenous and 2, 3 and 4 are exogenous. Let A_{ij} denote the matrix of normalized column coefficients obtained from T_{ij} and let υ_i denote that the incomes of groups *i=2,3,4* are taken as given exogenously. Then the income level of group 1 can be expressed by:

$$Y_1 = A_{11}Y_1 + A_{13}\,\upsilon_3 + A_{14}\,\upsilon_4$$

$$= (I - A_{11})^{-1}(A_{13}\,\upsilon_3 + A_{14}\,\upsilon_4) = M_{11}x \qquad (1)$$

where $M_{11} = (I-A_{11})^{-1}$ is the usual interindustry Leontief inverse and x is a vector of exogenous income levels. Since (1) implies $\Delta Y_1 = M_{11}\,\Delta x$, matrix M_{11} is also termed the multiplier matrix. Column *i* of M_{11} shows the global effects on all endogenous activity levels induced by an exogenous unit inflow accruing to *i*, after allowing for all interdependent feedbacks to run their course.

Consider now the dual case, where prices are responsive to costs but not to activity levels. The justifying assumption here, in addition to the usual excess capacity condition, is generalized homogeneity and fixed coefficients in activities. This is a situation where the classical dichotomy between prices and quantities holds true and prices can be computed independently of activity levels. Let now p_i denote a price index for the activity of group *i*.[iii]

With the same classification of endogenous and exogenous accounts and identical notational conventions, column 1 of the SAM yields:

$$p_1 = p_1A_{11} + \pi_2A_{21} + \pi_4A_{41}$$

$$= (\pi_2A_{21} + \pi_4A_{41})(I - A_{11})^{-1} = v_1M_{11} \tag{2}$$

where v_1 is a row vector of exogenous costs (factor payments, taxes, import costs) and M_{11} is the same multiplier matrix as in (1). Notice that from (2) we have $\Delta\ p_1 = \Delta v_1M_{11}$ so we can re-interpret the Leontief inverse by reading across rows. Row *j* of M_{11} displays the effects on prices triggered by a unitary exogenous change in sector *j* costs. This is a straightforward but seldom used interpretation of the Leontief multiplier matrix.

Starting from equation (1) of the basic linear model, SAM-based quantity models yield extensions to encompass a larger and more complete view of the income generating process. In the same way, we believe SAM-based price models departing from expression (2) may prove to be useful generalizations for evaluating the extensive cost linkages that pervade the relationships among households, factors, and producers.

To give content to this approach consider each one of these groups as undertaking an economic activity. Producers pay for raw materials (T_{11}) and factors (T_{21}) which are combined to generate output; factors make use of household endowments (T_{32}) to provide firms with labor and capital services. Finally, households purchase output (T_{13}) from production to obtain consumption.[iv] Additionally, each group is liable to pay taxes or import costs to the consolidated group 4. In terms of taxes, the government collects indirect production taxes from firms, taxes on the use of labor and capital from factors, and indirect consumption taxes and income taxes from households. Thus, each of these activities has an implicit cost or price index, which is linked to the rest of the price indices, through the coefficient sub matrices of the SAM. However, as it stands, price expression (2) omits these linkages and falls short of a satisfactory representation of interdependencies in the economy.

These links can be coherently integrated into a model by considering the three sets of accounts comprising producers, factors and households as endogenous and taking the consolidated account as exogenous. Using the column normalized expenditure coefficients and reading down the SAM columns for endogenous accounts yields:

$$p_1 = p_1A_{11} + p_2A_{21} + \pi_4A_{41}$$

$$p_2 = p_3A_{32} + \pi_4A_{42}$$

$$p_3 = p_1A_{13} + p_3A_{33} + \pi_4A_{43} \qquad (3)$$

Define a matrix A of normalized coefficients:

$$A = \begin{bmatrix} A_{11} & 0 & A_{13} \\ A_{21} & 0 & 0 \\ 0 & A_{32} & A_{33} \end{bmatrix}$$

Let $p = (p_1, p_2, p_3)$ be the vector of prices for the endogenous sectors of the SAM, and set the vector of exogenous costs (taxes, import costs) as $\nu = \pi_4A(4)$, where A(4) is the sub matrix of the SAM composed by column adjoining A_{41}, A_{42} and A_{43}. In matrix notation:

$$p = p\,A + \nu = \nu\,(I\text{-}A)^{-1} = \nu\,M \qquad (4)$$

where M is the multiplier matrix. For the same classification of endogenous and exogenous accounts, M is also the multiplier matrix of the endogenous income determination model:

$$Y = (I\text{-}A)^{-1}\,x = M\,x \qquad (5)$$

The interpretation of M is different, however, depending on whether we read its entries across the rows or down the columns. To clarify this distinction, M will be referred to as the (standard) multiplier matrix whereas its transpose M' will be termed the price-transmission matrix

3. DECOMPOSITION METHODS FOR PRICE TRANSMISSION ANALYSIS

In this section, we reformulate two of the leading methods of multiplier decomposition analysis for application to the study of price transmission.[v] As in the previous section, the approach is straightforward, but using it in the present context makes it possible to elucidate price linkages in economies such as Vietnam with extensive structural and institutional rigidities in a transparent and analytically useful manner.

3.1 Block-decomposition of Price Transmission

In a seminal series of contributions, Stone (1981), Pyatt and Round (1979), and Round (1985) showed how the multiplier matrix M can be decomposed into three economically meaningful additive (or multiplicative) components or sub-matrices. The first of these is a matrix that includes both the direct effects on the endogenous accounts of one-unit exogenous shocks (appearing as unit increases in the diagonal) and subsequent interaction effects among accounts within the same institutional group. The second component is a so-called open-loop matrix, which captures cross-effects between different institutional groups. These effects are transmitted from one category of endogenous institutions to other endogenous categories, and in turn, set in motion multiplier processes of within-category interaction effects, which amplify the initial stimulus. The third component is a closed-loop matrix, which detail the multiplier effects of an exogenous change on one institutional group, after it has traveled through the rest of endogenous accounts and returned to the original recipient. Thus, the closed-loop matrix captures the full circular multiplier effects *net* of own and open loop effects.

To decompose the price-transmission matrix M' take expression (4) and consider any matrix A*, which satisfy the standard algebraic requirements.[vi] It follows:

$$p = \nu M$$

$$= p\,A + p\,\tilde{A} - p\,\tilde{A} + \nu$$

$$= p\,(A - \tilde{A})(I - \tilde{A})^{-1} + \nu(I - \tilde{A})^{-1} = p\,A^* + \nu(I - \tilde{A})^{-1}$$

$$= [p\,A^* + \nu(I - \tilde{A})^{-1}]\,A^* + \nu(I - \tilde{A})^{-1}$$

$$= p\,A^{*2} + \nu(I - \tilde{A})^{-1}(I + A^*)$$

$$= [p\,A^* + \nu(I - \tilde{A})^{-1}]\,A^{*2} + \nu(I - \tilde{A})^{-1}(I + A^*)$$

$$= p\,A^{*3} + \nu(I - \tilde{A})^{-1}(I + A^* + A^{*2})$$

$$= \nu(I - \tilde{A})^{-1}(I + A^* + A^{*2})(I - A^{*3})^{-1}$$

$$= \nu M_1 M_2 M_3 \tag{6}$$

To obtain the corresponding decomposition M' = M_3'M_2'M_1', we extract from matrix A the blocks A_{11} and A_{33} and take:

$$\tilde{A} = \begin{bmatrix} A_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & A_{33} \end{bmatrix}$$

This yields:

$$M'_1 = \begin{bmatrix} (I\text{-}A_{11})^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & (I\text{-}A_{33})^{-1} \end{bmatrix}$$

$$M'_2 = \begin{bmatrix} I & A^*_{21} & A^*_{32}A^*_{21} \\ A^*_{13}A^*_{32} & I & A^*_{32} \\ A^*_{13} & A^*_{21}A^*_{13} & I \end{bmatrix} \tag{7}$$

$$M'_3 = \begin{bmatrix} (I - A^*_{13}A^*_{32}A^*_{21})^{-1} & 0 & 0 \\ 0 & (I - A^*_{21}A^*_{13}A^*_{32})^{-1} & 0 \\ 0 & 0 & (I - A^*_{32}A^*_{21}A^*_{13})^{-1} \end{bmatrix}$$

with $A^*_{13} = A_{13}(I - A_{33})^{-1}$, $A^*_{21} = A_{21}(I - A_{11})^{-1}$, $A^*_{32} = A_{32}$.

The first column of the multiplicative transfer matrix M'_1 shows how an exogenous cost increase affecting the production activities multiplies itself through the interindustry cost linkages (Leontief inverse) but exerts no effects on groups 2 and 3. In contrast, the first column of the open-loop matrix M'_2 indicates how the same exogenous cost increase ends up having an impact on factors (second entry, $A^*_{13}A^*_{32}$) after rebounding from households (third entry, A^*_{13}). Finally, the first column of the closed-loop matrix M'_3 captures the impact on production prices of the exogenous increase in producer costs after first affecting household cost indices (A^*_{13}), then moving onto factors ($A^*_{13}A^*_{32}$) and from these back to producers ($A^*_{13}A^*_{32}A^*_{21}$). The final figure shows the overall impact after this process has converged.

Any given element of the price-transmission matrix M' can be studied using either multiplicative or additive decomposition, which yield the same information in a different format. In a general disaggregate SAM, $n+m$ individualized sectors are detailed, n being taken as endogenous and m exogenous. Let I and J denote the indexing sets for the exogenous and endogenous accounts respectively. From (4) above, and $i \in I$, $j \in J$, the individual impact on price p_j of an exogenous cost change in sector i can be written as:

$$\frac{\partial p_j}{\partial v_i} = m_{ji} = 1 + n_{ji}^1 + n_{ji}^2 + n_{ji}^3 \tag{8}$$

where $m_{ji} \in M'$ and n_{ji}^1, n_{ji}^2, and n_{ji}^3 are elements of the additive component matrices:

$$N_1 = M_1$$

$$N_2 = (M_2 - I)\, M_1$$

$$N_3 = (M_3 - I)\, M_2 M_1$$

Note that $M = N_1 + N_2 + N_3$.

3.2 Path-decomposition of Price Transmission

The previous section shows that the SAM offers a convenient structure for detailed examination of price transmission. Prices can be computed and, furthermore, price changes can be decomposed according to three different categories of interdependence, which provide a detailed view of the extent system. To obtain a more comprehensive description of the effect of linkages on prices, however, we need to go one step further and analyze intersectoral linkages between individual accounts of the SAM by identifying the paths along which price/cost effects travel. The structural path analysis put forth by Defourny and Thorbecke (1984) showed the rich information structure that can be derived using this approach in a SAM framework. The use of path analysis to investigate cost-linkages is a natural extension and a promising way to enrich our understanding of the price formation mechanism.

Following the ideas in the contributions of these authors, we now use the concept of structural analysis in the SAM-based price model. Each pair *<i,j>* of indices in the SAM accounts is called an arc. A path is a sequence s of indices *s* = *<i,k,l,...,m,j>* which can be decomposed into consecutive arcs *<i,k>*, *<k,l>*,...,*<m,j>*. A path with non-repeating indices is termed an elementary path. A circuit of influence is a path *s* where the first and last indices coincide. The influence of account *i* on account *j* through a path *s* is represented by $(i \rightarrow j)_s$. To estimate the cost influence of account *i* on account *j* along *<i,j>*, notice from (4) above that, prior to any of the ensuing general equilibrium feedbacks, we have:

$$\frac{\partial p_j}{\partial p_i} = a_{ji} \tag{9}$$

Thus any exogenous price increase affecting p_i gives rise to a direct price increase in *j* measured by entry *(j,i)* of the transpose of the column normalized matrix A. Due to the linear structure of the model, the *direct* price influence along an elementary path *s*=*<i,k,...,m,j>* is the composite effect of the direct influences along the constituent arcs. So:

$$D^p_{(i \rightarrow j)s} = a_{ki} \ldots a_{jm} \tag{10}$$

In any given path *s* there may be feedback effects among its indices. Account *i* influences *k* but *k* in turn may influence *i*, either directly or through other intermediary indices. Accounts influence themselves through loops as well. All of these feedback effects taking place along circuits in the path work to amplify the magnitude of the direct influence being transmitted over the path. The expanded influence will be called *total* price influence, the ratio of total to direct price influence being the price path-multiplier:

$$T^p_{(i \rightarrow j)s} = D^p_{(i \rightarrow j)s} \, \mu^p_s \tag{11}$$

Notice, on the other hand, that more than one elementary path, each one with its respective feedback circuits, may span two indices *i,j*. Therefore the total price influence along a path does not capture the full or global price influence in the network of itineraries linking *i* and *j*. Let *S={s/i,j}* be the set of all elementary paths joining *i* and *j*. By additivity, the *global* price influence is defined as:

$$G^p_{(i\to j)s} = \sum_{s\in S} T^p_{(i\to j)s} = \sum_{s\in S} D^p_{(i\to j)s}\, \mu^p_s \qquad (12)$$

The last equality, where m_{ji} is the *(j,i)* entry in the price-transmission matrix M', follows from the fact that *S* includes all connecting paths between accounts *i* and *j*. Direct, total and global price influence are three distinct but related concepts of influence that supply precise information on the transmission mechanism underlying price formation.

4. ESTIMATES OF PRICE TRANSMISSION FOR VIETNAM

In this section, we apply our methodology to Vietnam, noting that there are at present significant disparities between domestic and external prices. For this reason, market based price transmission is likely to have extensive effects on future patterns of income and economic incentives, and policy makers need as already alluded to more detailed understanding of this kind of economic incidence to anticipate adjustment problems.

The SAM used here was developed in a separate research program. It is very detailed and was calibrated to data for 2000.[vii] Among other things, the SAM includes 97 activities/commodities, 14 factors of production (12 labor types, land, and capital), and 16 different household groups. Taken together, this table can elucidate very detailed incidence patterns, including a wide variety of producer, factor, and consumer price and cost index effects.[viii] To keep present discussion manageable, we illustrate our approach with a 30-sector, eight-factor, and eight-household aggregation of the Vietnam SAM.

The empirical interpretation of any given multiplier element m_{ji} in the matrix M' is quite straightforward if we take into account that benchmark prices are all calibrated to unity. Thus, m_{ji} gives us both the absolute and percentage variation of price *j* when the exogenous cost in sector *i* increases by one money unit, and the same considerations apply to any of the elements of a given decomposition. The multiplier matrix in itself is of some interest since it yields information on questions such as how a one Dong increase in taxes (production, consumption, and factor taxes, as well as tariffs) will raise producer prices, factor prices and individual cost-of-living indices. As such it provides useful information on the distortion effects of taxes but also, and equally important, about the welfare effects on individual consumers as measured by changes in their expenditures. Table 14-1 and Table 14-2 present a selection of the price decomposition results.[ix]

Table 14-1. [Block Decomposition Results for Price Transmission] all figures in percent

	Rice			*Oth. Processed Food*			*Vehicles*			*Manufactures (16)*			*Electricity, Gas, Water*		
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
	M	*N1+N2*	*%N3/M*	*M*	*N1+N2*	*%N3/M*	*M*	*N1+N2*	*%N3/M*	*M*	*N1+N2*	*%N3/M*	*M*	*N1+N2*	*%N3/M*
1 Rice	119	106	11	17	0	99	2	0	91	5	1	76	4	1	70
2 Rubber	8	0	100	11	0	100	2	1	59	3	1	75	4	2	43
3 Coffee Bean	10	0	99	13	0	99	1	0	87	3	1	80	3	1	72
4 Other Crops	14	0	100	19	0	100	2	0	97	4	1	88	4	1	85
5 Livestock	26	13	48	16	0	98	2	0	93	5	2	67	4	1	75
6 Other Agric	10	1	88	13	2	87	3	2	40	9	6	27	17	15	12
7 Forestry	17	0	100	22	0	100	2	0	89	6	2	71	4	0	89
8 Fishery	15	0	99	19	0	99	3	1	54	5	2	67	4	1	76
9 Coal	12	0	100	16	0	100	2	1	70	5	1	71	5	2	56
10 Oil and Gas	7	0	100	9	0	100	1	0	92	2	1	75	2	0	84
11 Other Minerals	9	1	87	12	2	84	2	1	53	5	3	44	10	9	17
12 Meat and Dairy	19	8	57	19	5	74	1	0	88	5	2	57	5	2	54
13 Bev. & Tobacco	10	2	82	13	3	79	1	0	80	6	4	37	4	2	46
14 Seafood	14	0	96	18	1	96	2	1	68	5	2	63	5	3	51
15 Oth Proc Food	74	62	16	119	104	13	2	0	87	5	1	68	5	2	57
16 Material Mfg	8	0	97	11	0	97	2	1	55	123	121	2	9	7	20
17 Chemicals	6	0	97	8	0	96	1	0	75	5	3	32	6	4	24
18 Technical Mfg	4	0	99	6	0	99	1	0	76	4	2	35	3	2	40
19 Vehicles	4	0	98	6	0	98	116	115	1	4	2	35	4	3	30
20 Oth Machinery	6	0	98	8	0	98	2	1	50	4	2	47	8	7	18
21 Text & Apparel	6	0	99	8	0	99	1	0	81	4	3	40	6	5	23
22 Other Industry	9	2	83	11	2	85	1	0	73	8	6	26	4	2	40
23 Elec, Gas, Water	5	0	99	6	0	99	1	0	64	2	1	66	112	111	1
24 Construction	8	0	97	10	0	97	2	1	48	28	26	7	6	4	29
25 Trade	11	0	96	14	0	96	2	0	74	4	2	64	4	2	57
26 Transport	6	0	99	8	0	99	5	4	15	4	2	41	3	2	47
27 Hotel & Rest	11	3	77	15	4	75	1	0	82	5	3	45	10	8	20
28 Pers Services	11	0	99	15	0	99	2	0	81	6	3	55	6	3	43
29 Commercial Srv	6	0	99	8	0	98	1	0	66	8	6	23	4	3	38
30 Public Services	12	0	98	16	0	97	2	0	79	8	5	41	6	4	43
31 Lab Rur Unsk	22	15	35	29	18	35	2	1	39	5	3	39	4	2	43
32 Lab Rur Mid Sk	22	14	35	28	18	35	2	2	38	5	3	39	4	2	42
33 Lab Rur Hi Sk	22	15	35	29	19	35	2	2	38	5	3	39	4	2	42
34 Lab Urb Unsk	15	9	42	20	12	40	2	1	35	5	3	34	5	3	31
35 Lab Urb Mid Sk	14	8	42	19	11	41	2	1	34	5	3	34	5	3	30
36 Lab Urb Hi Sk	14	8	43	19	11	41	2	2	33	5	3	33	5	3	30
37 Capital	1	1	37	1	1	37	0	0	38	0	0	37	0	0	37
38 Land	19	12	37	25	15	37	2	1	38	5	3	37	4	3	37
39 HH Rur Farm	23	15	35	29	19	36	2	1	41	5	3	39	4	2	44
40 HH Rur Self	21	13	35	27	18	35	3	2	34	5	3	40	4	3	39
41 HH Rur Wage	22	14	35	29	19	34	3	2	37	5	3	39	4	2	42
42 HH Rur Unemp	19	11	40	24	15	40	2	1	56	6	4	36	4	3	40
43 HH Urb Farm	19	12	36	26	16	36	2	1	42	5	3	38	4	3	37
44 HH Urb Self	15	9	41	20	12	40	2	1	36	5	3	34	5	3	29
45 HH Urb Wage	14	8	44	18	11	42	2	2	32	5	3	33	4	3	31
46 HH Urb Unemp	12	7	43	17	10	42	1	0	58	4	3	36	5	4	25
47 E01State	1	1	36	1	1	36	0	0	40	0	0	38	0	0	40
48 E02PrivDom	2	1	38	2	2	38	0	0	36	1	0	37	0	0	35
49 E03PrivFor	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Averages															
Producers	16	7	87	16	4	92	6	4	69	10	7	51	9	7	45
Labor	18	11	39	24	15	38	2	1	36	5	3	37	4	3	37
Capital and Land	10	6	37	13	8	37	1	1	38	3	2	37	2	1	37
Households	18	11	39	24	15	38	2	1	42	5	3	37	4	3	36
Total	15	8	67	17	7	70	4	3	57	7	5	45	7	5	41

Table 14-2. [Path Linkages from Products to Producers]

Path [x]	Global Effect	Total Effect	% of Global	Cum %
A15OtPrFd<-A01Rice	0.737	0.727	98.7	98.7
A15OtPrFd<-L01RU<-HH01RF<-A01Rice		0.001	0.2	98.8
A06OthAg<-A23ElGsWat	0.166	0.139	84.0	84.0
A06OthAg<-A16MatMfg<-A23ElGsWat		0.002	1.4	85.4
A06OthAg<-A20OthMach<-A23ElGsWat		0.001	0.7	86.1
A06OthAg<-L01RU<-HH01RF<-A23ElGsWat		0.001	0.7	86.8
A16MatMfg<-A23ElGsWat	0.086	0.059	68.0	68.0
A16MatMfg<-A11OthMin<-A23ElGsWat		0.006	7.2	75.2
A16MatMfg<-A25Trade<-A23ElGsWat		0.001	1.3	76.5

4.1 Block-decomposition of Price Multipliers

Table 14-1 focuses on the additive block-decomposition of the price multiplier matrix, which reveals the extent to which price effects can be traced from commodity prices across the economy, grouping effects by generic structural components (production, factors, and households). The examples show the decomposition of a unitary increase in the exogenous (column) commodity or production cost and the induced effects on downstream (row) producer, factor, and consumer price indexes.

We look at five sample sectors, Rice, Other Processed Food, Vehicles, Manufactures, and Utilities. For each sector, three block components are represented: total multiplier effect (M), effect due to direct (N_1) and open loop (N_2) linkages,[xi] and percent of closed loop effects in the total. The last item indicates something like a general equilibrium transmission effect, incorporating the longer chains of price linkage that might not be discernable to casual observers.

The most notable features of these results are three. Firstly, indirect price transmission is a very important component of the price adjustment process. Open loop effects average between 41 and 72% of total price effects, depending on the sector considered. This fact is hardly surprising, since direct linkages are relatively weak in this economy, but it certainly highlights the challenge facing policy makers who try to anticipate induced price and cost incidence of taxes, terms of trade shifts, and effects of exchange rate policy. Relying on intuition or rules of thumb alone in this context is very unlikely to achieve something approaching optimality.

The results for subsistence and processed foods are variegated. Outside the food sector, factor and consumer price indices, the transmission of food cost inflation is mostly indirect. Other sectors see producer prices rising only in response to induced factor cost indices and some longer chain intermediate linkages. For factors and households, the effects are of course more direct but, importantly, the implied inflation impact is relatively high. Households experience an (unweighted) average CPI increase equal to about 18% of any increase in rice prices, with the exact effect depending on the location and factor content of the household. The result for processed food is even more dramatic, stepping up household consumption costs by 24%. These results highlight the importance of food price stability to the economy. It is also noteworthy that about 40% of the inflation transmission from food prices to households is indirect, arising from open loop linkages.

A second feature of these results is the relatively low levels of inflation transmission in essential sectors associated with economic modernization: vehicles, manufactures, and utilities. In all three categories, the average impact of inflation is less than 7%. This result is due primarily to Vietnam's early stage of development, still primarily agrarian and highly subsistence oriented. For this reason, the proportions of domestic cost and expenditure allocated to manufactures and utilities are still relatively small, especially in the context of economy-wide household consumption.

4.2 Path Decomposition of Price Multipliers

The price-transmission matrix M' gives a tableau presentation of composite effects across the economy, but the true incidence of price transmission can only be ascertained by decomposing the detailed paths of influence that arise from expenditure linkages. This kind of detailed structural analysis is exactly the purpose of path decomposition, and we apply this analysis to the Vietnamese economy with a few examples from the 2000 SAM.

Tables 14-2, 14-3 and 14-4 show detailed chains of price transmission from products to producers, producers to households, and factors to households. In each case, we examine constituent chains of up to five institutions, contributing up to .001 of the total transmission effect m_{ji} in the price transmission matrix M'. Each table lists the global effect (m_{ji}), the component (total) effect along the constituent path specified, and the marginal and cumulative percentages of the total effect, respectively.

Table 14-3. [Path Linkages from Producers to Households]

Path	Global Effect	Total Effect	% of Global	Cum %
HH07UW<-A01Rice	0.136	0.007	5.2	5.2
HH07UW<-A05LivStk<-A01Rice		0.008	6.1	11.2
HH07UW<-A15OtPrFd<-A01Rice		0.076	55.7	66.9
HH07UW<-A12MtDary<-A05LivStk<-A01Rice		0.001	0.7	67.6
HH07UW<-A13BevTob<-A15OtPrFd<-A01Rice		0.001	0.6	68.2
HH07UW<-A27HotRest<-A15OtPrFd<-A01Rice		0.001	1.0	69.2
HH07UW<-A04OthCrp<-L01RU<-HH01RF<-A15OtPrFd<-A01Rice		0.001	0.7	69.9
HH07UW<-A05LivStk<-L01RU<-HH01RF<-A15OtPrFd<-A01Rice		0.001	0.6	70.5
HH07UW<-A30PubServ<-L01RU<-HH01RF<-A15OtPrFd<-A01Rice		0.002	1.7	72.2
HH07UW<-A15OtPrFd	0.184	0.125	68.1	68.1
HH07UW<-A12MtDary<-A15OtPrFd		0.001	0.5	68.5
HH07UW<-A13BevTob<-A15OtPrFd		0.001	0.7	69.3
HH07UW<-A27HotRest<-A15OtPrFd		0.002	1.2	70.5
HH07UW<-A04OthCrp<-L01RU<-HH01RF<-A15OtPrFd		0.002	0.8	71.3
HH07UW<-A05LivStk<-L01RU<-HH01RF<-A15OtPrFd		0.001	0.8	72.1
HH07UW<-A05LivStk<-L04UU<-HH06US<-A15OtPrFd		0.001	0.4	72.5
HH07UW<-A08Fishry<-L01RU<-HH01RF<-A15OtPrFd		0.001	0.8	73.3
HH07UW<-A13BevTob<-L01RU<-HH01RF<-A15OtPrFd		0.001	0.5	73.8
HH07UW<-A25Trade<-L01RU<-HH01RF<-A15OtPrFd		0.001	0.7	74.5
HH07UW<-A27HotRest<-L01RU<-HH01RF<-A15OtPrFd		0.001	0.5	75.0
HH07UW<-A29ComServ<-L01RU<-HH01RF<-A15OtPrFd		0.001	0.6	75.5
HH07UW<-A30PubServ<-L01RU<-HH01RF<-A15OtPrFd		0.004	2.2	77.7
HH07UW<-A30PubServ<-L01RU<-HH02RS<-A15OtPrFd		0.001	0.5	78.2

Table 14-4. [Path Linkages from Factors to Households]

Path	Global Effect	Total Effect	% of Global	Cum%
HH01RF<-A01Rice<-Capital	0.191	0.001	0.5	0.5
HH01RF<-A04OthCrp<-Capital		0.002	1.3	1.8
HH01RF<-A05LivStk<-Capital		0.003	1.6	3.4
HH01RF<-A08Fishry<-Capital		0.003	1.8	5.2
HH01RF<-A13BevTob<-Capita		0.015	7.7	12.9
HH01RF<-A15OtPrFd<-Capital		0.012	6.2	19.1
HH01RF<-A16MatMfg<-Capital		0.003	1.4	20.5
HH01RF<-A17Chemcl<-Capital		0.001	0.5	21.0
HH01RF<-A18TechMfg<-Capital		0.001	0.4	21.4
HH01RF<-A21TxtAprl<-Capital		0.003	1.6	23.0
HH01RF<-A22OthInd<-Capital		0.003	1.4	24.4
HH01RF<-A23ElGsWat<-Capital		0.003	1.5	25.8
HH01RF<-A25Trade<-Capital		0.009	5.0	30.8
HH01RF<-A26Transp<-Capital		0.005	2.5	33.4
HH01RF<-A27HotRest<-Capital		0.003	1.4	34.7
HH01RF<-A28PerServ<-Capital		0.004	2.2	37.0
HH01RF<-A29ComServ<-Capital		0.018	9.5	46.4
HH01RF<-A30PubServ<-Capital		0.005	2.4	48.8
HH01RF<-A05LivStk<-A22OthInd<-Capital		0.001	0.5	49.3
HH01RF<-A05LivStk<-A25Trade<-Capital		0.001	0.5	49.8
HH01RF<-A15OtPrFd<-A01Rice<-Capital		0.003	1.8	51.6
HH01RF<-A15OtPrFd<-A23ElGsWat<-Capital		0.001	0.4	52.0
HH01RF<-A15OtPrFd<-A25Trade<-Capital		0.006	3.2	55.2
HH01RF<-A25Trade<-A29ComServ<-Capital		0.001	0.4	55.6
HH01RF<-A30PubServ<-A23ElGsWat<-Capital		0.001	0.6	56.2
HH01RF<-A30PubServ<-A29ComServ<-Capital		0.002	1.1	57.3
HH01RF<-A15OtPrFd<-A01Rice<-A06OthAg<-Capital		0.001	0.4	57.7
HH01RF<-A15OtPrFd<-A01Rice<-A25Trade<-Capital		0.001	0.5	58.2
HH01RF<-A01Rice<-Land	0.060	0.006	10.5	10.5
HH01RF<-A04OthCrp<-Land		0.012	20.7	31.2
HH01RF<-A05LivStk<-Land		0.002	3.1	34.3
HH01RF<-A05LivStk<-A01Rice<-Land		0.002	3.4	37.8
HH01RF<-A05LivStk<-A04OthCrp<-Land		0.001	1.7	39.5
HH01RF<-A13BevTob<-A04OthCrp<-Land		0.001	1.4	40.8
HH01RF<-A15OtPrFd<-A01Rice<-Land		0.023	38.8	79.6
HH01RF<-A15OtPrFd<-A04OthCrp<-Land		0.002	3.2	82.8

Table 14-2 focuses on links between producer prices. Generally speaking, the current stage of Vietnam's economic development is characterized by fairly low levels of sectoral articulation, and thus producer to producer price transmission is fairly direct. To illustrate this, we give examples of the effects of rising Rice prices on Other Food Products and the effects of rising utility costs on Other Agriculture and heavy Manufacturing (MatMfg). In the first case, rice price increases are about 74% passed through to Other Processed Food. The vast majority of this effect comes directly from intermediate use of unprocessed rice, and represents the huge rice polishing mill industry. A very small contribution also arises from links through consumption (Rural Households) and factor payments (Rural Unskilled Labor).

Other Agriculture, which includes most irrigated specialty crops, is the most utility-intensive agricultural sector in Vietnam at present. For this reason, price increases in utilities are passed on directly (84% of global effects), but the impact is modest since global effects amount to only .166. About 3% more of the global effect is transmitted though intermediate goods (Manufactures and Machinery) and consumption linkages. In the case of Manufacturing, indirect linkages are somewhat more important because this relatively modern sector is better articulated to other sectors. Despite this fact, inflation transmission from utilities is very subdued, with a global effect of only .086. About three fourths of this effect is accounted for by direct links and trilateral links via two other upstream sectors (minerals and trade).

Table 14-3 gives examples of the impact of producer price changes on Vietnamese households. In this case, we focus on Urban Wage earning households, a group most sensitive to the consumption effects of changes in market prices. For the sake of comparison, we examine changes in two essential commodity prices, Rice and Processed Food. Because they are removed from the subsistence sector, this consumer group is more likely to feel the brunt of changing market prices and to be more import dependent. This makes it especially important for policy makers to understand and anticipate the incidence of price transmission to this group. Despite the very direct linkage between producer prices for these two goods (Table 14-2), we see more complex patterns emerging in consumer price index effects.

In light of the importance of this group to the economy in terms of growth and modernization, it is perhaps fortunate that the inflation impact for essential and basic foodstuffs is relatively low, .136 and .184 for Rice and Processed Food, respectively. Not only are these relatively modest, but they are propagated through fairly long price transmission chains. In the case of Rice, the main effect of course results after the rice is polished, and 55% of the global effect comes through the Other Food Processing sector.

Beyond this, the other prominent transmission paths include feed (Livestock), beverages, restaurants, and a variety of consumption linkages.

Other processed food hits the CPI of urban wage earners even more directly, with 68.1% of the global effect, followed by upstream links to other consumables (Meat and Dairy, Beverages and Tobacco, and Restaurants). Finally, an extended series of consumption linkages accounts for about 10% of the global effect. These generally extend from consumption of Other Food Processing goods by most household groups, and then to factor price inflation in urban consumption goods.

Final examples of path decomposition examine the composition of pass though effects from factor markets to households. In this case, we examine how the CPI can be affected by rising prices for nonhuman factors of production, capital and land. In an economy, undertaking a very dynamic transition from low-income levels, both these cost components can exert significant but complex influences on real living standards.

Table 14-4 indicates how the cost components affect the consumption weighted purchasing power of another essential constituency, the majority Vietnamese population of rural farmers. Capital price increases have a rather significant effect on these households, especially considering their relatively weak links to the market economy. A unitary increase in capital costs raises the market CPI of Rural Farmers by .20, a magnitude that should arouse concern in the minds of policy makers. This is particularly the case for Vietnam, which is at the earliest stages of agricultural mechanization and technology induced productivity growth.

Just as interesting, however, is the complexity of the transmission process. A total of 28 paths must be delineated to identify only 58% of the global effect, and individual path contributions are much more uniform than in previous examples. Having said this, it should be emphasized that these effects are coming mainly through the household consumption function, and assume a unitary increase in some abstract composite capital good. In reality, prices of different types of capital may change at different rates and even in different directions, changing the composition of these effects. Our estimates do indicate, however, that any significant shift in the aggregate factor terms of trade could have important consequences for rural households.

The case of land is somewhat simpler, and considerably more moderate in terms of inflation transmission. Land prices have substantial influence on Rural Households only by their effect on producer prices of agricultural goods and downstream processed foods. In total, the global effect of the Rural Farm CPI with respect to land cost is only .06, indicating that this factor is still relatively minor in its contribution to living costs.

5. CONCLUSION

In this paper we presented a novel way of examining the detailed structure of economy wide price transmission, which was applied to an interesting emerging Asian economy. Building on the seminal work of other contributors to the social accounting matrix literature, we used a SAM framework to elucidate the mechanism of price formation in the presence of endogenous factor prices and household cost-of-living adjustments. Despite its limitations compared to general equilibrium models with endogenous activity levels, the SAM-based price model has some distinct advantages, including transparency and the ability to estimate absolute price variations, providing information of immediate use to policy makers. Furthermore, price variations can be decomposed to reveal the underlying patterns of economic interdependence and price transmission. By partitioning the SAM accounts into blocks, and adapting decomposition techniques developed for SAMs, the analyst can distinguish the extent of price effects explained by interindustry linkages, the consumption expenditures of households, and factor prices.

The linear structure of the SAM price model also allows us to break down the price-transmission matrix with structural path decomposition techniques. Path analysis discloses in detail the network of path transmission paths and produces direct estimates of all the linkages connecting two SAM accounts. The information indicates what sectors are more cost responsive to changes taking place elsewhere in the economy. Not all sectors are equally responsive in magnitude and scope, and by identifying them in detail and understanding where the burdens of price distortions occurs more informed policies, such as changes in tax rates, could be designed to minimize undesirable welfare distortions.

In sum, the use of a price model of the kind presented in this paper is especially relevant to economies in transition, where price formation is heavily affected by rigidities in factor and other prices. In Vietnam, extensive state enterprise participation and systems of national, regional, and local price administration inhibit price transmission in ways that lead to significant distortions. Moreover, the key advantage of the methodology put forward is that it can help policy makers measure the adjustment burdens that such distortions imply and elucidate the detailed paths that eventual adjustments will take, helping to more clearly identify affected groups and implement mitigating policies.

NOTES

[*] David Roland-Holst is the James Irvine Professor of Economics at Mills College and Director of the Rural Development Research Consortium (RDRC) at the University of California, Berkeley (dwrh@rdrc.net). Finn Tarp is Professor of Development Economics and Coordinator of the Development Economics Research Group (DERG) at the University of Copenhagen (Finn.Tarp@econ.ku.dk). Opinions expressed here are those of the authors and should not be attributed to their affiliated institutions.

[i] SAMs have for example been used to study: (i) growth strategies in developing economies, Pyatt and Round (1985); (ii) income distribution, Pyatt and Alan Roe (1977), and Irma Adelman and Sherman Robinson (1978), and redistribution, David W. Roland-Holst and Ferran Sancho (1992); (iii) fiscal policy in national or regional settings, John Whalley and France St. Hilaire (1983, 1987); and (iv) decomposition of activity multipliers that shed light on the circuits comprising the circular flow of income, Richard Stone (1981), Pyatt and Round (1979), Defourny and Thorbecke (1984), and Robinson and Roland-Holst (1988).

[ii] Note that no flows are associated with cells (1,2), (2,2), (2,3) and (3,1).

[iii] The notion of price should be taken in the same broad sense as the notion of income of a sector or institution has in a SAM framework.

[iv] Transfers among households T_{33} can be thought of as distribution costs linked to consumption.

[v] Reference is made to respectively block (Pyatt and Round 1979; Round 1985; and Stone 1981), and path decomposition (Defourny and Thorbecke 1984).

[vi] $\tilde{A}$ must be conformal to A and $(I - \tilde{A})$ must be invertible.

[vii] See Finn Tarp, David W. Roland-Holst, John Rand and Henning Tarp Jensen (2002).

[viii] All the accounts are listed in the Annex.

[ix] The complete empirical results are available from the authors.

[x] Sector labels are defined in the Annex.

[xi] Note that nonzero elements in the matrices N_1 and N_2 do not overlap, so they can be added without loss of structural information.

ANNEX

Table 14-A1: [A MacroSAM with Four Accounts]

		I	II	III	IV	V
I.	Production	T_{11}	O	T_{13}	T_{14}	Y_1
II.	Factors	T_{21}	O	O	T_{24}	Y_2
III.	Households	O	T_{32}	T_{33}	T_{34}	Y_3
IV.	Rest	T_{41}	T_{42}	T_{43}	T_{44}	Y_4
V.	Total	Y_1	Y_2	Y_3	Y_4	

Table 14-A2: [Endogenous Institutions from the 2000 Vietnam SAM]

No.	SAM Label	Definition
1	A01Rice	Rice
2	A02Rubber	Rubber
3	A03Coffee	Coffee Bean
4	A04OthCrp	Other Crops
5	A05LivStk	Livestock
6	A06OthAg	Other Agric
7	A07Forest	Forestry
8	A08Fishry	Fishery
9	A09Coal	Coal
10	A10OilGas	Oil and Gas
11	A11OthMin	Other Minerals
12	A12MtDary	Meat and Dairy
13	A13BevTob	Bev. & Tobacco
14	A14Seafood	Seafood
15	A15OtPrFd	Oth Proc Food
16	A16MatMfg	Material Mfg
17	A17Chemcl	Chemicals
18	A18TechMfg	Technical Mfg
19	A19Vehicle	Vehicles
20	A20OthMach	Oth Machinery
21	A21TxtAprl	Text & Apparel
22	A22OthInd	Other Industry
23	A23ElGsWat	Elec, Gas, Water
24	A24Cnstrct	Construction
25	A25Trade	Trade
26	A26Transp	Transport
27	A27HotRest	Hotel & Rest
28	A28PerServ	Pers Services
29	A29ComServ	Commercial Srv
30	A30PubServ	Public Services
31	L01RU	Lab Rur Unsk
32	L02RM	Lab Rur Mid Sk
33	L03RH	Lab Rur Hi Sk
34	L04UU	Lab Urb Unsk
35	L05UM	Lab Urb Mid Sk
36	L06UH	Lab Urb Hi Sk
37	Capital	Capital
38	Land	Land
39	HH01RF	HH Rur Farm
40	HH02RS	HH Rur Self
41	HH03RW	HH Rur Wage
42	HH04RN	HH Rur Unemp
43	HH05UF	HH Urb Farm
44	HH06US	HH Urb Self
45	HH07UW	HH Urb Wage
46	HH08UN	HH Urb Unemp
47	E01State	E01State
48	E02PrivDom	E02PrivDom
49	E03PrivFor	E03PrivFor

REFERENCES

Adelman, Irma and Sherman Robinson. 1978. *Income Distribution Policy in Developing Countries: A Case Study of Korea.* Stanford, CA: Stanford University Press.

Defourny, Jacque and Erik Thorbecke. 1984. "Structural Path Analysis and Multiplier Decomposition within a Social Accounting Matrix." *Econ. J.*, 94:374, pp. 111-36.

Pyatt, Graham and Alan Roe. 1977. *Social Accounting for Development Planning with special reference to Sri Lanka.* Cambridge: Cambridge University Press.

Pyatt, Graham and Jeffery I. Round. 1979. "Accounting and Fixed Price Multipliers in a Social Accounting Matrix". *Econ. J.*, 89:356, pp. 850-73.

Pyatt, Graham and Jeffery I. Round. 1985. *Social Accounting Matrices: A Basis for Planning.* Washington, DC: The World Bank.

Robinson, Sherman and David W. Roland-Holst. 1988. "Macroeconomic Structure and Computable General Equilibrium Models." *J. Pol. Modeling*, 10:3, pp. 353-75.

Roland-Holst, David W. and Ferran Sancho. 1992. "Relative Income Determination in the United States: A Social Accounting Perspective." *Rev. Income Wealth*, 38:3, pp. 311-27.

Round, Jeffery I. 1985. "Decomposing Multipliers for Economic Systems Involving Regional and World Trade." *Econ. J.*, 95:378, pp. 383-99.

Stone, Richard. 1981. *Aspects of Economic and Social Modelling.* Geneva: Librairie Droz.

Tarp, Finn, David W. Roland-Holst, John Rand and Henning Tarp Jensen. 2002. A 2000 Social Accounting Matrix for Vietnam. Hanoi: CIEM, http://www.econ.ku.dk/rand/.

Whalley, John and France St. Hilaire. 1983. "A Microconsistent Equilibrium Data Set for Canada for Use in Tax Policy Analysis." *Rev. Income Wealth*, 29:2, pp. 175-204.

Whalley, John and France St. Hilaire. 1987. "A Microconsistent Data Set for Canada for Use in Regional General Equilibrium Policy Analysis." *Rev. Income Wealth*, 33:3, pp. 327-43.

Chapter 15

INSTITUTIONS, FACTOR ENDOWMENT AND INEQUALITY IN GHANA, KENYA AND SENEGAL

Christian Morrisson
Université de Paris I and Delta, morrisso@canoe.ens.fr

1. INSTITUTIONS AND INEQUALITY

Two seminal articles on this topic have been published recently by Engerman and Sokoloff (2000a, 2000b). The same ideas, but with different perspectives, are expressed by Acemoglu et al. (2001) and Easterly (2002).

Engerman and Sokoloff have shown the interactions in an endogenous process, assuming that:

- factor endowments explain for a large part income, human capital and political power inequalities
- these inequalities explain the structure and functioning of institutions which insure the persistence of inequalities
- such an institutional framework has a negative impact over time on growth.

The same line of argument proves that moderate inequalities result from different factor endowments. These inequalities lead to creating other institutions which offer opportunities to everybody. As all groups can realize their economic potential, inequalities remain moderate or decrease. Such an evolution of income distribution stimulates growth. Engerman and Sokoloff don't consider only income inequality, but also wealth, human capital, political power inequalities and they analyze the interactions between institutions and these inequalities.

Engerman and Sokoloff chose to compare two factor endowments in Latin America, one promoting inequality, the other equality. The first endowment included climate, soils, mineral resources well suited for growing sugar or extracting silver, with substantial economies of scale. The second endowment allowed grain agriculture without economies of scale (until the end of the 19th century and the introduction of mechanized agriculture). These different factor endowments explain extreme inequality in Brazil (sugar) and Mexico (silver) instead of moderate inequality in American colonies or USA after Independence (grain). Engerman and Sokoloff explain why high inequality has lasted several centuries. The rich elite are able to establish a legal framework (laws, political institutions) and to enforce land policies which contribute to persistence of inequality over time. The interactions between inequalities and institutions represent the core of their analysis. These inequalities are combined with ethnic heterogeneity. In Brazil and Mexico, the large estates (or mines) owners are always European migrants, whereas workers in plantations and mines are slaves from Africa or natives; the share of whites being below 25/30% instead of more than 80% in USA. If the South is excluded, the population is composed only of whites who have human capital whereas slaves and natives in Latin America are illiterate.

Since the end of the 19th century, several institutions changed in Latin America. Slavery disappeared as did the "encomiendas" institution. Today these countries are democracies and in Brazil the leader of the workers' party has been elected president. But large inequalities of income and human capital remain (in Brazil 55% of children of the households ranked in the bottom 40% don't reach the 5th year in primary school). Such inequalities are related more to past institutions than to present institutions.

Easterly, following the hypothesis of Engerman and Sokoloff, has confirmed with cross-country data the relation between institutions and inequality. Easterly assumes three hypotheses:

- income inequality depends on factor endowments, of climate and location, which were favourable or not to European settlers (which is also the hypothesis of Engerman and Sokoloff).
- institutions depend on income inequality, on geography (climate and location) and other variables. (Easterly added other variables to the equation of Engerman and Sokoloff).
- per capita income in 1990 depends on institutions, schooling (average years of schooling in the adult population) and openness (Easterly considers the impact of institutions on growth, a subject which is not analyzed in detail by Engerman and Sokoloff).

Easterly assumes an incidence of income inequality on institutions, but not the reverse. He takes into account income inequality only, whereas Engerman and Sokoloff consider also human capital and political power inequalities. But as Easterly assumes that growth depends on human capital, inequality of schooling is also included in his model. Vinod Thomas *et al.* (2000) have shown that the number of schooling years is negatively related to inequality of human capital. The hypotheses of Easterly are tested on a sample of 53 countries:

- the income shares of quintiles 2, 3 and 4 depend on land suitability for some crops, on mines and on settlers' mortality.
- the institutions depend on the shares of middle quintiles and settlers' mortality.
- GDP per capita in 1990 depends on institutions, openness and schooling.

The article of Acemoglu *et al.* analyzes also institutions, factor endowments and geography. But as income inequality is never taken into account, it is not a pertinent reference for our analysis.

We analyze only the relations between factor endowments, institutions and inequality, excluding growth. Following Engerman and Sokoloff's approach, we have chosen three African countries, with very different income distribution, in order to analyze the interactions between institutions, factor endowments and inequality since one century.

We want to check whether in Africa, as in America,

- inequality depends on factor endowment
- institutions depend on income distribution and factor endowment
- institutions have an impact on income distribution, according to Engerman and Sokoloff thesis.

The history of Africa and that of Latin America are very different. In some African countries, inequality is moderate, which is very scarce in Latin America. Most African countries have been independent for 40 years instead of almost two centuries. In African countries, the richest people are blacks like the poor, whereas the elite in Latin America are white.

The three African countries I propose to consider, Ghana, Kenya and Senegal, have nearly the same GDP per capita, expressed in Geary-Khamis 1990 US dollars (Angus Maddison, 2001), varying between 1100 and 1350 dollars (cf. Table 15-1). The GDP per capita has decreased in Ghana and Senegal since 1970, whereas the GDP per capita in Kenya, the lowest in 1970, is increasing. This convergence entails a reduction of the gap between the lowest and the highest value from 1.57/1 in 1970 to 1.25/1 in 1990. So in 1990, Ghana and Kenya have the same GDP per capita, 1100 dollars, whereas Senegal reaches 1350 dollars. In the three countries, more than 2/3 of population works in agriculture (including fishing in Senegal) and a large part of exports are agricultural products. After colonisation since the 19th century, independence was granted to the three countries between 1957 and 1963.

But the three income distributions are very different. The Gini coefficients in Kenya and Senegal are 50% higher than in Ghana (cf. Table

15-1). If we consider the variables which explain usually distribution, like GDP per capita (and its square), school enrollment, distribution of land, mineral resources... (François Bourguignon and Christian Morrisson, 1998), it is clear that they cannot explain such differences between Ghana and the other countries:

- the GDP per capita is nearly the same
- school enrollment in Ghana and Kenya is the same
- mineral resources are unimportant in Kenya and play a minor role in Ghana and Senegal.
- the land distribution in Ghana and Senegal is rather similar.

As inequality differences appear as a paradox, we have assumed that institutions in the past and their evolution since independence could explain partially income distribution.

We have chosen the following institutions and factor endowments as variables:

- the weight of public sector in GDP and the wages level and scale in this sector,
- the labor market legislation,
- the tariffs,
- the respect of property rights of the foreigners by the state,
- the climate and the environment favourable or not to European settlers,
- the land suitability for some crops and the availability of mineral resources.

Table 15-2 presents these variables in the three countries between circa 1960 and 1990 and eventually the changes since this date. We observe that each country has a specific institutional framework

2. INSTITUTIONS, FACTOR ENDOWMENT AND INCOME DISTRIBUTION IN SENEGAL

The situation of Senegal in 1960 is exceptional because at that time Dakar is not the capital of Senegal but that of the whole Afrique Occidentale Française (AOF). The GDP of Côte-d'Ivoire is nearly the same as Senegal's GDP (Maddison 2001). But according to the French tradition, all the administration of AOF is concentrated in Dakar. In 1958, 67% of current expenditures of AOF are spent in Senegal (16% in Côte-d'Ivoire). Two figures prove the concentration of high wages in Dakar. In 1957, one third of all professionals or managers in AOF administration work in Dakar (17% in Abidjan). Nearly 50% of foreign wage-earners (who receive the highest wages in AOF) work in Senegal.

Important public expenditures in Dakar entail a relatively important demand of consumer goods by households and, as a consequence, a private modern sector (manufacturing and services) has expanded. In 1957, 40% of employment in manufacturing and modern services in AOF are concentrated in Dakar. For all French colonies in AOF and AEF (Afrique Equatoriale Française), there is only one university, located in Dakar, and a third of secondary school students in AOF are in Dakar.

The independence of Senegal has excluded nearly 90% of the population which had relations with Dakar, a situation we can compare with that of Vienna, former capital of the Austro-Hungarian Empire, and after 1918 of Austria only. Despite the fact that French civil servants came back to France, Senegal inherited an overstaffed administration and a too large public sector which the private sector could not support. The weight of civil servants wages was in great excess because wages were too high and staffing too large. The wages resulted from a French wage scale which granted to civil servants higher wages (except for managers) than those received in the private sector. On the other hand, the French policy granted the same wage to Senegalese and French civil servants (except a bonus compensating expatriation for the latter). As a result, the wages of Senegalese civil servants were much higher than wages in modern enterprises. If we consider only the modern sector, civil servants (20% of wage-earners) received 50% of wages. In 1959, the average wage of civil servants was 10 times the average income of active population and 17 times the average income in agriculture (Table 15-5).

The most important point is that this imbalance lasted until 1990, 30 years after independence and the "départ" of French civil servants. In 1970, the civil servants (nearly 3% of active population) received more than 16% of total income and their average income amounted to 11 times the average income in agriculture (Table 15-5). As Jean-Claude Berthélemy et al. (1996) underline, the level of public wages was incompatible with state resources and average income in Senegal.

The legislation (the 1961 Code du Travail) concerning the labor market in the modern sector was also a factor of inequality. A state agency had the monopoly in hiring, until 1987. Direct hiring by the enterprise or private agency was forbidden. Each employment contract was, after two years, permanent. An enterprise could not lay off any employee without the agreement of the state (Ministère du Travail). Since 1981 the increases of the minimum or basic salary for each professsion were decided upon by a threefold commission including representatives of unions, employers and State administration. In fact basic wages, except for highly skilled workers, are determined by administration. Supplements and bonuses were either decreed by the government or negotiated between unions and employers, or offered by employers (Katherine Terrell and Jan Svejnar, 1989).

Under such regulations, the average wage in the modern sector is higher than the market level. The average wage of unskilled workers was twice in 1970 that of these workers in informal activities. In 1970, wage-earners in the modern sector (private) received 30% of income whereas they were less than 10% of active population. Their average wage was 6.5 times the average income in agriculture. Such a dualism of labor market combined with an overstaffed and overpaid administration entailed the following result: wage-earners of the modern sector (private and public), which amounted to less than 13 % of the active population, received 50% of total national income.

Contrary to Ghana, Senegal always respected the property rights of foreigners. In 1960, more than 80% of modern enterprises were owned by foreigners, in 1970, 60%. The percentage decreases to 50% in 1984 (Table 15-3), and afterward the majority of enterprises are owned by Senegaleses. So the number of foreign capitalists remained high during 25/30 years. As these persons receive very high incomes (four times the average wage of civil servants in 1970), their presence contributed to inequality.

We must also take into account Senegalese employers (plus independent professions such as medical doctors, lawyers...) of the modern sector. In 1970, the sum of the incomes of all employers (Senegalese and foreigners) amounted to 9% of household's income, whereas employers were only 1.4% of active population.

These employers share with wage-earners of the modern sector the benefits of high tariff barriers. Moreover, the import of some goods like sugar was forbidden during several years. Such prohibition allowed very high profits for businessmen producing these goods. After 1986, when government undertook liberalisation of foreign trade, a virulent opposition came from these entrepreneurs. Liberalisation was implemented only in 1990 with the structural adjustment programme. The decrease of tariffs since the 90's improved the living standard of farmers and workers in the informal sector. But as this programme included privatisations, the businessmen who accumulated a fortune due to protection or prohibition, could buy public enterprises at low prices and maintain a dominant position in Senegal (and a high share of total income).

The groundnut production belongs to the traditional agriculture and land distribution is not unequal (Table 15-4). But its transformation into oil and other products, the export of these products, are modern activities which provided important profits to private enterprises before and after independence. From 1968, a parastatal company (ONCAD) supplied inputs to farmers and bought groundnuts. The expansion of ONCAD entailed overstaffing and high wages at the expense of farmers. The producer price was constant whereas the margin of ONCAD increased continuously from 1973 to 1988. This margin amounts to a tax on farmers. Moreover, as ONCAD paid farmers with long delays, farmers sold groundnuts to private traders at a lower price.

The government established a low price for rice (lower than import price) which is consumed more and more by urban population, instead of stimulating national production by raising producer price. Berthélemy *et al.* (1996) estimated that this price policy (taxing farm income) less subsidies to farmers, amounts to a transfer of 100 billions F. CFA from farmers to the state for the 1967-1985 period. Such net tax increased the gap between average income in the modern sector and in agriculture, and increased inequality in Senegal.

At this point, we can answer the three following questions regarding Senegal.

a) Does factor endowment, including climate, have an incidence on inequality?

We could answer "no" because the distribution of farmers' income is not unequal (Table 15-4). Groundnuts are produced by small farmers and the gap between these farmers and those who produce only subsistence crops is not important.

But the choice of Dakar as capital of AOF, the main cause of inequality, is the consequence of factor endowment. Easterly stresses that living conditions for European settlers represent a factor endowment. The two advantages of Dakar were climate (Table 15-3) and location. Even if the climate of Dakar is relatively hot, the conditions for Europeans settlers are a lot more favourable than in Abidjan. Dakar is a port near of the river Senegal and this river was the only way inside the country two centuries ago. Dakar is also the nearest town from France. On the other hand, colonisation began here as early as the 18th century. The choice of Dakar was dictated by geography as well as by history.

b) Does inequality had an impact on institutions?

In 1960, Senegal is a very unequal country (foreigners who represent 2.2% of population receive 20/25% of income). Such inequality results from the colonial regime. This inequality entails a dualist consumption pattern. In the modern sector, households consume imported goods and some manufactured goods produced in Dakar, whereas 90% of the population lives in villages mainly from subsistence agriculture and consume very few manufactured goods. Senegal inherited at independence this dualist pattern for production, consumption and employment with 10% of households (in the modern sector) receiving more than 50% of income. Such dualism had an effect on the choice of institutions after independence. The state authorities established institutions which maintained inequality. The wage scale, the labor legislation, the schooling system were not changed. Priority was given to secondary school and university as was done before, whereas the enrollment rate in primary school remained below 50% until 1980 (Table 15-4). Education expenditures increased but inequality in schooling years between students was not reduced. The enrollment rates in secondary schools and university are above the average in Africa, whereas the rate in primary school is much below the African average (Berthélemy *et al.* 1996). The children of households working in the modern sector have access to secondary school and afterward can find a job in this sector, whereas the children of farmers have scarcely access to primary school and nearly never to secondary school.

The ruling classes thus maintained after independence the colonial system of education in which they had been raised. They were the "good pupils" and afterward the "good students" trained in colonial schools and university. They maintained a system which enabled them to achieve personal success and were therefore a model.

This unequal system entailed important consequences. The French riots in May 1968 gave rise to a long strike in Dakar University. The government yielded under such pressure. The university funding, including grants to students, was largely increased; the government appointed new civil servants in order to give jobs to students. So the size of the public sector, which was already overstaffed before 1968, expanded rapidly, which worsened the macroeconomic unbalances.

The history of Senegal is different from that of Brazil analysed by Engerman and Sokoloff. In Brazil, the elite established and maintained institutions that ensured the protection of its wealth. In Senegal, the majority of the elite before 1960 were French or Syro-Lebanese. The French civil servants left Senegal between 1960 and 1970 and Senegalese elite took their places. But these elite maintained the institutions and the dualism which allowed high inequality. So, the independence changed the citizenship of the elite, but not the institutions, as was acknowledged by President Senghor in 1963 (he said that the Senegalese elite have maintained all the privileges granted to French civil servants).

c) Do institutions have an incidence on income distribution?

More precisely, can institutions established in 1960 explain inequality in the 70's and the 80's?

The independent state expanded a parastatal sector which granted privileges to wage-earners. The number of public enterprises rapidly increased (86 in 1982 instead of 21 in 1962). All these enterprises are overstaffed and pay high wages. Moreover, the sum of wages paid to civil servants also increased (a consequence of higher wages and larger staff).

After the 1968 crisis, the minimum wage in the modern sector doubled in 5 years. Large food subsidies (for oil and rice) were granted in the 70's which have an impact only in urban zones. These subsidies are financed by taxes on cash crops, which increased the gap between average income in rural and urban areas and inequality.

These examples show that the institutions established in 1960 have benefited urban population and mainly people working in the modern sector. Such bias explains an inequality which is as high in 1990 as it was in 1960.

3. INSTITUTIONS AND INEQUALITY IN KENYA SINCE A CENTURY.

Colonial Kenya illustrates very well the pattern of Engerman and Sokoloff. But since independence was achieved in 1963, inequality remained nearly as high as in the past.

3.1 Colonial Kenya

Factor endowment explains inequality and institutions insuring dualism. Since the end of the 19th century, the English understood that the Highlands of Kenya allowed the development of a modern agriculture by European settlers because this region had a low population density, a temperate climate (Table 15-3) and fertile soils. In Subsaharan Africa, such conditions are an exception. Laws were rapidly introduced which established a dual agriculture (with large estates of settlers and African small holders growing subsistence crops). The 1901-1902 land laws divided the country between indigenous lands and Highlands which were Crown property and could be distributed to British settlers. The colonial government forbade Indians to own land and Africans to grow cash crops. So, the British settlers had the monopoly of cash crops. They grew coffee, tea, sugar, sisal, maize and grain and developed cattle raising. The white landowners dominated politics. British and Indians enjoyed autonomy whereas Africans were strictly controlled by the colonial government. The Kikuyus who live in the Highlands had to go into towns in order to find a job or remain in their village and work on settlers' estates because they had lost their land. They received very low wages because the subsistence income of smallholders was held down by the regulations which forbade cash crops. So in 20/30 years, an exceptional factor endowment had promoted inequality and institutions favouring dualism. As the government wielded all powers, the transition from the factor endowment to dualism, from dualism to institutions favouring inequality was very rapid. In 1914, Kenyan society was divided in 3 classes:

- the Europeans who owned large estates, all modern enterprises, held professional jobs and high-level administrative positions.
- the Indians who controlled trade and work in mid-level positions.
- The Africans who were confined to smallholder farming, unskilled jobs and petty trade.

The relative racial income levels for Africans, Indians and Europeans are estimated to have been 1 : 26 : 144 in 1914 (Arne Bigsten 1986). Although the share of Non-Africans in the population was very small (1%), their share in income was over 20%.

From the 1920's to 1950, this dualism expanded because it was facilitated by institutions, by the growth of modern agriculture and modern enterprises in towns. The income of Africans increased because they were allowed some cash crops and employment in estates and urban zones had expanded. But the government supported only the white population. Public investment benefited to the modern sector. The government aided landowners by technical assistance, subsidies to transports and high tariff barriers in order to protect domestic agriculture. The number of Europeans increased from 10,000 to 30,000 in 1950, 61,000 in 1960. In 1950, the share of Europeans and Indians in population was 2.5% (Table 15-3), whereas their share in income reached 51%. This evolution explains increasing inequality. The estimate of Bigsten is based on 13 income categories (Bigsten 1986). Between 1914 and 1950, the Gini coefficient increases from .50 to .70. The combination of colonial institutions established before 1914 with the growth of a modern sector explains such an evolution.

In 1925, the Kikuyus created an association in order to fight land expropriation in Highlands. In 1950, the Kenyan African Union, mainly composed of Kikuyus, demanded the distribution of these lands to Africans, and from 1952 onward the Mau-Mau (recruited among Kikuyus) were attacking or murdering European farmers. The colonial government had to change its policy. Smallholders were allowed to grow coffee, they received more financial support and minimum wage was pushed up. In spite of the Mau-Mau revolt, the economy boomed and the Non-African population reached 3% in 1960. Bigsten estimate of Gini coefficient in 1960 (.68) is nearly the same as in 1950. Inequality among Africans increased between 1950 and 1960 because only a minority improved rapidly their income position (the farmers who owned relatively large tracts of land where cash crops could be grown or, in towns, the skilled workers).

Even if inequality slightly decreased during the 1950's, it remains very high, higher indeed than in other African countries. But contrary to the elite in Brazil, who maintained its status over time, European landowners lost completely political power after independence in 1963. This difference with Brazil can be explained by several factors. The percentage of whites in Brazil was a lot higher than in Kenya (only 1%). The (Protestant) missions provided the young Africans with basic education whereas blacks or natives in Brazil were illiterate. The decolonisation in the 1960's is completely different from accession to independence of colonies in Latin America in the early 19th century.

3.2 Independence and inequality

The paradox of Kenyan history is the persistence of high inequality after and despite independence. All the privileges of European settlers have disappeared, twenty years after independence nearly all of them have left the country, all Kenyans can own land anywhere.

The state bought back the estates and in some cases established cooperatives, or gave land to small holders. But 80% of estates were bought by Kenyans with the support of public banks and a new class of African large landowners replaced the European class (Table 15-4). The rapid increase of enrollment rates in primary schools (which exceeds 80% since 1980) has given access to education to a large majority of young Kenyans (Table 15-4). The Indians were expelled from administration and private enterprises in order to supply jobs to Africans.

The government appointed new civil servants (employment reaches 380,000 in 1977 instead of 160,000 in 1963) and largely increased average wage. The government provided aid to small farmers in order to expand cash crops. In 1980, more than 50% of cash crops are grown in small holdings (Jennifer Sharpley, 1986). A direct tax has been established at the expense of the upper class. The average rate for the top decile reaches 13%. The last Presidential election in 2002 took place following democratic rules. These institutional changes and these policies have contributed to developing a middle class recruited among civil servants, as in Senegal, but also among skilled workers, small employers, middle size farmers. Around 20% of farmers produce coffee, rice, tea, maize, meat and milk. They hire workers and have a much higher income than other farmers (excluding large estates). This differentiation explains increasing inequality among small holders (less than 20 hectares): the Gini coefficient in 1973 reaches .52.

Despite these improvements, inequality remains nearly as important as in the past. Bigsten estimate of the Gini coefficient (active population) is the same in 1976 than in 1960 (.68). Following the same method, Vandermoortele (Wouter Van Ginneken and Jong-Goo-Park, 1984) estimates the Gini (total population) in 1977: .59; Morrisson (1972), using a similar method, estimates the coefficient in 1969: .61/62. Richard Anker and James Knowles (1983) with a macroeconomic model and a social accounting matrix estimate the coefficient for net income: .64. A household survey in 1992 gives a coefficient equal to .54 for expenditures, which means a coefficient of .61 for income distribution (assuming a ratio of .88 between the two coefficients). We observe that all these estimates are quite in the same range. All coefficients vary between .59 and .64 in 1969-1992 (Table 15-1). With such figures, Kenya is ranked among the most unequal countries in the world.

We cannot explain such inequality by racial differentiation. The percentages of Europeans and Asians in the population have rapidly decreased are non significant now. Almost all large estates are owned by Kenyans. The share of Non-Africans in total income has dropped from 51% in 1950 to less than 8% in 1990. As the share of the top quintile reaches 62% in 1980-1990, we conclude that more than 98% of people ranked in this quintile are Kenyans and they receive nearly 55% of total income. This elite (the top vintile) and this middle class (the vintiles 17, 18 and 19) have replaced the Non-African elite and middle class. There is now a high inequality among Kenyans instead of a high inequality between Non-Africans and Kenyans.

Several factors explain the persistence of such an inequality. The large estates are owned by Kenyans. After independence, the government respected the property rights of foreign capitalists (Table 15-3); it wanted to draw in foreign investments by low taxes on profits. In the 1960's and the 1970's, the expansion of the administration and the increase of average wage maintained a gap between the average income of civil servants and that of small farmers (Table 15-5) who produce subsistence crops in the East or the West, or practice cattle-raising in the North-East. As the population growth is very rapid (around 3.5%), the average size of small holdings decreases: in several regions the majority of holdings have less than 1 hectare. There is a "hard core of poor peasants which has not benefited from the opportunities" (Bigsten 1986). As a consequence of demography, the average income of people who have little and poor land, who are landless or pastoralists, has decreased. So the share of the poorest 50% is only 13% and the gap between this population and the elite has not been reduced.

The history of Kenya shows that if a country inherits at independence a dual economy with very high inequality, dualism and inequality can remain for a long time even if the colonial institutions have disappeared. The new elite and middle class promote institutions supporting their own interests. But if inequality has nearly not changed, social mobility has increased a lot. Before 1963, the European elite had the monopoly of the highest positions whereas now many Kenyans ranked in the elite have succeeded through education. Instead of a closed elite, the new elite is open which explains a very strong demand of education by families.

4. THE GHANEAN COUNTER-EXAMPLE

Income distribution in Ghana is clearly less unequal than in Kenya and Senegal. The three household expenditures surveys on living standards carried over in 1988-1989-1991 give nearly the same results: the Gini coefficients for expenditures distribution by individuals vary between .34 and .37. Assuming a ratio of .88 for expenditures Gini/ income Gini, we obtain the average value .405 instead of .56 in Senegal and .61 in Kenya (Table 15-1).

Such differences cannot be explained by the development of cash crops or openness. In 1965, exports reach 26% of GDP in Ghana, 31% in Kenya and 27% in Senegal. The structure of exports in Ghana and Senegal is rather similar with one cash crop as the main product (cocoa or groundnut), whereas phosphate or gold and wood have a secondary importance (Table 15-2). The development of cocoa in Ghana before independence has been a factor of growth: 1.5% per year for the GDP per capita between 1910 and 1960. In 1960, the GDP per capita exceeds the GDP per capita in Côte-d'Ivoire by 10% and is the double of the Kenyan figure.

Geography and history distinguish Ghana from Kenya and Senegal. Several factors explain the quasi absence of a European population around 1960:

- There is no region as the Highlands in Kenya with a climate suitable (Table 15-3) for European settlers.
- The factor endowment: good soils for cocoa do not entail large estates because there is no significant economies of scale.
- The history of Ghana has induced the British government to apply the "Indirect Rule" which minimizes the number of English civil servants. The Ashanti kingdom dominated the region since the 18th century by trading gold and ivory. In 1867, the Fanti established a confederation with a parliament and an assembly of tribal chiefs. In 1900, the British government dissolved the confederation and deposed the king. But in 1935, it reinstated the king and applied the "Indirect Rule". The existence of a pre-colonial state thus induced the British government to limit its interventions (in 1865, a commission sent by the British Parliament concluded that the best solution was self-government by the Africans).

Since the early 20th century, Africans have expanded cocoa production and a middle class of capitalist farmers has evolved. Between the two World

Wars, the Ghanean exports of cocoa amount to nearly to half of the world exports and cocoa forms 2/3 of Ghanean exports. The capitalist farmers produce 80% of national production and hire agricultural workers. They save and buy land which they give to sharecroppers to work. The expansion of cocoa has entailed a social differentiation with increasing inequality between this middle class and the sharecroppers, the small holders and the agricultural workers (Table 15-4).

In 1957, inequality among Africans is rather high, but there is no European elite as in Kenya and Senegal (Table 15-3). The history since independence can be summarized into four periods:

- 1957-1966: Nkrumah establishes a socialist regime with an overstaffed public administration and creates numerous state-owned enterprises
- 1966-1972: liberal policy
- 1972-1981: the socialist policy dominates again and the GDP per capita decreases (- 20%).
- Since 1981: an adjustment programme is undertaken in 1983 and applied until the early 1990

During the first period, the employment in administration reaches 600,000 persons in a country of 8.2 millions of habitants (more than 60% of active population in non-agricultural sector are civil servants). All foreigners in administration are expelled. Nkrumah eliminates private traders who collected cocoa and gives the monopoly of the collect and exportation to a public enterprise. This control allows taxation of cocoa: 2/3 of gate-price are thus confiscated by the state. This socialist policy reduced a lot the income of the middle class (capitalist farmers, cocoa traders, employers in the modern sector) and benefited the lower class in towns by creating hundred thousands jobs in administration and public enterprises (Table 15-5).

Between 1966 and 1972, the recovery in cocoa price increases the income of farmers and urban unemployment becomes more important because 12% of civil servants are laid off. But the new government in 1972 stops liberalisation and establishes a "Prices and Incomes Board" which fixed wages and all prices in the modern sector until 1986. The policy of cocoa taxation resumes at the expense of producers.

The results of these varied policies are:

- nearly all foreigners, whether Africans or Non-Africans, have been expelled,
- an important decrease in the income of cocoa producers has occurred (in 1966, it drops to 1/3 of the 1960 level), and likewise for that of cocoa traders and of employers in the modern sector,

- the appointment of hundred thousands civil servants benefit the urban population. But between 1970 and 1980, the average real wage of these people has been cut by half,
- The prices controls and the shortage of manufactured goods resulting from the lack of foreign exchange provide illegal incomes to small traders and some civil servants who practise corruption or buy these goods at official prices and retail them on the black market.

These effects of a socialist policy explain a moderate inequality at the end of the 1980's. With the public monopoly of cocoa and other products, the rich traders have disappeared. The large decrease of cocoa price and production has definitely impoverished the middle class. The civil servants who are a privileged elite in countries like Senegal have seen their real wage dramatically reduced, a consequence of inflation. This drop explains an average wage close to the average income in agriculture (Table 15-5). The only winners are the traffickers on informal or black markets whose income is perhaps underestimated by the household survey.

Such results are not surprising. The policy of Nkrumah has reduced inequality, entailed overstaffing in administration, but with low wages because the state could not finance a too large public sector with a falling GDP.

Ghana could have chosen another policy in 1957, a liberal policy in favour of the middle class and entailing more inequality (but less inequality than in Kenya because there was no large estates owned by foreigners). Between 1957 and 1961, Nkrumah increased rapidly public investment and established the cocoa monopoly. A socialist policy was applied after 1961 which entailed a complete change of the institutional framework. As the same policy was applied in 1972-1981, in the early 1980's Ghana was a socialist economy. The income distribution around 1990 is a legacy of this socialist framework (even if, with adjustment, the producer price of cocoa has been raised and the staff in administration reduced). The main characters of this framework are:

- the taxation of cocoa producers,
- an equalitarian wage scale in administration,
- a very large public sector whereas the modern private sector is very small,
- the absence of foreigners (Table 15-3).

Such factors reduce inequality, even if a small increase occurred after the adjustment programme. This impact of socialist institutions on income distribution confirms item c) (institutions have an impact on income distribution). For item a) (does inequality depend on factor endowment?), it is sure that very different endowments in Ghana and Kenya explain different income distribution.

The last item b) (do institutions depend on income distribution and factor endowment?) concerns the incidence of inequality on institutions. The Ghanean experience runs against this idea. At independence, Ghana had an important middle class (which was supported by the traditional chiefdom) and which could have gained control of political power. Inequality, although not low, was moderate and could have persisted in favour of this class which could have promoted institutions advantaging its interests. But this class was eliminated by Nkrumah. First, he created a new party in 1949 with a clientele of young people and civil servants in towns. Then, after 1961 he imposed a socialist regime. Neither the income distribution nor the structure of Ghanean society explains the choice of Nkrumah, but the fact that the charismatic leader of the struggle for independence was a Marxist. We can only add that Ghana expanded education before Kenya and Senegal. In the 1960's, a large group of young and educated Ghaneans had evolved and Nkrumah used this group as an instrument (and favoured their interests) at the expense of the middle class.

5. CONCLUSION

These three African experiences first allow us to qualify Engermann and Sokoloff's conclusions. The differences in the possible economies of scale (e.g. sugar vs. grain) are the key factor in their analysis but in the case of African countries climate has been the key factor. Easterly (2002) and Acemoglu *et al.* (2002) mention climatic conditions for settlement by Europeans as a factor endowment. Our analysis confirms the importance of these conditions. They explain the increasing inequality with the development, prior to 1960, of a modern agricultural sector in Kenya, of an administrative and manufacturing sector in Senegal. Corresponding appropriate institutions characterized by dualism were created at that time. Such institutions were maintained and continued after independence by the African elite for its benefit while Europeans were departing.

Relying on data collected (Tables 15-1 to 15-5) we can provide quantitative tests of the three linkages we have assumed.

a) inequality depends on factor endowments.

The average Gini coefficient around 1960 was:

.60 in Kenya and Senegal,

.40 in Ghana.

The key factor is climate:

Around 20° in Highlands and Dakar (except in the summer for Dakar) and dry climate in both;

28° in Accra and humidity.

The influence of climate on European settlements is clear:

2 to 2.5% of population in Kenya and Senegal;

0.2% in Ghana.

These data prove the impact of climate on European settlements which entail high inequality. With the development of a modern sector controlled by Europeans, only 2 to 2.5% of population received in 1960 more than 30 or 40% of total income in Kenya and Senegal.

b) institutions depend on income inequality

In Kenya and Senegal (Gini = .60) we observe:

- a large public sector, mainly in Senegal (wages = 17% of GDP), with overpaid civil servants.
- concentration of holdings in Kenya: 50% of land owned by less than 5% of farmers in 1960 and public aid to Europeans landowners with subsidies, technical assistance and high tariffs.
- in Senegal rigid labor market which entails high wages (the ratio unskilled worker wage/average income in agriculture reaches 2.4/1) and trade protection which allows important profits.

In Ghana (Gini = .40) there is no large public sector (wages < 5% of GDP) and no institutions in favour of Europeans landowners or entrepreneurs.

In 1960, Kenya and Senegal had institutions benefiting a rich minority of Europeans in agriculture, manufacturing or administration. The colonial regime allowed this minority to establish such institutions. On the contrary in Ghana in the absence of such a European minority there was no need for such institutions.

c) institutions have an impact on income distribution after independence.

Several institutions in the 70's and 80's in Kenya and Senegal are the same than in 1960 (land concentration, high tariffs, overpaid wage-earners in administration, with the following ratio average public wage/average income in agriculture 11/1 in Senegal, 4.4/1 in Kenya). Institutions in Ghana remain as non-inequalitarian as before (ratio 1.2/1 instead of 11/1 or 4.4/1; a land concentration which does not entail large inequality because cocoa is highly taxed).

This last relation insures in 1990 the persistence of high inequality in Kenya and Senegal (Gini = 0.575) instead of moderate inequality in Ghana (Gini = 0.40).

These African examples show two facts: the control of political power by a new national elite does not change inequalitarian institutions and the historical factor is important.

Once a highly dualist structure is established in a country according to the classical scheme (factor endowments influence income distribution, income distribution influences institutions which in turn influence distribution), even when the elite leaves the country because of its foreign origin, the dualist structure continues to prevail. The new national elite takes the foreigners' place and maintains or chooses institutions ensuring for its benefit the same inequality than the previous one. It is clear that the couple inequalities-institutions with reciprocal causality effects in both directions, which is linked to dualism concerning the distribution of assets, of production and consumption, resist to political changes, to external pressure (donor countries criticize the privileged status of civil servants in Senegal) and obtains for decades.

These three examples underline secondly, the importance of the historical factor. The choice of socialist institutions by Nkrumah is not accounted for by income distribution at independence. Once such a choice is made, a backward move and a change in institutions is very difficult. The liberal policies attempted in 1967-71 entailed high discontent of urban wage-earners and important strikes. The government decreed unions illegal; a putsch followed and socialist policies were resumed.

Although we deliberately did not include growth in our study, it should be kept in mind that GDP per capita decreased by10% in Ghana and Senegal between 1960 and 1998 while it increased by 50% in Kenya. Analyses of Ghana by Alan Roe and Hartmut Schneider (1992) and of Senegal by Berthélemy *et al.* (1996) show that the institutional framework, socialist in Ghana, statist in Senegal, had a negative impact on growth. As poverty depends on growth and income distribution, this relation between institutions and growth should be taken into account for any study of poverty trends in these countries.

Table 15-1.

	GDP per	Capita		Gini	Coefficient		
	1970	1980	1990	1960	1970	1980	1990
Ghana	1424	1172	1078				.40
Kenya	913	1029	1102	.62-64	.61-64	.59	.60-61
Senegal	1435	1301	1354	.56-57	.51		.54-55

Table 15-2.

	Ghana	Kenya	Senegal
Share of Public sector	Low in 1956, very high after 1961	Moderate	Always very high
Wages scale (public)	Equalitarian until the 80's	Further from market wages	Linked to French scale
Regulation of labor markets	Flexible before1961, rigid afterward	Flexible	Always rigid
Property rights	Not complied with	Complied with	Complied with
Tariffs	Very high until the early 1990's	Very high in 1963-73 and 1981-87; lower in 1974-81 and after 1987	Very high until the early 1990's
Climatic conditions for European settlers	Very unfavourable	Very favourable in Highlands	Slightly favourable in Dakar
Natural factor endowment	Cocoa, Diamonds, Timber	Tea, Coffee, Cotton, Cane Sugar, Sisal, Cattle raising	Groundnut, Fish, Phosphates

Table 15-3.

	Climate average temperature	Foreigners (% in total population)	Percentage of modern enterprises owned by foreigners
Ghana	28° (coastland) and relative humidity 80%/100%	1960: 0.2% Eur + 0.1% As 1968: 0.1% Eur + 0.1% As	1965 < 5%
Kenya	19° (Nairobi in Highlands)	1950: 2.5% (Eur + As) 1969: 1.1% (Eur + As) 1976: 1% (Eur + As)	1969 50% (percentage among executive and managers)
Senegal	22° in January	1956: 2.% (Non-Africans)	1966: 80% 1984: 50%
	29° in August (Dakar) and sea-breeze	1966: 1.5% (Non-Africans)	

Table 15-4.

	Enrollment rate in first and second degrees	% of holdings	% of land
Ghana	1960: 38%; 5%	1970: 55%	12.7%
	1978: 74%; 29%	22%	16%
	1987: 73%; 39%	12%	16.7%
		11%	54.6%
Kenya	1960: 49%; 4%	1969: 50%	12%
	1980: 85%; 20%	40%	37%
		10%	51%
Senegal	1960: 22%; 2%	1960: 60%	38%
	1980: 49%; 8%	30%	38%
	1990: 57%; 16%	10%	24%

Table 15-5.

	Employment as a percentage of active population (1)	Wages as a percentage of GDP	Ratio average wage/average income in agriculture
Ghana	1968: 11%	1968: 9%	1968: < 1.2/1
	1968: 20% (2)	1986-87: 5%	1986-87: around 1.2/1
	1986-87: 7%		
Kenya	1969: 4.5%	1970: 14.7% (2)	1970: 4.4/1
	1969: 6.6% (2)	1986: 10.3% (2)	(excluding modern
	1986: 8.4% (2)		agriculture)
Senegal	1959: 2.1%	1959: 16.7%	1959: 17/1
	1970: 2.8%	1970: 16.%	1970: 11/1
	1970: 4.2% (2)	1990: 8%	
	1989: 2.7% (2)		

1. We assume that active population amounts to 33% of total population.
2. Including parastatal corporations.

REFERENCES

Themes

Acemoglu, Daron, Simon Johnson and James Robinson. 2002. "Reversal of Fortune: Geography and Institutions in The Making of The Modern World Income Distribution" Quarterly Journal of Economics, 67:4; pp. 1231-94.

Bourguignon, François and Christian Morrisson. 1998. "Inequality and Development, The Role of Dualism" Journal of Development Economics, 57:2, pp. 233-58.

Easterly, William. 2002. Inequality Does Cause Underdevelopment: New Evidence. W.P. No. 1, Washington: Center for Global Development.

Engerman, Stanley, Stephen Haber and Kenneth Sokoloff. 2000a. "Inequality, Institutions and Differential Paths of Growth Among New World Economics", in Institutions, Contracts and Organizations. C. Menard ed. Cheltenham: Edward Elgar, pp. 108-134.

Maddison, Angus .2001. The World Economy, Paris: OECD.

Sokoloff, Kenneth and Stanley Engerman. 2000b. "History Lessons. Institutions, Factor Endowments and Paths of Development in The New World", Journal of Economic Perspectives, 14:3, pp. 217-232.

Thomas, Vinod, Yan Wang and Xibo Fan. 2000. Measuring Education Inequality: Gini Coefficients of Education. Mimeo, Washington: World Bank.

Thorbecke, Erik and Chutatong Charumilind. 2002. "Economic Inequality and its Socio-economic Impact", World Development, 30:9, pp. 1477-95.

Countries

Anker, Richard and James Knowles. 1983. Population Growth, Employment and Economic Demographic Interactions in Kenya: Bachue- Kenya. New-York: St Martin's Press.

Berthélemy, Jean.-Claude, Aboulaye Seck and Ann Vourc'h. 1996. Growth in Senegal: A Lost Opportunity? Paris: OECD.

Bigsten, Arne. 1986. "Welfare and Economic Growth in Kenya, 1914-76", World Development, 14:9, pp. 1151-60.

Morrisson, Christian. 1968. La Répartition des Revenus dans les pays du Tiers-Monde. Paris: Editions Cujas.

Morrisson, Christian. 1972. Income Distribution in Kenya; Income Distribution in Senegal. Development Research Centre. Washington: World Bank.

Republic of Ghana. 1985. Economic Survey 1982. Accra: Central Bureau of Statistics.

Roe, Alan and Hartmut Schneider. 1992. Adjustment and Equity in Ghana. Paris: OECD.

Sharpley, Jennifer. 1986. Economic Policies and Agricultural Performance, The Case of Kenya. Paris: OECD.

Terrell, Katherine. and Jan Svejnar. 1989. The Industrial Labour Market and Economic Performance in Senegal. Boulder: Westview Press.

Van Ginneken, Wouter and Jong-Goo Park. 1984. Generating International Comparable Income Distribution Estimates. Geneva: ILO.

Chapter 16

INCENTIVES, INEQUALITY AND THE ALLOCATION OF AID WHEN CONDITIONALITY DOESN'T WORK: AN OPTIMAL NONLINEAR TAXATION APPROACH

Ravi Kanbur[1] and Matti Tuomala[2]
[1]Cornell University, sk145@cornell.edu and [2]University of Tampere, matti.tuomala@ kolumbus.fi

1. INTRODUCTION

In his major overview of the relationship between the development doctrine and aid, Thorbecke (2000), states that "the decade of the 1990's was marked by a strong and lingering case of 'aid fatigue' influenced by a rising fear that foreign assistance was generating aid dependency relationships in poor countries." There seems no question that dissatisfaction with aid is at an all time high. In sharp contrast to the optimism of earlier evaluations, those of the 1990's have been almost uniformly negative (e.g. Boone, 1996, World Bank, 1998, Kanbur, Sandler and Morrison, 1999, Burnside and Dollar, 2000). Much of the discussion has focused on the failure of aid conditionality (see Kanbur, 2000, Killick, 1995, Mosley, Harrigan and Toye, 1995, Svensson, 2000a, World Bank, 1992). This literature has highlighted two inescapable facts. First, the record of aid in helping growth and poverty reduction is disappointing. And second, this poor record has been established despite aid conditionality, which was meant to channel resources to productive uses. On the latter, there is a growing realization that external donors cannot really change domestic political economy in the medium term (Devarajan, Dollar and

Holmgren, 2001). Some have argued that donors should accept a basic truth--the impact of their aid will come mainly through the effects of relaxing the resource constraints facing the recipient governments. In light of this, aid should be allocated on the basis of observable indicators that predict the impact of additional resources (for example, World Bank, 1998 and Burnside and Dollar, 2000).

This paper takes seriously the conditionality critique. While there is a developing literature on how conditionality could be "improved" (see for example Svensson 2000b or Kanbur, 2000), this paper instead asks two questions. What would be the consequences of aid for a country if there were no conditionality at all (which many would argue to be the case de facto, despite de jure conditionality)? And what would the optimal allocation of unconditional aid across countries look like?

The central assumption of this paper is that the effects of aid come through the relaxation of the recipient government's budget constraint—all other attempts to "influence" the government are ineffective. Given this increase in its resources, the government chooses its policy instruments to satisfy its own preferences. These instruments and preferences can be modeled in as complex a fashion as possible and necessary for the analysis. In this paper we model the net effects of the array of policy instruments at the government's disposal (e.g., direct and indirect taxes, trade interventions, different patterns of public expenditure) simply as the choice of a non-linear income tax and transfer schedule to maximize a welfare function defined on the distribution on incomes. This optimal non-linear income taxation approach, pioneered by James Mirrlees in his Nobel Prize winning paper (Mirrlees, 1971), captures the central features we are interested in highlighting: incentives, inequality, and the effects of different preferences over income distribution outcomes. It serves to shed new light on the literature on the consequences and optimal allocation of aid.

The plan of the paper is as follows: Section 2 introduces the basic model with a single recipient government and a single donor, with aid being modeled as an exogenous increase in the recipient government's budget constraint. In this framework, Section 3 considers the consequences of unconditional aid, including the possible "crowding out" effects of aid as relaxation of the resource constraint feeds through to incentive effects via the recipient government's (optimal choice) of the income tax and transfer schedule. It presents numerical simulations which give a quantitative handle on the magnitude and determinants of these effects. Section 4 turns to the case with two recipient governments which differ in their tastes and technologies, and characterizes the optimal allocation of aid where the donor is assumed to maximize a joint welfare function of the two countries in deciding on the allocation. Numerical simulations are presented on how inherent inequality in the two countries, and their tastes for inequality, affect the optimal allocation. Section 5 concludes the paper with a discussion of areas for further research.

2. THE BASIC MODEL

The simplest model in which incentives, inequality, preferences for equity, and aid without conditionality can be integrated in a coherent framework, and which can provide an engine of analysis for the questions we are interested in, turns out to be the Mirrlees (1971) model of optimal nonlinear income taxation. In this model there is inherent inequality because individuals differ in their labor productivities. The government chooses a nonlinear income tax and transfer schedule to maximize a welfare function which is in principle sensitive to inequality, but does so with the added constraint that individuals choose their labor supply in response to the tax function. The government must also satisfy the overall budget balance constraint, with tax revenues equal to outlays. Aid from the outside serves to relax this constraint, and therefore affects the whole equilibrium. Note that we view this model as the stylized representation of the myriad instruments the government has at its disposal. Each instrument has incentive, revenue and distributional effects. We start by laying out the basic model.

There is a continuum of individuals, each having the same preference ordering, which is represented by a utility function u(x,y) defined over consumption x and hours worked y, with $u_x > 0$ and $u_y < 0$ (subscripts indicating partial derivatives). Individuals differ only in the pre-tax wage n they can earn. There is a distribution of n on the interval (l,h) represented by the density function f(n). Writing gross income as z=ny and defining

$$s(x, z; n) = -u_y(x, z/n)/nu_x(x, z/n) > 0 \qquad (1)$$

preferences are taken to satisfy the further (standard) restriction that

$$\frac{\partial s}{\partial n} < 0. \qquad (2)$$

Condition (2) implies that indifference curves in (x,z)-space become flatter the higher is an individual's wage rate, which in turn ensures that both consumption and gross earnings increase with the wage rate.

Suppose that the aim of policy in a recipient country can be expressed as maximizing the following social welfare criterion

$$S = \int_l^h W(u(n))f(n)dn \ , \tag{3}$$

where W(.) is an increasing and concave function of utility. The government of a recipient country cannot observe individuals' productivities and thus is restricted to setting taxes and transfers as a function only of earnings, T[z(n)]. The government maximizes S subject to the revenue constraint

$$\int_l^h T(z(n))f(n)dn = R \tag{4}$$

where in the Mirrlees tradition R is interpreted as the required revenue for essential public goods. The more aid a government receives from external sources, the lower is R. In addition to the revenue constraint, the government faces incentive compatibility constraints. These in turn state that each n individual maximizes utility by choice of hours worked, solving

$$\max_{x,y} u(x,y) \text{ subject to } x = ny - T(ny), \tag{5}$$

where T(.) is the tax/transfer schedule. The problem (5) implies that

$$u_x(1-t(z)) + u_y = 0 \tag{6}$$

where t(z) = T'(z) denotes the marginal tax rate. Totally differentiating utility with respect to n, and making use of (6), we obtain the incentive compatibility constraints.[1]

$$\frac{du}{dn} = -\frac{yu_y}{n}. \tag{7}$$

Since T = ny-x, we can think of government as choosing schedules y(n) and x(n). In fact it is easier to think of it choosing a pair of functions, u(n) and y(n), which maximize welfare index (3) subject to the conditions of individual maximization (7) and the revenue requirement (4). Omitting details (for an exposition see Tuomala, 1990), the first order conditions of this problem imply a pattern of marginal rates satisfying

$$t(z(n)) = -\alpha(n)u_x s_n / \lambda f(n) \tag{8}$$

where λ is the multiplier on the revenue constraint and

$$\alpha(n) = \int_l^n (W'u_x - \lambda)(1/u_x)\exp(-\int_p^n (u_{nx}/u_x)dm)f(p)dp \tag{9}$$

is the multiplier on the incentive compatibility constraint. This latter satisfies the transversality conditions

$$\alpha(\mathrm{l}) = \alpha(\mathrm{h}) = 0. \tag{10}$$

Our objective is to trace through the consequences of unconditional aid. In this model, unconditional aid increases the resource envelope of the government. It reduces the revenue that the government needs to raise from the population. In other words, an increase in aid decreases R. The consequences of unconditional aid can in principle therefore be analyzed through comparative static of the optimum characterized by (8), (9) and (10), with respect to variations in R.

Unfortunately, however, as is well recognized in the non-linear taxation literature, closed form analytical results are few and far between.[ii] To see the complications that arise, consider equations (8) and (9), under the simplifying assumptions that $u_{xy} = 0$ and that the social welfare function is utilitarian i.e. W' = 1. Then (8) becomes

$$\frac{t}{1-t} = (e^{-1} + 1)\alpha(n)u_x / \lambda n f(n) \tag{11}$$

where $\alpha(n) = \int_l^n ((u_x - \lambda)(1/u_x)f(p)dp.$

Now (11) does however provide some insight. First, the term $(1+e^{-1})$, reflecting also conventional wisdom, (11) says that, other things equal, the marginal tax rate should be lower the larger is the compensated elasticity of labour supply, e. Governments fearing that disincentive effects are large will tend to set lower marginal rates.

Consider now the shape of the optimal tax schedule. Will the tax and transfer system become more or less progressive as aid increases? Equation (11) suggests that, other things equal, the marginal tax rate should be lower the denser the population at that point, i.e. higher f(n). In other words the more people affected, the higher is deadweight loss. On the other hand for the typical distribution the density weighted by n, nf(n), (e.g. lognormal distribution) is likely to decline with n above some point suggesting a higher marginal tax rate on high earners. These two factors may be seen as reflecting the efficiency side of the problem. The term $\alpha(n)$ in turn incorporates distributional concerns. It measures the social welfare gain from slightly increasing the marginal tax rate at n and using that extra revenue to loosen the government's budget constraint. This extra revenue can be used to soften the bite of the revenue requirement, and distributing as a lump sum subsidy to those below n. The implicit value of this loosening of the revenue constraint is just λ, the multiplier on the constraint. $\alpha(n)$ increases with n for low n and decreases with n for high n. The turning point depends on λ. The lower is λ, the higher is the n at which the turning point occurs. Thus as aid increases and the revenue requirement falls, and hence λ falls, the range over which $\alpha(n)$ is increasing stretches further. Since $\alpha(n)$ affects the marginal tax rate positively, this means that the range over which the latter increases also stretches further - at least for this reason. In this sense, therefore, more aid leads to a more progressive tax structure.

But this is about as far as we can get at this level of generality. And the above analysis says nothing about the consequences of increased aid for per capita income in the country, for pre tax and post tax income inequality, and for the well being of the poor. However, in the tradition of the non-linear taxation literature, we can provide better understanding of the form of optimal policy through numerical simulations. With these techniques, we can compute post tax income at each level of n, and thus calculate inequality of pre and post tax income as well as total income, for different values of key parameters. The next section takes up this task.

3. THE CONSEQUENCES OF UNCONDITIONAL AID

It should be clear from (11) that the variation of the optimal marginal tax rate with the level of income is a complex matter, and that comparative statics of inequality and averages as parameters vary will not be available in closed form. This is a general feature of the optimal nonlinear income taxation literature (see Tuomala, 1990) where, following the lead of Mirrlees (1971) numerical calculations have proved useful in generating useful results.[iii] We follow this route here. Our focus is on the consequences of varying R and varying the standard deviation of n. How do the consequences of aid play out in societies of differing degrees of "inherent" inequality?

We assume n to be distributed lognormally with parameters μ and σ (see Aitchison and Brown, 1957). This assumption is common in the literature, following Mirrlees (1971). For numerical simulations we choose σ = 0.39, 0.7 and 1 as a standard deviation of n and mean n = 0.4. The calculations were carried out for the following CES utility function

$$u = -\frac{1}{x} - \frac{1}{(1-y)} \tag{12}$$

where the elasticity of substitution between consumption and leisure, denoted by ε, is 0.5. The social welfare function of the recipient government is specified[iv] as

$$W(u) = -\frac{1}{\beta} e^{-\beta u}$$

so that ß measures the degree of inequality aversion in the social welfare function of the recipient government (in the case of ß = 0, we define W = u). R is specified as a fraction of national income, and is assumed to vary between -0.3 and 0.0. In other words, we consider external aid varying from zero to 30 percent of national income.

Tables 16-1 through 16-5 show comparative statics of the optimal tax schedule with respect to revenue requirement with different inherent inequality and inequality aversion. The tables give net income, x, gross income, z, and optimal marginal tax rates at various percentiles of the ability distribution, MTR. Moreover they give the net income, $x(n_0)=x_0$, and welfare of the poorest $u(n_0)=u_0$[v] and "crowding out" effect of unconditional aid. With no incentive effects, national income should increase by the same amount as aid. With incentive effects, this increase may be less—this is what we call "crowding out".

Table16-1. [(base run) Comparative static with respect to R with $\sigma = 0.39$.]

$\varepsilon = 0.5$ ß = 0 $\sigma = .39$

F(n)	R = 0.0 x	x_0=0.09 z	u_0=-11 MTR%	R=-.1 x	x_0=0.09 z	u_0=-10 MTR%	R=-.27 x	x_0=0.12 z	u_0=-8.3 MTR%
0.10	0.16	0.12	44	0.17	0.12	41	0.19	0.11	37
0.50	0.21	0.20	41	0. 22	0.22	39	0.24	0.19	35
0.90	0.29	0.33	35	0.30	0.30	33	0.32	0.31	31
0.99	0.39	0.48	29	0.41	0.47	23	0.42	0.44	25
Average	0.22	0.22		0.23	0.21		0.25	0.20	
Crowding out					4.5%			4.8%	
Effect									

x_0 is the net income of the poorest person, u_0 is the welfare of poorest person

Table 16-2. [Comparative static with respect to R with $\sigma = 0.7$.]

$\varepsilon = 0.5$ ß = 0 $\sigma = .7$

F(n)	R = 0.0	x_0=0.13	u_0=-8	R=-07	x_0=0.13	u_0=-7.7	R=-.17	x_0=0.15	u_0=-6.8
	x	z	MTR%	x	z	MTR%	x	z	MTR%
0.10	0.16	0.06	56	0.16	0.06	55	0.18	0.05	51
0.50	0.20	0.17	60	0.21	0.17	59	0.22	0.16	57
0.90	0.31	0.45	57	0.32	0.44	56	0.34	0.42	55
0.99	0.54	0.91	41	0.54	0.88	45	0.55	0.85	47
Average	0.22	0.22		0.23	0.22		0.25	0.21	
Crowding out Effect					1.4%			3.7%	

Table 16-3. [Comparative static with respect to R with $\sigma = 1.0$.]

$\varepsilon = 0.5$ $\beta = 0$ $\sigma = 1.0$

F(n)	R = 0.0	x_0=0.16	u_0=-6.2	R=-.1	x_0=0.17	u_0=-5.8	R=-.19	x_0=0.19	u_0=-5.4
	x	z	MTR%	x	z	MTR%	x	z	MTR%
0.10	0.17	0.02	55	0.18	0.02	53	0.19	0.02	50
0.50	0.21	0.14	68	0.23	0.13	67	0.24	0.12	65
0.90	0.35	0.55	71	0.36	0.58	69	0.39	0.56	68
0.99	0.70	1.61	58	0.75	1.60	52	0.75	1.54	55
Average	0.24	0.24		0.26	0.23		0.27	0.23	
Crowding out effect					4.1%			3%	

Table 16-4. [Comparative static with respect to R with $\beta = 1$.]

$\varepsilon = 0.5$ $\beta = 1$ $\sigma = 0.39$

F(n)	R = 0.0 x	x_0=0.12 z	u_0=-8.3 MTR%	R=-.1 x	x_0=0.13 z	u_0=-7.7 MTR%	R=-.28 x	x_0=0.15 z	u_0=-6.7 MTR%
0.10	0.16	0.10	65	0.17	0.09	62	0.19	0.09	56
0.50	0.19	0.19	59	0.20	0.18	56	0.23	0.17	52
0.90	0.26	0.33	47	0.27	0.32	45	0.29	0.30	43
0.99	0.36	0.50	29	0.38	0.49	28	0.39	0.46	32
Average	0.20	0.20		0.21	0.19		0.24	0.19	
Crowding out Effect					3%			2.6%	

Table 16-5 [Comparative static with respect to R with $\beta = \infty$ (maximin)]

$\varepsilon = 0.5$ maximin $\sigma=0.39$

F(n)	R = 0.0 x	x_0=0.14 z	u_0=-7.1 MTR%	R=-.1 x	x_0=0.15 z	u_0=-6.7 MTR%	R=-.19 x	x_0=0.15 z	u_0=-6.7 MTR%
0.10	0.14	0.05	85	0.15	0.04	84	0.16	0.03	83
0.50	0.17	0.18	69	0.19	0.17	67	0.20	0.17	66
0.90	0.24	0.24	51	0.25	0.33	50	0.26	0.32	48
0.99	0.34	0.51	32	0.35	0.50	30	0.37	0.49	26
Average	0.19	0.19		0.20	0.18		0.21	0.17	
Crowding out Effect					5.2%			5.5%	

Tables 16-1 through 16-5 show that there is indeed a crowding out effect as increased unconditional aid reduces labor supply incentives in the recipient country. This effect is always greater than 1 percent of total income, but can be as large as 5 percent. However, despite this crowding out, total consumption in the recipient country does indeed increase with aid. Moreover, net income and welfare of the poorest also increase with aid.

Consider now the progressivity of the domestic tax structure as a function of aid. Tables 16-1 through 16-5 show that optimal tax/transfer systems become more progressive when inequality increases, $\sigma = 0.7$ and 1.0, and when R becomes more negative (i.e. aid increases). To understand this, we can combine the results of two earlier studies. Kanbur-Tuomala (1994) shows that with greater inherent inequality optimal marginal tax rates increase with income over the majority of the population. On the other hand we know from Immonen-Kanbur-Keen-Tuomala (1998) that as the revenue requirement becomes negative so that for example foreign aid is available the minimum income requirement for the poor can be met without clawing back revenue with a high marginal tax rate. Thus we have low marginal tax rates on the poor. In other words, optimal progressivity, taking into account incentive effects, increases with higher inherent inequality and with additional aid resources.

4. OPTIMAL AID ALLOCATION

The previous section analyzed the consequences of unconditional aid for a given recipient. Suppose now that the donor is faced with two potential recipients, countries A and B, with their own specific characteristics. Any given aid allocation between the two recipients leads to consequences in each country along the lines laid out in Section 2—each recipient government chooses its policies in light of its technology, tastes and aid allocation. The donor now has the task of choosing the aid allocation from a fixed pool of aid resources, to optimize the donor's own welfare function. This is the problem that is set up and solved in this section.

Let the donor's welfare function be

$$S_D = \sum_i \theta_i \int W_i(v_i(n)) f(n) dn; \qquad (13)$$

where S_D is the donor's valuation function which is not necessary the same as W_i and θ_i is a fraction of the world's population in country i (= A and B). Thus we assume that the donor's preferences in inequality aversion may differ from preferences of recipient countries. The simplest representation of the donor's problem is to find b_i maximizing (13) subject to

$$\sum_i \theta_i b_i = b.$$

where b is a fixed pool of aid resources. Hence, the donor simply chooses the optimal allocation of the aggregate aid, b, over the two countries. This requires equating across countries the marginal social cost of raising an additional unit of aid, so that

$$\lambda_A = \lambda_B = \lambda, \tag{14}$$

where λ is the marginal social cost of donor funds.

We are interested in how the optimal allocation responds to key parameters of one the two countries, including (i) increase in the mean productivity, (ii) increase in inherent inequality and (iii) increase in inequality aversion. What will happen to the aid allocation? It should be clear from our discussion above that the comparative static is not available in closed form. Rather, we will have to rely on numerical calculations. The calculations are carried out using the same specification for the utility function, density function and social welfare as in Section 3.

We start with the case where the two countries are identical. In this case optimal allocation of aid is to divide the total aid equally between the two countries A and B. This is shown in Table 16-6. Then we conduct a sequence of experiments, one at a time. The computations are shown in Tables 16-6, 16-7, 16-8a, b, 16-9 and 16-10a, b. The tables give marginal and average tax rates at various percent points of the n-distribution, and the optimal distribution of aid across countries, for some variation from the base case of identical countries. To limit the number of variations, we keep population size of the two countries identical—it is in any case obvious that as the relative population size of one country increases, ceteris paribus its aid allocation should also increase.

Table 16-6. [Identical countries A and B]

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-47	42	0.50	0.50
0.50	-12	39		
0.90	7	34		
0.99	14	28		

Table 16-7. [Decrease in the mean productivity of country A]
$\lambda_A = \lambda_B = 18$, $f_A(\mu_A = -1.2, \sigma_A = 0.39)$ $f_B(\mu_B = -1.0, \sigma_B = 0.39)$, $\beta_A = \beta_B = 0$

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-100	34	0.82	0.18
0.50	-47	33		
0.90	-16	30		
0.99	-2	25		

Table 16-8a. [Increase in the inequality of country A]
$\lambda_A = \lambda_B = 18$, $f_A(\mu_A = -1.0, \sigma_A = 0.5)$ $f_B(\mu_B = -1.0, \sigma_B = 0.39)$, $\beta_A = \beta_B = 0$

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-77	48	0.46	0.54
0.50	-14	48		
0.90	15	43		
0.99	24	36		

Table 16-8b. [Increase in the inequality of country A]

$\lambda_A = \lambda_B = 18$, $f_A(\mu_A = -1.0, \sigma_A = 0.7)$ f_B $(\mu_B = -1.0, \sigma_B = 0.39)$, $\beta_A=\beta_B = 0$

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-100	56	0.02	0.98
0.50	-16	60		
0.90	30	58		
0.99	41	50		

Table 16-9. [Increase in the relative inequality aversion of country A (maximin)]

$\lambda_A = \lambda_B = 18$, f_A $(\mu_A = -1.0, \sigma_A = 0.39) = f_B$ $(\mu_B = -1.0, \sigma_B = 0.39)$, $\beta_B = 0$

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-100	81	0.83	0.17
0.50	-45	63		
0.90	4	46		
0.99	16	32		

Table 16-10a. [Decrease in the mean productivity and increase in the inequality of country A]

$\lambda_A = \lambda_B = 18$, $f_A(\mu_A = -1.2, \sigma_A = 0.5)$ f_B $(\mu_B = -1.0, \sigma_B = 0.39)$, $\beta_A=\beta_B=0$

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-100	41	0.82	0.18
0.50	-50	42		
0.90	-6	40		
0.99	9	35		

Table 16-10b. [Increase in the mean productivity and increase in the inequality of country A]
$\lambda_A = \lambda_B = 18$, $f_A (\mu_A = -1.2, \sigma_A = 0.7)$ $f_B (\mu_B = -1.0, \sigma_B = 0.39)$, $\beta_A=\beta_B= 0$

F(n)	ATR	MTR	Distribution of aid across countries	
			A	B
0.10	-100	49	0.76	0.24
0.50	-56	55		
0.90	13	54		
0.99	30	48		

From Table 16-7 we see that the optimal share of aid received by country A goes up when its mean productivity goes down. In other words the poorer country should get more aid. This is something we might expect. But the impact of inherent inequality on optimal aid allocation is not quite so obvious. Our numerical calculations produce a striking result. For the parameter values used here, the country with the greater inherent inequality should get less in the way of unconditional aid (see Tables 16-8a and b). What might be an explanation of this result? There are two forces pulling in opposite directions. For any given mean income, greater inequality means the poor are poorer so the need of the country is greater. However, with greater inequality, less of any unconditional aid will get to the poor. These effects are further mediated by the fact that the recipient government is choosing a tax and transfer schedule. Our numerical calculations resolve these forces sharply—the more unequal country should get less aid. However, a country with greater inequality *aversion* should get more aid, as might be expected—this is shown in Table 16-9. Finally, Tables 16-10a and b show cases in which mean productivity decreases and inequality increases simultaneously in country A. We see that for the parameter values used here the former effects dominates, and its share of aid goes up.

5. CONCLUSION

If aid conditionality worked, donors could influence recipients' use of funds to pursue more egalitarian objectives than they otherwise would. But aid conditionality does not seem to work, or at least does not work very well. Thus the only reliable lever donors have is that of augmenting the total resource envelope of recipients and work through the effects of this on their behavior. This paper conducts an analysis of such unconditional aid in the framework of optimal nonlinear income taxation, which takes into account incentive effects on governments of recipient countries in choosing policies, and the incentive effects of their citizens in responding to these policies.

We find that unconditional aid increases overall welfare of the recipient country, which is not surprising. It also increases the welfare of the poorest in the recipient country. More aid induces more progressive tax and transfer systems to be optimally chosen by the recipient government. There is, however, a "crowding out" effect as the extra resources reduce incentives to supply labor in the recipient country—this effect ranges from 1 to 5 percent of GDP.

When there are two potential recipient countries for a given pool of aid, the optimal allocation of aid depends on their relative characteristics. The poorer country should get more aid, as should the country which has higher inequality aversion. What about the effect of inherent inequality? There are two forces pulling in opposite directions. For any given mean income, greater inequality means the poor are poorer so the need of the country is greater. However, with greater inequality, less of any unconditional aid will get to the poor. Our numerical calculations produce striking result. For the parameter values used here, the country with the greater inequality should get less in the way of unconditional aid.

Further research suggests itself in bridging the gap between our assumption of totally unconditional aid, and the opposite assumption of perfect conditionality. In the framework developed here, imperfect conditionality might mean, for example, that the donor could enforce a minimum income level for the poor but could not control the entire tax-transfer schedule. Or we could think of the aid received being split into two categories, one with perfect conditionality and the other with total non-conditionality. The relative proportion of resources going to each could then be varied parametrically. In any event, the brute facts of the failure of conditionality mean that at least some aspects of unconditional aid will have to be introduced into the analysis and design of development assistance.

NOTES

[i] (6) is only a necessary condition for the individual's choice to be optimal, but we assume here that it is sufficient as well. Assumptions that assure sufficiency are provided by

Mirrlees (1976). Note also that while (6) presumes an internal solution for y, (7) remains valid even if individuals were bunched at y=0 since, for them, du/dn=0.

[ii] Equations (8) - (10) lead to the few qualitative conclusions available in this framework (see Tuomala, 1990). It can be shown that the marginal tax rate on income is nonnegative. This is more striking than it at first looks. It may very well be optimal to have the average tax rate less than zero, but it is never optimal to subsidize earnings at margin. An intuition is that it is cheaper to get people to given indifference curve by reducing average rate rather than by exacerbating deadweightloss through distorting their labour supply decisions. It can also be shown that the marginal tax rate is less than one. We also have the famous "end point" results. If wage distribution is bounded above, then the marginal tax rates at the top is zero. If it is optimal for least able individual to work then the marginal tax rate on least able is zero. An intuition behind these endpoint results is that only reason to have a marginal tax rate differing from zero is to raise an average tax rate above that point and lower it below i.e. equity considerations. But at the top is no one to take from and at the bottom there is no one to give to. So at the end points only efficiency considerations matter. Numerical solutions (Tuomala, 1990) have shown, however, that these results have very little practical relevance.

[iii] Tuomala (1990) gives details of the computational procedure.

[iv] For further discussion on the transformation of each individual's utility see Tuomala (1990).

[v] With the utility function we use, there is "bunching"—all those below a critical value of n choose not to work. Their pre tax income is thus zero and their post tax income is whatever the optimal tax and transfer regime gives them. This, then, is the net income of the poorest person, xo, as shown in the tables.

REFERENCES

Aitchison, J. & Brown, J.A.C. (1957): The lognormal distribution with special reference to its uses in economics, Cambridge University Press, 1957.

Berg, E. (2000): "Aid and Failed Reforms: The Case of Public Sector Management," in Finn Tarp (ed.).

Boone, P. (1996): "Politics and the Effectiveness of Foreign Aid," European Economic Review.

Burnside, C. and D. Dollar (2000): "Aid, Policies and Growth," American Economic Review.

Devarajan, S., D. Dollar and T. Holmgren, eds. (2001), Aid and Reform in Africa: Lessons from Ten Case Studies, World Bank, Washington, D.C.

Immonen, R. Kanbur, R., Keen, M. and Tuomala, M. (1998): Tagging and taxing: The optimal use of categorical and income information indesigning tax/transfer schemes, Economica 65, 179-92.

Kanbur, R., T. Sandler and K. Morrison (1999): The Future of Development Assistance, Washington: Johns Hopkins University Press.

Kanbur, R. & Tuomala, M.: Inherent inequality and the optimal graduation of marginal tax rates, Scandinavian Journal of Economics 96 (2), 275-282, 1994.

Kanbur, R. (2000): "Aid Conditionality and Debt in Africa," in Finn Tarp (ed.)

Killick, T. (1995): "Conditionality and the Adjustment-Development Connection," Pakistan Journal of Applied Economics, 11 (1-2), 17-36.

Mirrlees, J.A.: An exploration in the theory of optimum income taxation, Review of Economic Studies 38, 175-208, 1971.

Mirrlees, J.A (1976): Optimal tax theory: A synthesis, Journal of Public Economics 6, 327-58.

Mosley, P., J. Harrigan and J. Toye (1995): Aid and Power: The World Bank and Policy Based Lending, Vol. 1-2, London and New York: Routledge.

Svensson, J. (2000a): "When is Foreign Aid Policy Credible? Aid Dependence and Conditionality", Journal of Development Economics, Vol. 61 (1): 61-84.

Svensson (2000b): "Why Conditional Aid Doesn't Work and What Can be Done About it? Reforming Donor Institutions," Processed, Stockholm: Institute of International Economic Studies.

Tarp, F. (ed.) (2000): Foreign Aid and Development, London and New York: Routledge.

Thorbecke, E. (2000): "The Evolution of the Development Doctrine and the Role of Foreign Aid," in Finn Tarp (ed.).

Tuomala, M.: Optimal income tax and redistribution, Clarendon Press, Oxford, 1990

World Bank (1998): Assessing Aid: What Works, What Doesn't and Why, Oxford: Oxford University Press.

World Bank (1992): World Bank Structural and Sectoral Adjustment Operations: The Second OED Review, Operations Evaluation Department Report 10870, Washington, D.C: World Bank.

Chapter 17

AGRICULTURAL RESEARCH AND POLICY TO ACHIEVE NUTRITION GOALS[i]

Per Pinstrup-Andersen[ii]
Cornell University, pp94@cornell.edu

1. INTRODUCTION

It is a great pleasure to be part of this seminar to honor Professor Erik Thorbecke. Erik has made major contributions to the development and application of the analytical methodology needed to better understand the nature and magnitude of poverty and how economic policy may be applied to poverty alleviation. His work has enhanced our knowledge of the poverty problem, how to analyze the relevant economic issues, and how to help alleviate poverty. As H.E. Babcock Professor, Erik incorporated nutrition and food security concerns into the broader poverty analysis and debate as exemplified by much of his writings including Thorbecke 1982, 1985, 1986a, b, c and d, 1987, 1990, 1991, and 1995. This paper deals with one specific aspect of the malnutrition and poverty problem which has been part of Erik's research portfolio.

Poverty causes hunger and malnutrition and hunger and malnutrition contributes to poverty. While strong links exist with poverty and inequality, hunger and malnutrition are serious public health and development problems in their own right. Agriculture is essential to achieve nutrition goals in the most fundamental sense; it makes the food available without which nutrition goals could not be achieved. This paper goes beyond this fundamental relationship by exploring five related questions:

1. How could agricultural research and policy improve nutrition?
2. Should nutrition goals guide agricultural research and policy?
3. What policy measures are likely to be effective?
4. Are nutrition goals best achieved through pre- or post-harvest changes?

5. Would consumer behavior enhance or reduce the intended effect?

Each of these five questions will be reviewed in turn but first it may be useful to provide a brief overview of the nature and magnitude of the nutrition problems.

2. THE TRIPLE BURDEN OF HUNGER AND MALNUTRITION

Existing hunger and malnutrition may be classified into three related but distinctly different hunger and nutrition problems characterized by the nature of the problem:

1. Energy deficiencies resulting in hunger,
2. specific nutrient deficiencies resulting in hidden hunger, and
3. excessive net energy intake resulting in overweight, obesity, and chronic diseases.

In this paper, I shall refer to these as "the triple burden of hunger and malnutrition". If I had written this paper 20 years ago, I would have limited myself to the first. Fifteen years ago, I would have mentioned only the first two. In view of the epidemic proportions of the increase in overweight and obesity, particularly among low-income people, the time has come—in fact it is overdue—to consider all three sets of nutrition problems in food policy analyses and implementation, including analyses of how agriculture can help achieve nutrition goals.

Infectious diseases caused in part by unclean water, present a fourth burden in that it contributes to the nutrition problem. While this must be recognized in any discussion of the nutrition problem, this paper is limited to the "food side" of the problem. Similarly, the real concern associated with overweight and obesity is the increasing risk of chronic diseases such as diabetes, cancers, and cardiovascular diseases. Those risks are associated not only with overweight and obesity but also with exercise and the intake of specific dietary components such as trans fat, saturated fats, sugars, salt, fiber, fruits, and vegetables.

While individuals in both high- and low-income countries may be affected by one or two of these burdens, developing countries are increasingly being faced with all three. While all three are serious public health problems, they are also very important development problems, causing low productivity of labor, reduced economic growth, poverty, and high and increasing demand for public funds to deal with the resulting health problems. Furthermore, overweight and obesity are rapidly becoming an integral part of poverty—something governments should keep in mind when attempting to deal with poor people's problems. This does not imply that poverty causes obesity. Rather, both are caused by a low level of education and other factors.

About 800 million people suffer from hunger and food insecurity, while one-third of all preschool children in developing countries are stunted. In spite of the agreement entered into at the World Food Summit in 1996 and re-confirmed at the follow-up Summit four years later, by leaders from 186 countries, to reduce by half the number of hungry people from 800 million to 400 million by 2015, overall progress during the 1990's was dismal. Only one-third of the countries managed to reduce the number of hungry people, while about half of the countries saw an increase. China made tremendous progress, reducing its number of hungry people by 80 million during the decade of the 1990's. However, for the world as a whole (excluding China) there was an increase. Thus, excluding China, the world is moving towards more, rather than fewer, hungry people. This makes a mockery of the World Food Summit goal and even the easier attainable Millennium goal of reducing by half the proportion of the world population that suffers from hunger. The rhetoric of the world community and national governments has not been followed up with action.

Of the 800 million hungry people in the world, more than 300 million are found in South Asia and about 200 million in Sub-Saharan Africa. Stunting and underweight in preschool children are particularly widespread in South Asia and to a lesser but very significant degree in Sub-Saharan Africa.

Hidden hunger in the form of deficiencies in iron, vitamin A and zinc, affect close to 40 percent of the population in developing countries and overweight and obesity affect close to one-third of the world population. The number of overweight and obese people is increasing rapidly and more people are suffering from that problem than from hunger and food insecurity.

Increases in overweight and obesity, resulting primarily from excessive net intake of energy, are taking on epidemic proportions. Two-thirds of the United States population and more than half of the populations of several European countries are now overweight or obese. According to the American Obesity Association, there are now approximately 127 million adults in the United States who are either overweight or obese. Of these, 60 million are obese and 9 of them are severely obese, with a body mass index (BMI) above 40. Overweight and obesity is spreading even faster among children and adolescents. The estimated annual deaths in the United States due to overweight and obesity is 300,000 compared to 400,000 for deaths related to smoking (U.S. Department of Health and Human Services, 2001).

The high and rapidly increasing levels of overweight and obesity is not limited to industrialized countries. Middle-income developing countries such as Brazil, Colombia, Thailand, and China are also experiencing relatively high and increasing level. Recent rapid increases in overweight and obesity in China is projected to continue resulting in a situation where about one-third of the Chinese population will be overweight or obese by 2020. The impact on the public health expenditures and economic growth will be very significant. In contrast to hunger, which is closely related to poverty, overweight and obesity is found in all income groups, although the prevalence in most countries for which data are available tends to be higher among low-income people. Furthermore, recent evidence suggests that hunger and under nutrition and overweight and obesity coexist in a surprisingly large share of poor households, often reflected in under nourished children and an overweight or obese adult.

Evidence of negative impact of the triple burden of malnutrition on public health and economic growth is convincing although incomplete. Estimates by Susan Horton (1999) show productivity losses for both hunger and hidden hunger between 5 and 17 percent. The World Health Organization estimates that 15.9 percent of the global burden of child disease is due to under nutrition, the figure being 18 percent for developing countries (Stuart Gillespie and Lawrence Haddad, 2003). The World Bank estimate is 20-25 percent (World Bank 1993). According to research by David Pelletier, et al. (1994), the relative risks of infection increase exponentially as stunting prevalence increases (Gillespie and Haddad, 2003) and malnutrition is associated with an estimated 2.8 million child deaths annually, or 51 percent of all child deaths in nine Asian countries. Iron deficiency anemia is associated with 23 percent of all maternal deaths in these countries, corresponding to 65,000 deaths (Jay Ross and E.L. Thomas 1996). On the basis of findings from a number of studies of the effect of vitamin A supplementation in vitamin A-deficient populations, George Beaton et al. (1993) conclude that mortality in preschool children and pregnant women fell by 27 and 40 percent respectively and malaria attacks decreased by 30 percent. The impact of zinc supplementation on child growth and morbidity is assessed by K.H. Brown, et al. (2002) and R. Black, et al. (1999).

While the link between overweight and obesity on the one hand and chronic diseases and related health costs on the other is strong and well documented, estimates of the economic costs of overweight and obesity are still scarce. The U.S. National Institutes of Health (1998) estimated that the economic cost of health problems related to diet and exercise in the United States was $137 billion annually exceeding the economic costs of smoking ($90 billion) and alcohol abuse ($118 billion).

3. HOW COULD AGRICULTURAL RESEARCH AND POLICY IMPROVE NUTRITION?

There are essentially four ways in which agricultural research and policy could affect human nutrition (Figure 17-1). First, by changing incomes in households with malnourished members (at-risk households) through for example productivity increases in agricultural production, changing employment and wages. Of importance is not only the level of income but also risks of income loss, fluctuations in incomes, the form in which it comes to the household, and who controls it. Agricultural research aimed at the development of technology and production practices that increase productivity per unit of land, water, or labor and/or lower production risks such as drought-tolerant and insect and disease resistant crop varieties could help reduce both chronic and transitory hunger and malnutrition. Productivity increases resulting from research and policy changes are also likely to influence non-farm incomes due to increased demand by farmers.

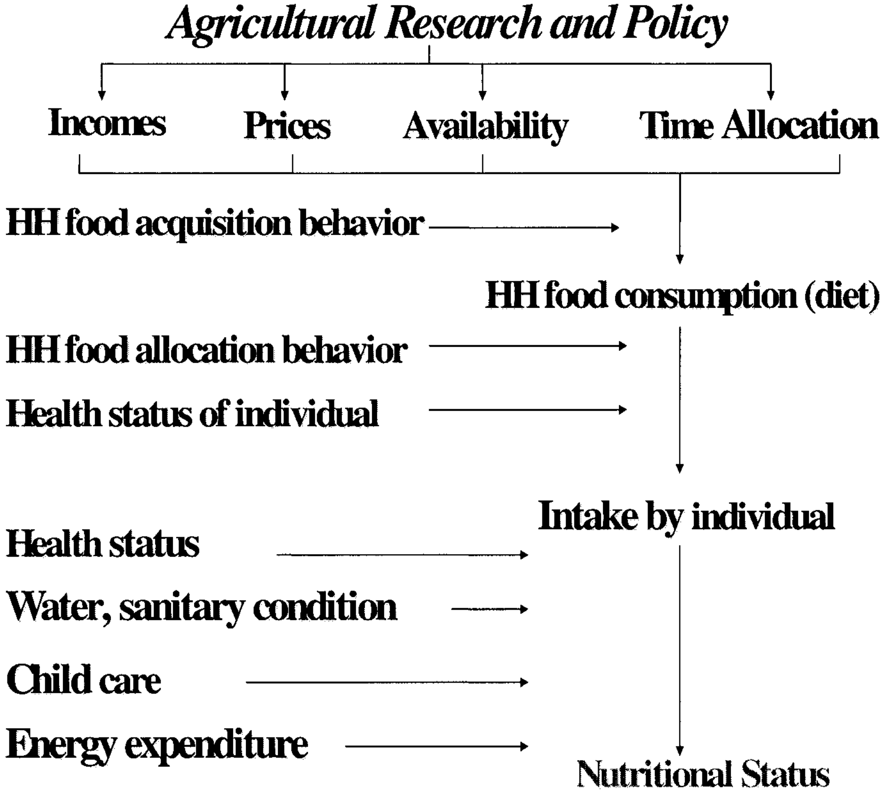

Figure 17-1. [Factors linking agricultural research and policy to nutritional status]

Second, by changing the prices at-risk households receive or pay for food. In addition to the price level, the price relative to other foods and non-foods available to the household and price fluctuations are important. In addition to direct price policy, food prices may be affected by a variety of agricultural and other policies.

Third, by changing the availability of various nutrients either through changes in the portfolio of foods available or by changing the content of individual foods or developing new foods. Commodity priorities in agricultural research and policies may greatly affect the relative quantities of each of the foods being produced. This, in turn, may affect nutrition. For example, the Green Revolution emphasized productivity increases in rice, wheat, and maize while relatively little research was done on livestock, grain legumes and other pulses. The resulting impact on production of each of the various commodities in selected countries and for the developing countries as a whole is shown in Figure 17-2. Research and other action often referred to as the Green Revolution were extremely successful in expanding both productivity and production of the basic cereals but the production of pulses and animal products increased considerably less than population growth during the period under consideration. As a result, cereal prices fell while prices for pulses and food of animal origin increased and low-income people had less access to iron and vitamin A-rich food. Many more people were able to get access to sufficient calories and protein while the intake of iron and vitamin A is likely to have decreased. Rapidly increasing anemia in South Asia during the period supports this notion.

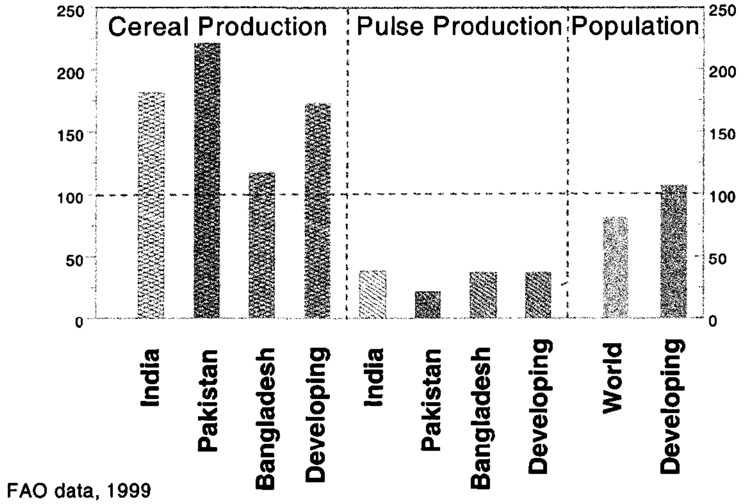

Figure 17-2. [Percent changes in cereal and pulse production and in population between 1965 and 1999]

While the success of the Green Revolution to avoid mass starvation in Asia is not questioned, the above illustrates the potential nutrition benefit of using a more comprehensive nutrition goal than energy as a guide to priority-setting in agricultural research. The nutrient content of the individual foods may be changes by agricultural research. For example, selection and breeding may produce crop varieties with higher content of iron, zinc, and vitamin A. While genetic engineering may be needed in some cases, such as the enhancement of rice with vitamin A, traditional selection and breeding may be the most appropriate approach in most cases.

Many other examples of how modification of the content of specific foods may affect nutrition and health could be mentioned. To mention only a few: orange-colored sweet potato with significantly higher content of vitamin A has been developed and is now on the market. Nutritionally superior oils are currently being bred into staple foods as for example, less saturated fatty acid in canola. Research is under way to improve vitamin E in maize, and various efforts are being made to increase the content of antioxidants in selected foods such as more flavanol and lycopene in tomatoes.

Fourth, agricultural research and policies may influence nutrition through changes in time allocation, child care, workload, gender-specific decision-making power, and energy expenditures of individual members of at-risk households (Figure 17-1). Agricultural projects and policies are often based on the assumption that members of at-risk households have lots of free time on their hands. However, labor-using changes are likely to draw household members away from other productive but poorly remunerated activities such as child care, which may have important effects on nutrition. Furthermore, additional work requires additional energy expenditures and the net increase in energy intake made possible by increasing incomes may be small. This would be particularly pronounced in cases of poorly paid strenuous work.

Changes in any of these four factors can, in turn, affect household food consumption. The actual effect will be mitigated by household food acquisition behavior, which varies across households and groups of households on the basis of preferences of household decision-makers, relative decision-making power of women and men, income levels, location of the household (e.g. urban or rural), and a variety of other factors. The impact on the intake of the at-risk household member will be mitigated by the health status of the individual and the household food allocation behavior, the latter being a function of preferences, level of knowledge, and distribution of decision-making power within the household.

The nutritional status of the at-risk individual will be affected not only by the change in food intake but by a variety of factors such as health status, access to clean water and good sanitation, child care, and energy expenditures. Thus, the final nutrition effect of changes in agricultural research and policy will depend on a series of factors and relationships outside the control of agriculture. It is the interactions between changes in agriculture and other factors such as level of knowledge by the households, health status, and access to clean water and good sanitation, that makes estimates of nutrition effect of changes in agriculture difficult. Focusing agricultural policies on achieving nutritional goals may be successful only if at-risk individuals also have access to clean drinking water and basic primary health care. Household decision-making may enhance potentially positive nutrition effects of agricultural policies, or it may prohibit the benefits from materializing.

4. SHOULD NUTRITION GOALS GUIDE AGRICULTURAL RESEARCH AND POLICY?

The answer depends on answers to several other questions including the following:

1. How important are nutrition goals to society relative to other goals that could be achieved fully or partially by agriculture? The answer to this question is highly subjective and will come from values and political priorities in each particular case. It is important to point out however, that improved nutrition is an investment in human resources that, in turn, generates increasing incomes not only for those whose nutrition improved but for society as a whole. A potential trade-off between food safety and nutrition has developed as an outcome of increasing globalization. High-income countries and high-income consumers demand increasingly higher levels of food safety. This is likely to increase food prices. To the high-income consumer, who may spend less than 10 percent of her income on food, such price increases are not important. For the poor, who may spend 60-80 percent of her incomes on food, even small price increases are important and will have negative nutritional effects. With globalization and associated international trade, identical food safety standards are sought across countries, and the standards of the non-poor are likely to dominate. For the low-income consumer, the increases in food safety may be insufficient to compensate for the higher food price and the fall in access to food. Higher levels of required food safety in rich countries may also harm poor countries by introducing trade barriers.
2. Could nutrition goals be achieved more cost-effectively by other means? In other words, is the cost of achieving a certain nutritional improvement through agriculture higher or lower than the cost of achieving the same

improvement by means of the best alternative? The answer will be case specific and no general answer can be given.

3. Are nutritional improvements achieved through agriculture more or less sustainable than solutions achieved through other means? In some cases, the answer is clear. For example, breeding a higher content of particular micronutrients such as iron, zinc or vitamin A into foods is likely to be more sustainable that a program to distribute vitamin pills.
4. Are there trade-offs between focusing agricultural research and policy on nutritional goals or other goals such as farm or consumer incomes and if so, how large are they? In other words, are nutrition goals compatible with other goals? If nutritional goals are added to or replacing other goals, what would be the benefits foregone and for whom? An illustration of that would be breeding for higher nutrient value in crops at the expense of higher yields. Such trade-offs between quantity of energy produced and the nutritional quality of food are widespread and result in part from consumer demand for inexpensive foods, with only limited willingness to pay more for higher nutritional quality. In some cases, such as current breeding efforts to increase the content of zinc in wheat grown on zinc-deficient soils, there may not be any trade-off—both the nutritional value and yields of wheat increase.
5. Are there tradeoffs between efforts to achieve improvements in hunger, hidden hunger, or overweight and obesity? If so, which of the three nutrition burdens should take priority? To answer this question objectively, a common denominator for improvements in the three would be needed. The only one with which I am familiar is the estimated disability-adjusted life years (DALY), which combines losses from premature death and loss of healthy life resulting from disability in an attempt to estimate the complete burden of a particular disease such that comparisons can be made across diseases (J.L. Murray and A.D. Lopez, 1996).

5. WHAT POLICY MEASURES ARE LIKELY TO BE EFFECTIVE?

As shown in Figure 17-1, policy measures that affect incomes, prices, availability and time allocation may affect nutrition. Furthermore, nutrition may be affected by policies that influence the mitigating factors such as household food acquisition and allocation behavior, health, child care, energy expenditures, and access to clean water and good sanitation.

5.1 Income policies

Poverty is correlated with all three burdens of malnutrition, with the strongest correlation occurring with hunger and to a lesser degree hidden hunger. However, even the prevalence of overweight and obesity tend to be negatively correlated with income level in most societies except for the poorest people and the poorest societies. Correlation does not necessarily imply causation. There is much evidence of a two-way causation between poverty on the one hand and hunger and hidden hunger on the other.

The evidence for causation between poverty and overweight/obesity is less clear. While available evidence indicates a higher prevalence in middle and lower-income households than in higher-income households in both industrialized and developing countries, the prevalence is low in the poorest households. Furthermore, a surprisingly large proportion of lower to middle-income households in developing countries contain one of more undernourished children and one or more overweight or obese adults. A strong correlation has been found between educational level of the head of the household and the prevalence and severity of overweight and obesity. More research is needed to separate effects of incomes from that of education. Relative prices, advertising, and promotion are likely to also play an important role. Again, little research has been done to separate these effects. Without a better understanding of the causation processes that lead to overweight and obesity, it is difficult to design appropriate policies.

Since about 70 percent of the people who suffer from hunger live in rural areas of developing countries, policies to improve rural incomes may offer great potential. To be most effective, specific policies should be tailored to the specific circumstances. In most middle and low-income developing countries, policies and public investment to improve rural infrastructure, domestic input and output markets, farmer access to land and other natural resources, inputs and credit and savings institutions, and non-farm income sources are likely to contribute to reduced hunger. Appropriate trade policies by both developing and industrialized countries are of paramount importance to promote the broad-based economic growth needed to fight hunger.

Existing trade and agricultural policies in the OECD countries make it difficult for developing countries to successfully pursue rapid rural development and broad-based economic growth. These policies are harming the poor and hungry in three ways. First, the agricultural subsidies in the OECD countries are supported by high import tariffs for agricultural commodities and processed agricultural products, which make it very difficult for developing countries to get into OECD markets. Second, surplus production resulting from OECD subsidies are exported to developing countries at prices below costs or dumped on these markets as food aid, with the outcome that poor rural households are unable to compete in their domestic markets, and third, excess production brought about by OECD subsidies linked to quantity produced depress international prices to the detriment of at-risk households in rural areas of developing countries who are trying to escape hunger by producing for the domestic or the international market.

The OECD agricultural and trade policies are also influencing nutrition in the OECD countries themselves although the net effect is not clear. The policies have altered relative prices of foods to both producers and consumers. This, in turn, has changes income distribution and diets. Half of the subsidy cost is paid through government revenues and the other half is paid by the consumers through higher prices. Two of the key culprits in excess net consumption of energy and resulting overweight and obesity are sugar and fats and oils. The agricultural subsidies have more than doubled sugar prices to the OECD consumer. Although the price elasticity may be low in absolute terms, it is safe to assume that the higher sugar prices have reduced sugar consumption with an expected positive nutrition effect. A similar case for a positive nutrition effect could be made for animal fat from beef and dairy products. On the other hand, subsidies tend to be linked to the quantity produced of a few major food commodities. This tends to reduce diet diversity and increase hidden hunger. As indicated by the above discussion, more research is needed to better understand how the OECD policies affect hunger, hidden hunger and overweight and obesity.

In addition to the impact of policies through changes in income levels, the nutritional status may also be affected by policies that alter the composition of income and income control within the household by, for example, increasing cash incomes relative to own production and by transferring more income control to women.

5.2 Price policies

Policies that change relative prices, such as those mentioned above, may be effective in altering diets in nutritionally at-risk households. Low-income households tend to be more responsive to price changes than higher-income households. Furthermore, the price elasticity tends to be larger in absolute value for non-stables such as meats and fruits than for staples such as basic grains. Thus, higher prices for calorie-dense foods such as those containing much sugar and fats imposed through excise taxes, may be more effective to reduce the risk of overweight and obesity among low-income households than among high-income ones. Similarly, consumer subsidies which lower the prices for foods with high density of iron, vitamin A and zinc may help reduce hidden hunger, particularly in low-income people. However, commodity-specific price policies to achieve nutrition goals have their limitations and may be internally contradictory. For example, while meats may contribute to the risk of obesity, they may also help eliminate iron deficiencies. The answer in this case would be to reduce the price for low-fat meats but price policies that specific are difficult to implement.

Reduced prices for basic food staples may be very effective in reducing hunger among the poor. This is so, not because of large price elasticities, but because the poor spend a large share of their income on staple foods. Cheaper staples could therefore increase purchasing power significantly. This is the argument for unit-cost saving productivity increases in agriculture which make it possible for consumer prices to fall without reducing farmers' incomes. As globalization proceeds and more countries open their markets for international prices, the economic gains from such unit-cost savings will be spread among a larger number of consumers and the domestic prices in the country where the savings are made will decrease much less than in a closed economy. This should be beneficial for the farmers but not for the domestic consumer. Since about 70 percent of the hungry are found in rural areas and most depend on agriculture directly or indirectly, the net effect on hunger eradication should be positive. Countries with a very large share of rural net buyers of food who may not benefit from increasing agricultural incomes, such as India and Brazil, may be an exception.

5.3 Policies to change availability

Relative prices may change in response to supply changes. As mentioned above, rapid increases in the productivity and production of wheat and rice and very limited increase in the production of pulses resulted in rather large changes in relative prices of the two groups of foods. Commodity priorities in agricultural research matters greatly for nutrition. A focus on achieving nutrition goals may result in priorities that are different from those most appropriate to achieve production goals. Agricultural research is a very powerful tool to reduce unit-costs of production and production risks and to change the nutrient composition in foods. As already discussed, economic gains from reduced costs may reduce hunger if captured by at-risk groups. Agricultural research offers exciting opportunities for reducing hidden hunger partly by prioritizing foods with high content of iron, vitamin A and zinc to improve the diet and partly by building into the foods a higher content of these micronutrients, the so-called "biofortification". Large variation in the context of micronutrients in staple foods, as illustrated by Welch and Graham (2000) in Table 17-1, provides an excellent foundation for biofortification based on traditional selection and breeding. The new international research program on biofortification led by IFPRI and CIAT offers exciting prospects for making significant nutrition improvements. This is an excellent example of how agricultural research can be focused on solving important nutrition problems.

Table 17-1 [Median and range of concentration of iron and zinc in various food staples]

	Iron		Zinc	
	Median	Range	Median	Range
Rice	3	2-10	16	10-22
Wheat	37	24-61	31	13-68
Maize	20	16-30	21	15-34
Soybean	70	48-110	45	36-70
Bean	-	33-80	-	19-65

Source: Welch, Ross M. and Graham, Robin, D. (2000). A new paradigm for world agriculture: Productive, sustainable, nutritious, healthful food systems. Food and Nutrition Bulletin, Vol. 21, No. 4, Dec 2000, pp. 361-366.

5.4 Policies to alter time allocation

Policies focused on or with implications for the allocation of time by at-risk household members are important for nutrition for at least two reasons. First, energy expenditures may be affected. This could have nutrition effects on individuals at risk of hunger, as well as those at risk of overweight and obesity. Second, time spent on activities particularly important for nutrition such as child care, breast feeding, fetching of water, or food preparation could be affected. While some people at risk of hunger are underemployed, the large majority of adults from at-risk households have little unused time. This is particularly pronounced among women. Therefore, labor-using policies and technologies are likely to reduce the time spent on other activities, some of which may be important for nutrition. Labor shortage in countries faced with high prevalence of HIV/Aids should be of particular concern in the design and implementation of policies and technologies. Strenuous work may result in larger energy expenditures, with negative nutrition effect on those with insufficient net energy balance (those at risk of hunger) and positive effect on those with excessive net energy intake (those at risk of overweight and obesity).

5.5 Policies to influence the mitigating factors

General as well as issue-focused education such as nutrition education may influence the mitigating factors including household behavior. This will be further discussed in a section to follow. Because the nutrition impact of agricultural research and policies will depend on the health of the at-risk individuals and their access to clean water and good sanitation, policies influencing these factors are of critical importance for how changes in agriculture translate into nutritional improvements. Children and adults who suffer from infectious diseases caused by contaminated water, poor sanitation and lack of health care may benefit little if at all from even the most appropriate policy change or technology in agriculture. Similarly, promotion of processed foods and drinks with high content of fats and sugar in populations at risk of overweight and obesity may influence consumer behavior more than well-designed agricultural and food policies. Public and private promotion and distribution of such foods and drinks in schools is a particularly devious activity from a nutrition point of view.

6. ARE NUTRITION GOALS BEST ACHIEVED THROUGH PRE- OR POST-HARVEST CHANGES?

Policies and research can help farmers increase productivity, reduce risks and fluctuations, and change the nutrient composition of the commodity portfolio or the individual commodity. As discussed, these changes are relevant for nutrition. At the post-harvest level, nutritional changes may be pursued through changes in storage, transport, processing, and fortification. Again, policies and research/technology can contribute to these changes. The choice between biofortification and industrialized fortification of food provides an illustration of the choice between pre- and post-harvest changes. Which of the two is likely to be most cost-effective? The answer depends on the specific case. In order for industrialized fortification to reach the at-risk individuals, they must purchase such foods. A large share of those who suffer from hunger and hidden hunger consume foods produced on their own farms or they buy foods from local producers and traders that are produced locally. Foods fortified by industry will not reach them. Biofortified food will, if it is grown locally.

Of course, the difference is not as sharp as outlined above. Post-harvest fortification can be undertaken on a small scale at the village level. The question then becomes one of relative costs and sustainability. If biofortified crops yield less and local post-harvest fortification is inexpensive, the latter may be preferable. Biofortification is likely to be more sustainable because once introduced into the seeds, it need not be replenished. Post-harvest fortification depends on on-going activity. If the electricity fails or funds for fortification run out, so does the fortified food. Consumer acceptance is also important. Golden rice, for example, which is a biofortified rice with high content of beta carotene, takes on a yellowish color. It is still unclear how that would influence consumer acceptance.

7. WOULD CONSUMER BEHAVIOR ENHANCE OR REDUCE THE INTENDED EFFECT?

One reason why people are malnourished is that they do not prioritize good nutrition above all else. A number of competing desires and preferences as well as the level of knowledge and the exposure to advertisement and promotion influence consumer behavior, the resulting food demand, and the diet that eventually materializes. If farmers organize their production in response to nutritional needs of consumers instead of the economic demand, they would probably go broke. This limits the extent to which agriculture can help solve the triple burden of malnutrition.

Consumers generally do not demand nutrients. They demand foods with the desired characteristics, including possibly nutritional qualities. Foods with better nutritional qualities may be undesirable to the consumer, particularly if the price is higher. But all three types of malnutrition carry high social costs in terms of low productivity and high health costs and the socially preferred policy approach would be to move economic demand and nutrition needs closer together. If social costs exceed private costs for those at risk, policies are needed to provide incentives for at-risk individuals to modify their behavior. Such policies could be product-specific price subsidies or excise taxes or public investment in agricultural research that would help combine characteristics desired by the consumers with nutritional qualities desired by society. School lunch programs should try to combine nutrition quality with food characteristics preferred by school children instead of serving as a dumping ground for surplus foods with high fat and sugar content and a place for the promotion of drinks with a large content of sugar. Strong economic interests, some of which may be in agriculture, may work against efforts to improve nutrition.

If, on the other hand, perceived private costs are below actual private costs, the relevant policy would be better education and information for the consumers. This might include clear messages about the health risks and labeling of foods for nutritional content. In countries where a large and increasing share of the food is consumed away from the home, such as the United States, such labeling should be enforced also in restaurants and canteens. Some fast food chains are beginning to do that and at the same time making available lower energy foods.

In high-income countries, such as those of the European Union and the United States, a dual consumer behavior is developing. An increasing, although still small, proportion of the consumers are questioning the wisdom of a continuation of further cost reductions by means of larger scale production and marketing units resulting in what they see as increasing risks, poor animal welfare, and lower quality of the food supply. Increasing demand for organically produced food, locally produced foods, and what is referred to as "natural foods" along with requests for identity preservation, traceability of food and direct trade between farmers and consumers through farmers' markets illustrate the developments in this area. At the same time, the large majority of consumers continue to seek low-priced food pushing farmers and the marketing sector to further reduce unit-costs. How these two different tracks in the food system are going to develop both in high- and low-income countries and how they influence future international trade and nutrition is not clear at this point but agriculture should be prepared to meet both sets of demand, whether they are justified on nutritional grounds or not.

8. CONCLUDING COMMENTS

The triple burden of malnutrition is causing very serious human misery to many millions of individuals and the economic and health costs to societies are high. Agricultural research and policy can be an important part of the solution but it can also be part of the problem. The key is to design policies and research priorities to fit the particular circumstances rather than seeking policies that fit all. Incorporating nutrition goals into the decision-making process can make these policies more effective in solving the nutrition problems but the impact will be limited by other barriers to good nutrition such as contaminated water, poor health and sanitation, vested economic interests, and consumer behavior that may not prioritize good nutrition. With 5-10 million preschool children dying of hunger and malnutrition every year, with 800 million people suffering from hunger, with every third preschool child in developing countries being either stunted or underweight or both, with almost one-third of the world population suffering from micronutrient deficiencies, and with the prevalence of overweight and obesity expanding at epidemic rates, already reaching about a billion people, it is time to act. Agriculture should do its part.

NOTES

[i] Paper prepared for "Poverty, Inequality and Development: A Conference in Honor of Erik Thorbecke, Cornell University, October 10-11, 2003.

[ii] H.E.Babcock Professor of Food, Nutrition, and Public Policy, Cornell University and Professor of Development Economics, The Royal Veterinary and Agricultural University, Copenhagen.

REFERENCES

Agricultural Biotechnology in Europe. 2003. "Future Developments in Crop Biotechnology", *Issue Paper 6.*

American Obesity Association. 2003. "Obesity in the U.S.". In *AOA Fact Sheets*, (www.obesity.org/subs/fastfacts/obesity-US.html)

Beaton, G., R. Martorell, K.J. Aronson, B. Edmonston, G. McCabe, A.C. Ross and B. Harvey. 1993. "Effective vitamin A supplementation in the control of young child morbidity and mortality in developing countries", *ACC/SCN State-of-the-art Series*, Nutrition Policy Discussion Paper No. 13, Geneva: UN

Black, R., Z. A. Bhutta, K. H. Brown, M. Gardner, S. Gore, A. Hidayat, F. Khatun, R. Martorell, N. X. Ninh, M. E. Penny, J. L. Rosado, S. K. Roy, M. Ruel, S. Sazawal and A. Shankar. 1999. "Prevention of Diarrhea and Pneumonia by Zinc Supplementation in Children in Developing Countries: Pooled Analysis of Randomized Controlled Trials." *Journal of Pediatrics*, 135, pp. 689-697.

Brown, K. H., J. M. Peerson, J. Rivera and L. H. Allen. 2002. "Effect of Supplemental Zinc on the Growth and Serum Zinc Concentrations of Prepubertal Children: A Meta-analysis of Randomized Controlled Trials." *American Journal of Clinical Nutrition*, 75, pp. 1062-1971.

FAO. 2003. "Assessment of the World Food Security Situation." Committee on World Food Security, 29th Session, Rome, Italy, May 12-16, 2003.

FAO. UNICEF, UN Population Fund and the Obesity Task Force.

Gillespie, S. and L. J. Haddad. 2003. *The Double Burden of Malnutrition in Asia: Causes, Consequences and Solutions.* New Delhi and London: Sage Publications.

Haddad, L. 2000. "A Conceptual Framework for Assessing Agriculture-nutrition Linkages." *Food and Nutrition Bulletin*, 21:4, pp. 367-373.

Horton, S. "Opportunities for Investments in Nutrition in Low-income Asia." *Asian Development Review*, 17:1-2, pp. 246 273.

Levin, C., J. Long, K. R. Simler and C. Johnson-Welch. 2003. "Cultivating Nutrition: A Survey of Viewpoints on Integrating Agriculture and Nutrition." IFPRI FCNP Discussion Paper No. 154, July. Washington D.C.: IFPRI

Murray, C. J.L. and A. D. Lopez. 1996. (Eds.), The Global Burden of Disease: Vol. I. Geneva: World Health Organization.

Pelletier, D. L., E. A. Frongillo, Jr., D. G. Schroeder and J. P. Habicht. 1994. "A Methodology for Estimating the Contributions of Malnutrition to Child Mortality in Developing Countries." *Journal of Nutrition*, 124. pp. 2106-2122.

Pinstrup-Andersen, P. and R. Pandya-Lorch, eds. 2001. "The Unfinished Agenda: Perspectives on Overcoming Hunger Poverty, and Environmental Degradation." Washington D.C.: IFPRI.

Ross, J. S. and S. Horton. 1998. *Economic Consequences of Iron Deficiency. Micronutrient Initiative*, Ottawa: Micronutrient Initiative

Ross, J. S. and E.L. Thomas. 1996. "Iron Deficiency Anemia and Maternal Mortality." *Profiles 3 (Working Notes Series No. 3).* Washington D.C.: Academy for Educational Development.

Suharno, D., C. E. West, Muhilal, D. Karyadi, J. GAJ Hautvast. 1993. "Supplementation with Vitamin A and Iron for Nutritional Anemia in Pregnant Women in West Java, Indonesia." *The Lancet*, 342:8883), pp. 1325-1328.

Thorbecke, E. 1995. "Health, Nutritional and Demographic Trends." *The Impact of Scientific Advances on Future Health* (WHO CIOMS Colloquium), pp. 75-87. World Health Organization, Advisory Committee on Health Research.

Thorbecke, E. and S. S. Kyereme. 1991. "Factors Affecting Food Poverty in Ghana." *Journal of Development Studies*, 28:1, pp. 39-52.

Thorbecke, E. and D. Berrian. 1990. "The Impact of Structural Adjustment Policies on Poverty and Nutrition Analyzed Within a General Equilibrium Framework," in *Policy Reforms, Poverty, and Nutrition.* Per Pinstrup-Anderson, Ed. The Cornell University Food and Nutrition Program Monograph 3, pp. 49-71.

Thorbecke, E. and S. S. Kyereme. 1987. "Food Poverty Profile and Decomposition Applied to Ghana." *World Development.* 15:9, pp. 1189-1199.

Thorbecke, E. and J. Greer. 1986a. "A Methodology For Measuring Food Poverty Applied to Kenya." *Journal of Development Economics*, 24:1, pp. 59-74.

Thorbecke, E. and J. Svejnar. 1986b. "Economic Policies and Agricultural Performance: The Case of Nepal 1960-1982. Paris: The OECD Development Center, p. 167.

Thorbecke, E. and J. Greer. 1986c. "Food Poverty and Consumption Patterns in Kenya," p.170. Geneva: International Labour Office, World Employment Programme.

Thorbecke, E. and J. Greer. 1986d. Food Poverty Profile Applied to Kenyan Smallholders." *Economic Development and Cultural Change*, 35:1, pp. 115-141.

Umeta, M., C. E. West, J. Haidar, P. Deurenberg, J. G.A.J. Hautvast. 2000. "Zinc Supplementation and Stunted Infants in Ethiopia: A Randomized Controlled Trial." Lancet, 355:9220, pp. 2021-2026.

UNICEF. 2003. "The State of the World's Children," http://www.unicef.org/sowc03.

U.S. Centers for Disease Control and Prevention. 2002. http://www.cdc.gov/nccdphp/Dnpa/obesity/epidemic.htm.

U.S. Department of Health and Human Services. 2001.

U.S. National Institutes of Health. 1998. Report submitted to the U.S. House of Representatives Committee on Appropriations.

Welch, R.M. and R.D. Graham. 2000. "A New Paradigm for World Agriculture: Productive, Sustainable, Nutritious, Healthful Food Systems." *Food and Nutrition Bulletin*, 21:4, pp. 361-366.

Wolf, A.M. and G.A Colditz. 1998. "Current Estimates of the Economic Cost of Obesity." *U.S. Obesity Research*, 6:2, pp. 97-106.

World Bank. 1993. "World Development Report: Investing in Health." New York: Oxford University Press.

World Health Organization. 2002. http://www.who.int/nut/db_bmi.htm.

Chapter 18

IS DUALISM WORTH REVISITING?

Gustav Ranis[*]
Yale University, gustav.ranis@yale.edu

1. INTRODUCTION

In the 1950's and 1960's, a neglected sub-field of economics was rediscovered: development economics, concerned with describing the economies of the world's poor countries, which even then accounted for more than half of the world's population. The economic problems facing these countries—many of them newly independent—were clearly acute and their development was seen as an urgent priority for both analysis and policy action.

Available economic models, however, seemed to offer only limited insights into the practical problems facing these countries. These were not modern industrial economies; most people worked not in factories but in subsistence agriculture or in urban informal jobs—from shoe shining to small-scale peddling to distributive trades and artisanal crafts. The dominant one-sector macro models of the day, from Keynesian to Harrod-Domar to Solow, seemed to have limited relevance for societies not primarily concerned with business cycle or steady state properties. Most contemporary growth models were seen as academic abstractions with little policy relevance. And the dominant assumptions of neoclassical micro theory—full employment, market clearing and perfect competition—seemed to have little relevance for the segmented commodity, labor, and credit markets of poor countries.

Against this backdrop, the concept of dualism attracted considerable attention. Sociological dualism, associated with the name of Boeke,[i] emphasized differences between Western economic and non-Western cultures and objectives. Technological dualism, emphasized by Higgins[ii] and Eckaus,[iii] focused on the difference between variable factor proportions in traditional and fixed coefficients in modern sectors. A third, and undoubtedly dominant strand, focused on the coexistence of sectors which are basically asymmetrical—and thus dualistic—in some key economic dimension.

Undoubtedly the first clear manifestation of this third version of dualism appeared in the tableau economique of the physiocrats. Although the emphasis there was on one preponderant "productive" sector, agriculture, the physiocrats also clearly envisioned its coexistence with a small non-agricultural (to them "non-productive") sector, providing services, artisanal goods and other requirements of the ruling nobility—if the "produit net" of the soil (read agricultural surplus) was large enough to permit some labor reallocation. This essentially circular flow mechanism may yield slow increases in real per capita income over time as the productivity-enhancing results of "father teaching son" plus inter-regional specialization and an enhanced division of labor lead to increases in agricultural productivity. But these represent limited amendments to what is basically a static situation, with the "serfs" in the system, mostly in agriculture, some allocated outside, continuing to maintain a virtually constant consumption standard.

Physiocratic dualism, which emphasized the fundamental primacy of agriculture, the importance of an agricultural surplus and the long run prognosis of stagnation, gave way to the concept of classical dualism, more or less coincident with the advent of the so-called industrial revolution in Western Europe. This classical concept á la Ricardo[iv] (1815) focused on the coexistence of still overwhelmingly dominant agricultural activities subject to diminishing returns to labor on the basically fixed land—and non-agricultural activities, growing as a consequence mainly of the accumulation of fixed capital. While the classical school did not model the interactions between these two sectors, it is clear that the main fuel for the reallocation of workers and the accumulation of industrial capital was seen as coming from the "profits" of agricultural capitalists, i.e., the agricultural surplus left over after agricultural workers and landlords (who were assumed to consume everything) had been paid off. It should be noted that the classicists also introduced the related assumption of the near fixity of land combined with Malthusian population pressures and that they retained the notion of an institutionally determined real wage in agriculture—even though the laboring class was now free and could bargain with the capitalists in setting the level of that wage.

While the classicists differed amongst themselves with respect to their overall prognosis for the dual economy as a whole, Ricardian-Malthusian pessimism with respect to the agricultural sector's ultimate stagnation was a dominant feature of their overall analytical work. In the absence of marked technology change, either generated within agriculture or via modern inputs from non-agriculture, agricultural stagnation and thus the drying up of the needed agricultural surplus represented the dominant long-term outlook. Whether technology change and the exploration of economies of scale within the industrial sector, reflecting Smith's optimism,[v] would be sufficiently strong to provide enough industrial profits to rescue the situation remained controversial.

It was, of course, Arthur Lewis who in his famous 1954 article[vi] built on some of the main ingredients of this classical tradition, leading him to emphasize dualism in labor markets, i.e., a competitive wage in non-agriculture but tied to a wage in excess of a very low, if not zero, marginal product in agriculture. Lewis, moreover, found himself allied with Smith, seeing the relatively small non-agricultural or commercialized sector as dynamic and expanding, fed by the mobilization of the "hidden rural savings" of Nurkse[vii] and Rosenstein-Rodan,[viii] enabling the reallocation of workers into higher productivity activities, while wage levels were kept relatively low, at a modest "hill" over the agricultural wage. This reallocation process would continue until all the "surplus labor," i.e., all those whose remuneration exceeded their marginal product, had been reallocated, marking a turning point at which dualism atrophies and the economy becomes fully commercialized or neo-classical.

In Section II, we will trace the further development of the dual economy model in the Lewis tradition. In Section III we will examine the criticisms which have been leveled against it by the neo-classical school, differentiating between what may be referred to as popular "red herring" attacks and more central critiques. Section IV will inquire into the current normative usefulness of the concept for both analytical and policy purposes. Section V concludes.

2. THE MODERN DIMENSIONS OF DUALISM

We will concentrate here on dualism in the labor market and, for reasons of convenience, restrict ourselves to the closed economy—relevant to all but very small economies.[ix] Dualism, of course, matters only when an economy's agricultural sector is initially relatively large and represented substantially by extended family or collective farm cultivation, often labeled "subsistence agriculture," in contrast to commercialized or plantation agriculture.

The critical organizational characteristic of this sector is that, at a given technology, the man/land ratio is such that marginal labor productivity is very low (if not zero or negative), while key decision makers such as heads of families, village elders or commune leaders feel obliged to share output with all members of these groups, even if such shares exceed the marginal product. In other words, income is shared, or an institutional wage is determined, based on bargaining rather than neo-classical principles. Moreover, the gap between this institutional wage in non-commercialized agriculture and the commercialized non-agricultural wage is likely to be in excess of the modest 50% "hill" asserted by Lewis and is materially affected by such institutional urban formal sector interventions as minimum wage legislation, union pressure and government wage setting. Extending the Lewis model further, there are likely to exist similar configurations in the so-called urban informal sector, with family enterprises lacking sufficient cooperating capital and forced to pursue service sector and distributive trade activities which yield low marginal labor productivities but are again characterized by a similar pooling of income in an extended family context. In both cases everyone "in the same kitchen" is fed and average, rather than marginal, product is relevant to the size of the shares.

In addition to the organizational dimensions of dualism, emphasized by Lewis, there is a product dimension, focused on the exchange between food produced by the peasant agricultural sector and the non-agricultural goods produced in the urban (and/or rural) commercialized non-agricultural sector. The inter-action between these two sectors, extending beyond the inter-sectoral labor market into the inter-sectoral commodity and financial markets, was fully analyzed by Fei and Ranis.[x] The key point here is that agricultural and non-agricultural products cannot readily be substituted for each other; in the closed economy food-producing agriculture becomes a necessary condition for industry, while the converse does not hold. Consequently, if agriculture lags behind non-agriculture during the labor reallocation process the deterioration of the non-agricultural sector's terms of trade may well cause a rise in the agricultural wage, as well as in the related unskilled non-agricultural real wage, and a consequent retardation if not collapse of the labor reallocation process long before the labor surplus has been eliminated. This "turning point," in contrast to Lewis', signals the relative inability of the system to follow a "balanced growth" path according to which agricultural productivity keeps pace with advances in non-agriculture and inter-sectoral commodity as well as financial and labor markets can clear at given terms of trade.

Other modifications of the basic Lewis model can be found in Harris-Todaro[xi] and Fields.[xii] While Harris-Todaro's main innovation was to introduce the notion that labor reallocation is affected not only by the inter-sectoral wage gap but also by the probability of obtaining a formal sector non-agricultural job, they accepted the notion of an institutional impact on the level of the non-agricultural urban wage but insisted on an equally competitive, i.e., neo-classical, agricultural wage. Fields, however, pointed out that the two choices for migrants offered by Harris-Todaro, a formal sector job or open urban unemployment, needed to be amended by introducing the urban informal sector. Indeed, just as in agriculture, very few urban residents can afford to be openly unemployed and rely on non-existent unemployment insurance. Instead, they fall back on the family, while working at very low levels of productivity, i.e., they are the urban underemployed.

Finally, Ranis and Stewart[xiii] further amended the basic dualistic model by suggesting the need to trace the interactions among four sectors once we further disaggregate the urban informal sector between a non-traditional dynamic sub-sector, with sub-contracting ties to the urban formal sector, and a traditional static sub-sector, serving as a low productivity sponge. The very meaning of "dualism" of course becomes somewhat tenuous as we move away from the basic two-sector Lewis model but, even though inter-sectoral relations become increasingly mind-blowing to trace, the basic asymmetries in labor market behavior remain critical for both analytical and policy purposes.

2.1 Neo-Classical Critiques and Responses

Dualism has been subjected too much criticism and attack over the past several decades. It has virtually disappeared from contemporary development discourse in the OECD countries, except via textbooks including intellectual history. Some of this may be termed not so much malign neglect as a consequence of the development sub-discipline, along with the rest of economics, moving away from grand macro-theorizing to a micro-econometric focus. But much of the attack has also focused quite specifically on the theory of dualism because of its "unacceptable" assumptions about labor market behavior. In spite of its apparent real world empirical relevance—on which more below—the critics reject out of hand any bargaining outcome which cannot be modeled precisely within a neo-classical framework.

Some of the specific critiques on record may be viewed as "red herrings" and can be readily responded to; others are more profound and require more careful consideration. Perhaps the leading attack in the "red herring" category was occasioned by the unfortunate choice of the "labor surplus" term deployed by Nurkse, Lewis, Fei-Ranis and others. This was widely interpreted as implying a zero marginal product in agriculture and led to the famous T.W. Schultz/Sen exchange[xiv] on whether or not a reallocation (or in this case demise) of part of the agricultural labor force could be expected to leave agricultural output unaffected. This misperception arose due to incautious asides in Lewis' 1954 paper and the resort to mathematical convenience in Fei/Ranis' 1961 and 1964 contributions. The basic point is that the marginal product is low, and sufficiently low to fall below the bargaining wage or income share. As Lewis put it in his 1972 retrospective piece[xv] "whether marginal productivity is zero or negligible is not at the core of fundamental importance to our analysis...this has led to an irrelevant and intemperate controversy." But the controversy has persisted. Otsuka in his review of Fei/Ranis' 1997 book[xvi] considers the zero marginal product notion as deeply embedded in the dual economy model but acknowledges that such a stark assumption is both empirically unlikely and theoretically unnecessary. What is necessary is that, during any short period of time, there exists an excess supply of labor at the going wage.

Does this mean that the reallocation of labor necessarily causes a food shortage, or the arrival of the Fei-Ranis "shortage point?" Again, as Fei-Ranis took pains to point out, a withdrawal of labor is highly likely to lead to a simultaneous reorganization of agricultural production, in effect an upward shift in agricultural labor productivity, permitting the maintenance, if not even an increase, of the agricultural surplus available for transfer to non-agriculture. Thus, the inter-sectoral terms of trade need not deteriorate against industry—unless, of course, there is a marked relative neglect of agriculture in the "balanced growth" context.

A second critique, again in the "red herring" category, dealt with the level of the exogenously given institutional wage. The neo-classical school, of course, fundamentally rejects the notion of an institutional or bargaining wage since it cannot be deduced from basic principles. But the fact that wages are, in fact, observed as not constant over time can be easily disposed of. Otsuka,[xvii] for example, claims he has "never encountered institutionally determined rigid wage rates in agrarian communities". This ignores the fact that it is the sharing rule, not the level of a wage, which may well vary over time, which is at stake. The dual economy assumption is that agricultural wages are related to, but not necessarily equal to, the average product of agricultural workers, since the head of the household, or whoever else commands the agricultural surplus, is bound to retain a portion for her own reinvestment purposes. As the average product rises with technical change, the agricultural bargaining wage is also likely to change. Thus, over time, we are likely to see a gently rising, not a horizontal, Lewis-type supply curve of labor. Over short periods of time it is horizontal; but what we have over longer periods is a step function made up of annual unlimited supply of labor segments, econometrically undistinguishable from a gently rising supply curve.

Let me cite, in passing, a few other criticisms of some assumptions of the dualism model, all of which have some validity but none of which are critical one way or the other. They include the adoption of the classical assumption that all wages are consumed and all profits are saved; that the system is savings-pushed, i.e., that Say's Law holds; and that all investment funds are allocated to the commercialized or non-agricultural sectors. There would be absolutely no problem in modifying any of these simplifying assumptions without any damage to the basic dualistic model.

Let us then turn to the crux of the critique, the rejection of a bargaining wage or consumption share exceeding the marginal product of labor at any point in time. I assume it is not difficult to see that in the kind of setting under discussion, i.e., extended family or other communal institutional arrangements, the unfavorable ratio of people to cooperating factors as part of the initial condition is not something under the control of decision makers, and that such decision makers cannot simply fire low productivity members to reach a neo-classical equilibrium or refuse to share much of the group's income with them. Fafchamps[xviii] provides an overview of the principles underlying the resulting "solidarity network" among peasants, as depicted in the anthropological evidence of Geertz[xix] and Scott.[xx] Ishikawa,[xxi] an astute long-time observer of Asian economic development, endorses the concept of a "minimum subsistence level of existence" (MSL), one version of the institutional real wage. His work indicates the prevalence of a "community principle of employment and income distribution which promises all families.....an income not less than MSL." Hayami and Kikuchi[xxii] find that in Indonesia "wages do not adjust on the basis of labor's marginal product, but according to the subsistence requirements of the time and social conventions." Only over time is there a tendency to adjust but even then it does not necessarily occur by altering wages to equal the marginal product, which could reduce the wage below subsistence. Instead, in Java harvest contracts began to include weeding duties without a complementary rise in the wage rate, thereby not threatening the MSL but moving institutionally towards equilibrium. Osmani[xxiii] presents a model of downward rigidity of the sharing rule insisted on by the workers themselves. Current work in what is called behavioral economics may also prove to be of help in developing a theoretical structure to rationalize cross-worker subsidization in the absence of assured reciprocity—especially as some members of the group are likely to be leaving agriculture over time.

Finally, we have ample historical evidence, e.g. for England from 1780 to 1840,[xxiv] for Japan from 1870 to 1920,[xxv] for Taiwan from 1950 to 1970,[xxvi] of labor abundant agriculture witnessing hefty increases in average agricultural labor productivity, while the agricultural wage rises only gently, i.e., lags substantially behind, until the commercialization or Lewis turning point is reached. As Sen[xxvii] has pointed out, even a horizontal supply curve of labor can be made consistent with a neo-classical explanation; but you have to work hard to make the pre-conceived theory fit the facts. And these facts are also fully consistent with an institutional wage which is gently rising as a result of the step function process previously described until, as in the country cases cited above, the turning point is reached and wages begin to rise steeply in concert with rising marginal productivity. But before that point is reached, a rising gap between agricultural productivity and wage levels is certainly not consistent with neo-classical assumptions about market clearance.

Rosenzweig[xxviii] and others have presented micro-econometric evidence of steeply rising labor supply curves in a cross-section in heavily populated agricultural settings such as India and claim that this puts the final nail into the coffin of the classical dualistic model. However, this is an expressly static relationship and does not address a developing economy's dynamic transition process. Moreover, as we have shown elsewhere[xxix] we would expect individual family labor responses to be quite inelastic at any point in time. Faced with hypothetical wage changes, hard-working, if not highly productive, agricultural family workers are not likely to have much room for trading off leisure for additional work. Rosenzweig's findings are inherently reasonable but they address a different issue. He is concerned with the cross-sectional labor/leisure decisions across agricultural households, while dualistic models are interested in the conditions governing inter-sectoral labor reallocation in a time series context.

3. DOES THE DUAL ECONOMY MODEL STILL SERVE A USEFUL PURPOSE?

Even if it were accepted that dualism is a useful construct for understanding the historical development experience of a number of countries, including England, Japan, and Taiwan, among others, does it have any empirical relevance for today's world; and, even more importantly, how does it relate to contemporary theorizing in the "new growth theory" or "new institutional economics" traditions?

We would contend that China, India, Bangladesh, as well as much of Central America and some portions of South America, comprising a majority of the people on earth, still meet the initial conditions for dualism relevance, i.e., a substantial food producing agricultural sector marked by heavy population pressure on scarce land, complemented by large urban informal sectors. And, while Sub-Saharan Africa was once described as "land surplus," there is increasing evidence, from high fertility rates, changing cultivation practices and a general reduction of fallow periods, that this pivotal region is also moving in the same direction. In these settings the issue addressed with the help of the dualism model, i.e., how to mobilize an agricultural surplus by reallocating an underemployed labor force into efficient non-agricultural pursuits, remains at the top of the development agenda.

The dualistic theory nexus, moreover, remains useful for a number of analytical reasons, including concerning the relationship between growth and the distribution of income, for the determination of the domestic inter-sectoral terms of trade, as well as for the choice of technology and of the direction of technology change. The effort to understand the relationship between growth and the distribution of income was first brought to our attention by Simon Kuznets in 1955.[xxx] While Kuznets emphasized the structural change an economy experiences as it shifts from initially dominant A (agriculture) to M (manufacturing) and S (services) sectors over time, his explanation of the famous inverted U-shaped pattern between growth and distribution embraced the dualistic model. The basic cause for the initial worsening of distribution was the reallocation of workers from a more equally distributed agricultural sector to a less equally distributed non-agricultural sector--with wages kept relatively low and savings rates rising--while the eventual improvement of equity is related to the upswing of real wages as full employment is reached everywhere. Bourguignon and Morrison[xxxi] see “the persistence of economic dualism as a powerful explanatory factor of cross-country differences in inequality.” While work by Fei, Ranis, and Kuo,[xxxii] Fields[xxxiii] and others has shown that no inevitability should be attached to the suggested inverse U-shaped path, it is clear that the nature of the growth pattern viewed in an expressly dualistic context determines its relationship to equity over time, differing markedly between the period before and after the commercialization point.

This concern about the relationship between growth and equity eventually spilled over into a focus on what is happening to poverty levels.[xxxiv] The notion of only gently rising real wages in both agriculture and non-agriculture during the pre-commercialization epoch affects not only labor, commodity and financial exchanges between the sectors but also has a definite impact on technology, favoring labor-intensive technology choices statically and labor-using technology choices dynamically. The reversals in these dimensions once a system enters the one-sector neo-classical world has also been documented.[xxxv] In all these instances the acceptance of initial asymmetry between sectors rather than the neo-classical assumptions of full employment and smooth homogeneity can prove helpful to open-minded analysts.

The pattern of the inter-sectoral terms of trade provides an indication of whether or not the dual economy has managed to maintain “balanced growth” and avoided the relative agricultural neglect much discussed in the literature. Finally, we should also note the relevance of dualism for contemporary mainstream development models. The "informal insurance" mechanism of Townsend[xxxvi] for example, by which farmers smooth consumption by insuring each other across space is not radically different from the aforementioned "moral peasant" of Scott[xxxvii] (1976) who is concerned with supporting others over time as well as space. Whether this can all still be forced into a comforting neo-classical bottle or approaches institutionalized altruism, of course, remains a point of contention. In the former, income is allocated *ex post* after neo-classical distribution rules are observed, whereas, in the latter, income is divided *ex ante* among the members of the extended family or wider community. It is unclear whether the policy implications for achieving a successful transition to modern growth differ widely depending on which concept is deployed. But what remains relevant is which model fits better the basic empirical reality of wage behavior relative to agricultural productivity change in successful countries; which is better suited to analyze agricultural neglect in failure cases; which provides a better explanation of the marked rise in a system's savings rate; which is more capable of explaining discontinuities in income distribution and technical choice--one that assumes full employment and smooth neo-classical equilibrium everywhere or one that recognizes initial underemployment and disequilibrium in the system en route to a one sector modern growth epoch.

4. CONCLUSION

The interaction between the agricultural and non-agricultural sectors resides at the heart of early stage development in many developing countries. We have in this paper argued that dualism, especially focused on its labor market dimension, continues to offer a theoretically valid, empirically relevant, and practically useful framework for dealing with this issue.

After briefly tracing the intellectual history of dualism, we cited and responded to various neo-classical critiques, distinguishing between those which can be characterized as of the "red herring" or straw man variety and those which are more fundamental and thus need to be seriously addressed.

The key assumption at issue is the acceptance or rejection of an institutional or bargaining real wage or income share as an imposed initial condition for an agricultural sector composed largely of owner/cultivators and, given very high man/land ratios, resulting in low levels of marginal productivity. A wage above this marginal product agreed to by the family, the community or the commune, whatever the organizational configuration--and whether in agriculture or in informal urban sector activities--is likely to be related but not equal to the average product--and that is difficult for adherents to the dominant neo-classical school to swallow. We have tried to respond to various criticisms, including the interpretation of surplus labor as zero marginal product labor, instead of simply indicating the existence of disguised unemployment, i.e., income or consumption shares exceeding the marginal product. We have characterized the supply curve of unskilled labor in agriculture and the informal urban sector as a step function composed of horizontal portions, each indicating the "unlimited" availability of labor for commercialized sector absorption over any short period of time. Concerning the more serious challenge to the dualism model, the finding of an inelastic supply curve of agricultural labor, our response was two-fold: that we see no inherent conflict between cross-sectional micro-econometric findings in the neo-classical tradition and dualism's effort to trace the dynamic time series of the inter-sectoral reallocation process at the macro level over time; second, that we find it not surprising that individual agricultural workers, even if underemployed due to the lack of cooperating factors, work long hours and have very little leisure left to surrender in response to a higher wage.

We, finally, inquired into the usefulness of the dualism concept for explaining both historical and prospective country development experience, as well as its relevance in general to contemporary economic modeling efforts. We found that the basic dualism model well fits the historical experience of such countries as England, Japan, and Taiwan and believe it is likely to continue to be relevant for China, India, parts of Africa and Latin America, among others. We believe, moreover, that there is ample room for further exploring the relationship between neo-classical insurance and classical altruism models as well as for forging better connections between the new behavioral economics and a revived application of dualism to both development theory and policy in large portions of the developing world.

NOTES

[*] Frank Altschul Professor of International Economics, Yale University
The author wishes to acknowledge the substantial contribution of Michael Boozer (Yale), some early inputs from Douglas Gollin (Williams) and the assistance of Tavneet Suri (Yale), Prateek Tandon (Yale), and Zaruhi Sahakyan (Yale).

[i] Boeke, J. H. (1953), Economics and Economic Policy in Dual Societies, Institute of Pacific Relations.

[ii] Higgins, B. (1956) "The 'Dualistic Theory' of Underdeveloped Areas," Economic Development and Cultural *Change*, 4(2), 99-115.

[iii] Eckaus, R. S. (1955), "The Factor Proportions Problem in Underdeveloped Countries," American Economic Review, 45:539-565.

[iv] Ricardo, D. (1815) Works and Correspondence of David Ricardo, Vols. 1-4. Cambridge: Cambridge University Press.

[v] Smith, A. (1880), The Wealth of Nations. Oxford: Clarendon Press.

[vi] Lewis, W. A. (1954), "Economic Development with Unlimited Supplies of Labour," Manchester School, 22:139-191.

[vii] Nurkse, R., (1953), Problems of Capital Formation in Underdeveloped Countries, New York: Oxford University Press.

[viii] Rosenstein-Rodan, (1943), "The Problem of Industrialization of Eastern and South-Eastern Europe," Economic Journal, 53, 202-211.

[ix] For the open economy implications of the Fei-Ranis dual economy model, see Ranis, Gustav (1988), "Analytics of Development: Dualism," in H.B. Chenery and T.N. Srinivasan, eds., Handbook of Development Economics, Vol. 1, Elsevier Publishers, Amsterdam: North Holland.

[x] Ranis, G. and John C. H. Fei (1961), "A Theory of Economic Development," American Economic Review, 51, 533-565 and Fei, John C.H. and Gustav Ranis (1964), Development of the Labor Surplus Economy: Theory and Policy. See also the emphasis on product dualism in R. Kanbur and J. McIntosh (1985), "Dual Economy Models," in The New Palgrave: A Dictionary of Economic Theory and Doctrine. New York: Macmillan.

[xi] Harris, J. and M. Todaro (1970), "Migration, Unemployment and Development: A Two Sector Analysis," American Economic Review, 40, 126-142.

[xii] Fields, G. S. (1975) "Rural-Urban Migration, Urban Unemployment and Underemployment and Job Search Activity in LDC's," Journal of Development Economics, Vol. 2, No. 2, pp. 165-188.

[xiii] Ranis, G. and F. Stewart (1999), "V-Goods and the Role of the Urban Informal Sector in Development," Economic Development and Cultural Change, 47(2), 259-288.

[xiv] Sen, A. K. (1967), "Surplus Labor in India: A Critique of Schultz' Statistical Test," Economic Journal, 77, 154-161.

[xv] Lewis, W.A. (1972), "Reflections on Unlimited Labor," in Luis DiMarco (ed.), International Economics and Development: Essays in Honor of Raul Prebisch, New York: Academic Press, pp. 75-96

[xvi] Otsuka, K. (2001), Book Review of "Growth and Development from an Evolutionary Perspective," Journal of Development Economics, 65, 237-241.

[xvii] Otsuka, K. (2001), op cit.

[xviii] Fafchamps, M. (1992), "Solidarity Networks in Preindustrial Societies: Rational Peasants with a Moral Economy," Economic Development and Cultural Change, 41, 147-174.

[xix] Geertz, C. (1963), Agricultural Involution: The Process of Ecological Change in Indonesia, University of California Press.

[xx] Scott, J. C. (1976), The Moral Economy of the Peasant, New Haven, CT: Yale University Press.

[xxi] Ishikawa, S. (1975), "Peasant Families and the Agrarian Community in the Process of Economic Development," in L. Reynolds, (ed.), Agriculture in Development Theory, Yale University Press, New Haven, CT.

[xxii] Hayami, Y. and M. Kikuchi (1982), Asian Village Economy at the Crossroads, Baltimore, MD: Johns Hopkins University Press, p. 217

[xxiii] Osmani, S. R., (1991), "Wage Determination in Rural Labor Markets: The Theory of Implicit Cooperation," Journal of Development Economics, 34, 3-23.

[xxiv] Williamson cites such evidence of rising agricultural productivity and nearly constant real wages after the enclosure movement created a "labor surplus" condition.

[xxv] Fei, John C. H. and G. Ranis, (1997) Growth and Development From an Evolutionary Perspective, Blackwell.

[xxvi] Fei and Ranis, ibid.

[xxvii] Sen, A. K. (1966), "Peasants and Dualism With and Without Surplus Labor," Journal of Political Economy, 74, 425-450

[xxviii] Rosenzweig, M. (1988), "Labor Markets in Low Income Countries," in H. Chenery and T.N. Srinivasan, eds., Handbook of Development Economics, Vol. 1. Amsterdam: North Holland.45.

[xxix] Fei and Ranis (1997), op.cit., Appendix to Chapter 3.

[xxx] Kuznets, S. (1955), "Economic Growth and Income Inequality," American Economic Review, Vol. 45, No. 1.

[xxxi] Bourguignon, F and C. Morrison (1995), "Inequality and Development: The Role of Dualism" DELTA, Document #95-32, p. 21.

[xxxii] Fei, John C.H., G. Ranis and Shirley W.Y. Kuo (1979), Growth with Equity: The Taiwan Case, Oxford University Press.

[xxxiii] Fields, G. (1980), Poverty, Inequality and Development, Cambridge: Cambridge University Press.

[xxxiv] e.g. Fields (1980), op. cit

[xxxv] See Fei and Ranis (1997), op. cit

[xxxvi] Townsend, R. (1994), "Risk and Insurance in Village India", Econometrica, Vol. 62, No. 3, pp 539-591.

[xxxvii] Scott (1976) op. cit.

REFERENCES

Boeke, J. H. (1953), Economics and Economic Policy in Dual Societies, Institute of Pacific Relations.

Bourguignon, F. and C. Morrison (1995), "Inequality and Development: The Role of Dualism" DELTA, Document #95-32, p. 21.

Eckaus, R. S. (1955), "The Factor Proportions Problem in Underdeveloped Countries," American Economic Review, 45:539-565.

Fafchamps, M. (1992) "Solidarity Networks in Preindustrial Societies: Rational Peasants With a Moral Economy," Economic Development and Cultural Change, 41, 147-174.

Fei, John C. H. and G. Ranis, (1964), Development of the Labor Surplus Economy: Theory and Policy, Homewood, Illinois: Richard A. Irwin, Inc.

Fei, John C. H. and G. Ranis, (1997), Growth and Development From an Evolutionary Perspective, Blackwell.

Fei, John C. H., G. Ranis and S. W.Y. Kuo (1970), Growth with Equity: The Taiwan Case, Oxford University Press.

Fields, G. S. (1980), Poverty, Inequality and Development, Cambridge: Cambridge University Press.

Fields, G. S. (1975), "Rural-Urban Migration, Urban Unemployment and Underemployment and Job Search Activity in LDC's," Journal of Development Economics, Vol. 2, No. 2, pp. 165-188.

Geertz, C. (1963), Agricultural Involution: The Process of Ecological Change in Indonesia, University of California Press.

Harris, J. and M. Todaro (1970), "Migration, Unemployment and Development: A Two Sector Analysis," American Economic Review, 40, 126-142.

Higgins, B. (1956) "The 'Dualistic Theory' of Underdeveloped Areas," Economic Development and Cultural Change, 4(2), 99-115.

Hayami, Y. and M. Kikuchi (1982), Asian Village Economy at the Crossroads. Baltimore, MD: Johns Hopkins University Press, p. 217

Ishikawa, S. (1975), "Peasant Families and the Agrarian Community in the Process of Economic Development," in L. Reynolds, (ed.), Agriculture in Development Theory, (Yale University Press, New Haven, CT).

Kanbur, R. and J. McIntoch (1985), "Dual Economy Models," in The New Palgrave: A Dictionary of Economic Theory and Doctrine, New York, MacMillan.

Kuznets, S. 1955, "Economic Growth and Income Inequality," American Economic Review, Vol. 45, No. 1, March 1955.

Lewis, A. (1954), "Economic Development with Unlimited Supplies of Labour," Manchester School, 22: 139-191.

Lewis, W. A. (1972), "Reflections on Unlimited Labor," in Luis DiMarco (ed.), International Economics and Development: Essays in Honor of Raul Prebisch, New York: Academic Press, pp. 75-96.

Nurkse, R. (1953), Problems of Capital Formation in Underdeveloped Countries, New York: Oxford University Press.

Osmani, S. R. (1991), "Wage Determination in Rural Labor Markets: The Theory of Implicit Cooperation," Journal of Development Economics, 34, 3-23.

Otsuka, K. (2001), Book Review of "Growth and Development from an Evolutionary Perspective," Journal of Development Economics, 65, 237-241.

Ranis, G. (1988), "Analytics of Development: Dualism," in H.B. Chenery and T.N. Srinivasan, eds, Handbook of Development Economics, Vol. 1, Elsevier Publishers, Amsterdam: North Holland.

Ranis, G. and F. Stewart (1999), "V-Goods and the Role of the Urban Informal Sector in Development," Economic Development and Cultural Change, 47(2), 259-288.

Ricardo, D. (1815) Works and Correspondence of David Ricardo, Vols. 1-4. Cambridge: Cambridge University Press.

Rosenstein-Rodan (1943), "The Problem of Industrialization of Eastern and South-Eastern Europe," Economic Journal, 53, 202-211. Reprinted in A. N. Agarwala and S.P. Singh (eds), The Economics of Underdevelopment, Bombay: Oxford University Press, 1958.

Rosenzweig, M. (1988), "Labor Markets in Low Income Countries," in H. Chenery and T.N. Srinivasan, eds., Handbook of Development Economics, Vol. 1. Amsterdam: North Holland.45.

Scott, J. C. (1976), The Moral Economy of the Peasant, New Haven, CT: Yale University Press.

Sen, A. K. (1966), "Peasants and Dualism With and Without Surplus Labor," Journal of Political Economy, 74, 425-450

Sen, A. K. (1967), "Surplus Labor in India: A Critique of Schultz' Statistical Test," Economic Journal, 77, 154-161.

Smith, A. (1880), The Wealth of Nations. Oxford: Clarendon Press.

Townsend, R. (1994), ``Risk and Insurance in Village India", Econometrica, Vol. 62, No. 3, pp 539-591.

Williamson, J. (1989), "Inequality, Poverty, and the Industrial Revolution;" "Migration and Wage Gaps: An Escape from Poverty?" and "Accumulation and Inequality: Making the

Connection." Third Simon Kuznets Memorial Lectures, Economic Growth Center, Yale University.