Natural Resource Management and Policy
Series Editors: David Zilberman · Renan Goetz · Alberto Garrido

Siwa Msangi
Duncan MacEwan *Editors*

Applied Methods for Agriculture and Natural Resource Management

A Festschrift in Honor of Richard E. Howitt

Springer

Natural Resource Management and Policy

Volume 50

There is a growing awareness to the role that natural resources, such as water, land, forests and environmental amenities, play in our lives. There are many competing uses for natural resources, and society is challenged to manage them for improving social well-being. Furthermore, there may be dire consequences to natural resources mismanagement. Renewable resources, such as water, land and the environment are linked, and decisions made with regard to one may affect the others. Policy and management of natural resources now require interdisciplinary approaches including natural and social sciences to correctly address our society preferences.

This series provides a collection of works containing most recent findings on economics, management and policy of renewable biological resources, such as water, land, crop protection, sustainable agriculture, technology, and environmental health. It incorporates modern thinking and techniques of economics and management. Books in this series will incorporate knowledge and models of natural phenomena with economics and managerial decision frameworks to assess alternative options for managing natural resources and environment.

More information about this series at http://www.springer.com/series/6360

Siwa Msangi · Duncan MacEwan
Editors

Applied Methods for Agriculture and Natural Resource Management

A Festschrift in Honor of Richard E. Howitt

Springer

Editors
Siwa Msangi
Silver Spring, MD, USA

Duncan MacEwan
ERA Economics LLC
Davis, CA, USA

ISSN 0929-127X ISSN 2511-8560 (electronic)
Natural Resource Management and Policy
ISBN 978-3-030-13486-0 ISBN 978-3-030-13487-7 (eBook)
https://doi.org/10.1007/978-3-030-13487-7

Library of Congress Control Number: 2019931839

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

I first met Richard E. Howitt while we were both graduate students; I was hired on the faculty and became Richard's Ph.D. Advisor for his dissertation. At the time, my life was very complicated: traveling back and forth to University of Chicago from Davis and running a nearby family farm located some 25 miles from Davis. Richard, along with five other Ph.D. students who had selected me as Director, kept me grounded. Richard kept me particularly grounded because he had a passion for evaluating, measuring, and designing policies to address environmental externalities. In fact, at that point in time, I believe that Richard knew more about environmental externalities and their practical measurement than any other Ph.D. student, or for that matter, UC Davis faculty member.

At the University of Chicago, I had arranged for a postdoc for Richard following the completion of his Ph.D. dissertation. An offer had been made to Richard, by U Chicago, and he had accepted it. I shortly thereafter decided to leave UC Davis, and an offer was made to Richard to become a permanent faculty member in the department. In fact, based on my recollection, there are only three Ph.D. students who have completed their dissertations at UC Davis that have been hired on the faculty: Richard, me, and more recently Pierre Mérel. At least with regard to my departure, a win–win situation emerged. Richard has been an extraordinary faculty member at UC Davis, and I pursued positions at other universities, landing eventually at UC Berkeley.

Shortly after my arrival at Harvard University, I was doing work with the NBER, and Richard and I published a paper, entitled "Stochastic Control of Environmental Externalities." I presented this paper on three separate occasions at NBER conferences. There was much interest among Ph.D. students at UC Davis and elsewhere in the country, including myself, in stochastic, adaptive, and dual control methodologies. Richard continued to pursue this methodological passion over the course of his academic career, focusing largely on practical applications, not theoretic constructs. Our paper was submitted for consideration and was awarded the Outstanding Publication from the AAEA in 1976. This award provided some assistance in launching both Richard's and my academic career.

Among the many attributes and characteristics of Richard, I most admire that he has been one of the major intellectual leaders in combining outreach (cooperative extension) with sound, practical academic research. Richard has advanced his research through a variety of outreach projects. He has served on numerous CalFed Cost Allocation Advisory Committees, developed economic modeling workshops for the US Bureau of Reclamation, and frequently worked on the planning process for the California Department of Water Resources. I can recall no other faculty member at any of the three University of California campuses composing the Giannini Foundation who has been more successful in combining these two major activities. For those of you who have not been engaged in outreach, it can be an unbelievable time sink. Because of Richards's unique characteristics and contributions, he has been able to capture what complementarities exist between these two activities. Time is too short to enumerate and give my assessments of all of Richard's remarkable contributions to our profession. He is certainly one faculty member at either UC Berkeley or UC Davis that no one questions their contribution to the original mandate of the Giannini Foundation. His many seminal research contributions have focused largely on California.

He has also combined fundamental science with his economic analysis. His interface between water science and water politics in California, the West, as well as internationally has influenced the development of comprehensive water marketing institutions in many regions. There are many claimed fathers of the water marketing success, but Richard, without question, is one of its major intellectual leaders. He has also been instrumental in forging multi-disciplinary collaborations with hydrologists, engineers, and agronomists.

In the case of California water, he has covered every critically important issue and offered sound advice and counsel. He has been an intellectual leader in all aspects of water resources research, including:

- The California drought
- Groundwater sustainability
- Contamination of drinking water resources (particularly nitrates)
- Salinity
- Water recycling
- California's Sacramento-San Joaquin Delta evaluation and some of the most important economic analysis that has been done on this critically important question
- California land use and cropping patterns
- Development of both aggregate and disaggregate frameworks to explain the response of California farmers and land use to changes in prices, property rights, as well as evaluation of environmental constraints on resource use.

He has exercised sound judgement in the use of various data sources, most recently his work on disaggregating land and resource use, exploiting remote sensing methods in lieu of his earlier work utilizing economic surveys. He is among the early intellectual leaders to recognize that such approaches provide a much more efficient basis for physical land use and responses at a fraction of the cost of economic surveys. He has always been interested in the use of modern methodologies to develop calibrated models based on, for example, maximum entropy estimators. He has done so recognizing the inherent heterogeneity that exists for the utilization of California agricultural lands.

For many years, Richard was engaged in developing *positive mathematical programming*. The value that is potentially captured by this methodology is the flexibility introduced to conventional mathematical programming, recognizing the inherent nonlinearities that exist and are implicitly observed in land allocation decisions at either a regional or farm level. His methods for calibrating the model in terms of output/input use, objective function values, and dual values are particularly insightful. His unique justification for his ultimate publication in the *AJAE* of the positive mathematical programming approach offers an unusual motivation for the paper: *Sometimes new methodologies are published but not implemented. Positive mathematical programming is a methodology that has been implemented but not published*.

In conclusion, during the early 1970s, a number of us who were working closely at UC Davis (which included some other remarkable Ph.D. students that I had the pleasure of mentoring during my brief time on the faculty at UC Davis) achieved much success within and beyond our agriculture and resource economics profession. No one stands taller than Richard E. Howitt with regard to his seminal contributions to understanding and generating insights on California cropping patterns, resource use patterns, and the externalities that arise from private actions for which there has historically been very little in the way of incentives orchestrated by governmental policy to move closer to the proper measures of social cost.

Berkeley, USA

Gordon Rausser
Robert Gordon Sproul Distinguished Professor
University of California, Berkeley
rausser@berkeley.edu

Preface

This book covers a range of topics which touch on these important issues and range from agricultural production and sector modeling, to natural resource management and—ultimately—to analytically guided decision support to assess various policy options. The examples laid out in the chapters range beyond California and illustrate the wide-ranging applicability of robust quantitative methods that can be applied to a variety of agricultural policy and resource management problems. The range of work contained in this book represents the breadth of Richard E. Howitt's academic and professional interests, and the immense influence that his research and teaching career have had on the agricultural and resource economics profession, and on the professionals themselves, who are represented by the authors in each chapter.

This book came out of a *Festschrift* conference that was held for Richard E. Howitt in late May 2016, in honor of his 40-year career as Professor of Agricultural and Resource Economics at the University of California at Davis. The presentations at that conference followed the thematic structure of this book, covered the major areas of research that Richard engaged in during his UC Davis career, and discussed innovative analytical methods that he pioneered, in collaboration with his students, mentors, and colleagues. These approaches still stand, today, at the cutting edge of the field and cover applied production analysis for agriculture, water resource management, and information-theoretic methods for doing robust estimation with limited data.

What has made Richard E. Howitt's career remarkable is not only his length of service and dedication to the same institution—the University of California at Davis—but also his determination to fulfill the mission of an "agricultural experiment station" that faculty appointments at Davis are supposed to support. Rather than be wholly absorbed in the pursuit of publishing in academic journals that are seldom read by agricultural producers and policy-makers, Richard E. Howitt has made tremendous efforts over the years to make his work relevant and directly applicable to solving agricultural policy and resource management issues faced by the state of California. His ongoing partnerships and collaboration with staff in the California Department of Water Resources, both through the UC Davis-based Center for

Watershed Sciences as well as through direct outreach, are evidence of this—as is the large number of staff within that institution (and others) that have been trained by him. In the past, he has advised state water resource managers on how best to set up drought water banks to deal with periods of severe scarcity, and to take advantage of the increased willingness to sell/buy of agricultural and urban water users during those times. His current work through the consulting firm *ERA Economics* to advise local water districts on how best to comply with the recently passed California Groundwater Management Act and set up economically sound mechanisms for managing resource extraction and potential trading is further evidence of the applicability of his work (and his failure to fully retire!). His contributions to the water policy work of the Public Policy Institute of California as Collaborator and Visiting Fellow are captured in a number of publications on their website (https://www.ppic.org/water/) and can serve as a useful reference and inspiration to those researchers interested in deepening the policy relevance and impact of their work.

The foreword to this book (by Prof. Gordon Rausser) will refer to many remarkable aspects of Richard E. Howitt's career and work. Some of those highlights were captured in the comments of those who attended and presented at the 2016 *Festschrift* conference and could not be documented in this book. Howitt's longtime colleague, neighbor, and friend—Prof. James E. Wilen—pointed out to the conference attendees that the remarkable number of dissertations Richard chaired during his faculty tenure (28 in total) understates the crucial empirical contributions he made to many others, of which he might have been the second or third member on the committee. As Prof. Wilen mentioned, there is usually one committee member who actually directly helps and enables the student to get down to the nuts-and-bolts, nitty-gritty of making the empirical model-based analysis work—and that person was often Richard E. Howitt, in many of the dissertations involving agricultural production analysis and water resource management. His ability to remain completely hands-on with any modeling analysis taken on by himself or his students throughout his entire career is rare among faculty and has been a source of personal inspiration for myself, in the never-ending effort to keep up-to-date with modeling approaches and methods.

It was not possible to include contributions from a number of key collaborators that Richard E. Howitt has worked with over his 40-year career—but we feel that this collection of papers captures the essence of the kind of applied work that Richard has led during his professional life. It is our hope that the readers of this book will find useful ideas and examples embedded within these chapters and that they will inspire their further use toward other applications and case studies.

Silver Spring, USA — Siwa Msangi
Davis, USA — Duncan MacEwan

Acknowledgements

The editors extend their sincere thanks to all of the participants of the May 2016 *Festschrift* conference in honor of Richard E. Howitt, whose tributes and thoughtful reflections and presentations underscored the significance of his research achievements, and which laid the foundation for this book. We also extend our deep thanks to the authors of these chapters, and their willingness to commit their thoughts to paper, despite the many pressures and professional obligations that they face. Johannes Glaeser of Springer was instrumental in enabling this book project to take place, and the inputs of anonymous reviewers helped to strengthen the clarity and readability of these chapters, and enhance the overall quality of the book. We acknowledge the support of funds from the CGIAR research program on Policies, Markets and Institutions that allowed the *Festschrift* conference to take place, and which supported the foundational work that launched the book project.

Contents

Contributors

George Ayaga Kenya Agricultural and Livestock Research Organisation (KALRO), Food Crops Research Institute, Nairobi, Kenya

Peter Berck Department of Agricultural and Resource Economics, University of California, Berkeley, CA, USA

Leonidas Dusengemungu Rwanda Agriculture Board, Rubilizi, Kigali, Rwanda

George Frisvold Department of Agricultural and Resource Economics, The University of Arizona, Tucson, USA

Bruno Henry de Frahan Earth and Life Institute, Université catholique de Louvain, Louvain-La-Neuve, Belgium

Richard E. Howitt Department of Agricultural and Resource Economics, University of California, Davis, CA, USA

Qiuqiong Huang Department of Agricultural Economics & Agribusiness, The University of Arkansas, Fayetteville, AR, USA

Geophrey Kajiru Lake Zone Agricultural Research Development Institute (LZARDI), Mwanza, Tanzania

Kazim Konyar Department of Economics, California State University at San Bernardino, San Bernardino, USA

Birungi Korutaro Kilimo Trust, Bugolobi, Kampala, Uganda

Duncan MacEwan ERA Economics, Davis, USA

Siwa Msangi International Food Policy Research Institute, Washington, DC, USA; Independent Consultant, Silver Spring, MD, USA

Joseph Mudiope Kilimo Trust, Bugolobi, Kampala, Uganda

Jean-Claude Muhutu Rwanda Agriculture Board, Rubilizi, Kigali, Rwanda

Bernard Musana College of Agriculture, Animal Sciences and Veterinary Medicine, University of Rwanda, Kigali, Musanze, Rwanda

Qianyao Pan Department of Agricultural and Resource Economics, University of California Agricultural Issues Center, University of California, Davis, Davis, USA

Quirino Paris Department of Agricultural and Resource Economics, University of California, Davis, Davis, USA

Scott D. Rozelle Stanford University, Stanford, CA, USA

Daniel A. Sumner Department of Agricultural and Resource Economics, University of California Agricultural Issues Center, University of California, Davis, Davis, USA

Lucas Tanui Kenya Agricultural and Livestock Research Organisation (KALRO), Food Crops Research Institute, Nairobi, Kenya

Marcelo Torres Department of Economics, University of Brasília, Brasília, DF, Brazil

Kennedy Were Kenya Agricultural and Livestock Research Organisation (KALRO), Food Crops Research Institute, Nairobi, Kenya

James E. Wilen University of California, Davis, Davis, CA, USA

Lunyu Xie Department of Energy Economics, School of Applied Economics, Renmin University of China, Beijing, China

Chapter 1
Introduction

Siwa Msangi

In a complex world, facing increasing environmental stresses and socioeconomic challenges, policy-makers have a tremendous need for reliable and robust decision-support tools to evaluate options and their inherent trade-offs. This describes the world food situation, fairly accurately, especially when put within the context of the additional challenge of meeting environmental quality standards and ensuring long-term sustainability. Within the North American context, the state of California serves as a good case example that is representative of the larger 'balancing act' that many regions try and achieve between sustained profitability and productivity in the agricultural sector while maintaining environmental quality and ecosystem health, over a varied and biologically rich landscape. The impressive landscape of California's agriculture is supported by critical surface and groundwater resources whose unevenly distribution must be managed by a complex infrastructure network that must also accommodate the ever-increasing levels of urbanization and settlement that have occurred throughout California's history availability, and which are biased toward wealthy, coastal population centers. The challenges of water management facing northern China are quite similar to this and are addressed in later chapters of the book. These competing pressures, in addition to the increasing climatic variability now facing the region, pose challenges and critical trade-offs that require robust tools of analysis to assess them and to provide decision support to policy-makers.

Figure 1 captures some key linkages between firm- or household-level agricultural production, markets, the environment, and consumption. In the case of rural agricultural farm households, they may source their consumption directly from own production or might choose to use the market for marketing their produce (or to purchase other goods).

To simulate the decisions of the producers, and to evaluate the market-mediated response to environmental or socioeconomic influences, the analyst needs to resort

S. Msangi (✉)
International Food Policy Research Institute, Washington, DC, USA
e-mail: siwamsangi@gmail.com

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_1

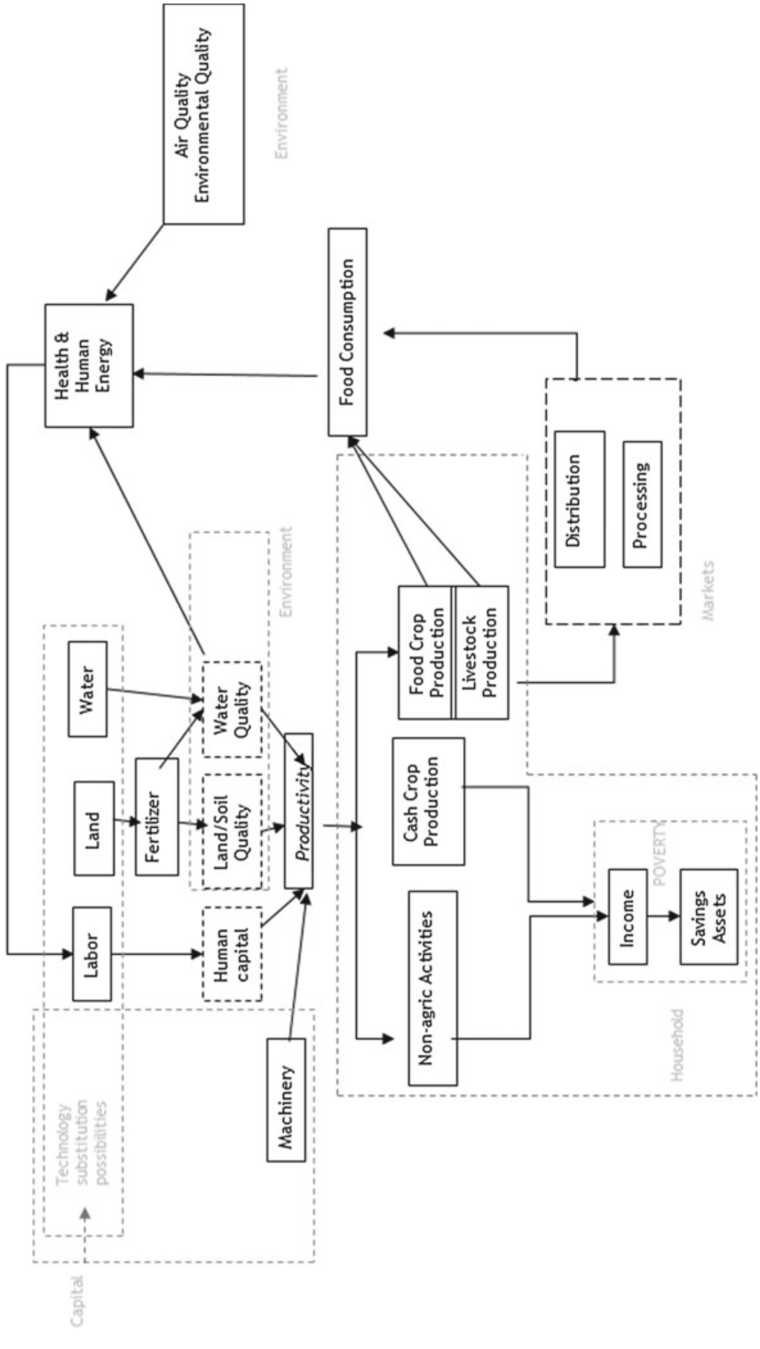

Fig. 1 Key linkages between agricultural production, consumption, markets, and the environment

to model-based approaches that can reasonably capture the essential components. This is expanded up on in the following subsections.

1 Capturing Key Behavioral Dimensions at Micro- and Macro-Level

At the micro-level, the farmer faces complex decisions that depend on prevailing economic conditions as well as on the physical environment in which she finds herself—which largely define the resource constraints on production. The behavioral dimensions of the farm manager must also confront the biophysical realities of agronomy that determine crop yields, and any analytical approach that is used to study the farmer's behavior must be able to take these realities into account in some way (albeit simplified or reduced in form). Following in the rich economic tradition of studying the producer's decision problem, which have been established in methodological approaches taught to most economists (Chambers 1988), reduced-form approaches to studying farm-level behavior have been successful in exploiting the dual relationships of the producer's fundamental economic decision problem. Adopting a primal approach, however, can allow the analyst to handle the input/output linkages and other biophysical aspects of agricultural production more explicitly, and make better use of the wealth of information and data that can be obtained from agronomic or environmental studies. Studies where outputs from agronomic models are incorporated into an economic production analysis (Merel et al. 2014), are a good example of this.

At a more macro-level, the aggregate behavior of producers and consumers are more amenable to reduced-form representations that can abstract further away from the biophysical realities that constrain the day-to-day decisions of both households and agricultural producers. Careful consideration of how micro-level behavior should be properly aggregated up is warranted and has been explored by some studies (Önal and McCarl 1991). Given the more systematic collection of available information on prices and costs that often occurs at the market level, over time, dual-based approaches can be applied more widely and robustly when used to characterize consumer and producer behavior. Nonetheless, primal approaches can still be applicable if the analyst wants to exploit mathematical programming-based model calibration methods. This can be done either on the supply side of the market, as has been doing in numerous examples looking at the European agricultural landscape under the Common Agricultural Policy (Henry de Frahan et al. 2007; Thomas Heckelei and Britz 2000; Jansson and Heckelei 2011; Júdez et al. 2001), where the rich FADN dataset provides researchers and analysts with detailed information on various types of farm enterprises across Europe. PMP-based methods also allow one to calibrate directly on trade flows occurring in the market, as the chapter of Paris (in this book) illustrates. Given that agricultural policies affect a wide range of farmers and consumers, analytical methods that can account for market-level impacts and which can also address the distributional impacts across producer and consumer types are important and useful to employ in policy analysis.

2 Robust Analysis Under Limited Information

All researchers and analysts acknowledge the importance of high-quality data in carrying out robust and credible analysis, and a situation in which more observations are available is clearly superior to one in which limited observations can be obtained. Even in a period, like now, where researchers and analysts find themselves having more and more access to better, more abundant, and more disaggregated data than ever before, one is still confronted with situations where some key information might be limited, poorly or infrequently collected, or missing altogether. Our analytical methods, therefore, need to be amenable to such situations and provide us with a means of obtaining robust parameter estimates—or allow us an internally consistent framework for introducing 'expert judgement' or other non-sample information into the analysis. Bayesian methods provide a well-formalized way of incorporating the researcher's 'priors' into the process of parameter estimation and offer a useful alternative to classical methods of statistical and econometric analysis. Information-theoretic approaches operate in a similar spirit and allow the researcher to extend the range of econometric problems that can be solved.

Since the early 2000s, the use of information-theoretic approaches has become more widely incorporated into the range of econometric methods that economists use, and several important and comprehensive textbooks have helped to mainstream this approach (Golan et al. 1996; Mittelhammer et al. 2000). Aside from estimation of parameters for reduced-form functions (Marsh et al. 2014), information-theoretic approaches can also be applied to the estimation and calibration of behavioral models of production, in which the number of desired parameters might exceed the number of observations, such that the analyst faces an inference problem with negative degrees of freedom (Heckelei and Wolff 2003; Paris and Howitt 1998). This can be particularly important within the context of developing countries, where national agricultural information and data-collection systems still have large gaps in coverage, or where field-level observations may be limited or not freely available from the development (or development aid-funded) agencies and entities that collect them.

The chapters in Part 3 of this book illustrate the application of the maximum entropy principle to various empirical problems—one purely econometric and another focusing on field-level agricultural production decisions. The chapter of Msangi et al. in Part 4 of the book also uses entropy-based methods to construct an internally consistent database for the economic market model that is employed, using the best expert knowledge available and allowing for as much spatial disaggregation as possible.

3 Guiding Policy Decisions in Agriculture and Resource Management with Useful Tools

The ultimate goal of any good analysis of agriculture, environment, and natural resource management is to be useful to the formation of policy and to help guide decision-making by those responsible for administration and resource allocation. Decision-makers often do not want to delve deeply into the underlying complexities of the analysis, but still appreciate an objective treatment of whatever uncertainty may exist, and would want to know the range of alternative outcomes that could arise from making different choices or key decisions. The ability of the analyst to calibrate a model well and have it reproduce a range of observed (or at least plausible) behavior goes a long way toward enhancing the confidence that a policy-maker would have in the results from any given tool or analytical approach. It should, therefore, be the goal of every policy analysis to provide such a basis for confidence for decision-makers and to draw upon the best methods that can achieve that objective.

The chapters within this book are aimed at bringing out the importance of utilizing robust applied methods for agricultural policy analysis. The problems that researchers and policy analysts working on agriculture face are the complexities that characterize agricultural economies and the biophysical realities of the agricultural landscape. The analysis of policy issues related to agricultural production and markets cannot be entirely divorced from the complexities of agriculture itself—although efforts are always made to simplify and focus. Given that agriculture depends so critically upon the resource base of production—namely land and water—the degradation and depletion of those resources also figure prominently in the policy analysis of an agricultural economy. In this book, there are compelling examples of robust applied methods that need to be used in order to convincingly address this range of issues.

In the first part of the book, the authors will focus on the empirical modeling of the agricultural sector, with a heavy focus on production technologies and behavior, agricultural markets, and evaluating the importance of key resources such as soil and water availability on the production sector. Many of the examples in this chapter draw upon the analytical methods agricultural sector modeling calibration that Richard E. Howitt pioneered—Positive Mathematical Programming (hereafter referred to by its popular acronym "PMP"). Applications of PMP also show up in other parts of the book, but are best explained in this first section, which begins with a chapter by Bruno Henry de Frahan which takes a detailed view of various extensions to the original PMP methodologies. The paper of Quirino Paris, in this first part, applies the methodology to the calibration of a spatial equilibrium model of trade.

The second part of the book focuses on water resource management issues, which Richard E. Howitt has made significant research contributions toward, especially with regard to the problems facing water management in California. Here, the authors discuss a range of examples, ranging from Brazil to China—which draw inspiration from the work that Howitt and his colleagues have done over the years on applying production modeling to the estimation of water demands, so that the derived demand functions can guide the efficient allocation of scarce water resources around the state of California.

The third part of the book deals with applications of information theory both to problems of pure statistical inference and to the empirical analysis of farm-level crop rotations, using highly disaggregated data. This section reflects the interest that Richard E. Howitt and his colleagues have shown in the application of robust inference methods to the analysis of agricultural production, beginning from the highly influential article of Paris and Howitt which expanded PMP calibration to the ill-posed case (Paris and Howitt 1998).

The final section of the book addresses a range of agricultural policy and resource management issues in a variety of settings (not limited to California) and highlights the applicability of robust empirical methods such as those that have been covered in earlier parts of the book. The messages in this chapter focus on providing decision- and policy-makers with robust decision-support tools that make good use of the existing data and which can address the complexities of producer behavior and response, within the context of highly constrained natural resources or restrictive policy regimes. Although much of the book focuses on the technical aspects of modeling agricultural production and resource use—this section highlights the ultimate relevance and importance of applying these methods to address urgent issues that are relevant to policy and the key trade-offs that decision-makers face when assessing investments and alternative strategies and interventions to strengthen the agricultural sector and the natural resource base upon which it critically depends.

References

Chambers, R. G. (1988). *Applied production analysis: A dual approach*. Cambridge: University Press.

Golan, A., Judge, G. G., & Miller, D. (1996). *Financial Economics and Quantitative Analysis Series: Maximum entropy econometrics: Robust estimation with limited data.*. Chichester: Wiley.

Heckelei, T., & Britz, W. (2000). Positive mathematical programming with multiple data points: A cross-sectional estimation procedure. *Cahiers d'Economie et de Sociologie Rurales (CESR)*, Institut National de la Recherche Agronomique (INRA), *57*, 28–50. Retrieved from https://ideas.repec.org/a/ags/inrace/206148.html.

Heckelei, T., & Wolff, H. (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy. *European Review of Agriculture Economics, 30*(1), 27–50. https://doi.org/10.1093/erae/30.1.27.

Henry de Frahan, B., Buysse, J., Polomé, P., Fernagut, B., Harmignie, O., Lauwers, L., … Meensel, J. (2007). Positive mathematical programming for agricultural and environmental policy analysis: Review and practice. In *Handbook of operations research in natural resources* (pp. 129–154). Boston, MA, US: Springer. https://doi.org/10.1007/978-0-387-71815-6_8.

Jansson, T., & Heckelei, T. (2011). Estimating a primal model of regional crop supply in the European Union. *Journal of Agricultural Economics, 62*(1), 137–152. https://doi.org/10.1111/j.1477-9552.2010.00270.x.

Júdez, L., Chaya, C., Martínez, S., & González, A. A. (2001). Effects of the measures envisaged in "Agenda 2000" on arable crop producers and beef and veal producers: An application of positive mathematical programming to representative farms of a Spanish region. *Agricultural Systems, 67*(2), 121–138. https://doi.org/10.1016/S0308-521X(00)00051-2.

Marsh, T., Mittelhammer, R., & Cardell, N. (2014). Generalized maximum entropy analysis of the linear simultaneous equations model. *Entropy, 16*(2), 825–853. https://doi.org/10.3390/e16020825.

Merel, P., Yi, F., Lee, J., & Six, J. (2014). A regional bio-economic model of nitrogen use in cropping. *American Journal of Agricultural Economics, 96*(1), 67–91. https://doi.org/10.1093/ajae/aat053.

Mittelhammer, R., Judge, G. G., & Miller, D. (2000). *Econometric foundations*. Cambridge: University Press.

Önal, H., & McCarl, B. A. (1991). Exact aggregation in mathematical programming sector models. *Canadian Journal of Agricultural Economics/Revue Canadienne d'agroeconomie, 39*(2), 319–334. https://doi.org/10.1111/j.1744-7976.1991.tb03575.x.

Paris, Q., & Howitt, R. E. (1998). An analysis of ill-posed production problems using maximum entropy. *American Journal of Agricultural Economics, 80*(1), 124–138. https://doi.org/10.2307/3180275.

Part I
Applied Methods for Agricultural Production and Sector Modeling

Chapter 2
Towards Econometric Mathematical Programming for Policy Analysis

Bruno Henry de Frahan

Abstract This contribution focuses in reviewing the development of positive mathematical programming towards econometric mathematical programming. Starting with the entropy approach it reviews alternative approaches and model specifications that appeared in the recent PMP-related literature for estimating those nonlinear terms that achieve the accurate calibration of optimisation programmes and guide the simulation response to policy scenarios. Combining recent contributions from this literature, it then proposes a possible framework to estimate and calibrate simultaneously model parameters ready to use for performing policy simulations.

1 Introduction

The formulation of the concept of positive mathematical programming (PMP) is due to Richard Howitt (1995a) in the *American Journal of Agricultural Economics* (AJAE). The core concept of PMP consists in achieving an accurate calibration of an optimisation programme thanks to the addition of specific nonlinear terms in the objective function such as to satisfy the optimality conditions of the programme precisely at the observed levels of the decision variables. Because these nonlinear terms directly control the response of the optimisation programme to changes in variables exogenous to the programme, their determination has been subject to close examination in the PMP literature.

In his seminal paper, Howitt (1995a) relied in the use of information contained in just one behavioural observation of the economic agent to establish the PMP concept. Unless to resort to ad hoc specification rules (e.g., arbitrary assumptions on some calibration parameters) or exogenous parameters (typically rental rates of limiting inputs and supply elasticities to price) to overcome the under-determination of the whole system of the first-order optimal conditions to calibrate, it became apparent that the calibration of any nonlinear objective function in PMP modelling would

B. Henry de Frahan (✉)
Earth and Life Institute, Université catholique de Louvain, Louvain-La-Neuve, Belgium
e-mail: Bruno.henrydefrahan@uclouvain.be

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_2

need more than just one observation on decision variables. Several mathematical programming applications were therefore prompt to extend the PMP concept to the use of a set of several observations. Some modellers used cross-sectional data across similar economic units (e.g., Heckelei and Britz 2000; Paris 2001a, 2015, 2017; Arfini et al. 2008; Arfini and Donati 2011; Louhichi et al. 2016; Arata et al. 2017); while some others used time series of the same unit (e.g., Paris 2001b; Heckelei and Wolff 2003; Jansson and Heckelei 2011; Cortignani and Severini 2012). However, fewer applications made use of panel data, thereby crossing the two dimensions (e.g., Buysse et al. 2007a; Henry de Frahan et al. 2011; Britz and Arata 2019).

With several observations to use for determining these nonlinear PMP terms, different estimation techniques appeared in the related literature. Starting first with entropy estimators, continuing with Bayesian estimators and nowadays into standard econometric estimators and bi-level programmes, the PMP-related literature developed itself into being progressively closer to econometrics with the trick of still providing a calibrated model ready for simulation. Exploiting together the advantages of mathematical programming and econometric approaches lead to the emergence of a new field of empirical investigation, which was first named "econometric programming" (Henry de Frahan et al. 2007) and then labelled afterwards "econometric mathematical programming" (Buysse et al. 2007b), which could be viewed as the econometric estimation of programming model parameters based on multiple observations.

This chapter reviews the development of this new field of investigation. It takes a different angle than in recent reviews on PMP (Heckelei et al. 2012; Mérel and Howitt 2014) because it is interested in showing how the econometric estimation of those PMP parameters evolved since the pioneering contribution of Paris and Howitt (1998). To create an even closer bridge between mathematical programming and econometrics, this chapter also proposes a methodological contribution waiting for an empirical application, before concluding.

2 From Positive Mathematical Programming to Econometric Mathematical Programming

As announced in the introduction, the passage of positive mathematical programming (PMP) to econometric mathematical programming (EMP) is captured in this section through the successive use of alternative estimation approaches. First, the maximum entropy approach started to be in fashion since the publication of the book of Golan et al. in 1996 until reservations for this approach accumulated later on in the 2000s. Then, the Bayesian approach was suggested as an alternative to the entropy approach with the paper of Heckelei et al. (2008). In particular, this approach is currently implemented into an impressive effort of the European Commission to model every farm contained in its rich Farm Accounting Data Network (FADN) dataset for policy simulations (Louhichi et al. 2016). More recently, the bi-level programming approach

(Jansson and Heckelei 2009) has started to be applied in a few applications while the use of standard econometrics is still in its infancy (Henry de Frahan et al. 2011). In parallel to these estimation approaches, different model specifications have been used and, hence, reviewed as well. In particular, the book of Paris (2011) proposes several of them.

2.1 The Maximum Entropy Approach

The move forward from typical programming models into econometric programming models was instigated by the influential paper of Paris and Howitt (1998) in the *American Journal of Agricultural Economics* and was thereafter relayed by another influential paper but of Heckelei and Wolff (2003) in the *European Review of Agricultural Economics*. Although the paper of Paris and Howitt (1998) kept using just one single observation with its associated variables, it opened the path of recovering econometrically PMP parameters by resorting to the maximum entropy (ME) criterion (see Golan et al. 1996: 8) to address first the under-determination problem of standard calibration of PMP models. Paris and Howitt (1998) and subsequent programming modellers (e.g., Heckelei and Britz 2000; Paris 2001a, b; Arfini et al. 2008; Graveline and Mérel 2014; Petsakos and Rozakis 2015) actually used the generalised maximum entropy (GME) formalism for defining discrete probability distributions over the parameter space to recover by sets of discrete support points with corresponding probabilities (Golan et al. 1996: 67). These modellers applied the GME estimator on the so-called pure linear inverse ME formulations of the under-determinacy problem, therefore implicitly assuming a deterministic underlying relation between observations and their associated variables and an absence of measurement errors or other disturbances on observations (see Golan et al. 1996: 67–69). None used the generalised cross-entropy (GCE) formalism to include prior probabilities over the same sets of discrete points as initial hypotheses (see Golan et al. 1996: 67–69). Instead, some used prior information on elasticities and shadow prices in the GME estimation process as shown below.

However, if maximum entropy or any other econometric criterion is really to use several observations, the paper of Heckelei and Wolff (2003) highlighted a fundamental inconsistency in the estimation process of the parameters used in the ultimate calibrated model, seen as representing adequately the true data generating process and the initial models simulated in the first stage in the PMP procedure. In particular, the shadow values of the binding constraints implied by the ultimate model in the last PMP stage are determined differently than those implied by the initial models in the first PMP stage when several observations are used. As a result, most of the following PMP applications (e.g., Henry de Frahan et al. 2007, 2011; Buysse et al. 2007a; Kanellopoulos et al. 2010; Mérel and Bucaram 2010; Jansson and Heckelei 2011; Mérel et al. 2011, 2014; Cortignani and Severini 2012; Graveline and Mérel 2014; Louhichi et al. 2016; Garnache et al. 2017; Britz and Arata 2019) skipped this first PMP stage and simultaneously recovered PMP parameters and shadow prices directly from the optimality conditions of the desired programming model. Most likely, applications relying on shadow values of the binding and calibrating constraints delivered by the first PMP stage to define the support points in a GME formalism (e.g., Paris and Howitt 1998; Paris 2001a, b; Arfini et al. 2008) were prone to recovering biased calibrated parameters and, hence, subject to producing questionable simulation behaviours as explained by Heckelei and Britz (2000) and Britz et al. (2003).

By the same juncture, the paper of Heckelei and Wolff (2003) turned the recovery process of PMP parameters in a different direction by still using the GME formalism but applying it to the stochastic version of the linear inverse ME formulations of the under-determinacy problem, recognising in this way the stochastic nature of economic processes as well as measurement errors and other disturbances among observations (see Golan et al. 1996: 85–88). Adding error terms to decision variables in GME applications resolutely contributed to move mathematical programming towards econometric modelling. It made the GME estimator a smooth channel into the estimation process of PMP parameters, starting from ill-posed problems but also covering well-posed problems when the number of available observations on decision variables exceeds the number of unknown parameters to estimate as in standard econometrics. A couple of subsequent papers (e.g., Buysse et al. 2007a; Cortignani and Severini 2012) took that direction as, for instance, the rich FADN dataset of the European Commission became more commonly used. More recent papers (e.g., Jansson and Heckelei 2011; Paris 2015, 2017) have also capitalised on that direction, introducing measurement errors in observed decision variables as well as observed prices and input requirements.

The GME estimator has been used to estimate directly the parameters of the first-order conditions of the optimisation model at hand. This had a serious advantage compared with estimating a system of derived behavioural functions as in a standard duality-based econometric estimation (e.g., Chambers and Just 1989; Guyomard et al. 1996; Moro and Sckokai 1999; Gorddard 2013) for several reasons outlined in Heckelei and Wolff (2003). First, using directly the optimality conditions as data constraints in a GME approach avoids deriving closed-form solutions for these duality-based behavioural functions and, hence, widens the choice of functional forms and technological or institutional constraints. Second, it allows the use of a more complex structure of the optimisation model by means of several limiting inputs and constraints. With the inclusion of complementary slackness conditions, it is possible to address the additional problem that some of these constraints may not necessarily be binding for every observation in the sample. Third, the resulting simulation model is as explicit as the initial optimisation model to which possible additional constraints can be incorporated without jeopardising the validity of the estimated structural parameters. Gocht (2005), however, reported computational difficulties in finding solutions for estimation problems that rely on a gradient-based solver. The direct estimation of first-order conditions may raise additional computational difficulties when these conditions include a large number of inequality constraints, as in complex agricultural trade simulation models. As explained below, the so-called bi-level optimisation programme can eventually cope with these difficulties.

The GME approach can also easily incorporate out-of-sample information on key parameters or shadow prices of limiting inputs in the estimation process. For example, externally estimated supply elasticities to prices have been used within the GME estimation process (e.g., Heckelei and Wolff 2003; Cortignani and Severini 2012; Graveline and Mérel 2014; Petsakos and Rozakis 2015). The use of prior information on elasticities in the GME estimation process, however, depends on the availability of their analytical expression. Heckelei and Wolff (2003) derived an analytical expression of the land allocation elasticities with respect to own gross margins, that are a function of the unknown parameters to be estimated for the simple Leontief production function with a quadratic cost adjustment, subject to a single land constraint. They showed that the use of such external elasticities improves the convergence of the GME estimator towards the true parameter values, particularly with small sample sizes.

Some other analytical expressions of supply elasticities were derived in the recent literature for specific programming models. For instance, Mérel and Bucaram (2010) provided compatible closed-form expressions for own-price supply elasticities implied by a quadratic programming model characterised with a production technology that is of either Leontief or Constant-Elasticity-of-Substitution (CES) type, subject to a single binding constraint. Mérel et al. (2011) extended these closed-form expressions for the case where the production technology is of generalised CES type with decreasing returns to scale but still in the context of one single binding constraint. Mérel et al. (2014) used this closed-form expression to precisely calibrate their generalised CES programming model against acreages, yields, and exogenous crop supply elasticities to evaluate nitrogen use in cropping. Finally, Garnache and Mérel (2015) stretched these closed-form expressions to the general case of multiple constraints but with fixed proportions among inputs and decreasing returns to scale. All these recently available closed-form expressions of elasticity equations can therefore also be added as constraints to a GME type of estimation procedure, in order to help estimate parameters of Leontief-quadratic, Leontief decreasing returns to scale, CES-quadratic or generalised CES programming models of agricultural supply against a set of exogenous own-price supply elasticities that are taken to specify the range of their support points. Closed-form solutions for more complex programming models may, however, not be always derived.

Available out-of-sample elasticities are, however, most often estimated in a different context and, hence, may not reflect the entire set of resource, technological and institutional constraints present in the programming model to be estimated. Therefore, it is a good practice to provide a sensitivity analysis on those external elasticity estimates when specifying the elasticity support range in the GME estimation process as in Arndt et al. (2002) and Graveline and Mérel (2014). The solution suggested by Mérel and Bucaram (2010) and Garnache and Mérel (2015) is to include into the calibration process of the second-order parameters only those constraints implicit in the external estimates and to introduce the other constraints when calibrating first-order parameters and using the ultimate simulation model. This procedure could eventually be imitated in the GME estimation process.

Exogenous rental rates of limiting resources were also used within the GME estimation process (e.g., Heckelei and Wolff 2003; Cortignani and Severini 2012; Graveline and Mérel 2014) when confronted with limited data information. These rental rates may, however, not be always available for the situation under study at the appropriate disaggregation level when information on markets for those resources is not fully transparent to the modeller. For such a situation, Garnache et al. (2017) proposed a method to calibrate those shadow values of any binding constraints that are not observable to the reference conditions, subject to the replication of exogenous supply responses. They applied this method for the fixed-input proportions, quadratic-cost and power models, the CES-quadratic model and the generalised CES model. This method of calibrating shadow values could also be eventually transposed when specifying the range of the support points in a GME estimation process.

From practitioners of GME and GCE, influences of prior information on parameter estimates can be synthesized as follows. On the one hand, parameter estimates are indeed sensitive to the design of the sets of their support points, especially their support end-point values (Paris and Howitt 1998; Léon et al. 1999; Oude Lansink 1999; Heckelei and Britz 2000; Paris 2001a, b; Arndt et al. 2002; Mittelhammer et al. 2013). When there is little or no reliable prior information about the plausible values of these parameters, it is therefore recommended to specify relatively wide bounds on model parameters. On the other hand, parameter estimates begin to be relatively insensitive to bounds on error terms when these bounds are specified wider than three standard deviations from the expected value (Preckel 2001). However, widening the error bounds reduces the fit of the parameter estimates (Lence and Miller 1998). Finally, as the support for either or both the parameters and the errors widens, parameter values obtained from the entropy estimation approach those obtained from the least-squares estimation of linear regression (Preckel 2001). Where it is possible, increasing the number of observations also helps reduce the influence of the selected support points on the estimation outcomes (Heckelei and Wolff 2003; Mittelhammer et al. 2013). As the sample size increases, such influence can be further modulated by decreasing progressively the weight of the prior-related probabilities.

Although statistical inference on estimated parameters is not the prime objective of the GME estimation process for calibrating programming models, there are ways to perform conventional asymptotic tests for GME estimates. Mittelhammer et al. (2013) developed formulae to perform those statistical tests on model parameters for the data-constrained GME and GCE estimators of the general linear model. Marsh et al. (2014) defined asymptotically tests capable of performing extended asymptotic tests for the data-constrained GME estimator of the linear simultaneous equations model. So far, these statistical tests have not been applied in the available empirical literature. Actually once these parameters are estimated, information theory suggests to use all parameter estimates, significant or not, in the simulation model since all available information has already been used in the estimation process. Doing otherwise would imply the existence of additional information, a possibility that has already been ruled out according to Arndt et al. (2002).

In addition to the difficulty in deriving closed-form expressions for elasticities from complex programming models if we want to use them as priors, Heckelei et al.

(2008) explained some other difficulties of the GME and GCE estimation processes. First, the prior information employed in those entropy estimation processes actually results not only from interactions between the chosen discrete support points and the corresponding reference prior probability distribution on these support points, but also from the final probability distributions on the support points implied by either the maximum or minimum cross-entropy criterion. These combined effects cause poor transparency on prior information. Even in the case that the reference probability distribution over the sets of support points is taken as uniform as in the standard GME estimation process, prior information is not necessarily uninformative on the parameters of interest as it has been incorrectly interpreted in some of the GME applications. Second, the nature of the estimation objective that is used to combine data information with prior information cannot be characterised easily. This could make the statistical evaluation of the resulting parameter estimates difficult. So far, applications of the statistical tests proposed by Mittelhammer et al. (2013) and Marsh et al. (2014) have not yet been seen in the available empirical literature. Third, the introduction of additional variables and equations into the estimation process increases the computational demand on solving complex optimisation problems. To improve the transparency of the estimation process and relieve the computational demand found in typical GME applications, Heckelei et al. (2008) proposed an alternative Bayesian approach to the solution of under-determined systems of equations.

2.2 Bayesian Approach

The Bayesian approach to estimate unknown values of parameters in cases of a simulation model or unknown values of variables in cases of a data reconciliation exercise treats these unknowns as stochastic. It joins the posterior density on these stochastic unknowns to their prior density multiplied by the likelihood function representing information obtained from the data in conjunction with the assumed model. Applied to an under-determined system of structural equations with some prior distribution weights on potential solution values of the unknowns, the Bayesian approach consists in selecting those unknown values that maximise the pre-selected prior probability distribution of those values on the condition that these unknown values are also solutions to the structural system of equations. These optimal values are hence said to provide the Bayesian highest posterior density (HPD) solutions of the equation system. In cases where the prior weighting on those unknown values is not sufficiently informative in the sense that the number of prior probability distributions that are uniform exceeds the number of independent equations, Heckelei et al. (2008) showed that it is still possible to solve for the posterior means of the unknowns as long as the prior uniform distributions integrate to unity. In cases of no informative prior information at all, the posterior mean solutions to the unknowns become the means of these unknowns from among all equally likely values within the prior uniform distributions.

Practically, the Bayesian HPD estimator results in maximising an objective function that corresponds to the posterior density function conditional to the prior density function and the data in conjunction with the structural model. For posterior density functions being the product of independent and normally distributed densities of the unknowns, their natural logarithmic transformation gives a sum of squares of deviations between the observed and unknown values to minimise, each deviation being normalised by their estimated variance. Typically prior normal densities include those of measurement errors of key random parameters and variables, the first- and second-order conditions of the optimisation model, the definitions of the error terms and, possibly, mathematical expressions of prior supply elasticities to prices and dual values of some constraints as seen in the estimation of regional crop supplies by Jansson and Heckelei (2011). Louhichi et al. (2016) similarly applied the HPD estimator but to derive parameters of EU-wide individual farm models with yield-in-value risks in an expected utility framework (following the mean-variance approach with a constant absolute risk aversion specification). Gocht and Britz (2011) used the HPD estimator to disaggregate sector models into farm-type models and Jansson et al. (2014) to estimate a farm-level simulation model with yield-in-value risks in an expected utility framework (following the mean-variance approach with a relative risk aversion specification). However, to render the estimation of these large estimation exercises feasible, these modellers still had to assume binding constraints (Heckelei et al. 2012). Also, an absence of statistical inference could also question the empirical reliability of the parameter estimates and, hence, the reliance on the various simulation results. Aside from these examples, not many studies have followed the Bayesian approach so far.

2.3 Bi-Level Programming Approach

The bi-level optimisation programme consists in solving an optimisation problem, called the outer problem, using the solutions of another optimisation problem called the inner problem, as its domains (Heckelei and Britz 2005). The outer problem, for example, optimises a statistical benchmark such as a weighted least-squares, entropy or Bayesian estimator, while the inner problem includes the first- and second-order conditions of the optimisation problem. The bi-level optimisation programme ends up identifying the optimal values of the parameters or variables with respect to the statistical criterion such that these values also satisfy the optimality conditions of the inner problem. When optimality conditions comprise complementarity slackness, numerical difficulties, however, start to appear in obtaining a solution to the bi-level programme, and call for specific algorithms to be employed (Heckelei and Britz 2005).

Jansson and Heckelei (2009) used a bi-level mathematical programme to estimate the parameters of a transport programming model using observations of regional prices in addition to observations of transports costs and trade flows as well as complementary slackness conditions for zero trade flows. In this case, the inner problem is composed of the optimality conditions of the transport problem, while the outer problem minimises the weighted sum of squared measurement errors on prices and trade costs. They showed that their resulting bi-level programme estimates regional prices as well as trade costs more precisely than the traditional calibration method applied in transport models. They used two complementary algorithms based on smooth approximations from Ferris et al. (2002) to solve their bi-level programme with complementary slackness conditions.

Britz and Arata (2019) applied the concept of bi-level optimisation with complementary slackness conditions to estimate individual farm models with yield and price risks in an expected utility framework (following the mean-variance approach with a relative risk aversion coefficient). In their case, the inner problem maximises a per hectare (ha) expected utility that is composed of a per ha expected revenue, a per ha cost function and a variance function with the relative risk aversion coefficient, where the decision variables are crop shares. The outer problem minimises the weighted sum of squared measurement errors on the allocated crop shares and per ha total costs. The per ha cost function that is used is, however, not in line with the standard definition of a per ha dual cost function and fails to meet all theoretical properties. The bi-level programming model achieves its purpose, in particular in dealing with inequality constraints such as sale quotas and set-aside obligations, for estimating the model parameters. They used the so-called extended mathematical programming (Ferris et al. 2009) package of the general algebraic modelling system (GAMS) software to formulate automatically the first-order conditions of their bi-level mathematical programme and solve it (see Vicente and Calamai 1994). They also successfully derived confidence intervals around some of the parameter estimates based on the Cramer-Rao bound and numerical estimation of the Fisher information matrix. Goodness of fit seems satisfactory for the estimated model. Sinha et al. (2017) provided an extensive recent review on bi-level optimisation and applications.

2.4 Standard Econometric Approach

Obviously with a sufficiently large number of degrees of freedom, the estimation problem becomes well-posed. Estimators such as least-squares (LS) or generalised methods of moments (GMM) can therefore more appropriately be used in place of entropy-based and Bayesian methods, and statistical inference on parameter estimates can be more straightforwardly performed. This is what Henry de Frahan et al. (2011) undertook for estimating parameters of individual farm models, but in two steps. First, they estimated a multi-output, multi-input flexible cost function for each individual sample farm. Second, they embedded each individual farm cost function into the objective function of a profit-maximisation programming model with several constraints that is designed and calibrated for each farm of the sample as well as with a regional land constraint. But any other regional constraint that restrict the volume of output sales or pollution emissions can be added into the structure of the individual programming models.

There are several advantages to this approach—namely: theoretical restrictions can be easily verified or imposed in the estimation step, estimated parameters encapsulate constraints whether they are binding or not, full statistical inference can be easily performed on the estimated parameters, and resource, technological and institutional constraints can still stay explicit in the programming model. One disadvantage, however, is that the programming model still needs to be augmented with a calibration term as in any other simulation model.

Arfini and Donati (2011) adopted a similar approach for estimating and calibrating parameters of individual farm models, but collapsed the two steps into a single one. Here, however, the error terms whose squares are to be minimised are actually the linear terms of the activity quadratic cost function. Additional restrictions are also included in the combined estimation and calibration process: (1) a relationship between variable marginal costs derived from a linear function and marginal costs derived from an activity quadratic cost function; (2) a relationship between the total estimate of the explicit specific costs and the total farm accounting variable costs; (3) a relationship between total marginal costs, including the marginal use cost of the fixed factors of production, and total marginal revenues, and; (4) an equality between the values of the objective function of the primal problem and the objective function of the corresponding dual problem. This last equality actually imposes a long-run equalisation between total costs (including rents to fixed factors of production), and total revenues upon the parameter estimation problem. Such a constraint was not necessary imposed upon previous attempts at structural parameter estimations. It is, however, questionable whether individual farms observed in the sample have really reached their long-run equilibrium, meeting the zero-profit condition.

This specification has the advantage that estimation and calibration are completed simultaneously. The disadvantages of it are that estimated parameters reflect a long-run equilibrium of the farm and are not prone to straightforward statistical inference. Paris (2011: 400) showed that the result of this least-squares minimisation problem precisely corresponds to the principal set-up of the original PMP specification. However, the minimisation problem here no longer includes, explicitly, the much discussed 'tautological' relation of the first PMP stage stating that optimal levels of activities be less or equal to their observed levels. But it implicitly includes the first-order conditions of the final model specification, hence removing the alleged disconnect between the first and second PMP stages (Paris 2011: 399).

Using a similar primal–dual approach, Arata et al. (2017) estimated individual farm models with price risks in an expected utility framework (following the mean-variance approach with a constant absolute risk aversion specification). Their approach consists of merging the primal model of the first PMP stage with its corresponding dual model and adding the nonlinear equation of the second PMP stage as a constraint. The combination of the objective functions of the primal and dual models of the farmer's expected utility maximisation problem is included in the objective function of the final specification of the programming model to estimate as in Paris (2011: 398). Here again, the estimated parameters of the expected utility maximisation problem reflect a long-run equilibrium of the individual sample farms. This estimation procedure is also not amenable to straightforward statistical inference.

Paris (2017) also merged the formulation of the primal model of the first PMP stage with its corresponding dual formulation. Here, however, he assumed that the observed values of the activity levels as well as the observed values of the proxies for limiting input duals contain some measurement error as initiated by Heckelei and Wolff (2003). First, the sum of squared deviations of both activity levels and limiting input duals is minimised, as in a weighted least-squares estimation procedure subject to the duality relations of the initial PMP model. Whether the observed values are strictly positive or null, Paris (2017) demonstrated that the least-squares solution is unique. Second, the farm-specific parameters of the cost function are estimated by importing the optimal least-squares solutions of the activity levels and limiting input duals into the output marginal cost and limiting input demand functions. If some priors on regional output supply and limiting input demand elasticities are available, it is possible to insert their mathematical expressions in the estimation of the cost parameters with their priors. The final model to estimate can be highly nonlinear in the constraints. In this case, Paris (2017) reported that the Branch-And-Reduce Optimisation Navigator (BARON) solver of the GAMS software achieved an equilibrium solution. Note that here there is no assumption that the sample farms are achieving a long-run equilibrium with zero profit. The estimated parameters are still not amenable to statistical inference.

Paris (2015) applied the same approach to estimate individual farm models with price risks within a mean-standard deviation utility function framework, with any combination of risk preferences represented by absolute risk aversion and relative risk aversion. What is interesting in both cases is that the dual formulation of the profit-maximisation problem or mean-standard utility maximisation problem is used together with its primal formulation to estimate the measurement error terms on activity levels and limiting input duals. Still, the first two PMP stages are used: the first one to obtain the calibrating least-squares values of the activity levels and limiting input duals; the second to estimate the parameters of the cost function of interest. Our methodological proposal developed in the next section consists of merging both stages together so as to avoid the inconsistency highlighted between the traditional first and third PMP stages, as recommended by studies since the seminal paper of Heckelei and Wolff (2003).

2.5 Alternative Model Specifications

Mérel and Howitt (2014) already reviewed some model specifications and have shown the necessity for those specifications to produce globally convex models for calibration and simulation purposes. In the PMP-related literature, this necessity has typically been translated into a nonlinear, concave objective function subject to a convex set of constraints in the case of a profit-maximisation problem. Beyond this practical programming necessity, justifications for the nonlinearity of the concave profit function differed in the same literature. It is, however, possible to group them in two distinct strands. A first strand justified it on empirical grounds arguing that it captures any type of model mis-specification, data errors, aggregation bias, land heterogeneity, fixed factors, capacity constraint, risk behaviour and price expectations (Howitt 1995a, b; Heckelei and Britz 2000; Helming et al. 2001; Kanellopoulos et al. 2010; Louhichi et al. 2010; Gocht and Britz 2011; Jansson and Heckelei 2011; Heckelei et al. 2012; Louhichi et al. 2016; Arata et al. 2017). A second strand used a more standard theoretical view, arguing that it represents the second-order approximation of a dual cost function that depends on output quantity levels but also in some instances on input prices and fixed inputs (Paris and Howitt 1998; Paris 2001a, b, 2011, 2015, 2017; Buysse et al. 2007a, b; Arfini et al. 2008; Arfini and Donati 2011; Henry de Frahan et al. 2011; Britz and Arata 2019). Heckelei and Britz (2005), Mérel and Howitt (2014), and Heckelei et al. (2012), however, concluded that those justifications from the first strand offer limited rationalisation of the nonlinear terms of the ultimate objective function.

PMP-related model specifications mostly varied according to how the nonlinearity of the final objective function is actually implemented. Some applications placed it on the revenue side, whereas some others on the cost side of the profit function. In practice, few applications (7 out of 36 recorded) aimed to recover parameters of marginal revenues (Howitt 1995a; Heckelei and Wolff 2003; Mérel et al. 2011, 2014; Graveline and Mérel 2014; Garnache and Mérel 2015; Garnache et al. 2017), while most applications (29 out of 36 recorded) attempted to recover parameters of marginal costs (Howitt 1995b; Paris and Howitt 1998; Heckelei and Britz 2000; Graindorge et al. 2001; Júdez et al. 2001; Paris 2001a, b, 2015, 2017; Heckelei and Wolff 2003; Henry de Frahan et al. 2007, 2011; Buysse et al. 2007a, b; Arfini et al. 2008; Cortignani and Severini 2009; Kanellopoulos et al. 2010; Louhichi et al. 2010; Mérel and Bucaram 2010; Arfini and Donati 2011; Jansson and Heckelei 2011; Cortignani and Severini 2012; Frisvold and Konyar 2012; Medellín-Azura et al. 2012; Jansson et al. 2014; Louhichi et al. 2016; Britz and Arata 2019; Arata et al. 2017). Some of these applications also added the recovery of parameters of input-derived demands (Howitt 1995b; Graindorge et al. 2001; Helming et al. 2001; Heckelei and Wolff 2003; Henry de Frahan et al. 2011; Medellín-Azura et al. 2012; Graveline and Mérel 2014; Mérel et al. 2014; Garnache et al. 2017), limiting input-derived demands (Paris 2001a, b, 2015, 2017), revenue (Arfini et al. 2008) and risk-aversion behaviours (Louhichi et al. 2010; Cortignani and Severini 2012; Paris 2015; Petsakos and Rozakis 2015; Britz and Arata 2019; Arata et al. 2017).

When the nonlinearity was placed on the cost side, the most widely used functional form was the quadratic form. In some instances, the generalised Leontief or the weighted-entropy functional forms were used as well (Paris and Howitt 1998; Paris 2001a, b, 2015, 2017). When the nonlinearity was placed on the revenue side, functional forms included the yield function, the Cobb-Douglas, the CES or the power forms. As long as the functional form meets the standard theoretical properties of a cost or a revenue function, any functional form can be used. Its choice then mostly depends on the data and the problem at hand.

PMP-related model specifications also varied according to the identification of the decision variables. Following the tradition of mathematical programming models designed for the agricultural sector, most applications (23 out of 35 recorded) defined their decision variables in terms of allocable fixed resources (i.e., farm land or livestock) while fewer (12 out of 35 recorded) were defined in terms of output quantities. Three quarters of the applications selecting allocable fixed resources for decision variables also recovered the parameters of a cost function, the other quarter recovering the parameters of a revenue function. All applications, except a few, selecting output quantities for decision variables recovered the parameters of a cost function. These applications selecting output quantities for decision quantities may have the advantage of implicitly endogenizing yield formation. Some applications selecting allocable fixed resources as the decision variables, endogenized yield formation by adding a production function with input substitution (Howitt 1995b; Graindorge et al. 2001; Heckelei and Wolff 2003; Mérel et al. 2011, 2014; Frisvold and Konyar 2012; Medellín-Azura et al. 2012; Graveline and Mérel 2014; Garnache et al. 2017). These specifications remove the need to use a discrete approach specifying different activity variants to capture variations in intensive margin as used in Röhm and Dabbert (2003), Cortignani and Severini (2009) and Louhichi et al. (2010). As long as the revenue function or the cost function respects the standard theoretical properties, there is no strong argument to privilege one specification over another. What should determine the choice of one specification over the other is, rather, the data and the problem at hand.

As already discussed above, PMP-related model specifications also varied according to whether decision variables or parameters are deterministic or random. As noted in the review of Heckelei et al. (2012), discussing which variables or parameters ought to be treated as deterministic or random is not yet well established in the PMP-related literature. But this is actually the new trend when using econometric estimation methods such as the LS, GMM, ME and Bayesian estimators as well as the bi-level programming approaches.

With the exception of a few more recent papers, all papers reviewed above relied on the primal set-up of the profit-maximisation problem to derive the first- and second-order conditions to calibrate or estimate the parameters of interest. Instead, Arfini and Donati (2011), Paris (2015, 2017), and Arata et al. (2017) (reviewed above) relied on both the primal and the dual set-up of the profit-maximisation problem. This primal–dual approach was actually instigated by a previous paper of Paris (2001a) further developed in his book (Paris 2011). Paris (2011: 357) motivated his extension of PMP for avoiding the risk of degeneracy of some of the dual variables associated with the presence of multiple structural constraints. First, he constructed a symmetric structure of the primal and dual constraints of the maximisation problem and formulated it as an equilibrium model by resorting to the market price of all limiting inputs. Information on prices that is generally available at least at the regional level for land, water, capital, labour, and production or sale rights can, therefore, be useful. The symmetry of the new structure was judiciously obtained by inserting the possibility to rent out limiting inputs at market prices. The solution of this equilibrium model generated estimates of activity levels, effective supplies of limiting inputs, total marginal costs of activities and marginal costs of limiting inputs. Second, using a flexible cost function, he estimated the parameters of the marginal cost function of activities and the derived demand function for limiting inputs. Third, he expressed the final model as an equilibrium problem between marginal costs and marginal revenues of the output activities, on the one hand, and derived demand and supply functions of the limiting inputs, on the other hand. Because of this equilibrium and symmetry, Paris named his novel formulation as a symmetric positive equilibrium problem (SPEP). Britz et al. (2003), however, raised some doubts about Paris's (2001a) novel procedure. Their fundamental critique was again the inconsistency between the outcomes of the first and third stages of the procedure. Some other critiques were later addressed in the book of Paris (2011: 357–361).

In the same paper, Paris (2001a) also proposed a solution to the self-selection process with respect to the choice of activities that is common in farm surveys, where different farms of the survey sample may select to be actively engaged in different subsets of the available activities for unclear reasons. In such a case, parameters for latent activities cannot be calibrated or estimated at the individual level. To overcome this self-selection problem during the first and second PMP stages, Paris (2001a) added a supplementary PMP model for the overall sample to the individual farm models and calibrated a frontier cost function for every activity observed in the overall sample. In the third PMP stage, the final calibrated individual farm models can eventually resort to parameters calibrated for every possible activity observed in the overall sample to simulate the emergence of activities that were not necessarily observed in the baseline situation of the individual farms. This procedure was also further developed in Paris's (2011: 348–353) book. The problem raised by Britz et al. (2003) for this solution concerned the distribution of costs and prices across the sample that might bias estimates that are recovered from the overall constructed sample. It seems that this procedure would not be appropriate when there are not too many unobserved activities across the individual farms of the sample.

Finally, for the sake of completeness, Paris (2001b) extended SPEP to include a dynamic structure in it. In the dynamic positive equilibrium problem (DPEP) that he set up for this purpose, also developed in his book (Paris 2011: 361–369), the underlying dynamic connection of activity levels through time was modelled by a process of adaptive expectations for output prices. The objective function to maximise in this problem consisted of the discounted stream of profits over a time horizon as well as the discounted value of profits from this time horizon to infinity that is represented as a salvage function. From this set-up, Paris (2001b) derived an equilibrium problem with structural relations similar to those of the SPEP but to solve backward in time. Stages 2 and 3 of DPEP were set up and solved as their corresponding SPEP stages with some alterations for taking care of the dynamic character of the new equilibrium model. Again, this novel procedure went through the traditional three PMP stages, which results in the same inconsistency already reported between the first and the third stages.

A common problem where Paris (2001a, b, 2011, 2015, 2017) merged the primal model of the first PMP stage with its corresponding dual formulation is also the use of a cost functional form that is not from standard production economics since it includes the prices, not the levels, of the limiting inputs. With the use of a standard cost function expressed in limiting input levels, then the derivative of such a function with respect to a limiting input quantity must equal minus the price of the associated limiting input.

Consequently, the derived demand for a limiting input should follow the schedule of minus this derivative over its associated limiting input level, as the derived demand for a variable input follows the schedule of its marginal value productivity over its associated variable-input level.

2.6 Preliminary Conclusions

In sum, over the last two decades, the PMP-related literature enriched itself with alternative approaches and model specifications in the aim of estimating the nonlinear terms that are much critical for achieving calibrated optimisation programmes and obtaining sound simulation responses. While the entropy approach has dominated the first decade since the 1998 AJAE paper of Paris and Howitt, the Bayesian approach as well as more standard econometric approaches that can, in turn, be embedded into bi-level programmes have started to take off thanks mainly to contributions from Heckelei and his close collaborators. Notably due to contributions from Paris, model set-up diversified itself, some of them being still questioned in the academic literature. Methodological solutions started to emerge to tackle this challenging objective.

From this comprehensive review, it is possible to retain several key elements for motivating the methodological contribution developed in the next section. This contribution fits into this general aim of estimating and calibrating simultaneously optimisation model parameters ready to use for performing policy simulation. These elements are part of the following list:

(1) estimating directly the first- and second-order conditions of the ultimate optimisation model, skipping therefore the first PMP stage;
(2) using a flexible functional form for the cost function that fulfils every standard theoretical property;
(3) setting up both the primal and dual of the problem at hand, i.e., adding the dual formulation of the profit-maximisation problem to its primal formulation;
(4) defining the decision variables as random variables;
(5) using all available information and observations, including prices and quantities of limiting inputs and, possibly, prices and quantities of variable inputs;
(6) having the possibility to use priors of elasticities for output supplies and derived input demands;
(7) relying on standard econometric methods to estimate parameters with the possibility to resort to a bi-level programme if the presence of complementary slackness conditions results in numerical difficulties; and
(8) having the opportunity to accommodate the simulation model with additional constraints not yet included into the estimation process.

3 Bridging Positive Mathematical Programming and Econometric Mathematical Programming

Let us start immediately with the ultimate specification of the programming model to optimise as advocated in Heckelei and Wolff (2003), but express its specification as in Paris (2017) with an additional constraint expressing the true activity level being equal to the observed activity level augmented with an error term as in Heckelei and Wolff (2003). This error term reflects some measurement error that either understates or overstates the level of the activity level consistent with the technical and economic information of a given producing unit as stated in both papers. Let us further assume that this error term is stochastic with mean zero and a standard deviation σ.

The following programming model is set up as for one producing unit f for every time period t in vector notation where lower-case, bold-faced letters represent items that are column vectors for each time period t, upper-case bold-faced letters represent matrices, and upper-case italic letters to represent scalars. When needed, the dimensions of the vectors and matrices are denoted by upper-case letters and the indices of the elements of these vectors or matrices by a lower-case version of the same letter. As usual, the prime character ($'$) denotes the ordinary transpose of a vector or a matrix and the (-1) exponent the inverse of a matrix. Since all individual models have identical structure and no cross-unit constraints or relationships are assumed here to keep the presentation simple, indices for producing units can be omitted in this generic presentation. The individual programming model for each period t can then be written as:

$$MAX_{\mathbf{x_t},\mathbf{h_t}} TNR_t = \mathbf{p}_\mathbf{t}'\mathbf{x_t} - \mathbf{c}'\mathbf{x_t} - 1/2\mathbf{x}_\mathbf{t}'\mathbf{Q}\mathbf{x_t} \tag{1}$$

subject to

$$\mathbf{A_t}\mathbf{x_t} \leq \mathbf{b_t} \quad [\mathbf{y_t}] \quad \text{dual variables } \mathbf{y_t} \tag{2}$$

$$\mathbf{x_t} = \mathbf{x}_\mathbf{t}^\mathbf{o} + \mathbf{h_t} \quad [\boldsymbol{\lambda}_\mathbf{t}] \quad \text{dual variables } \boldsymbol{\lambda}_\mathbf{t} \tag{3}$$

with the vectors $\mathbf{x_t} \geq 0$ and $\mathbf{h_t}$ free, where the scalar TNR_t represents the total net revenue of the producing unit f to maximise in period t; the vector $\mathbf{x_t}$ represents the true activity levels that are preferably expressed in output quantities for each of the J production activities to conform to a standard cost function; the vector $\mathbf{p_t}$ represents the J vector of output prices when activity levels are expressed in output quantities; the vector $\mathbf{c}$ represents the J vector of linear parameters of the quadratic cost function; the matrix $\mathbf{Q}$ represents the $J \times J$ symmetric positive definite matrix of quadratic parameters of the quadratic cost function; the matrix $\mathbf{A_t}$ represents the $K \times J$ matrix of limiting input coefficients for K limiting inputs; the vector $\mathbf{b_t}$ represents the K vector of total availability of limiting inputs; the vector $\mathbf{x}_\mathbf{t}^\mathbf{o}$ represents the observed activity levels; the vector $\mathbf{h_t}$ represents the error terms due to measurement on these activity levels; and the two vectors $\mathbf{y_t}$ and $\boldsymbol{\lambda}_\mathbf{t}$ represent the dual variables corresponding to the constraints for limiting inputs and the equalisations of the true and observed activity levels thanks to the error terms, respectively. The dual variable $\boldsymbol{\lambda_t}$ can be interpreted as being the marginal effect of the error term about the true respective activity level on the true total net revenue.

Note that the quadratic cost function here is expressed in its simplest form, i.e., depending solely on the activity levels, not on variable-input prices and limiting input levels. This functional form can therefore be extended by first adding to it the variable-input prices as is done with the symmetric generalised McFadden (SGM) functional form in Henry de Frahan et al. (2011) or with the generalised Leontief functional form in Paris (2017). For this extension, both quantities and prices of variable inputs need to be available since the Shephard's lemma applied to this extended cost function provides the variable-input demand functions. Variable-input quantities are most often available from farm data, in particular the FADN dataset, and variable-input prices are also often observable at least at the regional level, if not in the form of an input price index. As for the activity levels, let us then express the true variable-input price level as being equal to the observed variable-input price level augmented with an error term. In such a case, it is possible to add a set of constraints to the previous programming model, such as the true levels of variable-input prices are equal to the observed levels of variable-input prices augmented with error terms due to measurement error on these prices. A system of two sets of equations to estimate is then formed, which consist of the derivatives of total costs with respect to outputs and derivatives of total costs with respect to variable-input prices. This extension is not further developed here so as to focus on the methodological approach, rather than on the detailed specification of the programming model.

This functional form can also be further extended by adding to it the limiting input quantities with measurement errors on these quantities and, hence, removing the corresponding limiting input binding constraints. For this extension, quantities and prices of limiting inputs need to be available since the derivatives of this extended cost function with respect to their limiting input quantities must equal minus the prices of their associated limiting inputs. Limiting input quantities are most often available from farm data, in particular the FADN dataset, and limiting input prices are also often the case at least at the regional level, if not in the form of an input price index. As for the activity levels and the variable-input prices, let us then express the true limiting input levels being equal to the observed limiting input levels augmented with an error term. In such a case, it is possible to add a set of constraints to the previous programming model such as the true levels of limiting inputs are equal to the observed levels of limiting inputs augmented with error terms due to measurement error on these levels. A system of three sets of equations to estimate is then formed: derivatives of total costs with respect to outputs, derivatives of total costs with respect to variable-input prices and derivatives of total costs with respect to limiting input quantities. This fully-fledged cost function removes the last restriction of fixed production coefficients that could have been still present in the limiting input binding constraints. This extension is not further developed here to focus on the methodological approach, not on the detailed specification of the programming model.

Furthermore, as in Heckelei and Wolff (2003) in the context of an entropy estimator, in Jansson and Heckelei (2011) in the context of a Bayesian estimator or, in particular, in Paris (2017) in the context of a least-squares estimator, let us add in the objective function (Eq. 1) the sum of squared deviations $\mathbf{h_t'W_th_t}$ to minimise where the matrix $\mathbf{W_t}$ is diagonal with elements $p_{jt} > 0$ on the main diagonal, $j = 1, \ldots, J$ to harmonise the units of measurement in the new objective function. The Lagrange function and the corresponding necessary Kuhn–Tucker conditions become as the following expressions:

$$LMAX_t(\mathbf{x_t}, \mathbf{h_t}, \mathbf{y_t}, \boldsymbol{\lambda}_\mathbf{t}) = \mathbf{p_t'x_t} - \mathbf{c'x_t} - \frac{1}{2}\mathbf{x_t'Qx_t} - \frac{1}{2}\mathbf{h_t'W_th_t} + \mathbf{y_t'}[\mathbf{b_t} - \mathbf{A_tx_t}] + \boldsymbol{\lambda}_\mathbf{t}'\left[\mathbf{x_t^o} + \mathbf{h_t} - \mathbf{x_t}\right] \tag{4}$$

$$\frac{\partial LMAX_t}{\partial \mathbf{x_t}} = \mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} - \mathbf{A_t'y_t} - \boldsymbol{\lambda}_\mathbf{t} \le 0 \tag{5}$$

$$\frac{\partial LMAX_t}{\partial \mathbf{h_t}} = -\mathbf{W_th_t} + \boldsymbol{\lambda}_\mathbf{t} = 0 \tag{6}$$

$$\frac{\partial LMAX_t}{\partial \mathbf{y_t}} = \mathbf{b_t} - \mathbf{A_tx_t} \ge 0 \tag{7}$$

$$\frac{\partial LMAX_t}{\partial \boldsymbol{\lambda}_\mathbf{t}} = \mathbf{x_t^o} + \mathbf{h_t} - \mathbf{x_t} = 0 \tag{8}$$

$$\mathbf{x_t}\frac{\partial LMAX_t}{\partial \mathbf{x_t}} = \mathbf{x_t}\left(\mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} - \mathbf{A_t'}\,\mathbf{y_t} - \boldsymbol{\lambda}_\mathbf{t}\right) = 0 \tag{9}$$

$$\mathbf{y_t}\frac{\partial LMAX_t}{\partial \mathbf{y_t}} = \mathbf{y_t}(\mathbf{b_t} - \mathbf{A_tx_t}) = 0 \tag{10}$$

$$\mathbf{x_t} \ge 0 \tag{11}$$

$$\mathbf{y_t} \ge 0 \tag{12}$$

Equations 5 and 6 provide together:

$$\mathbf{A_t'y_t} + \mathbf{W_th_t} \ge \mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} \tag{13}$$

Let us now state the dual of model Eqs. (1)–(3) as in Paris (2017):

$$MIN_{\mathbf{y_t},\mathbf{u_t}} TC_t = \mathbf{b_t'}y_\mathbf{t} + \frac{1}{2}\mathbf{x_t'Qx_t} + \left(\mathbf{x_t^o} + \mathbf{h_t}\right)'\boldsymbol{\lambda}_\mathbf{t} \tag{14}$$

subject to

$$\mathbf{A_t'y_t} + \boldsymbol{\lambda}_\mathbf{t} \ge \mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} \quad [\mathbf{x_t}] \quad \text{dual variables } \mathbf{x_t} \tag{15}$$

$$\mathbf{y_t} = \mathbf{y_t^o} + \mathbf{u_t} \quad \left[\boldsymbol{\psi}_\mathbf{t}\right] \quad \text{dual variables } \boldsymbol{\psi}_\mathbf{t} \tag{16}$$

with the vectors $\mathbf{y_t}$ and $\mathbf{x_t} \geq 0$ and the vectors $\mathbf{h_t}$ and $\mathbf{u_t}$ free, where the scalar TC_t represents the total cost of the producing unit f to minimise in period t; the vector $\mathbf{y_t}$ represents the shadow price levels for each of the K limiting inputs; the vector $\mathbf{y_t^o}$ represents the observed price levels for each of the K limiting inputs; the vector $\mathbf{u_t}$ represents the error terms due to measurement on these price levels; the two vectors $\mathbf{x_t}$ and $\boldsymbol{\psi}_\mathbf{t}$ represent the dual variables corresponding to the constraints for total marginal costs being greater or equal to output prices and the equalisations of the shadow and observed limiting input price levels thanks to the error terms, respectively; and the other vectors and matrices are defined as in the primal model (1)–(3). The dual variable $\boldsymbol{\psi}_\mathbf{t}$ can be interpreted as being the marginal effect of the error term about the true shadow price level of the respective limiting input on the true total cost.

As in Paris (2017) in the context of a least-squares estimator, let us also add in the objective function (Eq. 14) the sum of squared deviations $\mathbf{u_t'V_tu_t}$ to minimise where the matrix $\mathbf{V_t}$ is diagonal with elements $b_{kt}/y_{kt}^o > 0$ on the main diagonal, $k = 1, \ldots, K$ to harmonise the units of measurement in the new objective function. The Lagrange function and the corresponding necessary Kuhn–Tucker conditions become as the following expressions:

$$\begin{aligned} LMIN_t(\mathbf{y_t}, \mathbf{x_t}, \mathbf{u_t}, \boldsymbol{\psi}_\mathbf{t}) = {} & \mathbf{b_t'y_t} + \frac{1}{2}\mathbf{x_t'Qx_t} + (\mathbf{x_t^o} + \mathbf{h_t})'\boldsymbol{\lambda}_\mathbf{t} + \frac{1}{2}\mathbf{u_t'V_tu_t} \\ & + \mathbf{x_t'}[\mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} - \mathbf{A_t'y_t} - \boldsymbol{\lambda}_\mathbf{t}] + \boldsymbol{\psi}_\mathbf{t}'[\mathbf{y_t^o} + \mathbf{u_t} - \mathbf{y_t}] \end{aligned} \tag{17}$$

$$\frac{\partial LMIN_t}{\partial \mathbf{y_t}} = \mathbf{b_t} - \mathbf{A_tx_t} - \boldsymbol{\psi}_\mathbf{t} \geq 0 \tag{18}$$

$$\frac{\partial LMIN_t}{\partial \mathbf{x_t}} = \mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} - \mathbf{A_t'y_t} - \boldsymbol{\lambda}_\mathbf{t} \geq 0 \tag{19}$$

$$\frac{\partial LMIN_t}{\partial \mathbf{u_t}} = \mathbf{V_tu_t} + \boldsymbol{\psi}_\mathbf{t} = 0 \tag{20}$$

$$\frac{\partial LMIN_t}{\partial \boldsymbol{\psi}_\mathbf{t}} = \mathbf{y_t^o} + \mathbf{u_t} - \mathbf{y_t} = 0 \tag{21}$$

$$\mathbf{y_t}\frac{\partial LMIN_t}{\partial \mathbf{y_t}} = \mathbf{y_t}(\mathbf{b_t} - \mathbf{A_tx_t} - \boldsymbol{\psi}_\mathbf{t}) = 0 \tag{22}$$

$$\mathbf{x_t}\frac{\partial LMIN_t}{\partial \mathbf{x_t}} = \mathbf{x_t}(\mathbf{p_t} - \mathbf{c} - \mathbf{Qx_t} - \mathbf{A_t'y_t} - \boldsymbol{\lambda}_\mathbf{t}) = 0 \tag{23}$$

$$\mathbf{y_t} \geq 0 \tag{24}$$

$$\mathbf{x_t} \geq 0 \tag{25}$$

Equations 18 and 20 provide together:

$$\mathbf{A_tx_t} \leq \mathbf{b_t} + \mathbf{V_tu_t} \tag{26}$$

By the substitution of constraints (8) and (21) into constraints (13) and (26) and the rearrangement of terms, a system of two sets of equations subject to constraints (11) and (12), or constraints (24) and (25) is obtained:

$$\mathbf{A}_\mathbf{t}'\left(\mathbf{y}_\mathbf{t}^\mathbf{o} + \mathbf{u}_\mathbf{t}\right) + \mathbf{W}_\mathbf{t}\mathbf{h}_\mathbf{t} + \mathbf{c} + \mathbf{Q}\left(\mathbf{x}_\mathbf{t}^\mathbf{o} + \mathbf{h}_\mathbf{t}\right) \geq \mathbf{p}_\mathbf{t} \tag{27}$$

$$\mathbf{A}_\mathbf{t}\left(\mathbf{x}_\mathbf{t}^\mathbf{o} + \mathbf{h}_\mathbf{t}\right) - \mathbf{V}_\mathbf{t}\mathbf{u}_\mathbf{t} \leq \mathbf{b}_\mathbf{t} \tag{28}$$

$$\mathbf{x}_\mathbf{t}^\mathbf{o} + \mathbf{h}_\mathbf{t} \geq 0 \tag{29}$$

$$\mathbf{y}_\mathbf{t}^\mathbf{o} + \mathbf{u}_\mathbf{t} \geq 0 \tag{30}$$

With these four relationships, it is possible to proceed with the estimation of cost parameters $\mathbf{c}$ and $\mathbf{Q}$ for one individual producing unit over a time frame of $t = 1, \ldots, T$ using the following weighted least-squares specification:

$$MIN_{\mathbf{h}_\mathbf{t},\mathbf{u}_\mathbf{t},\mathbf{c},\mathbf{Q}} LS = \sum_{t=1}^{T}\left(\mathbf{h}_\mathbf{t}'\mathbf{W}\mathbf{h}_\mathbf{t} + \mathbf{u}_\mathbf{t}'\mathbf{V}\mathbf{u}_\mathbf{t}\right)/2 \tag{31}$$

subject to Eqs. (27)–(30) and the standard theoretical restrictions on the cost parameters.

If those inequalities (27)–(30) generate computational difficulties, then it is possible to resort to a bi-level programme where those four inequalities form the inner problem together with the theoretical restrictions and the weighted sum of least-squares deviations forms the outer problem to minimise as in Britz and Arata (2019).

Otherwise, if it is reasonable to observe that vectors $\mathbf{x}_\mathbf{t}^\mathbf{o}$ and $\mathbf{y}_\mathbf{t}^\mathbf{o}$ have all positive components, then it is also reasonable to assume that vector $\mathbf{x}_\mathbf{t} > 0$ and vector $\boldsymbol{y}_\mathbf{t} > 0$. Via the complementary slackness conditions (9) and (22), the previous least-squares estimation is computationally simplified with the constraints (27)–(30) in equality mode. If some components of vectors $\mathbf{x}_\mathbf{t}^\mathbf{o}$ and $\mathbf{y}_\mathbf{t}^\mathbf{o}$ are not observed, then the solution to the self-selection process proposed by Paris (2001a) and developed further in Paris (2011: 348–353) can be applied with, however, the caveat raised by Britz et al. (2003). In such a case, the weighted least-squares estimator needs to be applied together for the individual producing units and the constructed regional producing unit.

To increase the number of degrees of freedom, the estimation process can benefit from a resolution over a group of producing units f if it is reasonable to assume that those producing units share the same technology and, hence, the cost parameters $\mathbf{c}$ and $\mathbf{Q}$. If the assumption of fixed technology over time is too restrictive to impose upon the panel estimation, then the inclusion of a time-varying component in the cost function can be done.

If it is reasonable to use constraints (27–30) in equality mode, then they can be rearranged to be used with a standard econometric estimation method such as the nonlinear seemingly unrelated regression estimator over the panel data. In this case, one error term is defined over the two dimensions, e.g., the vector $\mathbf{h_{ft}}$, and the other error term is defined over one dimension, e.g., the vector $\mathbf{u_t}$:

$$\mathbf{x_{ft}^o} = \mathbf{Q}^{-1}\mathbf{p_t} - \mathbf{Q}^{-1}\mathbf{c} - \mathbf{Q}^{-1}\mathbf{A'_{ft}}\mathbf{y_t^o} - \mathbf{Q}^{-1}\mathbf{A'_{ft}}\mathbf{u_t} - (1 + \mathbf{Q}^{-1}\mathbf{W_{ft}})\mathbf{h_{ft}} \tag{32}$$

$$\mathbf{x_{ft}^o} = \mathbf{A_{ft}^{-1}}\mathbf{b_{ft}} + \mathbf{A_{ft}^{-1}}\mathbf{V_t}\mathbf{u_t} - \mathbf{h_{ft}} \tag{33}$$

subject to the theoretical restrictions on the cost parameters. Statistical inference on the estimated parameters becomes here straightforward. Statistical tests and remedies for endogeneity and other risks also become available.

If appropriate supply elasticities to output prices are available, then it is also possible to use their values and add their mathematical expressions into the estimation procedure as in Paris (2017).

If interested in calibrating exactly to a base year for simulation purposes, then it is furthermore possible to impose that the vectors $\mathbf{h_t}$ and $\mathbf{u_t}$ be equal to a vector of zero for t corresponding to the base year as in Arndt et al. (2002). As a result, the model parameters endogenously calibrate themselves at the base year t.

With the estimates of the cost parameters $\hat{\mathbf{c}}$ and $\widehat{\mathbf{Q}}$ and, possibly, of the additional parameters for a cost function also defined on variable-input prices, it is finally possible to formulate a calibrated equilibrium model for each producing unit that is ready for simulations as in Paris (2017). With the estimation procedure imposing the vectors $\mathbf{h_t}$ and $\mathbf{u_t}$ to zero at base year t, then the simulation model for one single producing unit f has the following structure:

$$MIN_{\mathbf{z_{pf}},\mathbf{z_{df}},\mathbf{x_f},\mathbf{y_f}} CSC_f = \mathbf{z'_{df}}\mathbf{x_f} + \mathbf{z'_{pf}}\mathbf{y_f} \tag{34}$$

subject to

$$\mathbf{A'_f}\mathbf{y_f} + \hat{\mathbf{c}} + \widehat{\mathbf{Q}}\mathbf{x_f} = \mathbf{p_f} + \mathbf{z_{df}} \tag{35}$$

$$\mathbf{A_f}\mathbf{x_f} + \mathbf{z_{pf}} = \mathbf{b_f} \tag{36}$$

with the vectors $\mathbf{y_t} \geq 0_\mathbf{t}, \mathbf{x_f} \geq 0, \mathbf{z_{pf}} \geq 0$, and $\mathbf{z_{df}} \geq 0$, where the scalar CSC_f includes the complementary slackness conditions of the producing unit f; and the variables $\mathbf{z_{pf}}$ and $\mathbf{z_{df}}$ are slack-surplus variables of the primal and dual constraints, respectively. If the simulation scenarios contain some regional constraints, then it is possible to solve the individual simulation models over these regional constraints as in Henry de Frahan et al. (2011). As it is set-up, the model (34)–(36) is ready to evaluate responses to changes in output prices, direct subsidies and limiting inputs, but can be accommodated to evaluate responses to changes in variable-input prices if a cost function that is also defined on variable-input prices is used.

4 Conclusions

The PMP concept was formalised by Richard E. Howitt in 1995 within the context of scarcity of disaggregated bio-physical-agronomic and socio-economic data to calibrate exactly mathematical programming models for policy simulations. At that time for that context, this concept was an ingenious leap forward in this empirical field. Since then, many modellers have examined different econometric methods to integrate the information contained in more than one observation because this information most often does exist even within developing and emerging countries, in terms of either time or cross-sectional series or panel data such as the rich FADN dataset of the European Commission. Some of these modellers have also questioned implicit assumptions underlying the original PMP approach and developed, in turn, methodological advances.

There are, nowadays, a range of estimation methods available to deal with the information contained in more than one single observation and obtain estimated model parameters that not only calibrate the simulation model at hand but also provide more reliable responses to technological, institutional or economic changes in the variables of interest. These estimation methods span from the entropy approach to the Bayesian approach and the standard econometric approach with or without bi-level programming. What is striking about this review of estimation methods is the realisation that the development of this field mainly rested on a few scientists and research institutes, be it within the group of Cloé Garnache (nowadays at University of Oslo), Richard E. Howitt, Pierre Mérel and Quirino Paris at the University of California, Davis, the group of Wolfgang Britz, Alexander Gocht (nowadays at Thünen Institute for Rural Studies), Thomas Heckelei, and Torbjörn Jansson (nowadays at the Swedish University of Agricultural Sciences) at the University of Bonn, and Filippo Arfini and Michele Donati at the University of Parma. This contribution has been to trace down the development of these estimation methods and propose a methodological framework integrating together contributions from these scientists, in particular Heckelei and Wolff (2003) and Paris (2017).

Of course, the main relevance of this methodological proposal for an application depends on the availability of sufficiently reliable observations, which was taken for granted in the previous section. Notwithstanding its interest, this methodological proposal calls for more careful scrutiny and refinement and requires that it be put to the test of real-world data. The recent methodological developments in the PMP-related literature have helped definitively to bridge the gap between mathematical programming and econometric methods into a new empirical field captured in the "econometric mathematical programming" denomination.

References

Arata, L., Donati, M., Sckokai, P., & Arfini, F. (2017). Incorporating risk in a positive mathematical programming framework: A dual approach. *Australian Journal of Agricultural and Resource Economics, 61*(2), 265–284.

Arfini, F., & Donati, M. (2011). The impact of the Health Check on structural change and farm efficiency: A comparative assessment of three European agricultural regions. In C. Moreddu (Ed.), *Disaggregated impacts of CAP reforms: Proceedings of an OECD Workshop*. Paris, France: OECD Publishing.

Arfini, F., Donati, M., Grossi L., & Paris, Q. (2008). Revenue and cost functions in PMP: A methodological integration for a territorial analysis of CAP. Paper presented at the 107th EAAE Seminar 'Modelling of Agricultural and Rural Development Policies'. Sevilla, Spain, January 29–February 1st 2008.

Arndt, C., Robinson, S., & Tarp, F. (2002). Parameter estimation for a computable general equilibrium model: A maximum entropy approach. *Economic Modelling, 19,* 375–398.

Britz, W., & Arata, L. (2019). Econometric mathematical programming: an application to the estimation of costs and risk preferences at farm level. Agricultural Economics, 50.

Britz, W., Heckelei, T., & Wolff, H. (2003). Symmetric positive equilibrium problem: A framework for rationalizing economic behaviour with limited information: Comment. *American Journal of Agricultural Economics, 85*(4), 1078–1081.

Buysse, J., Fernagut, B., Harmignie, O., Henry de Frahan, B., Lauwers, L., Polomé, P., et al. (2007a). Farm-based modelling of the EU sugar reform: Impact on Belgian sugar beet suppliers. *European Review of Agricultural Economics, 34*(1), 21–52.

Buysse, J., Van Huylenbroeck, G., & Lauwers, L. (2007b). Normative, positive and econometric mathematical programming as tools for incorporation of multifunctionality in agricultural policy modelling. *Agriculture, Ecosystems & Environment, 120*(1), 70–81.

Chambers, R. G., & Just, R. E. (1989). Estimating multioutput technologies. *American Journal of Agricultural Economics, 71*(4), 980–995.

Cortignani, R., & Severini, S. (2009). Modelling farm-level adoption of deficit irrigation using positive mathematical programming. *Agricultural Water Management, 96*(12), 1785–1791.

Cortignani, R., & Severini, S. (2012). Modelling farmer participation to a revenue insurance scheme by means of the positive mathematical programming. *Agricultural Economics—Czech, 58*(7), 324–331.

Ferris, M. C., Dirkse, S. P., Jagla, J.-H., & Meeraus, A. (2009). An extended mathematical programming framework. *Computers & Chemical Engineering, 33*(12), 1973–1982.

Ferris, M. C., Dirkse, S. P., & Meeraus, A. (2002). *Mathematical programs with equilibrium constraints: Automatic reformulation and solution via constrained optimization* (Computing Laboratory Report No. 02/11). Oxford: England: Oxford University.

Frisvold, G. B., & Konyar, K. (2012). Less water: How will agriculture in Southern Mountain states adapt? *Water Resources Research, 48,* W05534. https://doi.org/10.1029/2011WR011057.

Garnache, C., & Mérel, P. (2015). What can acreage allocations say about supply elasticities? A convex programming approach to supply response disaggregation. *Journal of Agricultural Economics, 66*(1), 236–256.

Garnache, C., Mérel, P. R., Howitt, R. E., Howitt, R. E., & Lee, J. (2017). Calibration of shadow values in constrained optimization models of agricultural supply. *European Review of Agricultural Economics, 44*(3), 363–397.

Gocht, A. (2005). Assessment of simulation behaviour of different mathematical programming approaches. In F. Arfini (Ed.), *Modelling agricultural policies: State of the art and new challenges*. *Proceedings of the 89th European Seminar of the European Association of Agricultural Economists* (pp. 48–73). Italy: University of Parma.

Gocht, A., & Britz, W. (2011). EU-wide farm type supply models in CAPRI-How to consistently disaggregate sector models into farm type models. *Journal of Policy Modeling, 33*(1), 146–167.

Golan, A., Judge, G., & Miller, D. (1996). *Maximum entropy econometrics*. Chichester: Wiley.

Gorddard, R. (2013). Profit-maximizing land-use revisited: The testable implications of non-joint crop production under land constraint. *American Journal of Agricultural Economics, 95*(5), 1109–1121.

Graindorge, C., Henry de Frahan, B., & Howitt, R. E. (2001). Analysing the effects of Agenda 2000 using a CES calibrated model of Belgian agriculture. In T. Heckelei, H. P. Witzke, & W. Henrichsmeyer (Eds.), *Agricultural sector modelling and policy information systems* (pp. 177–186). Kiel: Vauk Verlag.

Graveline, N., & Mérel, P. (2014). Intensive and extensive margin adjustments to water scarcity in France's cereal belt. *European Review of Agricultural Economics, 41*(5), 707–743.

Guyomard, H., Baudry, M., & Carpentier, A. (1996). Estimating crop supply response in the presence of farm programmes: Application to the CAP. *European Review of Agricultural Economics, 23,* 401–420.

Heckelei, T., & Britz, W. (2000). Positive mathematical programming with multiple data points: A cross-sectional estimation procedure. *Cahiers d'Economie et Sociologie Rurales, 57*(4), 28–50.

Heckelei, T., & Britz, W. (2005). Models based on positive mathematical programming: State of the art and further extensions. In F. Arfini (Ed.), *Modelling agricultural policies: state of the art and new challenges*. *Proceedings of the 89th European Seminar of the European Association of Agricultural Economists* (pp. 48–73). Italy: University of Parma.

Heckelei, T., & Wolff, H. (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy. *European Review of Agricultural Economics, 30*(1), 27–50.

Heckelei, T., Britz, W., & Zhang, Y. (2012). Positive mathematical programming approaches—recent developments in literature and applied modelling. *Bio-based and Applied Economics, 1*(1), 109–124.

Heckelei, T., Mittelhammer, R., & Jansson T. (2008). A Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models. In Food and Resource Economics Discussion Paper 2008:2, Institute for Food and Resource Economics, University of Bonn.

Helming, J. F. M., Peeters, L., & Veendendaal, P. J. J. (2001). Assessing the consequences of environmental policy scenarios in Flemish agriculture. In T. Heckelei, H. P. Witzke, & W. Henrichsmeyer (Eds.), *Agricultural sector modelling and policy information systems*. *Proceedings of the 65th EAAE Seminar, March 29–31, 2000 at Bonn University* (pp. 237–245). Kiel: Vauk Verlag.

Henry de Frahan, B., Buysse, J., Polomé, P., Fernagut, B., Harmignie, O., Lauwers, L., et al. (2007). Positive mathematical programming for agricultural and environmental policy analysis: Review and practice. In A. Weintraub, C. Romero, T. Bjørndal, R. Epstein, & J. Miranda (Eds.), *Handbook of operations research in natural resources, international series in operations research and management science* (pp. 129–154). New York: Springer.

Henry de Frahan, B., Baudry, A., De Blander, R., Polomé, P., & Howitt, R. (2011). Dairy farms without quotas in Belgium: Estimations and simulations with a flexible cost function. *European review of agricultural economics, 38*(4), 1–27.

Howitt, R. E. (1995a). Positive mathematical programming. *American Journal of Agricultural Economics, 77*(2), 329–342.

Howitt, R. E. (1995b). A calibration method for agricultural economic production models. *Journal of Agricultural Economics, 46*(2), 147–159.

Jansson, T., & Heckelei, T. (2011). Estimating a primal model of regional crop supply in the European Union. *Journal of Agricultural Economics, 62*(1), 137–152.

Jansson, T., & Heckelei, T. (2009). A new estimator for trade costs and its small sample properties. *Economic Modelling, 26,* 489–498.

Jansson, T., Heckelei, T., Gocht, A., Basnet, S. K., Zhang, Y., & Neuenfeldt, S. (2014). Analysing impacts of changing price variability with estimated farm risk-programming models. Paper presented at the EAAE 2014 Congress 'Agri-Food and Rural Innovations for Healthier Societies'. Slovenia, Ljubljana, 26–29 Aug 2014.

Júdez, L., Chaya, C., Martinez, S., & Gonsalez, A. A. (2001). Effects of the measures envisaged in “Agenda 2000” on Arable Crop Producers and Beef and Veal Producers: An application of positive mathematical programming to representative farms of a Spanish region. *Agricultural Systems, 67,* 121–138.

Kanellopoulos, A., Berentsen, P., Heckelei, T., van Ittersum, M., & Oude Lansink, A. (2010). Assessing the forecasting performance of a generic bio-economic farm model calibrated with two different PMP variants. *Journal of Agricultural Economics, 61*(2), 137–152.

Lence, H. L., & Miller, D. (1998). Recovering output specific inputs from aggregate input data: A generalized cross-entropy approach. *American Journal of Agricultural Economics, 80*(4), 852–867.

Léon, Y., Peeters, L., Quinqu, M., & Surry, Y. (1999). The use of maximum entropy to estimate input-output coefficients from regional farm accounting data. *Journal of Agricultural Economics, 50,* 425–439.

Louhichi, K., Ciaian, P., Espinosa, M., Perni, A., Vosough Ahmadi, B., Colen, L., & Gomez y Paloma, S. (2016). *An EU-Wide Individual farm model for common agricultural policy analysis (IFM-CAP 1.0) (technical documentation)* (JRC Science and Policy Reports). Sevilla, Spain: European Commission, Joint Research Centre, Institute for Prospective Technological Studies.

Louhichi, K., Kanellopoulos, A., Janssen, S., Flichman, G., Blanco, M., Hengsdijk, H., et al. (2010). FSSIM, a bio-economic farm model for simulating the response of EU farming systems to agricultural and environmental policies. *Agricultural Systems, 103*(8), 585–597.

Marsh, T. L., Mittelhammer, R., & Cardell, N. S. (2014). Generalized maximum entropy analysis of the linear simultaneous equations model. *Entropy, 16,* 825–853.

Medellin-Azuara, J., Howitt, R., & Harou, J. (2012). Predicting farmer responses to water pricing, rationing and subsidies assuming profit maximizing investment in irrigation technology. *Agricultural Water Management, 108,* 73–82.

Mérel, P., & Bucaram, S. (2010). Exact calibration of programming models of agricultural supply against exogenous supply elasticities. *European Review of Agricultural Economics, 37*(3), 395–418.

Mérel, P., & Howitt, R. E. (2014). Theory and application of positive mathematical programming in agriculture and the environment. *Annual Review of Resource Economics, 6,* 451–470.

Mérel, P., Simon, L. K., & Yi, F. (2011). A fully calibrated generalized constant-elasticity-of-substitution programming model of agricultural supply. *American Journal of Agricultural Economics, 93*(4), 936–948.

Mérel, P., Yi, F., Lee, J., & Six, J. (2014). A regional bio-economic model of nitrogen use in cropping. *American Journal of Agricultural Economics, 96*(1), 67–91.

Mittelhammer, R. C., Cardell, N. S., & Marsh, T. L. (2013). The data-constrained generalized maximum entropy estimator of the GLM: Asymptotic theory and inference. *Entropy, 15,* 1756–1775.

Moro, D., & Sckokai, P. (1999). Modelling the CAP arable crop regime in Italy: Degree of decoupling and impact of Agenda 2000. *Cahiers d’Economie et Sociologie Rurales, 53,* 50–73.

Oude Lansink, A. (1999). Generalised maximum entropy and heterogeneous technologies. *European Review of Agricultural Economics, 26,* 101–115.

Paris, Q. (2001a). Symmetric positive equilibrium problem: A framework for rationalizing economic behavior with limited information. *American Journal of Agricultural Economics, 83*(4), 1049–1061.

Paris, Q. (2001b). Dynamic positive equilibrium problem. Working Paper. Davis: Department of Agricultural and Resource Economics. University of California.

Paris, Q. (2011). *Economic foundations of symmetric programming*. Cambridge: Cambridge University Press.

Paris, Q. (2015). Positive mathematical programming with generalized risk: A revision. Working Paper. Davis: Department of Agricultural and Resource Economics. University of California, April 2015.

Paris, Q. (2017). Cost function and positive mathematical programming. *Bio-based and Applied Economics, 6*(1), 19–35.

Paris, Q., & Howitt, R. E. (1998). An analysis of ill-posed production problems using maximum entropy. *American Journal of Agricultural Economics, 80*(1), 124–138.

Petsakos, A., & Rozakis, S. (2015). Calibration of agricultural risk programming models. *European Journal of Operational Research, 24,* 536–545.

Preckel, P. V. (2001). Least squares and entropy: A penalty function perspective. *American Journal of Agricultural Economics, 83*(2), 366–377.

Röhm, O., & Dabbert, S. (2003). Integrating agri-Environmental programs into regional production models: An extension of positive mathematical programming. *American Journal of Agricultural Economics, 85*(1), 254–265.

Sinha, A., Malo, P., & Deb, K. (2017). A review on bilevel optimization: From classical to evolutionary approaches and applications. arXiv:1705.06270v1 [math.OC]. 17 May 2017.

Vicente, L. N., & Calamai, P. H. (1994). Bilevel and multilevel programming: A bibliography review. *Journal of Global Optimization, 5*(3), 291–306.

Chapter 3
Soil and Crop Choice

Peter Berck and Lunyu Xie

Abstract This contribution uses econometric analysis to uncover the various factors driving crop choice in six states along the Mississippi River. Aside from temperature and precipitation, soil characteristics are also included as explanatory factors—which is a factor often omitted from many studies. The analysis shows soil to be a key determinant of corn and soybean area in the regions studied.

1 Introduction

Crop choice ultimately depends on market factors, climate, and soil. In this contribution, we will explain the choice of major crops in parts of six states along the Mississippi River and show how that choice depends upon weather and soil. Using these regressions, we will then show how changes in climate would affect the choice of crops.

The recent and classic work of Schlenker and Roberts (2009) shows that very hot weather leads to large losses in crop yield. This finding implies that even modest global warming leads to large losses in grain yield. One consequence of large yield losses would be that land now devoted to corn and soy would be devoted to something else. This chapter looks at how much land would be diverted from corn and soy in response to marginal changes in precipitation and extreme temperature. Because the data on land use are inherently measured at a much finer level than the county level, which is the usual aggregate to measure crop yield, this contribution examines

Peter Berck is now deceased.

P. Berck
Department of Agricultural and Resource Economics, University of California, Berkeley, CA, USA

L. Xie (✉)
Department of Energy Economics, School of Applied Economics, Renmin University of China, Beijing, China
e-mail: lunyuxie@ruc.edu.cn

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_3

how high temperature, low precipitation, and soil quality interact to determine the response of acreage to climate change. Looking at the role of soil and looking at the interactions with other factors gives much finer evidence on how and where climate change will matter, compared to just looking at temperature and precipitation. The key interaction is between high heat and low moisture, which go often together. In this chapter, we add an interaction term for high heat and low precipitation and show that it is a very significant variable, even when added to a regression that has high heat alone. This finding is in line with the use of drought indices, which are statistics based on temperature and precipitation.

This work differs from the extensive literature on adaptation to climate change and agricultural effects of climate change in its very reduced form approach. There are neither lagged variables nor spatial fixed effects. By avoiding lagged variables, used in the classic Nerlove (1956) type regressions, we do not make crop choice dependent on prior crop choice. Particularly in our 11-year-long panel, serial dependence would attribute the use of land for soy and corn to its prior use for soy and corn. There is no external validity to such a model, as applying it to a new place requires knowing how much corn and soy there was in the new place, which is of course what one wanted to find out, not what one knew. The loss in not depending on lagged variables is that the dynamics of the shift between corn and soy will be lost. The argument is the same for eschewing fixed land effects. Fixed effects in regressions, such as the alternative specific constant of logit, make it impossible to extend the regressions to new places or to new alternatives. To do so requires the heroic assumption that one knows the new fixed effect.

An alternative to the fixed effects paradigm of crop choice is a hedonic attributes approach (Anderson et al. 2012). Crops, including new crops, have characteristics such as water demand and ability to tolerate acidity. Crops are grown when the crop characteristics match the soil and weather characteristics. This approach allows for new crops and also allows for changing crops in response to climate change. Climate change means a different set of crops better matches the soil and weather, and as a result, different crops are grown as an adaptation. This model properly elaborates the mechanism behind changing crops; however, it is not parsimonious nor are the needed data readily available.

The regressions in this chapter take market conditions for granted. That is, a fixed effect controls for prices. Of course, if climate change reduces the acreage in crops and that increases price, the regressions will not account for that second-order effect. The regressions show the shift in the crop supply curve, not the movement from one equilibrium to another. In the context of devoting cropland to biofuels, there are many models that close the loop with changing prices. See for example, Searchinger (2008) or Taheripour et al. (2011).

The plan of the work is to use a panel of 4 km^2 that covers parts of six states from 2000 to 2010. In this study, we use the fraction of land devoted to each crop in each 4 km^2. This data is matched to climate and soil data. We then run a linear probability system for the major uses. The regression coefficients are then used to find how changes in soil and weather change the amount of land dedicated to corn and soy, the two major crops that are grown throughout the study area.

Table 1 Variable definitions

S_{int}	Coverage share of crop i planted at grid cell n in year t
Soil_n	Land classification code (LCC), grouped as I and II, III and IV, V and VI, and VII and VIII
GDD_{nt-1}	A vector of degree-days by month group in the last growing season (April through November in year $t-1$). The data are binned at 10, 15, 20, 25, 29, 30, and 32 °C, where 10 °C is degree-days ≥10 °C, etc. Months are grouped into (4, 5), (6, 7, 8), and (9, 10, 11)
PDD_{nt}	A vector of degree-days in the current planting season (April and May in year t). The data are binned at 10 and 15 °C
GP_{nt-1}	A vector of precipitation by month group (see above) in the last growing season
PP_{nt}	A vector of precipitation by month in the current planting season
DD32LCC_{nt-1}	A vector of interactions of the soil type groups and the degree-days at 32 °C and above in the previous summer months (6, 7, 8)
PRECLCC_{nt-1}	A vector of interactions of the inverse ($1/x$) of the precipitation in the previous summer months (6, 7, 8) and the LCC group
$\text{PRECDD32LCC}_{nt-1}$	For the previous summer months (6, 7, 8), a vector of interactions of the inverse ($1/x$) of the precipitation, the degree-days at 32 °C and above, and the LCC group
μ_t	Year fixed effects

2 Data

The data come from three sources: the crop data layer (CDL), the PRISM project, reprocessed by Schlenker and Roberts (2009), and STATSGO2, the USDA soil system.

The states included in the analysis are those along the Mississippi–Missouri River corridor for which there are land cover data from 2000: Wisconsin, Iowa, Illinois, and part of Missouri, Arkansas, and Mississippi. Table 1 presents the definition of all variables.

Land cover data are derived from the Cropland Data Layer (CDL) of the National Agricultural Statistics Service, available annually from 2000 to 2010 for the six states. Mueller and Seffrin (2006) provide a good description of this data. We divide land cover into major crops, other crops, non-crop and wildland, urban, and water bodies. The major crops include corn and soybean for Iowa, Wisconsin, and Illinois; and corn, soybean, rice, and cotton for Missouri, Arkansas, and Mississippi. In this chapter, we focus on the results for corn and soybeans. The category of non-crop and wildland includes pasture, forest, improved pasture, and conservation reserve land. We define agricultural land as the sum of major crops, other crops, and non-crop and wildland. We exclude urban areas and water bodies. Therefore, we define the share of major crops as the area of major crops divided by the area of agricultural land. The limiting factor in determining land use is the number of plots observed on the ground ("ground truth"). It is for this reason that the standard errors in classification in minor crops, such as oats, are quite large. Similarly, it is why pasture, forest, and improved pasture are not well distinguished in the data.

For soil data, we focus on the land capability class (LCC) variables, which come from the USDA's US General Soil Map (STATSGO2). The underlying soil data include percent clay, sand, and silt, water-holding capacity, pH value, electrical conductivity, slope, frost-free days, depth to water table, and depth to restrictive layer. These data, as well as soil type, were used by the USDA to construct LCC codes. A LCC value of one defines the best soil with the fewest limitations for production. A LCC score of one or two is highly suitable for crops; those beyond four are suitable for pasture and other extensive use; and an LCC of eight means that agricultural planting is nearly impossible.

Figure 1 shows the LCCs across the study area. Each square has one or more LCCs. The LCC variables are the percent of the square in that LCC. The coloring in the figure corresponds to the average LCC in a 4 km^2. LCC I and II are heavily concentrated in Northern Iowa and in Illinois, parts of Eastern Wisconsin, and near the Mississippi and other rivers in the southern states. This pattern is a result of glaciation in the last ice age and subsequent erosion in the river valleys.

For weather data, we use PRISM data processed by Schlenker and Roberts (2009) to a 4 km by 4 km spatial resolution, with a daily level of temporal resolution. The data include daily highs and lows, which are then processed to temperature by hour using the sine curve interpolation method (Baskerville and Emin 1969). The innovation in Schlenker and Roberts was to count the degree-days in bins, for instance, degree-days between 20 and 25 °C. In this way, the effect of temperatures over a critical temperature, such as 29 °C for corn, can be isolated from the effect of all other temperatures. The data also include precipitation. Unlike modeling yield, where current summer weather is known, modeling planting requires an estimate of coming summer weather, since planting happens before the summer season. The planting decision depends upon the weather observed during the spring in the year the decision is made and on past weather during the growing season because past weather is a good predictor of coming summer weather. Therefore, weather variables are computed separately for the planting season and growing season. The weather is then further disaggregated by summer and fall, to account for the different stages in crop growth. In all, this gives three seasons of weather: current spring, last fall, and last summer.

Figure 2 shows the percent of the landscape covered by corn and soy. The maps are composed of the 4 km^2, each of which has some percentage of corn and soy coverage.

There is a striking (and well-known) correspondence between Figs. 1 and 2. The corn–soy complex is heavily concentrated on the better soils. It is this soil determinism that makes it unlikely that we will find that a warmer climate will result in much higher corn–soy coverage in places like central Wisconsin. They are north of the corn belt and so should benefit from some warming, but they are also north of the better soils.

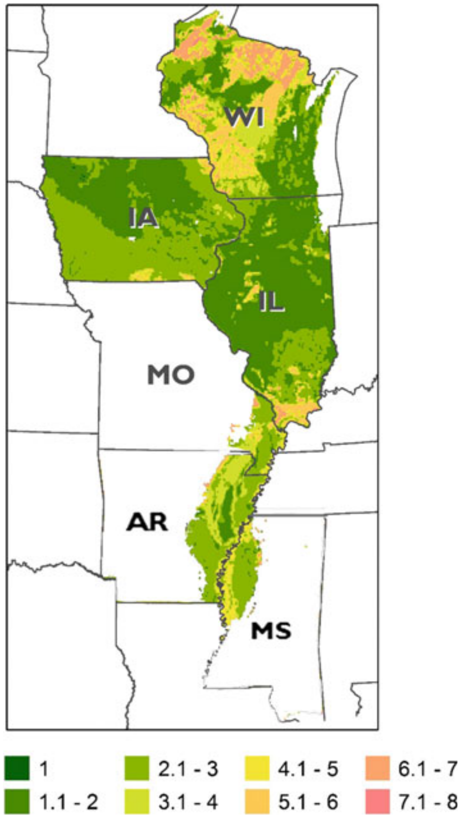

Fig. 1 Distribution of land capability classification (LCC) levels. Originally published in Xie et al. (2018). Reprinted with kind permission of © Springer Nature Netherlands 2018.

Next, we consider what other explanatory variables might be included. Nerlove's adaptive price expectations model (Nerlove, 1956) assumed that farmers have rational price expectations based on their information set and described the agricultural system in three equations. Braulke (1982) derived a reduced form from the three equations by removing the unobserved variables. Choi and Helmberger(1993) combined this reduced form and farmer's demand functions. Based on their work, Huang and Khanna (2010) described crop share as a function of lagged shares, climate variables, economic variables, risk variables, population density, and time trend. Hausman (2012) included most of these explanatory variables, as well as futures prices, substitute crop shares, and crop yield. We depart from the literature in presenting the reduced form, without dependent lagged variables. Given that the interest in this study is the effect of weather and soil, rather than the more common price elasticity, we are able to use fixed effects to account for many variables that are common to the observations across space. (1) In many countries (e.g., the USA and the European Union), government payments are part of the incentive to grow crops. As these programs change from year to year and have different marginal effects for different farmers, it is not possible to have a fully satisfactory treatment of the

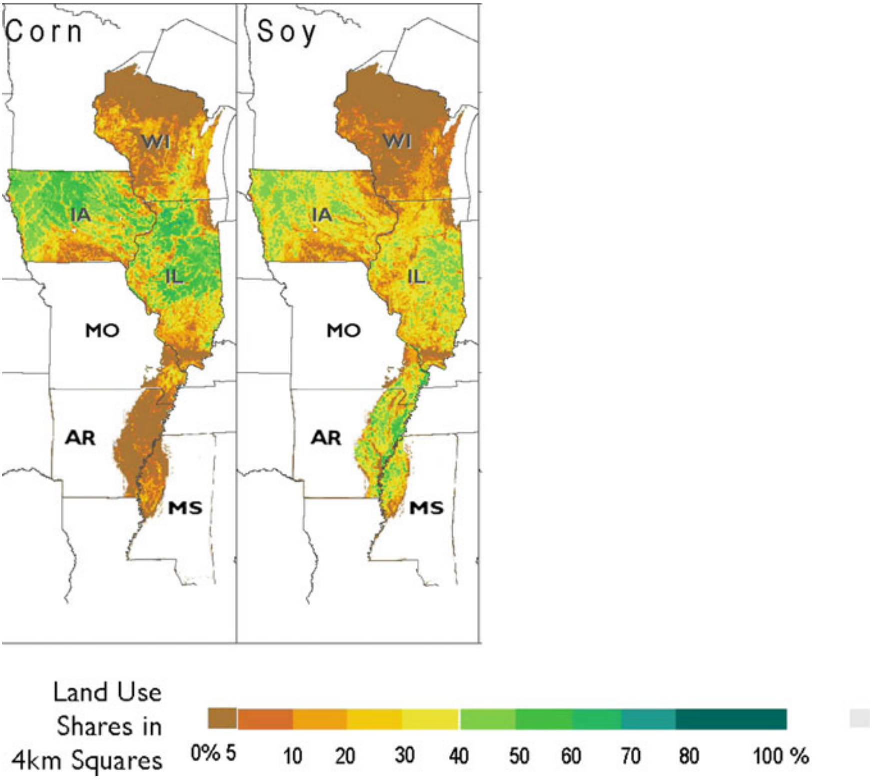

Fig. 2 Percentage of land area in corn and soybean. Originally published in Xie et al. (2018).

payments variable. We use year fixed effects to account for both prices and government programs. The year fixed effects also account for differences in input prices. (2) Many authors (Just 1974; Chavas and Holt 1990; Lin and Dismukes 2007) argue that the risk of growing a crop, perhaps the variance or lower semi-variance, is an important determinant of crop choices. So long as the risk of growing a crop is taken as constant, which is a good approximation in a short time series, crop fixed effects account for this factor.

In addition to these basic weather and soil variables, we include the interaction term of heat and moisture to account for the possibility that dry warming is much more harmful than warming with moisture (Lobell et al. 2011). The landscape we consider has a very wide range of precipitation and temperature. In the easternmost part of Iowa, for instance, the average precipitation during the growing season was less than 7 cm per month, while in Mississippi there was as much as 12 cm of precipitation. For temperature, the average in northern Wisconsin was close to 10 °C during the growing season, while it was as high as 24 °C in Mississippi. While this landscape has hot places and dry places, it generally does not have hot and dry in the same place. The wet and hot south can be very hot, but the soil may never become very dry. So temperature alone, or even temperature and precipitation, is not enough to predict where crops will be grown. It is hot and dry together that discourages growing of corn or soy. Even accounting for the interaction of moisture and temperature is not enough for these purposes. Sandy soils do not hold moisture well and so are more sensitive to dry spells. Interacting the temperature, precipitation, and soil type provide a fuller story.

The interaction variables used are made from interacting the LCC group with degree-days in the summer at the hottest temperature bin (32 °C+), and with the inverse (1/Precip) of precipitation, a measure of dryness. This produces three vectors of variables: hot temperature by LCC group, the inverse of precipitation by LCC group, and the interaction of all three.

The regressions also include time fixed effects. These fixed effects account for a multitude of variables that have been found important by other authors: prices of outputs, prices of alternative crops, prices of inputs, price risk, yield risk, support programs, and so on. While the CDL is an ideal dataset for examining weather, which varies considerably across space, it does not yet have a long enough time span to be good at capturing the effects of prices, which vary little across space and are a yearly phenomenon.

3 Regression Model

The regression model is a linear probability model. For each 4 km^2, the dependent variables are the percent of the agricultural land in that square that is covered by corn and soy. These are the two crops that are grown throughout most of the study area. Rice and cotton are grown in substantial quantities in the southern states. The remaining agricultural area is composed of forest and various types of pasture, as well as minor crops, such as oats. The time fixed effects account for the prices or values of the other uses, so there is no need to explicitly include them in the regression. The included regressors are all exogenous, as they are the outcome of the weather and very slow to change soil characteristics. There is very little irrigated agriculture in this landscape, largely because the better parts of the landscape are quite wet for the purposes of growing corn and soy.

In the study area, corn occupies 18%, and soy occupies 17% of the agricultural land. The next largest crops by percent coverage are rice and cotton, at about 1% each. Less than 1% of the land is classified as LCC I, while nearly half (49%) is LCC II. LCC III occupies a quarter (24%) of the land and the remaining classes have 14% or less each. Extreme degree-days are rare. For instance, in July, on average over the landscape, the temperature was above 32 °C only 2.2% of the time.

Again referring to Table 1, for the definitions of the variables, the estimating equations are

Table 2 Marginal effects of temperature and precipitation on land use of corn and soy

Change in percent for	Corn	Soy
One more degree-day over 32 during the summer at average precipitation and soil	−0.003	0.008
One more degree-day over 32 during the summer at minimum precipitation and average soil	−0.009	0.005
More LCC 1–2 soil at av. degree-day and precipitation	0.143	0.146
More LCC 3–4 soil at average degree-day and precipitation	0.024	0.09
One more degree-day over 32 LCC 1–2 and average precipitation	−0.012	0.001
One more degree-day over 32 at LCC 3–4 and average precipitation	0.004	0.014
One more degree-day over 32 at LCC 5–6 and average precipitation	0.009	0.018
One more cm summer precipitation at average soil and temperature	0.009	0.003
One more cm summer precipitation at average soil and maximum temperature	0.015	0.007
All marginal values significantly different from zero at the 99% level		

$$\begin{aligned} S_{int} = \alpha &+ \boldsymbol{\varphi}_i^{\prime}\mathbf{Soil}_n + \boldsymbol{\theta}_{1i}^{\prime}\mathbf{GDD}_{nt-1} + \boldsymbol{\theta}_{2i}^{\prime}\mathbf{PDD}_{nt} + \boldsymbol{\theta}_{3i}{}^{\prime}\mathbf{GP}_{nt-1} + \boldsymbol{\theta}_{4i}^{\prime}\mathbf{PP}_{nt} \\ &+ \boldsymbol{\theta}_{5i}^{\prime}\mathbf{DD32LCC}_{nt-1} + \boldsymbol{\theta}_{6i}^{\prime}\mathbf{PRECLCC}_{nt-1} \\ &+ \boldsymbol{\theta}_{7i}^{\prime}\mathbf{PRECDD32LCC}_{nt-1} + \mu_t + \varepsilon_{int} \end{aligned}$$

The equations are fit by ordinary least squares, and the reported standard errors are the errors from ordinary least squares. The equations for corn and soy both had 367,000 observations, based on 10 years of data on 36,700 4 km^2. The R^2 for the corn and soy equation are 0.63 and 0.47, respectively. All but two coefficients are significant at the 99% level or higher. The two nonsignificant coefficients come from the same source. In both corn and soy, the high-temperature effect in September, October, and November cannot be separated into an over-30 and over-32 effect.

4 Results

The parameters of interest in these equations are the partial derivatives with respect to temperature and precipitation. They are evaluated at different levels of the other interaction variables. Table 2 presents these marginal results.

The land use changes for corn and soy are given in change in percent. For reference, corn and soy average 18 and 17% of the agricultural land in the base year. So, one more high degree-day at average temperature and soil (first row of the table) leads to a decrease of 0.3% or a change from 18 to 17.7% coverage. Looking across both soy and corn, the land allocated to these major crops goes up by 0.5% when temperature increases (0.008 for soy less 0.003 for corn). When evaluated at minimum precipitation, land usage decreases by 0.4% (0.005 for soy less 0.009 for corn).

Comparing these two results shows that it is the combination of low precipitation and higher temperature that leads to a decrease in the area in major crops.

Changing the soil leads to a large change in the coverage of the major crops. The experiment is to increase the percentage of soil that is LCC I and II from the sample mean to the sample max (and decrease the LCC VII and VII to compensate). Both crops respond with large changes in crop coverage: 14% for corn and 15% for soy. Adding more LCC III and IV land leads to about half this effect. This corroborates the visual impression that soil quality and corn–soy coverage go together.

The next two results further explore increased temperature. The overall effect of a higher temperature on coverage was positive. Looking at it by land class, there is a large positive effect on LCC III and LCC IV. This must be offset by large negative effects for the higher classes because the effect on LCC I and LCC II is small. Further research could break down the LCCs in terms of water-holding ability. Heavy clay soils presumably would be best at resisting or even benefiting from a short period of higher heat.

Finally, summer precipitation is uniformly positive for the growing of corn and soy. Thinking in terms of a climate that is less wet, looking at the effect at maximum temperature, each centimeter of lowered rainfall leads to a 2.2% change in corn and soy coverage. If climate change were to mean a hotter Midwest, then even marginal drying would have a dramatic effect on the extent of corn and soy plantings.

5 Conclusion

Soil is the biggest determinant of where corn and soy grow. It is high LCC class soils that lead to the corn and soy complex in Iowa and Illinois. Given that soil determines crop coverage to such a large degree, it is not surprising that temperature and precipitation play a secondary role. At average precipitation, an additional degree-day over 32 °C has little effect on planting, but at minimum precipitation the effect is about one-half percent coverage change per degree. Comparing these results to Schlenker and Roberts (2009), land use responds more to hot and dry changes than to heat alone.

References

Anderson, S., Wang, C. & Zhao, J. (2012). Let them eat switchgrass? Modeling the displacement of existing food crops by new bioenergy feedstocks. s.l.:s.n.

Baskerville, G., & Emin, P. (1969). Rapid estimation of heat accumulation from maximum and minimum temperatures. *Ecology, 50,* 514–517.

Braulke, M. (1982). A note on the Nerlove model of agricultural supply response. *International Economic Review, 23*(1), 241–244.

Chavas, J.-P., & Holt, M. T. (1990). Acreage decisions under risk: The case of corn and soybeans. *American Journal of Agricultural Economics, 72*(3), 529–538.

Choi, J.-S., & Helmberger, P. G. (1993). How sensitive are crop yields to price changes and Farm programs? *Journal of Agricultural and Applied Economics, 25,* 237–244.
Hausman, C. (2012). Biofuels and land use change: Sugarcane and soybean acreage response in Brazil. *Environmental and Resource Economics, 51*(2), 163–187.
Huang, H., & Khanna, M. (2010). *An econometric analysis of US crop yield and cropland acreage: Implications for the impact of climate change* (pp. 25–27). Denver, Colorado, s.n.
Just, R. E. (1974). An investigation of the importance of risk in farmers' decisions. *American Journal of Agricultural Economics, 56*(1), 14–25.
Lin, W., & Dismukes, R. (2007). Supply response under risk: Implications for counter-cyclical payments' production impact. *Applied Economic Perspectives and Policy, 29*(1), 64–86.
Lobell, D. B., Banziger, M., Magorokosho, C., & Vivek, B. (2011). Nonlinear heat effects on African maize as evidenced by historical yield trials. *Nature Climate Change, 1*(1), 42–45.
Mueller, R., & Seffrin, R. (2006). New methods and satellites: A program update on the NASS cropland data layer acreage program. *Remote Sensing Support to Crop Yield Forecast and Area Estimates, ISPRS Archives, 36*(8), 48.
Nerlove, M. (1956). Estimates of the elasticities of supply of selected agricultural commodities. *Journal of Farm Economics, 38*(2), 496–509.
Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of Sciences, 106*(37), 15594–15598.
Searchinger, T., et al. (2008). Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science, 319*(5867), 1238–1240.
Taheripour, F., Tyner, W. E., & Wang, M. Q. (2011). Global land use changes due to the US cellulosic biofuel program simulated with the GTAP model. Argonne National Laboratory. http://greet.es.anl.gov/files/luc_ethanol.
Xie, L., Lewis, S. M., Auffhammer, M. et al. (2018). Environmental and Resource Economics. https://doi.org/10.1007/s10640-018-0271-7.

Chapter 4
Spatial Equilibrium, Imperfect Competition, and Calibrating Models

Quirino Paris

Abstract Spatial models of trade among regions require a burdensome series of information: Commodity demand and supply functions for each region and bilateral unit transaction costs. Even when this formidable amount of information is available, the trade flow matrix resulting from the model solution is typically very different from the exchanged trade flow that was realized in a previous economic cycle. This discrepancy may be attributed to two sources: incorrect measurement of transaction costs and imprecise knowledge of demand and supply function parameters. To remedy the undesirable result, we assume that the matrix of bilateral trade exchanges is observed—by the researcher—together with the realized demand and supply prices. With this additional information, we discuss the calibration of three categories of spatial models—(a) cartel behavior on the supply and export markets: This model corresponds to monopsony and monopoly behavior; (b) Nash-Cournot behavior on the supply and export markets: This model corresponds to oligopsony and oligopoly behavior; (c) perfect competition on both markets. The calibrating approach presented in this contribution is in the spirit of positive mathematical programming and its prescription: To achieve satisfactory results, it is important to use all the available information. The empirical part of the contribution is divided into two sections. First, we use only the observed matrix of bilateral trade flows to reveal the necessary adjustments to the unit transaction costs and achieve a calibrating model. Second, the observed demand and supply prices are used to reveal the adjustments to the intercepts of the demand and supply functions that correspond to a more general calibrating model.

Q. Paris (✉)
Department of Agricultural and Resource Economics,
University of California at Davis, Davis, USA
e-mail: paris@primal.ucdavis.edu

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_4

1 Introduction

Spatial equilibrium deals with a section of economics that attempts to explain the trade flow of commodities and their price formation among producing and consuming regions. In general, it involves three categories of economic agents: consumers, producers, and traders. From a behavioral perspective, consumers are considered as price takers who express their demand for a commodity by means of an aggregate demand function. Producers can be considered either price takers or agents who may behave according to imperfect competition rules. Their decisions are aggregated into regional supply functions. Traders may behave as oligopsonists (monopsonists in the limit) on the regional supply markets and either as oligopolists or monopolists (cartels) on the regional demand markets. Often, producers and traders are considered as the same economic agents. The limit case of perfect competition among all regions implies that there are no identifiable traders: Commodities are transferred from producing to consuming regions by the action of the "invisible hand".

Given the vast range of behaviors characterizing spatial equilibrium, we will limit the analysis to three behavioral rules: (a) cartel behavior (monopsony) on the supply market and cartel behavior (monopoly) on the export/consumption market; (b) oligopsony (Nash-Cournot equilibrium) on the supply market and oligopoly on the export/consumption market; (a) perfect competition on both the supply and consumption markets. The Nash-Cournot equilibrium refers to non-cooperative oligopoly and oligopsony firms: Each Nash oligopoly (oligopsony) firm makes production and profit-maximizing decisions assuming that its choices do not affect oligopolists' (oligopsonists') decisions in other regions. There is only one oligopoly–oligopsony firm in each region.

We consider the exchange of only one commodity among R regions. The extension to more than one commodity is straightforward. We assume knowledge of a linear inverse demand function for each region

$$p_j^D = a_j - D_j x_j^D \quad j = 1, \ldots, R \tag{1}$$

where p_j^D and x_j^D are price and quantity demanded in the jth region. The known coefficients $a_j > 0$ and $D_j > 0$ are the intercept and slope of the demand function, respectively. We assume knowledge also of a linear supply function for each region. This function can also be regarded as the marginal cost (MC_i) function for each region

$$p_i^S = b_i + S_i x_i^S = \mathrm{MC}_i \quad i = 1, \ldots, R \tag{2}$$

where p_i^S and x_i^S are price and quantity supplied in the ith region. The known coefficients b_i and $S_i > 0$ are the intercept and slope of the supply function, respectively. Bilateral unit transaction costs are also known for all pairs of regions and are stated as t_{ij}.

In spite of this formidable amount of initial information, it is very likely that the matrix of bilateral trade flows resulting from the solution of any spatial equilibrium model will be very different from the trade exchanges realized in the last economic cycle. To remedy this undesirable result, we will construct calibrating models using the last cycle trade flows among all regions and the corresponding demand and supply prices. These categories of information will be indicated as x_{ij}^{obs}, $p_j^{D,\text{obs}}$, and $p_i^{S,\text{obs}}$.

The chapter is organized as follows. Sections 2 and 3 will develop the cartel and Nash-Cournot models without the use of the observed trade flows and prices. The perfect competition model will be obtained as a special case of these models. Section 4 will illustrate numerically the solution of these three categories of spatial models. Sections 5, 6 and 7 will discuss calibrating models in the spirit of positive mathematical programming (PMP). In these sections, the source of discrepancy between optimal and observed trade flows is assumed to be caused by imprecision in the measurement of unit transaction costs. Section 8 will illustrate, numerically, the results of the calibrating models. Section 9 will assume that imprecision has affected also the intercepts of the regional demand and supply functions. Calibrating models will use the observed demand and supply prices. Section 10 will illustrate these models with numerical examples.

2 Spatial Cartel Equilibrium: Monopoly–Monopsony

When exporters collude, a cartel is formed. The intent of a cartel is to maximize total aggregate profit for the cartel members. The behavior of cartel members, therefore, is to maximize the joint profit by selling the monopoly output in each region at the monopoly price. Furthermore, they acquire the commodity on the supply market at monopsony prices. Hence, the spatial monopoly–monopsony model assumes that, in all regions, the output is controlled by one agent identified as the cartel.

The primal structure of this model is stated as

$$\begin{aligned}\max\ \text{Cartel}\,\pi &= \sum_{j=1}^{R} p_j^D x_j^D - \sum_{i=1}^{R} p_i^S x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij} \\ &= \sum_{j=1}^{R}(a_j - D_j x_j^D)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij} \qquad (3)\end{aligned}$$

subject to $\qquad\qquad$ dual variables

$$x_j^D \le \sum_{i=1}^{R} x_{ij} \quad \text{regional demand} \qquad \rho_j \ge 0 \qquad (4)$$

$$\sum_{j=1}^{R} x_{ij} \le x_i^S \quad \text{regional supply} \qquad \phi_i \ge 0 \qquad (5)$$

with all nonnegative variables. The objective function expresses the cartel profit with the first term represented by the monopolist total revenue, the second term as the monopsonist total cost due to purchase of the requisite supply and the third term as the total transaction cost. Constraint (4) specifies that, for equilibrium, the regional demand must be less than or equal to total supply from all regions. Constraint (5) states that the demand from all regions must be less than or equal to the supply of each region.

The Karush–Kuhn–Tucker (KKT) conditions reveal the equilibrium necessary relations in terms of marginal revenue and marginal cost. From the Lagrange function of model [(3)–(5)]

$$L = \sum_{j=1}^{R}(a_j - D_j x_j^D)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij} + \sum_{j=1}^{R}\rho_j\left(\sum_{i=1}^{R} x_{ij} - x_j^D\right) + \sum_{j=1}^{R}\phi_i\left(x_i^S - \sum_{j=1}^{R} x_{ij}\right) \tag{6}$$

the relevant KKT conditions are

$$\frac{\partial L}{\partial x_j^D} = a_j - 2D_j x_j^D - \rho_j \le 0 \tag{7}$$

$$\frac{\partial L}{\partial x_i^S} = -b_i - 2S_i x_i^S + \phi_i \le 0 \tag{8}$$

$$\frac{\partial L}{\partial x_{ij}} = \rho_j - \phi_i - t_{ij} \le 0 \tag{9}$$

The vertical line originating at the quantity level $x_j^D = x_i^S$ (see Fig. 1) verifies the marginal revenue-marginal cost relation expressed by $p_j^D - D_j x_j^D \le p_i^S + t_{ij} + S_i x_i^S$ as discussed in (10).

Assuming that each region will have a positive demand, $x_j^D > 0$, and a positive supply, $x_i^S > 0$, relations (7) and (8) will turn into equations (by complementary slackness conditions) and, thus, $\rho_j = a_j - 2D_j x_j^D$ and $\phi_i = b_i + 2S_i x_i^S$ which, in turn, will induce relation (9) to take on the following structure

$$\begin{aligned}
&(a_j - 2D_j x_j^D) - (b_i + 2S_i x_i^S) - t_{ij} \le 0\\
&(a_j - D_j x_j^D) - D_j x_j^D - (b_i + S_i x_i^S) - S_i x_i^S - t_{ij} \le 0\\
&p_j^D - D_j x_j^D \le p_i^S + t_{ij} + S_i x_i^S\\
&\mathrm{MR} \ \le \ \mathrm{MC}
\end{aligned} \tag{10}$$

In relation (10), $p_j^D - D_j x_j^D \le p_i^S + t_{ij} + S_i x_i^S$, the terms $D_j x_j^D$ and $S_i x_i^S$ constitute the markups (over the perfect competition marginal cost $p_i^S + t_{ij}$) of the monopolist

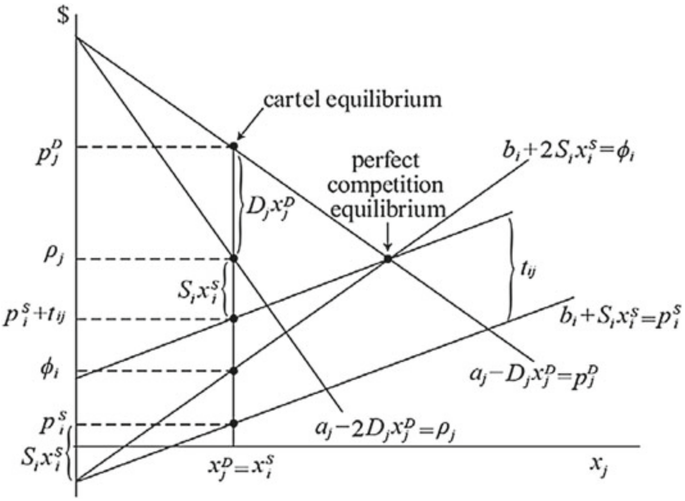

Fig. 1 Cartel behavior: monopoly–monopsony

and the monopsonist, respectively. These terms are also regarded as measures of monopoly and monopsony power. Figure 1 illustrates this cartel spatial model.

The perfect competition solution, originally formulated by Samuelson (1952) and further elaborated by Takayama and Judge (1964), is easily obtained by setting the value markups $D_j x_j^D$ and $S_i x_i^S$ equal to zero. In this case, the equilibrium condition (9) reduces to $p_j^D \leq p_i^S + t_{ij}$, as illustrated in Fig. 1. In the perfect competition case, $\rho_j = p_j^D$ and $\phi_i = p_i^S$.

3 Spatial Nash-Cournot Equilibrium: Oligopoly–Oligopsony

The perfect competition and the cartel (monopoly–monopsony) models represent limiting specifications of spatial equilibrium. In between these two cases, there exists a wide series of behavioral performances classified under the two categories of non-cooperative and cooperative imperfect competition rules. We consider here an imperfect competition hypothesis that goes under the name of non-cooperative Nash equilibrium (sometimes, the same hypothesis is called Cournot equilibrium). The analytical structure of this spatial model was developed by Hashimoto (1985). In this context, suppliers (who operate under a perfect competition market) are also exporters. Consumers are price takers, as usual. Each region has one supplier–exporter who makes profit-maximizing decisions about output quantities assuming that his choices do not affect the decisions of supplier–exporters in other regions. This is the non-cooperative feature of the model.

In the process toward a complete Nash-Cournot model, the ith supplying region primal problem states the maximization of profit, π_i, subject to the supply constraint of the ith region

$$\begin{aligned}\max\ \pi_i &= \sum_{j=1}^{R} p_j^D x_{ij} - (b_i + S_i x_i^S/2)x_i^S - \sum_{j=1}^{R} t_{ij}x_{ij} \\ &= \sum_{j=1}^{R} (a_j - D_j x_j^D)x_{ij} - (b_i + S_i x_i^S/2)x_i^S - \sum_{j=1}^{R} t_{ij}x_{ij} \\ &= \sum_{j=1}^{R} (a_j - D_j \sum_{k=1}^{R} x_{kj})x_{ij} - (b_i + S_i x_i^S/2)x_i^S - \sum_{j=1}^{R} t_{ij}x_{ij} \qquad (11)\end{aligned}$$

subject to

$$\sum_{j=1}^{R} x_{ij} \leq x_i^S \qquad (12)$$

The first term of the profit equation is total revenue. The second term is the cost of purchasing the commodity supply, and the third term is the transaction costs. The non-cooperative hypothesis is expressed by the equation $x_j^D = \sum_k x_{kj}$ which is simply the sum of the supply quantities of all the regions satisfying the demand of the jth region. The relevant KKT conditions of problem [(11)–(12)] are derived from the Lagrange function

$$\begin{aligned}L_i = &\sum_{j=1}^{R}\left(a_j - D_j \sum_{k=1}^{R} x_{kj}\right)x_{ij} - (b_i + S_i x_i^S/2)x_i^S - \sum_{j=1}^{R} t_{ij}x_{ij} \\ &+ \sum_{i=1}^{R} p_i^S\left(x_i^S - \sum_{j=1}^{R} x_{ij}\right) \qquad (13)\end{aligned}$$

and KKT conditions

$$\frac{\partial L_i}{\partial x_i^S} = -b_i - S_i x_i^S + p_i^S \leq 0 \qquad (14)$$

$$\begin{aligned}\frac{\partial L_i}{\partial x_{ij}} &= (a_j - D_j \sum_{k=1}^{R} x_{kj}) - D_j x_{ij} - p_i^S - t_{ij} \leq 0 \\ &= p_j^D - D_j x_{ij} - p_i^S - t_{ij} \leq 0 \qquad (15)\end{aligned}$$

Assuming a positive trade flow on the $i-j$ route, relation (15) becomes an equation $p_j^D = (p_i^S + t_{ij}) + D_j x_{ij} = \mathrm{MC}_{ij} + D_j x_{ij}$ (by complementary slackness condition).

In other words, the Nash-Cournot demand price of the ith oligopolistic firm in the jth region is equal to the marginal cost plus the segment (markup) $D_j x_{ij}$ (oligopoly power).

The above discussion pertaining to the non-cooperative behavior of the ith region (oligopoly firm) guides the specification of the overall spatial Nash-Cournot equilibrium model that must be expressed as a mathematical programming structure capable of reproducing the necessary conditions (KKT conditions) of each oligopoly firm (region) as stated in relations (14) and (15). Such a model assumes the following specification

$$\max \ \text{Nash} = \sum_{j=1}^{R} (a_j - D_j x_j^D/2) x_j^D - \sum_{i=1}^{R} (b_i + S_i x_i^S/2) x_i^S \\ - \sum_{i=1}^{R} \sum_{j=1}^{R} t_{ij} x_{ij} - \sum_{i=1}^{R} \sum_{j=1}^{R} D_j x_{ij}^2/2 \tag{16}$$

subject to

$$x_j^D \leq \sum_{i=1}^{R} x_{ij} \tag{17}$$

$$\sum_{j=1}^{R} x_{ij} \leq x_i^S \tag{18}$$

with all nonnegative variables. The term $\sum_i \sum_j D_j x_{ij}^2/2$ is required for deriving the correct KKT conditions of each non-cooperative Nash-Cournot firm as demonstrated in [(20)–(21)]. The Lagrange function is stated as

$$L = \sum_{j=1}^{R} (a_j - D_j x_j^D/2) x_j^D - \sum_{i=1}^{R} (b_i + S_i x_i^S/2) x_i^S - \sum_{i=1}^{R} \sum_{j=1}^{R} t_{ij} x_{ij} \\ - \sum_{i=1}^{R} \sum_{j=1}^{R} D_j x_{ij}^2/2 + \sum_{j=1}^{R} p_j^D \left(\sum_{i=1}^{R} x_{ij} - x_j^D \right) + \sum_{i=1}^{R} p_i^S \left(x_i^S - \sum_{j=1}^{R} x_{ij} \right) \tag{19}$$

with relevant KKT conditions

$$\frac{\partial L}{\partial x_j^D} = a_j - D_j x_j^D - p_j^D \leq 0 \tag{20}$$

$$\frac{\partial L}{\partial x_i^S} = -b_i - S_i x_i^S + p_i^S \leq 0 \tag{21}$$

$$\frac{\partial L}{\partial x_{ij}} = p_j^D - t_{ij} - D_j x_{ij} - p_i^S \leq 0 \tag{22}$$

Relations (21) and (22) are identical to relations (14) and (15) which characterize the Nash-Cournot structure of the spatial problem for the ith oligopoly firm.

Analogous specification can be formulated for the jth oligopsony firm with the final result that the overall oligopoly–oligopsony Nash-Cournot equilibrium can be specified as the following primal problem

$$\max \text{ Nash } \pi = \sum_{j=1}^{R}(a_j - D_j x_j^D/2)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S/2)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij}$$
$$- \sum_{i=1}^{R}\sum_{j=1}^{R} D_j x_{ij}^2/2 - \sum_{i=1}^{R}\sum_{j=1}^{R} S_i x_{ij}^2/2 \tag{23}$$

subject to

$$x_j^D \leq \sum_{i=1}^{R} x_{ij} \tag{24}$$
$$\sum_{j=1}^{R} x_{ij} \leq x_i^S \tag{25}$$

The terms $\sum_{i=1}^{R}\sum_{j=1}^{R} D_j x_{ij}^2/2$ and $\sum_{i=1}^{R}\sum_{j=1}^{R} S_i x_{ij}^2/2$ are required to obtain the correct KKT conditions, which are

$$\frac{\partial L}{\partial x_j^D} = a_j - D_j x_j^D - p_j^D \leq 0 \tag{26}$$
$$\frac{\partial L}{\partial x_i^S} = -b_i - S_i x_i^S + p_i^S \leq 0 \tag{27}$$
$$\frac{\partial L}{\partial x_{ij}} = p_j^D - t_{ij} - D_j x_{ij} - p_i^S - S_i x_{ij} \leq 0 \tag{28}$$

Relation (28), in particular, expresses the behavioral guidelines of this oligopoly–oligopsony hypothesis that is reflected in the fundamental MR $\leq$ MC relation with the following structure $p_j^D - D_j x_{ij} \leq p_i^S + t_{ij} + S_i x_{ij}$. The markups $D_j x_{ij}$ and $S_i x_{ij}$ express the oligopoly and the oligopsony power of the corresponding exporters and suppliers. Figure 2 illustrates this Nash-Cournot (oligopoly–oligopsony) hypothesis.

From Fig. 2, we can deduce that, in general, the cartel quantity will be smaller than the Nash-Cournot quantity which, in turn, will be smaller than the perfect competition quantity. Correspondingly, the cartel monopoly price will be higher than the Nash-Cournot oligopoly price that, in turn, will be higher than the perfect competition price. An inverse relation characterizes the supply prices.

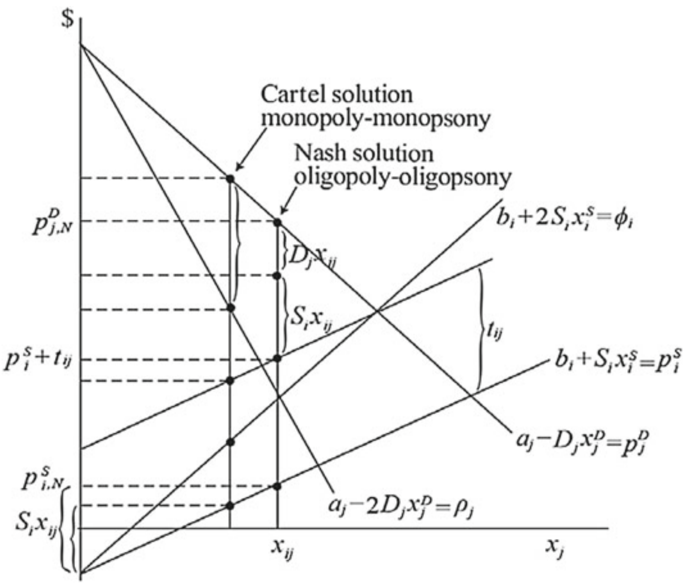

Fig. 2 Nash-Cournot equilibrium: oligopoly–oligopsony

4 Numerical Examples of Spatial Equilibria

We present a numerical example of four regions that produce and exchange one commodity through one of the three behavioral hypotheses discussed in previous sections. These results are exhibited from Tables 3, 4, 5, 6, 7 and 8.

We begin with the given information common to the three models. Table 1 presents the intercepts and slopes of the demand and supply functions for the four regions.

Table 2 presents the unit transaction costs. Notice that the nominal transaction cost within each region is equal to zero. It could be positive, for example, if we were to consider a commodity priced at farm gate that is sold at a retail store within the same region.

Tables 3, 4, 5, 6, 7 and 8 present the solutions of the three behavioral models discussed in the previous sections. For ease of comparison, we group the results of the

Table 1 Demand and supply functions

Regions	Demand intercept a_j	Demand slope D_j	Supply intercept b_i	Supply slope S_i
A	40.0	1.2	0.4	1.3
B	32.0	0.8	0.2	2.0
U	25.0	0.8	−0.6	1.9
E	38.0	1.1	−0.5	0.6

various optimal quantities and prices according to the order: cartel, non-cooperative Nash-Cournot, and perfect competition.

The obvious comment is that the cartel and the perfect competition models exhibit positive trade flows in very few locations. In contrast, the Nash-Cournot trade flows present positive trade in all the locations. Further comments concerning each region total demand and supply are presented in connection with Tables 4 and 5.

As indicated by Fig. 2, the cartel model chooses the smallest quantity, the non-cooperative Nash-Cournot firms choose an intermediate quantity, and the perfect competition model chooses the largest quantity.

Each region follows the trend where the cartel, the non-cooperative Nash-Cournot firms and the perfect competitive model exhibit, respectively, an increasing quantity of commodity supplied. Total supply is obviously equal to total demand.

Tables 6 and 7 present the equilibrium prices for the five models.

As indicated in Fig. 2, demand prices are higher for the cartel followed by the Nash-Cournot and perfect competition behaviors.

Table 2 Unit transaction costs t_{ij}

Regions	A	B	U	E
A	0.000	2.050	1.620	10.800
B	2.050	0.000	3.240	10.800
U	1.620	3.240	0.000	9.990
E	10.800	10.800	9.990	0.000

Table 3 Trade flow of the five behavioral models

Cartel (monopoly-monopsony) trade flow, x_{ij}				
A	7.699			
B		5.459		
U	0.460	0.767	3.877	
E				11.324
Non-coop Nash (oligopoly-oligopsony) trade flow, x_{ij}				
A	4.601	2.434	2.026	0.819
B	2.524	3.464	0.994	0.133
U	3.362	2.196	2.079	0.427
E	2.223	1.766	0.733	10.042
Perfect competition trade flow, x_{ij}				
A	15.398			
B		10.919		
U	0.921	1.535	7.753	
E				22.647

Table 4 Demand of each region, x_j^D

A	B	U	E	Total
Cartel (monopoly–monopsony) quantity demanded				
8.159	6.227	3.877	11.324	29.586
Non-cooperative Nash (oligopoly–oligopsony) quantity demanded				
12.711	9.861	5.832	11.421	39.825
Perfect competition quantity demanded				
16.319	12.453	7.753	22.647	59.173

Table 5 Supply of each region, x_i^S

A	B	U	E	Total
Cartel (monopoly–monopsony) quantity supplied				
7.699	5.459	5.105	11.324	29.586
Non-cooperative Nash (oligopoly–oligopsony) quantity supplied				
9.880	7.106	8.064	14.775	39.825
Perfect competition quantity supplied				
15.398	10.919	10.209	22.647	59.173

Table 6 Demand price for each region, p_j^D

A	B	U	E
Cartel (monopoly–monopsony) demand prices			
30.209	27.019	21.899	25.544
Non-cooperative (oligopoly–oligopsony) Nash demand prices			
24.747	24.111	20.335	25.437
Perfect competition demand prices			
20.417	22.037	18.797	13.088

Table 7 Supply price for each region, p_i^S

A	B	U	E
Cartel (monopoly–monopsony) supply prices			
10.409	11.119	9.099	6.294
Non-cooperative (oligopoly–oligopsony) Nash supply prices			
13.245	14.411	14.722	8.365
Perfect competition supply prices			
20.417	22.037	18.797	13.088

Table 8 Profit of each region

A	B	U	E	Total
Cartel (monopoly–monopsony) profit				
156.948	90.626	61.530	217.978	527.082
Non-cooperative (oligopoly–oligopsony) Nash profit				
81.740	52.107	53.745	194.096	381.688
Perfect competition profit				
0.000	0.000	0.000	0.000	0.000

Supply prices follow the inverse relation of demand prices: Cartel supply prices are the lowest ones followed by the Nash-Cournot prices and finishing with the perfect competition prices.

Table 8 presents the profit for each region. The cartel acquires the highest level of profit in each region followed by the profit of the non-cooperative Nash-Cournot firms and by the perfectly competitive firms whose profit is equal to zero by construction.

5 Spatial Equilibrium Under Imprecise Transaction Costs

In Sect. 2, we assumed that the realized and observed trade flow among regions, x_{ij}^{obs}, is known to the researcher. In general, the optimal trade flow matrix resulted from the solution of model [(3)–(5)], $\left[x_{ij}^{*}\right]$, (or any other model discussed in previous sections) diverges substantially from the observed trade flow matrix $\left[x_{ij}^{\mathrm{obs}}\right]$. This discrepancy may be attributed to the imprecision of the given information and especially to the imprecision of the unit transaction costs t_{ij}.

To remedy this undesirable result, model [(3)–(5)] may be augmented with a tautological constraint such as

$$x_{ij} = x_{ij}^{\mathrm{obs}} \tag{29}$$

whose purpose is to uncover the additional marginal transaction cost of producing a trade flow equal to the observed traded quantities. This additional marginal cost corresponds to the shadow price of constraint (29) (see Paris et al. 2010).

In order to approach the calibration issue gradually, we will discuss first the calibration of a perfect competition spatial equilibrium model. The calibration procedure develops in two phases. Phase I is concerned about an estimate of the adjustments to the observed unit transaction costs. Phase II uses these adjustments to define a calibrating model. In phase I, the relevant Lagrange function for a perfect competition spatial equilibrium model is specified as follows:

$$L = \sum_{j=1}^{R}(a_j - D_j x_j^D/2)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S/2)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij} + \sum_{j=1}^{R} p_j^D\left(\sum_{i=1}^{R} x_{ij} - x_j^D\right) + \sum_{j=1}^{R} p_i^S\left(x_i^S - \sum_{j=1}^{R} x_{ij}\right) + \sum_{i=1}^{R}\sum_{j=1}^{R}\lambda_{ij}\left(x_{ij}^{\text{obs}} - x_{ij}\right) \tag{30}$$

while the relevant KKT conditions show that

$$\frac{\partial L}{\partial x_j^D} = a_j - D_j x_j^D - p_j^D \le 0 \tag{31}$$

$$\frac{\partial L}{\partial x_i^S} = -b_i - S_i x_i^S + p_i^S \le 0 \tag{32}$$

$$\frac{\partial L}{\partial x_{ij}} = p_j^D - p_i^S - t_{ij} - \lambda_{ij} \le 0 \tag{33}$$

Relation (33) shows that, for a perfect competition equilibrium, $p_j^D \le p_i^S + (t_{ij} + \lambda_{ij})$ which corresponds to $\text{MR}_j \le \text{MC}_{ij}$, the demand price (marginal revenue) in the jth region must be less than or equal to the marginal cost of supplying the jth region by way of the $i - j$ route. Given the equality sign of constraints (29), the corresponding dual variables λ_{ij} are unrestricted. Therefore, using the estimate $\hat{\lambda}_{ij}$ of the additional marginal cost of supplying and transporting the commodity over the $i - j$ route, it is possible to specify a phase II calibrating spatial equilibrium model that does not include the tautological constraints (29) by simply augmenting (or reducing) the original measure of the unit transaction costs as in the following objective quasi-welfare function (QWF)

$$\max\ \text{QWF} = \sum_{j=1}^{R}(a_j - D_j x_j^D/2)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S/2)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R}(t_{ij} + \hat{\lambda}_{ij})x_{ij} \tag{34}$$

The solution of model (34) with constraints (4) and (5) exhibits an optimal trade flow that is identical to the observed trade flow; that is, $\left[x_{ij}^{\text{cal PC}}\right] = \left[x_{ij}^{\text{obs}}\right]$. We call this model a perfectly calibrated (cal) model for perfectly competitive (PC) markets.

6 Calibrated Cartel Spatial Model: Monopoly–Monopsony

It is very likely that also the optimal trade flow matrix in the spatial cartel model (monopoly–monopsony) [(3)–(5)] will be different from the observed trade flow matrix. In order to approach the calibration of cartel models gradually, we discuss first the monopoly–perfect competition rule. Therefore, a calibrated monopoly–perfect competition (cartel) model can be specified in the same two-phase way as the perfect competitive spatial model discussed in Sect. 5, with an important qualification. The use of the same observed trade matrix $\left[x_{ij}^{\mathrm{obs}}\right]$ for any behavioral hypothesis implies that the demand quantity of the cartel is equal to the demand quantity of the perfectly competitive model. This last quantity is determined by the intersection of the demand and supply functions (plus transaction costs). Therefore, in this calibrating cartel model—where the demand quantity is equal for both the monopolist and the perfectly competitive firms—it is as if the given demand function were to act as a marginal revenue function for the cartel (see Fig. 3). This means that a fictitious demand function for the cartel becomes $p_j^{D,F} = a_j - D_j x_j^D/2$ (where the superscript F stands for fictitious). The calibrating objective function of this spatial cartel model is thus

$$\begin{aligned} \max\ \text{Cartel}\pi &= \sum_{j=1}^{R} p_j^{D,F} x_j^D - \sum_{i=1}^{R} (b_i + S_i x_i^S/2) x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} (t_{ij} + \hat{\lambda}_{ij}) x_{ij} \\ &= \sum_{j=1}^{R} (a_j - D_j x_j^D/2) x_j^D - \sum_{i=1}^{R} (b_i + S_i x_i^S/2) x_i^S \\ &\quad - \sum_{i=1}^{R}\sum_{j=1}^{R} (t_{ij} + \hat{\lambda}_{ij}) x_{ij} \end{aligned} \tag{35}$$

We remark that—in this case—the revenue term $(a_j - D_j x_j^D/2)x_j^D$ is not the integral under the original demand function. It is just the monopoly revenue where

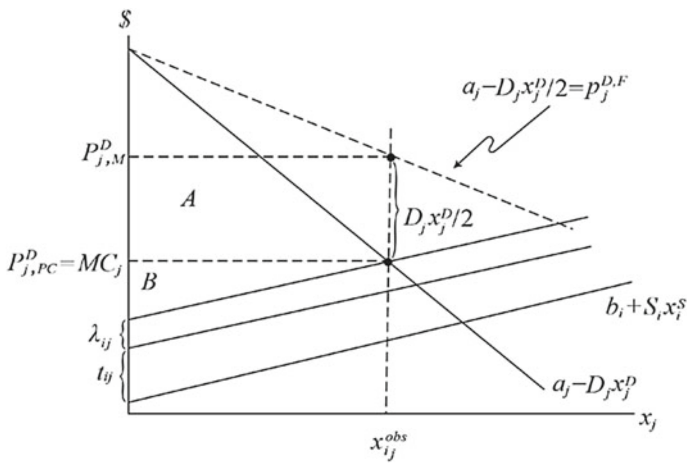

Fig. 3 Calibrated cartel (M) and perfect competition (PC) models

the monopoly price is $p_j^{D,F} = a_j - D_j x_j^D/2$: Linear and quadratic specifications play this trick.

The objective function (35), together with constraints (4) and (5), will produce an optimal trade flow that is identical to the observed trade flow. The cartel demand prices, however, will be different from the demand prices of the perfect competition model as shown in Fig. 3.

We now tackle the cartel (monopoly–monopsony) specification. The fictitious demand function is the same as for the previous cartel (monopoly–perfect competition) model. The input prices are given by the familiar function $p_i^S = b_i + S_i x_i^S$. Then, the calibrating (phase II) objective function of the cartel (monopoly–monopsony) assume the following specification

$$
\begin{aligned}
\max \ \text{Cartel } \pi &= \sum_{j=1}^{R} p_j^{D,F} x_j^D - \sum_{i=1}^{R} p_i^S x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R}(t_{ij} + \hat{\lambda}_{ij})x_{ij} \\
&= \sum_{j=1}^{R}(a_j - D_j x_j^D/2)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R}(t_{ij} + \hat{\lambda}_{ij})x_{ij}
\end{aligned} \tag{36}
$$

We arrive at this specification by means of the phase I Lagrange function

$$
\begin{aligned}
\max \ L &= \sum_{j=1}^{R} p_j^{D,F} x_j^D - \sum_{i=1}^{R} p_i^S x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij} + \text{constraints} \\
&= \sum_{j=1}^{R}(a_j - D_j x_j^D/2)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R} t_{ij}x_{ij} \\
&\quad + \sum_{j=1}^{R}\rho_j\left(\sum_{i=1}^{R} x_{ij} - x_j^D\right) + \sum_{i=1}^{R}\phi_i\left(x_i^S - \sum_{j=1}^{R} x_{ij}\right) \\
&\quad + \sum_{i=1}^{R}\sum_{j=1}^{R}\lambda_{ij}(x_{ij}^{\text{obs}} - x_{ij})
\end{aligned} \tag{37}
$$

and the relevant KKT conditions that show the structure of the marginal revenue and marginal cost associated with the calibrating trade flow:

$$
\begin{aligned}
\frac{\partial L}{\partial x_j^D} &= a_j - D_j x_j^D - \rho_j \le 0 \ \Rightarrow \ \rho_j = a_j - D_j x_j^D \text{ for } x_j^D > 0 \\
\frac{\partial L}{\partial x_i^S} &= -(b_i + 2S_i x_i^S) + \phi_i \le 0 \ \Rightarrow \ \phi_i = (b_i + 2S_i x_i^S) \text{ for } x_i^S > 0 \\
\frac{\partial L}{\partial x_{ij}} &= -t_{ij} + \rho_j - \phi_i - \lambda_{ij} \le 0
\end{aligned}
$$

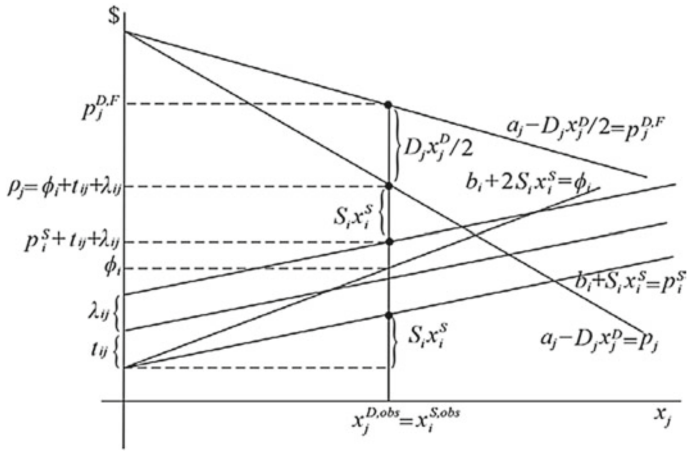

Fig. 4 Calibrated cartel: monopoly–monopsony

$$
\begin{aligned}
&-t_{ij} + (a_j - D_j x_j^D/2) - D_j x_j^D/2 - (b_i + S_i x_i^S) - S_i x_i^S - \lambda_{ij} \leq 0 \\
&p_j^{D,F} - D_j x_j^D/2 \leq p_i^S + (t_{ij} + \lambda_{ij}) + S_i x_i^S \\
&\text{MR} \leq \text{MC}
\end{aligned}
\tag{38}
$$

This cartel (monopoly–monopsony) specification is illustrated in Fig. 4. A comparison between Fig. 3 [cartel (monopoly–perfect competition)] and Fig. 4 requires that the two quantity solutions be aligned on the same vertical line. In Fig. 4, therefore, the cartel (monopoly–perfect competition) solution would require setting $S_i x_i^S = 0$ and a different adjustment λ_{ij} that will intercept the monopoly marginal revenue according to the relation $p_j^{D,F} - D_j x_j^D/2 \leq p_i^S + (t_{ij} + \lambda_{ij})$.

7 Calibrated Nash-Cournot Equilibrium: Oligopoly–Oligopsony

Also in this case, the optimal trade flow resulting from the solution of model [(23)–(25)] is likely to be different from the observed trade flow, x_{ij}^{obs}. Again, we proceed gradually and discuss first a Nash-Cournot model where firms behave as oligopolies on the export market and as perfect competitors on the supply market. By adopting the two-phase procedure described above, it is possible to define a calibrating Nash-Cournot equilibrium model that reproduces an optimal trade flow equal to the observed commodity exchange. In this case, and keeping in mind that the fictitious demand function is defined as $p_j^{D,F} = a_j - D_j x_j^D/2$, the calibrating objective function for a Nash-Cournot equilibrium assumes the following specification

$$\max\ \text{Calibrated} - \text{Nash} = \sum_{j=1}^{R}(a_j - D_j x_j^D/4)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S/2)x_i^S \\ - \sum_{i=1}^{R}\sum_{j=1}^{R}(t_{ij} + \hat{\lambda}_{ij})x_{ij} - \sum_{i=1}^{R}\sum_{j=1}^{R} D_j x_{ij}^2/4 \tag{39}$$

where $\hat{\lambda}_{ij}$ are—as before—the optimal shadow prices of the tautological constraints (29). The objective function (39) together with constraints (4) and (5) reproduces an optimal trade flow that is identical to the observed exchanges. The relevant KKT conditions are derived from the following Lagrange function

$$L = \sum_{j=1}^{R}(a_j - D_j x_j^D/4)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S/2)x_i^S - \sum_{i=1}^{R}\sum_{j=1}^{R}(t_{ij} + \hat{\lambda}_{ij})x_{ij} \\ - \sum_{i=1}^{R}\sum_{j=1}^{R} D_j x_{ij}^2/4 + \sum_{j=1}^{R} p_j^D\left(\sum_{i=1}^{R} x_{ij} - x_j^D\right) + \sum_{i=1}^{R} p_i^S\left(x_i^S - \sum_{j=1}^{R} x_{ij}\right) \tag{40}$$

with KKT conditions

$$\frac{\partial L}{\partial x_j^D} = a_j - D_j x_j^D/2 - p_j^D \le 0 \tag{41}$$

$$\frac{\partial L}{\partial x_i^S} = -b_i - S_i x_i^S + p_i^S \le 0 \tag{42}$$

$$\frac{\partial L}{\partial x_{ij}} = p_j^D - (t_{ij} + \hat{\lambda}_{ij}) - D_j x_{ij}/2 - p_i^S \le 0 \tag{43}$$

Although quantities and prices of the calibrated non-cooperative Nash-Cournot equilibrium are equal to the quantities and prices of the cartel problem, the overall calibrated Nash-Cournot profit is lower than the calibrated cartel profit because of the higher adjustments of the transaction costs, as exemplified in Fig. 5. Algebraically, the lower profit of the non-cooperative Nash-Cournot model is established by recalling that all the quantities are the same for the cartel and the Nash-Cournot models. Furthermore, the demand prices of the two models are also identical. Hence,

$$p_{j,M}^D = p_i^S + (t_{ij} + \lambda_{ij,M}) + D_j x_j^D/2 = p_{j,N}^D = p_i^S + (t_{ij} + \lambda_{ij,N}) \\ + D_j x_{ij}/2\lambda_{ij,M} + D_j x_j^D/2 = \lambda_{ij,N} + D_j x_{ij}/2 \tag{44}$$

Notice that, in general, $D_j x_j^D/2 > D_j x_{ij}/2$ with the consequence that $(\lambda_{ij,N} > \lambda_{ij,M})$. This conclusion reduces the profit of the non-cooperative Nash-Cournot equilibrium firms (regions) below the profit of the cartel, as illustrated in Fig. 5.

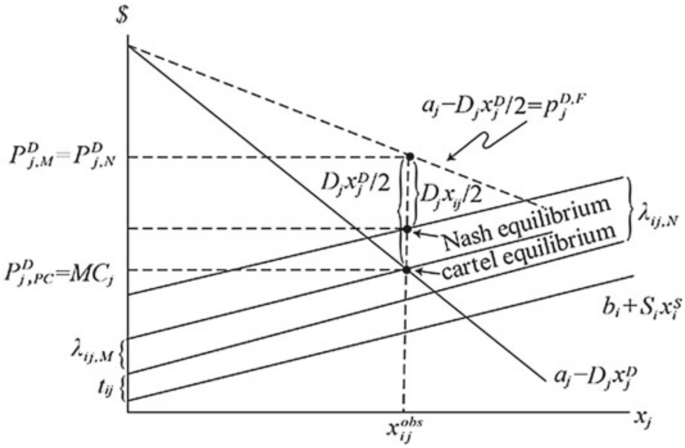

Fig. 5 Calibrated cartel (M) and Nash (N) equilibria

To deal with the Nash-Cournot case where firms behave as oligopolies on the export market and oligopsonies on the supply market, we modify the preceding objective function and integrate it with the oligopsony specification to exhibit

$$\max \text{ Calibrated} - \text{ Nash} = \sum_{j=1}^{R}(a_j - D_j x_j^D/4)x_j^D - \sum_{i=1}^{R}(b_i + S_i x_i^S)x_i^S$$
$$- \sum_{i=1}^{R}\sum_{j=1}^{R}(t_{ij} + \hat{\lambda}_{ij})x_{ij} - \sum_{i=1}^{R}\sum_{j=1}^{R} D_j x_{ij}^2/4 - \sum_{i=1}^{R}\sum_{j=1}^{R} S_i x_{ij}^2/2 \qquad (45)$$

By now, we know how to state the relevant KKT conditions that, in this case, take on the following structure

$$\frac{\partial L}{\partial x_{ij}} = p_j^D - (t_{ij} + \hat{\lambda}_{ij}) - D_j x_{ij}/2 - p_i^S - S_i x_{ij} \leq 0 \qquad (46)$$

that, when rearranged as MR $\leq$ MC, assumes the analytical form of

$$p_j^D - D_j x_{ij}/2 \leq p_i^S + (t_{ij} + \hat{\lambda}_{ij}) + S_i x_{ij} \qquad (47)$$

We remind the reader that the term $D_j x_{ij}/2$ is the markup (from the perfect competition marginal cost) due to oligopoly power and $S_i x_{ij}$ is the markup due to oligopsony power.

8 Example of Spatial Equilibria with Imprecise Transaction Costs

The "observed" trade matrix is presented in Table 9.

This table will apply to each of the three behavioral models. This means that the regional total demand and supply quantities are the same for the three behavioral hypotheses. In turn, the demand prices of the three behavioral hypotheses must be aligned on the same vertical line, as in Figs. 4 and 5.

Given the application of the tautological constraint (29), all three models calibrate precisely the trade flow matrix. Therefore, we do not report the achieved calibration of the trade flow matrix. One must remark that, given the transportation structure of the trade flow matrix, it would be possible for optimal solutions to exist that are different from the trade flow matrix. However, if the initial point of the nonlinear problem is chosen as the observed trade flow, it is very likely that the optimal solution will correspond to the observed trade flow.

Table 10 presents the adjustment to the transaction costs and the total transaction costs.

Negative transaction costs can be interpreted as subsidies. The complex pattern of adjustments and the many negative entries may be due to the subjective choice of parameter values.

Tables 11 and 12 exhibit the demand and supply prices under calibration of the observed trade flow.

Table 13 presents the profit under calibration.

The calibrated total profit follows the trend discussed in a previous section. The highest profit level accrues to the cartel. The non-cooperative Nash-Cournot firms (regions) exhibit a lower profit level even though the demand and supply prices are identical to those of the cartel model. Within regions U and E, this trend does not apply.

Given the observed trade flow, x_{ij}^{obs}, the question arises as to which of the three behavioral models would better fit (interpret economically) the observed empirical scenario. In general, this type of question requires a statistical test that—in this case—is precluded by the single observation of the trade flow. We can make the following remarks. First, the observed trade flow matrix indicates the routes that are activated between two regions. Hence, the model that approximates more closely the pattern of activated routes may have an edge over the others spatial rules. Second,

Table 9 Observed trade flow matrix, x_{ij}^{obs}

Regions	A	B	U	E	Total supply
A	3.0	2.5	2.0		7.5
B	0.5	5.0	1.0		6.5
U	1.0		6.0	4.0	11.0
E	10.0	3.0		5.0	18.0
Total demand	14.5	10.5	9.0	9.0	

Table 10 Transaction cost adjustments ($\hat{\lambda}_{ij}$) and total transaction costs (TTC)

Transaction cost adjustments					Total transaction costs				
Cartel (monopoly–monopsony) transaction costs									
	A	B	U	E		A	B	U	E
A	2.70	−0.35	−5.07	−2.60	A	2.700	3.700	2.100	8.200
B	−7.65	−2.60	−11.6	−8.90	B	−3.600	−2.600	−8.40	1.900
U	−21.6	−20.8	−23.4	−23.1	U	−18.6	−17.6	−23.4	−13.1
E	−9.30	−8.30	−13.3	7.00	E	1.500	2.500	−3.30	7.000
Non-cooperative Nash (oligopoly–oligopsony) transaction costs									
	A	B	U	E		A	B	U	E
A	5.700	−2.15	−6.07	2.350	A	5.700	1.900	−3.15	13.150
B	0.100	−10.4	−10.3	−3.95	B	4.150	−10.40	−7.10	6.850
U	−17.70	−16.64	−33.6	−22.7	U	−11.80	−13.40	−33.6	−12.75
E	−19.60	−11.30	−9.69	6.200	E	−8.800	−0.500	0.300	6.200
Perfect competition transaction costs									
	A	B	U	E		A	B	U	E
A	12.450	9.400	4.680	7.150	A	12.450	13.450	7.650	17.950
B	5.350	10.400	1.360	4.100	B	9.400	10.400	4.600	14.900
U	−0.670	0.060	−2.50	−2.190	U	2.300	3.300	−2.50	7.800
E	1.500	2.500	−2.49	17.800	E	12.300	13.300	7.500	17.800

Table 11 Demand price for each region (with calibration), p_j^D

A	B	U	E
Cartel (monopoly–monopsony) demand prices			
31.300	27.800	21.400	33.050
Non-cooperative Nash (oligopoly–oligopsony) demand prices			
31.300	27.800	21.400	33.050
Perfect competition demand prices			
22.600	23.600	17.800	28.100

Table 12 Supply price for each region (with calibration), p_i^S

A	B	U	E
Cartel (monopoly–monopsony) supply prices			
10.150	13.200	20.300	10.300
Non-coop Nash (oligopoly–oligopsony) supply prices			
10.150	13.200	20.300	10.300
Perfect competition supply prices			
10.150	13.200	20.300	10.300

Table 13 Regional total profit (with calibration)

A	B	U	E	Total
Cartel (monopoly–monopsony) profit				
199.275	128.600	262.300	238.950	829.125
Non-cooperative Nash (oligopoly–oligopsony) profit				
70.625	118.325	282.900	310.400	782.250
Perfect competition profit				
0.000	0.000	0.000	0.000	0.000

the level of traded commodity is another important factor. To gauge a measure of discrepancy between observed levels of the traded commodity and the levels generated by the three behavioral models, we compute the average squared deviation for those matching routes that are activated in both the models and the observed trade matrix.

For this numerical example, those results are as follows:

Activated routes in the observed trade flow matrix	12
Activated routes in the cartel (monopoly–monopsony) matrix	6
Matching activated routes in the cartel (monopoly–monopsony) matrix	5
Activated routes in the Nash (oligopoly–oligopsony) matrix	16
Matching activated routes in the Nash (oligopoly–oligopsony) matrix	12
Activated routes in the perfect competition matrix	6
Matching activated routes in the perfect competition matrix	6

On the basis of these matching results, it appears that the non-cooperative Nash-Cournot model (oligopoly–oligopsony) would be preferred. The average squared deviation for the quantities transported on the matching routes is as follows:

Cartel (monopoly–monopsony)	40.834
Non-cooperative Nash (oligopoly–oligopsony)	11.335
Perfect competition	104.442

The Nash-Cournot model (oligopoly–oligopsony) is preferred also on the basis of this measure.

9 Imprecise Intercepts of Demand and Supply Functions

Let us suppose, again, that the trade flow between regions, x_{ij}^{obs}, is known. As a consequence, also the regional total demand and supply, $x_j^{D,\text{obs}}$ and $x_i^{S,\text{obs}}$, are known. Furthermore, when the trade flow is observed, also demand and supply prices may be known. We assume, therefore, that also these prices are observed: $p_j^{D,\text{obs}}$ and $p_i^{S,\text{obs}}$. Often, the demand elasticities for the various regions are assumed known, say η_j^{obs}. From the above available information, it is possible to reconstruct the implied demand function, albeit with some imprecision of the intercept:

$$\eta_j^{\text{obs}} = \frac{\partial x_j^D}{\partial p_j^D} \frac{p_j^{D,\text{obs}}}{x_j^{D,\text{obs}}} < 0 \rightarrow \frac{x_j^{D,\text{obs}}}{p_j^{D,\text{obs}}} \eta_j^{\text{obs}} = \frac{\partial x_j^D}{\partial p_j^D} \tag{48}$$

and the demand function can be stated as

$$x_j^D = a + \left(\frac{x_j^{D,\text{obs}}}{p_j^{D,\text{obs}}} \eta_j^{\text{obs}} \right) p_j^D \tag{49}$$

with the inverse demand function as

$$p_j^D = -a \left(\frac{p_j^{D,\text{obs}}}{x_j^{D,\text{obs}}} \frac{1}{\eta_j^{\text{obs}}} \right) + \left(\frac{p_j^{D,\text{obs}}}{x_j^{D,\text{obs}}} \frac{1}{\eta_j^{\text{obs}}} \right) x_j^D \tag{50}$$

It is clear that the choice of the parameter a may likely lead to imprecision in the intercept. With the given observed quantities and prices, x_{ij}^{obs}, $p_j^{D,\text{obs}}$ and $p_i^{S,\text{obs}}$, we deal first with the perfect competition specification. This time, because of the presence of observed prices, both phases I and II must take on the structure of a linear complementarity problem (LCP) that combines primal and dual constraints.

Phase I—Perfect Competition

$$\min \text{ LS}_{\text{PC}} = \sum_{j=1}^{R} e_j^2/2 + \sum_{i=1}^{R} v_i^2/2 + \sum_{j=1}^{R} \varepsilon_j^2/2 + \sum_{i=1}^{R} \psi_i^2/2 \tag{51}$$

subject to

(primal and dual constraints) Dual variables

$$x_j^D \leq \sum_{j=1}^{R} x_{ij} \qquad p_j^D \tag{52}$$

$$\sum_{i=1}^{R} x_{ij} \leq x_i^S \qquad p_i^S \tag{53}$$

$$x_{ij} = x_{ij}^{\text{obs}} \qquad \lambda_{ij} \tag{54}$$

$$(a_j + e_j) - D_j x_j^D \le p_j^D \qquad x_j^D \tag{55}$$

$$p_i^S \le (b_i + v_i) + S_i x_i^S \qquad x_i^S \tag{56}$$

$$p_j^D \le p_i^S + (t_{ij} + \lambda_{ij}) \qquad x_{ij} \tag{57}$$

$$p_j^D = p_j^{D,\text{obs}} + \varepsilon_j \qquad \varepsilon_j \tag{58}$$

$$p_i^S = p_i^{D,\text{obs}} + \psi_i \qquad \psi_j \tag{59}$$

(complementary slackness conditions)

$$p_j^D \left[\sum_i^R x_{ij} - x_j^D \right] = 0 \tag{60}$$

$$p_i^S \left[x_i^S - \sum_j^R x_{ij} \right] = 0 \tag{61}$$

$$x_j^D \left[p_j^D - (a_j + e_j) + D_j x_j^D \right] = 0 \tag{62}$$

$$x_i^S \left[(b_i + v_i) + S_i x_i^S - p_i^S \right] = 0 \tag{63}$$

$$x_{ij} \left[p_i^S + (t_{ij} + \lambda_{ij}) - p_j^D \right] = 0 \tag{64}$$

The least squares approach minimizes the squared adjustments of the intercepts and the squared deviations from the observed prices. The structure of the LCP exhibits the primal and dual constraints in relations (52)–(59) and the corresponding complementary slackness conditions in relations (60)–(64).

Phase II uses the least squares estimates of the intercept deviations, $\hat{e}_j$ and $\hat{v}_i$, and the dual variables of the tautological constraints (54), $\hat{\lambda}_{ij}$, to minimize the sum of all the complementary slackness conditions (CSC) which, by definition, must equal to zero. This specification uses slack variables in the primal, $z_j^{P1} \ge 0$, $z_i^{P2} \ge 0$, and the dual constraints, $z_j^{D1} \ge 0$, $z_i^{D2} \ge 0$, $z_{ij}^{D3} \ge 0$:

$$\min \ \text{CSC} = \sum_{j=1}^R z_j^{P1} p_j^D + \sum_{i=1}^R z_i^{P2} p_i^S + \sum_{j=1}^R z_j^{D1} x_j^D + \sum_{i=1}^R z_i^{D2} x_i^S + \sum_{i=1}^R \sum_{j=1}^R z_{ij}^{D3} x_{ij} = 0 \tag{65}$$

$$\text{subject to} \qquad x_j^D + z_j^{P1} = \sum_{j=1}^R x_{ij} \tag{66}$$

$$\sum_{i=1}^{R} x_{ij} + z_i^{P2} = x_i^S \tag{67}$$

$$(a_j + \hat{e}_j) - D_j x_j^D + z_j^{D1} = p_j^D \tag{68}$$

$$p_i^S + z_i^{D2} = (b_i + \hat{v}_i) + S_i x_i^S \tag{69}$$

$$p_j^D + z_{ij}^{D3} = p_i^S + (t_{ij} + \hat{\lambda}_{ij}) \tag{70}$$

Model (65)–(70) calibrates the solution to be equal to the observed trade flow, while the equilibrium demand and supply prices are within the smallest possible deviation from the observed demand and supply prices.

The same observed information, when applied to a cartel (monopoly–monopsony) specification, induces a modification of three dual constraints [(55), (56) and (57)] and the corresponding complementary slackness conditions, namely

Phase I—Cartel (monopoly–monopsony)

$$\begin{aligned}
&(a_j + e_j) - D_j x_j^D \le \rho_j \\
&(a_j + e_j) - D_j x_j^D \le p_j^D - D_j x_j^D/2 \\
&(a_j + e_j) - D_j x_j^D/2 \le p_j^D
\end{aligned} \tag{71}$$

and

$$\begin{aligned}
&\phi_i \le (b_i + v_i) + 2S_i x_i^S \\
&p_i^S + S_i x_i^S \le (b_i + v_i) + 2S_i x_i^S \\
&p_i^S \le (b_i + v_i) + S_i x_i^S
\end{aligned} \tag{72}$$

and assuming that each region will have positive demand and supply, $x_j^D > 0, x_i^S > 0$

$$\begin{aligned}
&\rho_j \le \phi_i + (t_{ij} + \lambda_{ij}) \\
&(a_j + e_j) - D_j x_j^D/2 - D_j x_j^D/2 \le (b_i + \nu_i) + S_i x_i^S + S_i x_i^S + (t_{ij} + \lambda_{ij}) \\
&p_j^D \le p_i^S + (t_{ij} + \lambda_{ij}) + D_j x_j^D/2 + S_i x_i^S
\end{aligned} \tag{73}$$

Finally, the same observed information, when applied to a non-cooperative Nash-Cournot (oligopoly–oligopsony) specification, induces a modification of the same three dual constraints [(55), (56) and (57)] and the corresponding complementary slackness conditions, namely

Phase I—Non-cooperative Nash (oligopoly–oligopsony)

$$(a_j + e_j) - D_j x_j^D/2 \leq p_j^D \tag{74}$$

and

$$p_j^D \leq p_i^S + (t_{ij} + \lambda_{ij}) + D_j x_{ij}/2 + S_i x_{ij} \tag{75}$$

The solution of the calibrated models for the cartel (monopoly–monopoly) and the non-cooperative Nash-Cournot (oligopoly–oligopsony) rules requires solving the same set of relations stated in [(51)–(64)] with the replacement of relations (55), (56) and (57) (together with the corresponding complementary slackness conditions) with relations (71), (72) and (73) for the cartel model and with (74) and (75) for the Nash-Cournot model.

10 Numerical Example of Spatial Equilibria with Imprecise Trade Flow and Imprecise Intercepts of the Demand and Supply Functions

Table 14 exhibits the observed information dealing with the trade flow as well as the demand and supply prices, x_{ij}^{obs}, $p_j^{D,\text{obs}}$, $p_i^{S,\text{obs}}$. The transaction costs are the same as in Table 2.

The three behavioral models calibrate precisely the observed trade flow matrices which, therefore, are not reported again. Table 15 presents the adjustments to the transaction costs and the total transaction costs.

The adjustments to the intercepts of the demand functions are given in Table 16.

The adjustments to the intercepts of the supply functions are presented in Table 17.

The demand prices of the three behavioral hypotheses are presented in Table 18 together with the observed prices.

The supply prices of the three behavioral hypotheses are presented in Table 19 together with the observed prices.

Table 20 presents the profit under this general calibration.

Table 14 Observed trade flow, demand, and supply prices, x_{ij}^{obs}, $p_j^{D,\text{obs}}$, $p_i^{S,\text{obs}}$

Regions	A	B	U	E	Total supply	Demand prices	Supply prices
A	3.0	2.5	2.0		7.5	24.0	16.0
B	0.5	5.0	1.0		6.5	22.5	19.0
U	1.0		6.0	4.0	11.0	19.1	15.0
E	10.0	3.0		5.0	18.0	18.0	11.0
Total demand	14.5	10.5	9.0	9.0			

Table 15 Transaction cost adjustments ($\hat{\lambda}_{ij}$) and total transaction costs (TTC)

Transaction cost adjustments					Total transaction costs				
Cartel (monopoly–monopoly) transaction costs									
	A	B	U	E		A	B	U	E
A	−3.875	−5.93	−9.15	10.450	A	−3.875	−1.875	−6.175	21.250
B	−14.20	−8.15	−15.7	10.425	B	−10.15	−8.150	−12.45	21.225
U	−22.57	4.260	−21.9	−27.96	U	−19.60	7.500	−21.90	−17.97
E	−13.30	−11.3	1.500	−0.875	E	−2.500	−0.500	11.490	0.875
Non-cooperative Nash (oligopoly–oligopsony) transaction costs									
	A	B	U	E		A	B	U	E
A	8.875	3.775	0.805	20.900	A	8.875	7.825	3.775	31.700
B	6.200	−2.95	−1.49	20.850	B	10.250	−2.950	1.750	31.650
U	4.530	4.260	−11.2	−11.91	U	7.500	7.500	−11.20	−1.925
E	−5.800	0.700	3.000	9.125	E	5.000	11.500	12.990	9.125
Perfect competition transaction costs									
	A	B	U	E		A	B	U	E
A	10.225	5.925	2.405	7.975	A	10.225	9.975	9.650	18.775
B	3.150	6.950	−0.89	7.950	B	7.200	6.950	2.350	18.750
U	2.680	2.160	0.800	−4.590	U	5.650	5.400	0.800	5.400
E	1.850	1.600	−0.30	12.400	E	12.650	12.400	9.690	12.400

Table 16 Adjustments to the intercept of the demand functions, $\hat{e}_j$

A	B	U	E
Cartel (monopoly–monopsony)			
−3.650	−2.650	−1.150	−7.525
Non-cooperative Nash (oligopoly–oligopsony)			
−3.650	−2.650	−1.150	−7.525
Perfect competition			
0.700	−0.550	0.650	−5.050

Table 17 Adjustments to the intercept of the supply functions, $\hat{v}_j$

A	B	U	E
Cartel (monopoly–monopsony)			
2.925	2.900	−2.650	0.350
Non-cooperative Nash (oligopoly–oligopsony)			
2.925	2.900	−2.650	0.350
Perfect competition			
2.925	2.900	−2.650	0.350

Table 18 Observed and calibrated demand prices, p_j^D

A	B	U	E
Observed demand prices			
24.000	22.500	19.100	18.000
Cartel (monopoly–monopsony)			
27.650	25.150	20.250	25.525
Non-cooperative Nash (oligopoly–oligopsony)			
27.650	25.150	20.250	25.525
Perfect competition			
23.300	23.050	18.450	23.050

Table 19 Observed and calibrated supply prices, p_i^S

A	B	U	E
Observed demand prices			
16.000	19.000	15.000	11.000
Cartel (monopoly–monopsony)			
13.075	16.100	17.650	10.650
Non-cooperative Nash (oligopoly–oligopsony)			
13.075	16.100	17.650	10.650
Perfect competition			
13.075	16.100	17.650	10.650

Table 20 Regional and total profit (with calibration of quantities and prices)

A	B	U	E	Total
Cartel (monopoly–monopsony) profit				
331.813	218.238	212.415	66.659	829.125
Non-cooperative Nash (oligopoly–oligopsony) profit				
34.525	63.050	124.500	157.750	379.825
Perfect competition profit				
0.000	0.000	0.000	0.000	0.000

The profit trend corresponds to the predicted outcome: The cartel exhibits the highest level of profit followed by the non-cooperative Nash behavior. Perfect competition has always zero profit.

11 Conclusion

The analysis of three combinations of behavioral rules and the use of all the available information have given insight into the structure of spatial equilibrium specifications suitable for trade models. PMP has shown how to calibrate all those models to reproduce the realized and observed matrix of trade flows. Imprecisions in the transaction costs and in the intercepts of demand and supply functions have been easily taken care by means of a PMP approach.

References

Hashimoto, H. (1985). Spatial nash equilibrium model. In P. T. Harker (Ed.), *Spatial price equilibrium: Advances in theory, computation and application* (pp. 20–40). Berlin: Springer.

Paris, Q., Drogué, S., & Anania, G. (2010). Calibrating spatial models of trade. *Economic Modelling, 28,* 2509–2516.

Samuelson, P. A. (1952). Spatial price equilibrium and linear programming. *American Economic Review, 42*(3), 283–303.

Takayama, T., & Judge, J. J. (1964). Equilibrium among spatially separated markets: A reformulation. *Econometrica, 32*(4), 510–529.

Part II
Applied Methods for Water Resource Management

Chapter 5
Payment for Environmental Services: How Big Must Be the Check to Multiproduct Farmers?

Marcelo Torres and Richard E. Howitt

Abstract Environmental conservation policies based on the payment for environmental services (PES) are being increasingly adopted around the world. Among several factors that may dictate the success of a PES conservation program is the payment or monetary compensation level. As participation in the program is voluntary, if payments are not enough to compensate for the eventual economic losses faced by the users, users will not participate, and conservation goals will not be achieved. Also if payments are set significantly higher than the users' opportunity costs, conservation goals are more likely to be achieved, but the program will not be cost-effective. In this context, by using primary data from a watershed in the Brazilian Savannah, this contribution calculates opportunity costs using an agricultural net-revenue multiproduct model, parameterized with the use of positive mathematical programming (PMP) method and coupled with a hydrological model. It is shown that land and water opportunity costs not only vary from farmer to farmer due to differences in crop and input mix but also the variations in water supply affect land use opportunity costs. And, in turn, land supply affects water use opportunity costs. Given this, researchers and policy makers should not be surprised that agricultural PES programs that rely on a flat, crop-and-farmer invariant compensation value per hectare often result in failure and cost-ineffectiveness.

1 Introduction

Around the world, several environmental protection programs based on the principal of 'payment for environmental services' (PES) have been implemented (World Bank 2014; Pagiola et al. 2013; Wunder and Alban 2008; Grima et al. 2016; Ly and Nam

M. Torres (✉)
Department of Economics, University of Brasília, Campus Darcy Ribeiro—Prédio da FACE, 70910-900 Brasília, DF, Brazil
e-mail: motorres@hotmail.com.br

R. E. Howitt
Department of Agricultural and Resource Economics, University of California, Davis, USA

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_5

2014; Porras et al. 2008). Although scheme and design mechanisms differ, the idea (in principle) is rather simple. Users of a natural resource, such as land, air, and water, are viewed as potential providers of ecosystem services and goods, either because they can use more or less of it, or, given the quantity used, they can adopt more or less environmentally sound practices. For example, by using less (or not using more) of their forested land, landowners supply hectares of standing forest with its biodiversity and water resources that benefit humans and wildlife. This implies that natural resource users have to adopt costly changes in production and consumption patterns, which must then be compensated.

Since well-functioning markets for most environmental goods and services are unlikely to evolve naturally due to high transaction costs, the 'public good' nature of environmental services, and their inherent free-rider and asymmetric information problems, they must be accounted for and regulated and that is what PES programs ultimately do. Once property rights are defined and enforced, beneficiaries and suppliers of the environmental services are linked through a formally designed payment system that may be directly funded by public institutions (e.g., national governments, World Bank, etc.) that use monetary resources to compensate the suppliers of the environmental good service and/or by the beneficiaries themselves.

Among several factors that may dictate the success of a PES conservation program is the payment or monetary compensation level. If the payment corresponds to the benefit of the services provided, it must also be enough to compensate for the eventual economic losses faced by the users. As participation in the program is voluntary, if payments are less than the opportunity cost of the natural resource preservation, users will not participate, and conservation goals will not be achieved. For example, in a farming and deforestation context, payments must be at least equal to the net-revenue obtained from converting forested land to agriculture minus any foregone economic benefits that could be extracted from the standing forest, Börner and Wunder (2008). Also if payments are set significantly higher than the users' opportunity costs, conservation goals are more likely to be achieved, but the program will not be cost-effective.

Payment levels have been a subject of evaluation and discussion in several PES programs and studies (Alix-Garcia et al. 2008; Pagiola and Platais 2007; Garcia 2015; Muñoz-Piña et al. 2005; Börner and Wunder 2008; Chomitz 2004; Wunder et al. 2008; Serra and Russman 2006; Pagiola et al. 2008, 2013). Among other factors, these papers highlight the need to accurately estimate opportunity costs and the importance of using them in the design process of PES programs.

Opportunity costs evaluation is generally challenging, particularly in the context of multiproduct and irrigated agriculture. Farmers' opportunity costs of land and water vary with agricultural net-revenues, which in turn vary due to highly volatile input and output prices. Also, as net-revenues vary across crops, opportunity costs vary with the type of crop farmers decide to grow. An additional complication is that since land and water are spatially linked, the opportunity cost associated with land is affected by water scarcity, and the opportunity cost of water is affected by land scarcity. Given this, researchers and policy makers should not be surprised that agricultural PES programs that rely on a flat, crop-and-farmer invariant compensation value per hectare often result in failure and cost-ineffectiveness.

In this chapter, we calculate opportunity costs using an agricultural net-revenue multi-product model based on Torres et al. (2016), parameterized with the use of positive mathematical programming (PMP) method and coupled with a hydrological model. We show that land and water opportunity costs not only vary from farmer to farmer due to the differences in crop and input mix but also that variations in water supply affect land use opportunity costs, and in turn, land supply affects water use opportunity costs. The model is applied, using primary data, to a community of farmers that belong to the Buriti Vermelho subwatershed near Brasília, Brazil.

2 Methodology

Farmers are assumed to be net-revenue maximizers in a multi-product and multi-input context, where crops may be irrigated or rainfed. More specifically, each farmer g chooses amounts of inputs in order to maximize their annual net-revenues, which is defined as

$$\begin{aligned}\sum_{i,j} [&p_i q_i^{\mathrm{ir}}(x_{\mathrm{land}}, x_{\mathrm{aw}}, x_{\mathrm{m}}, x_{\mathrm{hl}}, x_{\mathrm{l}}) + p_j q_j^{\mathrm{r}}(x_{\mathrm{land}}, x_{\mathrm{m}}, x_{\mathrm{hl}}, x_{\mathrm{l}}) \\ &- p_{\mathrm{land}}(x_{\mathrm{land}_i} + x_{\mathrm{land}_j}) - p_{\mathrm{sw}} x_{\mathrm{sw}_i} - MatCost_i + MatCost_j \\ &- p_{\mathrm{hl}}(x_{\mathrm{hl}_i} + x_{\mathrm{hl}_j}) - p_{\mathrm{fl}}(x_{\mathrm{fl}_i} + x_{\mathrm{fl}_j}) - lc_i^{\mathrm{ir}}(x_{\mathrm{land}}) - lc_j^{\mathrm{r}}(x_{\mathrm{land}})]\end{aligned} \tag{1}$$

Where $x_{\mathrm{land}}, x_{\mathrm{aw}}, x_{\mathrm{m}}, x_{\mathrm{hl}},$ and x_{fl} are the amounts of land (land), applied water (aw), materials (m), hired labor (hl), and family labor (fl). p_i and p_j are the received crop prices of irrigated crop i and rainfed crop j, respectively. $q_i^{\mathrm{ir}}(\bullet)$ and $q_j^{\mathrm{r}}(\bullet)$ are irrigated (ir) and rainfed (r) production functions, respectively. $lc_i^{\mathrm{ir}}(x_{\mathrm{land}})$ and $lc_j^{\mathrm{r}}(x_{\mathrm{land}})$ are implicit land cost functions associated with land allocation to irrigated and rainfed crops. Both production functions are specified as a CES function with the difference that in the rainfed production function, precipitation enters as a shifter parameter (Precip_j) defined as the ratio of the actual (x_p^a) and expected (x_p^e) amounts of precipitation to fall onto crop j, that is, $\mathrm{Precip}_j = \frac{x_p^a}{x_p^e}$. Since rainfall is exogenous it is not part of a farmer's decision set, although the amount of rain affects optimal input allocation through the first-order conditions. $lc_i^{\mathrm{ir}}(x_{\mathrm{land}})$ and $lc_j^{\mathrm{r}}(x_{\mathrm{land}})$ follow exponential functional forms as in Medellín-Azuara et al. (2010).

Applied water used on irrigated crop i (x_{aw_i}) is defined as the sum of the amount of water used from surface water (x_{sw_i}) and actual precipitation ($x_{p_i}^a$). That is, $x_{\mathrm{aw}_i} = x_{\mathrm{sw}_i} + x_{p_i}^a$. The price of surface ($p_{\mathrm{sw}}$) water for farmer g is composed of the sum of the costs of hired labor used in irrigation, pumping electricity, and irrigation capital over all irrigated crops divided by the number of crops irrigated. The cost of hired labor is the price of a man-hour of work (p_{hl}) multiplied by the number of man-hours used (x_{hl}). The price of a man-hour of family labor (p_{fl}) is set to be equal to p_{hl}. For materials, prices are not modeled as separate from quantities, so the expenditures (unitary price paid times input quantity used) with each input in the materials category are summed up by crop. In the case of a irrigated crop i, total expenditures are called $MatCost_i$ and for a rainfed crop j, $MatCost_j$.

By using (1) and its related first-order conditions for a maximum, in addition to the assumption of constant returns to scale, the expressions for each parameter of the production functions can be then derived. These expressions are functions of the input quantities, market input prices, scarcity values for the limited availability inputs (land, surface water, and family labor), and the crop- and farm-specific marginal implicit cost of land. A prior on the elasticity of substitution is also required and assumed to be 0.3 for rainfed crops and 0.7 for irrigated crops. These values are based on studies by Salhofer (2001), Gomez et al. (2004), Boyd and Newman (1991), and Seung et al. (1998). For implicit land cost functions, parameter estimation requires another prior on the crop supply elasticity, which is assumed to be 0.7 for all crops. All scarcity values and the marginal implicit cost of land are calculated through a mathematical linear programming optimization model.

The parameters are calculated from the primary data collected on input quantities by crop and farmer, input and output prices faced by each farmer, and the other information described above. These parameters are then reintroduced in a regional net-revenue model built as the sum of Eq. (1) over all farmers subject to a set of constraints. The regional model chooses x_{land}, x_{sw}, x_{m}, x_{hl}, and x_{fl} for each farmer such that the regional annual net-income is maximized. The constraints establish that the annual amount of land and family labor farmer g can use is restricted by the annual amount of land and family labor available. The annual amount of surface water a farmer g can use is restricted by the sum of the monthly surface water available. Also the amount of precipitation farmer g uses as applied water is restricted by the monthly precipitation that falls onto each crop area. Lastly, a water stress constraint is added to put an upper limit on the amount of water stress.

All estimates of farmer-specific surface water availability are calculated with the use of a natural flow model (Rodrigues et al. 2012), which was parameterized with estimates of river discharges in the Buriti Vermelho subwatershed calculated by a hydrological model based on Liebe et al. (2009).

3 Data

The data used for the model calibration describe the agricultural input (prices paid and amounts used) and output (prices received and amounts produced) of 23 farmers located within the Buriti Vermelho watershed near Brasília, Brazil, during the agronomic year of 2007/2008. The estimates of surface water used by each farmer and crop were calculated with data collected on the frequency and duration of irrigation, irrigation technology, and pump characteristics. Also, data on daily millimeters of precipitation that fell onto the subwatershed and a farmer- and crop-specific planting and harvesting calendar are used as the basis for the calculation of the amount of precipitation used as part of the total applied water.

The vast majority of farmers located in the watershed can be considered small with an average annual cropping land of 3.5 ha and average annual net-income of 20,000 reais (US$ 13,000.00 approximately).[1] Most of their crops (vegetables and fruits) are irrigated with annual average water use of 1500 m^3/ha. A larger farmer, specialized in grain crops, that uses center-pivot irrigation technology, is also located in the watershed.

4 Results

With the regional net-income model calibrated, simulations of restrictions on the amount of land and water were used to derive the estimates of opportunity costs. Farmers were divided into groups according to their access to surface water. Farmers in group 1 and group 2 withdraw water from the most upstream reservoir. Water is carried to these groups through a channel that splits in two: one with a larger diameter that brings water to group 1 and another with a smaller diameter that brings water to group 2. What remains in this upstream reservoir goes to another midstream reservoir that is used by the farmers in group 3. The larger farmer in group 4 can use up the water that remains in the downstream reservoir. Assuming that proportional cuts in surface water are proportional to cuts in precipitation, Tables 1 and 2 present the estimated opportunity costs in terms of foregone net-revenue associated with cuts in the amount of water available for irrigation, keeping the amount of arable land available as in the base year. Not only do the opportunity costs vary by farmer but also by crop.

Table 1 presents that averaging over all farmers, predicted regional net-income would fall by 3.5% in the event of a 10% cut in water availability. Successively, cuts are followed by predicted nonlinear decreases in net-revenue. The same pattern is similar for the smaller farmer groups. Table 1 also presents that the impact is not uniform across farmer types. Farmers in group 1 have the highest opportunity costs and farmers in group 2 have the lowest.

The differences between farmer groups probably reflect differences in crop mix, as we can see in Table 2 that opportunity costs vary by crop type. Grains have the highest opportunity costs, followed by fruits and then vegetables.

Tables 3 and 4 present the impact of cuts in land availability keeping surface water and precipitation amounts at base year levels. Land opportunity costs are also not homogenous across farmers groups, being highest for group 1 and lowest on average for farmers in group 2. Table 4 presents that the land opportunity costs also vary significantly across crops, being highest for fruits, then grains, and finally vegetables.

[1] Using the 2008 yearly average exchange rate of 1US$ = 1.5BR$ provided by the Brazilian Central Bank.

Tables 1 through 4 present either the predicted impacts of cuts in water availability on farmers' net-revenue by keeping the amount of arable land constant at base year levels, or the impacts of arable land availability on net-revenue, keeping water availability fixed. Alternatively, however, cuts in surface water use often result in cuts in arable land in cases where water conservation implies land conservation, and vice versa. In this case, cuts in land availability may affect the magnitude of opportunity costs associated with reduced water availability. This is, in fact, presented in Table 5.

For example, we saw that a 10% cut in water availability, holding arable land constant at base year levels, would cause a 3.5% decrease in regional net-revenue (see Table 1). If, however, this cut happens when arable land must also be reduced by 10%, regional net-revenue would decrease in average by 7.1% (Table 5). This percentage ranges from 4.5 to 8.3% depending on the farmer groups. If arable land availability is reduced by 20 or 30%, the impacts on net-revenue of a 10% cut in water availability would be much higher (13.6 and 20.1%, respectively).

Table 1 Farmers net-revenue impacts due to cuts in surface water availability

% cuts in water availability			Farmers groups							
	All farmers		1		2		3		4	
	Value[a]	(%)	Value	(%)	Value	(%)	Value	(%)	Value	(%)
0	734	–	149	–	66	–	191	–	382	–
10	708	−3.5	145	−2.7	65	−1.5	187	−2.1	365	−4.5
20	670	−8.7	140	−6.0	62	−6.1	181	−5.2	340	−11.0
30	624	−15.0	131	−12.1	60	−9.1	175	−8.4	311	−18.6
40	581	−20.8	122	−18.1	57	−13.6	166	−13.1	286	−25.1
50	540	−26.4	110	−26.2	53	−19.7	153	−19.9	270	−29.3
60	498	−32.2	95	−36.2	48	−27.3	133	−30.4	261	−31.7

[a]Values in thousands of Brazilian reais as of 2008

Table 2 Net-revenue impacts due to cuts in water availability by crop groups

% cuts in water availability	Grains		Vegetables		Fruits	
	Value[a]	(%)	Value	(%)	Value	(%)
0	386	–	259	–	136	–
10	370	−4.1	254	−1.9	135	−0.7
20	343	−11.1	246	−5.0	132	−2.9
30	313	−18.9	238	−8.1	127	−6.6
40	287	−25.6	226	−12.7	117	−14.0
50	270	−30.1	210	−18.9	104	−23.5
60	262	−32.1	187	−27.8	89	−34.6

[a]Values in thousands of Brazilian reais as of 2008

Table 3 Farmers net-revenue impacts due to cuts in arable land

% cuts in land availability	Farmers groups									
	All Farmers		1		2		3		4	
	Value[a]	(%)	Value	(%)	Value	(%)	Value	(%)	Value	(%)
0	734	–	149	–	66	–	191	–	382	–
10	719	−2.0	144	−3.4	66	0.0	182	−4.7	372	−2.6
20	686	−6.5	135	−9.4	64	−3.0	175	−8.4	353	−7.6
30	642	−12.5	123	−17.4	61	−7.6	166	−13.1	329	−13.9
40	590	−19.6	110	−26.2	58	−12.1	154	−19.4	301	−21.2
50	533	−27.4	96	−35.6	53	−19.7	140	−26.7	273	−28.5
60	469	−36.1	80	−46.3	48	−27.3	121	−36.6	242	−36.6

[a]Values in thousands of Brazilian reais as of 2008

Table 4 Net-revenue impacts due to cuts in arable land availability by crop groups

% cuts in land availability	Grains		Vegetables		Fruits	
	Value[a]	(%)	Value	(%)	Value	(%)
0	386	–	259	–	136	–
10	377	−2.3	252	−2.7	130	−4.4
20	357	−7.5	241	−6.9	124	−8.8
30	331	−14.2	226	−12.7	115	−15.4
40	304	−21.2	209	−19.3	104	−23.5
50	274	−29.0	190	−26.6	91	−33.1
60	243	−37.0	167	−35.5	76	−44.1

[a]Values in thousands of Brazilian reais as of 2008

5 How Fat Should the Check Be?

As we can see from the results displayed above, opportunity costs of land and water in a multi-product, agricultural context vary across crops and farmers. Also, since land and water are spatially linked, the amount of forgone net-revenue associated with lower water uses depends on the amount of arable land and vice versa. This suggests that PES programs that focus on agricultural water and/or land conservation based on flat rate compensations will either fail in achieving a given regional goal, or succeed to achieve a given goal but not at a minimum cost. For example, suppose the objective of the environmental planner in the Buriti Vermelho watershed is to induce each farmer to reduce its use of arable land by 20%. Once intact, the unused land would then be converted into a non-primary cerrado forest. Table 6 presents the foregone net-revenue by farmer.

Assuming that farmers would have to leave the non-used area intact and that participation in the PES program is voluntary, numbers in Table 6 reveal that to leave

Table 5 Farmers net-revenue impacts due to cuts in surface water availability and a 10, 20, and 30% cut in arable land

10% cut in land	All farmers value[a] (%)		Farmers groups							
% cuts in water availability			1		2		3		4	
			Value	(%)	Value	(%)	Value	(%)	Value	(%)
0	719	–	144	–	66	–	182	–	373	–
10	668	−7.1	132	−8.3	63	−4.5	171	−6.0	346	−7.2
20	605	−15.9	118	−18.1	59	−10.6	159	−12.6	307	−17.7
30	554	−22.9	106	−26.4	55	−16.7	146	−19.8	278	−25.5
40	520	−27.7	96	−33.3	52	−21.2	135	−25.8	266	−28.7
50	488	−32.1	86	−40.3	49	−25.8	123	−32.4	255	−31.6
20% cut in land	**All farmers value (%)**		**Farmers groups**							
% cuts in water availability			1		2		3		4	
			Value	(%)	Value	(%)	Value	(%)	Value	(%)
0	686	–	135	–	64	–	175	–	353	–
10	593	−13.6	112	−17.0	58	−9.4	154	−12.0	301	−14.7
20	508	−25.9	90	−33.3	51	−20.3	134	−23.4	259	−26.6
30	445	−35.1	72	−46.7	46	−28.1	113	−35.4	234	−33.7
40	388	−43.4	58	−57.0	41	−35.9	95	−45.7	211	−40.2
50	340	−50.4	46	−65.9	37	−42.2	79	−54.9	192	−45.6

(continued)

Table 5 (continued)

30% cut in land	All farmers value (%)		Farmers groups							
% cuts in water availability			1		2		3		4	
			Value	(%)	Value	(%)	Value	(%)	Value	(%)
0	642	–	123	–	61	–	166	–	329	–
10	512	−20.1	91	−26.0	52	−14.8	134	−19.3	261	−20.7
20	411	−35.9	64	−48.0	43	−29.5	105	−36.7	218	−33.7
30	337	−47.4	45	−63.4	37	−39.3	81	−51.2	188	−42.9
40	269	−58.0	31	−74.8	32	−47.5	63	−62.0	156	−52.6
50	200	−68.8	20	−83.7	28	−54.1	48	−71.1	113	−65.7

[a]Values in thousands of Brazilian reais as of 2008

Table 6 Annual forgone net-revenue from a 20% reduction in arable land use

Farmers	Forgone annual net-revenue	Farmers	Forgone annual net-revenue	Farmers	Forgone annual net-revenue	Farmers	Forgone annual net-revenue
v35	28,400	v32	1906	v27	465	v10	219
v31	7348	v17	1448	v18	390	v12	128
v22	6909	v25	1072	v26	350	v30	97
v20	3526	v28	732	v13	296	v15	30
v23	3276	v24	729	v19	244	v29	2
v21	1974	v16	687	v14	226		

Table 7 Annual forgone net-revenue from a 10% reduction in surface water availability in the event a uniform 20% drop in monthly rainfall

Farmers	Forgone annual net-revenue	Farmers	Forgone annual net-revenue	Farmers	Forgone annual net-revenue	Farmers	Forgone annual net-revenue
v35	41,498	v27	1565	v21	726	v14	536
v32	4693	v13	1537	v28	712	v18	498
v20	2407	v17	1132	v26	687	v23	326
v25	2280	v12	870	v31	683	v10	308
v19	1898	v22	841	v30	640	v15	19
v16	1672	v24	739	v29	626		

20% of the arable land within each farm intact at a minimum cost, each farm should receive a compensation per year at least equal to the annual foregone net-revenue. Summing it over all farmers, that means that the PES program would cost, in terms of compensation, around 60,000.00 Brazilian reais. The same goal could be achieved off course if the program sets an annual flat compensation per farmer at the highest level (28,400), but that would not be cost-effective. If set at the mean value (≈2700), only farmers v35, v31, v22, v20, and v23 would adhere to the program and the goal would not be achieved.

Suppose now that the policy goal is to keep a minimum environmental flow in the Buriti Vermelho River during droughts. Table 7 presents the foregone revenue when the surface water available for irrigation is reduced by 10% in the event of a drop of monthly rainfall by 20%. Again the PES program would be cost-effective at a much lower cost (around 67,000.00 Brazilian reais) if based on a compensation rate that reflects the foregone net-revenue for each farmer.

6 Conclusions

This contribution shows, by using a multi-product agricultural production model calibrated with positive mathematical programming methods, that the opportunity costs associated with land and water, in the context of multi-product farming, vary considerably from farmer to farmer and across crops. More importantly, restrictions on surface water use have significant impacts on the magnitude of the land opportunity costs, and vice versa. For example, given the level of arable land available in the base year, a 10% cut in surface water available for irrigation would imply a 3.5% decrease in the farmers' net-revenues. Or given the level of water availability in the base year, a 10% cut in arable land would, on average, reduce the regional net-revenue by 2.0%. However, the negative impacts of a 10% cut in water availability, in a situation where 10% of the total arable land must be preserved, would more than double (from 3.5 to 7.1%). In short, the results show that the water opportunity costs increase non-linearly with restrictions on arable land and vice versa.

These results have clear implications for market-based environmental conservation policies and in particular policies based on the payment for environmental services (PES). Since the biggest users of land and water are farmers, and participation in PES programs are voluntary, the results show that a successful PES program requires the use of a more flexible system for farmer compensation that can vary across multi-crop farmers. In fact, the simulations show that compensation rates that vary across farmers and that are based on their foregone net-revenue would bring down PES program costs and increase efficacy considerably. While the spatial extent of the current empirical example is too small to be scaled up, results do indicate that PES programs applied in the context of multi-product farmers are likely to fail if farmer heterogeneity in crop and input mixes is not taken into account.

References

Alix-Garcia, J., de Janvry, A., & Sadoulet, E. (2008). The role of deforestation risk and calibrated compensation in designing payments for environmental services. *Environment and Development Economics, 13*(03), 375–394.

Börner, J., & Wunder, S. (2008). Paying for avoided deforestation in the Brazilian Amazon: From cost assessment to scheme design. *International Forestry Review, 10*(3), 496–511.

Boyd, R., & Newman, D. H. (1991). Tax reform and land-using sectors in the US economy: A general equilibrium analysis. *American Journal of Agricultural Economics, 73*(2), 398–409.

Chomitz, K. (2004). Transferable development rights and forest protection: An exploratory analysis. *International Regional Science Review, 27*(3), 348–373.

Garcia, M. A. B. (2015). Effectiveness of payment for environmental services in Mexico. *All Dissertations*. Paper 1484. 5-2015. Clemson University TigerPrints. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.873.6366&rep=rep1&type=pdf.

Gomez, C. M., Tirado, D., & Rey Maquieira, J. (2004). Water exchanges versus waterworks: Insights from a computable general equilibrium model for the Balearic Islands. *Water Resources Research, 4,* W10502. https://doi.org/10.1029/2004WR003235.

Grima, N., Singh, S. J., Smetschka, B., & Ringhofer, L. (2016). Payment for ecosystem services (PES) in Latin America: Analysing the performance of 40 case studies. *Econsystem Services, 17,* 24–32.

Liebe, J., van de Giesen, N., Andreini, M. S., Walter, M. T., & Steenhuis, T. (2009). Determining watershed response in data poor environments with remotely sensed small reservoirs as runoff gauges. *Water Resources Research, 45,* W07410. https://doi.org/10.1029/2008WR007369.

Ly, N. T. Y., & Nam, P. T. (2014). Payment for environmental services in Southeast Asia: A regional review of policy implementation, EEPSEA, 2014-RR10.

Medellín-Azuara, J., Harou, J. J., & Howitt, R. (2010). Estimating economic value of agricultural water under changing conditions and the effects of spatial aggregation. *Science of the Total Environment, 408*(23), 5638–5648.

Muñoz-Piña, C., Guevara, A., & Torres, J. M. (2005). Payment for the hydrological services of Mexico's forests: Analysis, negotiations and results. Instituto Nacional de Ecología.

Pagiola, S., Carrascosa von Glehn, H., & Taffarello, D. (2013). Brazil's experience with payments for environmental services, PES Learning Paper 2013-1, World Bank, Washington, D.C., USA.

Pagiola, S., & Platais, G. (2007). *Payments for environmental services: From theory to practice*. Washington, D.C.: World Bank.

Pagiola, S., Rios, A., & Arcenas, A. (2008). Can the poor participate in payments for environmental services? Lessons from the Silvopastoral Project in Nicaragua. *Environment and Development Economics, 13,* 299–325.

Porras, I., Grieg-Gran, M., & Neves, N. (2008). *All that glitters: A review of payments for watershed services in developing countries*. London: The International Institute for Environment and Development (IIED).

Rodrigues, L. N., Sano, E. E., Steenhuis, T. S., & Passo, D. P. (2012). Estimation of small reservoir storage capacities with remote sensing in the Brazilian Savannah Region. *Water Resources Management, 26*(4), 873–882.

Salhofer, K. (2001). Elasticities of substitution and factor to supply elasticities in European agriculture: A review of past studies. *Market Effects of Crop Support Measures* September 12, OECD, Paris.

Serra, R., & Russman, E. (2006). On the efficiency of environmental service payments: A forest conservation assessment in the OSA Peninsula, Costa Rica. *Ecological Economics, 59*(1), 131–141.

Seung, C. K., Harris, T. R., McDiarmid, T. R., & Shaw, W. D. (1998). Economic impacts of water reallocation: CGE analysis for the Walker River basin of the Nevada and California. *Journal of Regional Analysis and Policy, 28*(2), 13–34.

Torres, M. O., Howitt, R., & Rodrigues, L. N. (2016). Modeling the economic benefits and distributional impacts of supplemental irrigation. *Water Resources and Economics, 14,* 1–12.

World Bank, (2014). *GEF investments on payment for ecosystem services schemes*. Global Environment Facility (GEF). World Bank, Washington, D.C., USA. http://documents.worldbank.org/curated/en/674691468155116353/GEF-investments-on-payment-for-ecosystem-services-schemes.

Wunder, S., & Alban, M. (2008). Decentralized payments for environmental services: The cases of Pimampiro and PROFAFOR in Ecuador. *Ecological Economics, 65*(4), 685–698.

Wunder, S., Engel, S., & Pagiola, S. (2008). Taking stock: A comparative analysis of payments for environmental services programs in developed and developing countries. *Ecological Economics, 65*(4), 834–852.

Chapter 6
Optimal Allocation of Groundwater Resources: Managing Water Quantity and Quality

Qiuqiong Huang, Scott D. Rozelle, Richard E. Howitt and James E. Wilen

Abstract Despite the importance of groundwater in the economy of the Hai River Basin (HRB), falling water tables and ***salinization*** of aquifers are both occurring in the region. Hydrological and hydrogeological studies have shown that increases in the salinization of parts of the freshwater aquifers are closely related to the extraction of groundwater. This study uses a framework that considers the interaction between water quantity and quality to examine how the presence of the prehistoric saline water layer affects groundwater management. Simulation results show that in a region where there is a salinization problem like in the HRB, it is optimal to pump at high rates in the early stage of extraction when the quality of groundwater is high. It is then optimal to reduce the pumping rate rapidly as the quality of groundwater deteriorates. Given this characteristic of the optimal pumping path, the heavy extraction currently observed in the HRB does not necessarily indicate that groundwater resources are being overused. However, unregulated extraction by non-cooperative users would eventually cause both the depletion of the water resource and the deterioration of water quality. Hence, joint quantity–quality management is required in the HRB. The study also shows that benefits to groundwater management are higher and costs are lower in regions with salinization problems.

Q. Huang (✉)
Department of Agricultural Economics & Agribusiness,
The University of Arkansas, Fayetteville, AR 72701, USA
e-mail: qqhuang@uark.edu

S. D. Rozelle
Stanford University, Stanford, CA, USA
e-mail: rozelle@stanford.edu

R. E. Howitt · J. E. Wilen
University of California, Davis, Davis, CA, USA
e-mail: howitt@primal.ucdavis.edu

J. E. Wilen
e-mail: wilen@primal.ucdavis.edu

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_6

1 Introduction

Despite the importance of groundwater resources in stimulating the rapid development of the Hai River Basin (HRB), one of the main economic and political centers of China, the resource base is diminishing. Between 1958 and 1998, the level of the groundwater in the HRB fell by up to 50 m in certain shallow aquifers and by more than 95 m for some deep aquifers (Ministry of Water Resource et al. 2001). During a field trip to Cang County, Hebei province (the province in which most of the HRB resides) in July 2004, the authors observed farmers extracting deep aquifer freshwater from tubewells that were sunk to a level of more than 400 m. Given the fact that the depth to the bottom of the deep aquifers is between 500 and 600 m in most places in the HRB and that the current level of extraction of deep aquifer water far exceeds the rate of recharge (Chen 1999; Hebei Bureau of Geology Reconnaissance 2003), many policymakers are worried that China is using its water resources too rapidly.

In addition to the declining water levels, another potential problem is arising in some places in the middle and eastern parts of the HRB, namely the increased salinization of some aquifers. The salinity level of freshwater in certain parts of the deep aquifers, measured by the level of total dissolved solutes (TDS), is known to have increased by 14.3 mg/L annually in Hengshui County (Song and He 1996). In some places in Cang County, the TDS level increased from less than 2000 mg/L to more than 5000 mg/L, a level above which water is considered to be saline (Hebei Bureau of Geology Reconnaissance 2003).

Hydrological and hydrogeological studies have shown that increases in the salinization of parts of the freshwater aquifers are likely to be closely related to the extraction of groundwater (Mu and Zhang 2002; Zhu et al. 2002). Large layers of prehistoric saline water exist between the shallow and deep freshwater aquifers in the middle and eastern parts of the HRB. When groundwater is extracted and the stock of groundwater declines, the pressure difference between the freshwater and the overlying saline water layers increases. Although the saline water and freshwater are separated by a layer of clay (an aquitard), the pressure difference, according to hydrologists, can push saline water past the clay layer and into the freshwater layer—a process that can increase the salinity level of the freshwater.

Since the level of the groundwater and the degree of salinization are both related to the rate of extraction of groundwater, when determining how to use groundwater optimally, a framework that considers the interaction between water quantity and quality is required for complete and more efficient management of groundwater resources. When groundwater is pumped and the level of water changes in the HRB, water users incur two costs. First, pumping costs rise as groundwater levels fall—even when the quality of water remains the same. Second, when groundwater is extracted, the intrusion of saline water into the freshwater, as seen above, increases the salinity level of the groundwater. When this happens, the application of saline water in irrigation can cause salt accumulation in the soil. Agronomic studies have shown that crop productivity falls when there is an excessive level of salinity in the soil [e.g., Maas and Hoffman (1977)]. Because of these dual effects, it is likely that the way to optimally extract groundwater in the HRB is different from what it would be in an area without a salinization problem.

Given the importance of the North China Plain in the nation's economy, and given the changes that will continue in the future, one of the key issues facing policymakers is how to manage the quantity of water and maintain water quality. As an attempt to start addressing this issue, the overall goal of this contribution is to examine how the presence of the prehistoric saline water layer affects groundwater management. Specifically, we will address three questions: (1) How does the optimal allocation of groundwater resource (both the pumping path and the level of pumping lift and salinity level at the steady state) differ between the regions with salinization problems and regions without such problems? (2) In regions with salinization problems, how does the impact of pumping on groundwater quantity and quality differ when water users extract in a cooperative way as if they are managed by a social planner and when water users extract in a non-cooperative way? (3) What are the implications for policies that we can draw for managing groundwater in the HRB?

This study will make several contributions. It will be one of the first studies that take into account the interaction between the quantity and quality of groundwater to analyze the optimal use of groundwater resources in North China. Lessons learned from this study will also help tackle salinity problems that relate to groundwater extraction in other countries such as Australia, India, and Bangladesh. Few studies have utilized joint quantity–quality framework outside of China. Some of the exceptions are Roseta-Palma (2003, 2002). However, these papers only have a general model of quantity–quality problem. Scholars also have worked on salinity issues extensively (e.g., Dinar et al. 1993; Kan et al. 2002; Knapp 1992a, b, c). However, to our knowledge there have been none inside China, this study will contribute to the resource economics literature by providing an example of a resource problem where the interlinks between the quantity and quality are considered.

There are several subjects, however, that are not addressed in this study. First, the conjunctive use of groundwater and surface water is not considered. Water users in many places in the HRB depend solely on groundwater resources from deep confined aquifers. Unlike shallow aquifers that are recharged directly by surface water supply such as precipitation, deep aquifers in the HRB are recharged by a much slower horizontal flow from the mountain area, which is far less stochastic than surface water supply. Hence, a deterministic framework is used since the stochastic surface water is not included. Second, the irrigation salinity problem that occurs when irrigation water causes the water table to rise and brings salt to the surface is not addressed. In fact, this irrigation salinity problem ceased to exist after the heavy extraction of groundwater resources started in the 1970s in the HRB (Nickum 1988). This study only analyzes the increase in salinity due to the intrusion of saline water.

2 Analytical Framework and Propositions

We begin with the standard groundwater extraction problem, focusing on the decision of the extraction quantity that balances the current and future benefits from extraction versus the current and future costs of doing so. In the first instance, we ignore salinization (Model 1). Second, we then consider the case when there is a saline water layer and use a cooperative extraction model in which water users cooperate with each other and act as if their extraction is managed by a social planner (Model 2). Since the social planner is maximizing the total benefits of society, this solution will provide the optimal solution to groundwater extraction in the presence of a saline water layer in the aquifer. We use this model to establish a baseline against which we can compare the results when there is no social planner and there is inevitably a less optimal allocation of water. In the third step, we use a non-cooperative extraction model to more accurately reflect the real-world situation in the HRB where water users are not regulated (Model 3). Comparing the case with salinization when there is a social planner (Model 2) and without one (Model 3), we can show the differences that occur when there is no effective regulation of water use.

2.1 Cooperative Extraction Model

Model 1
In most studies in the literature, groundwater is pumped in an environment in which the pumping of groundwater does not have any impacts on its quality (e.g., Burt 1964). We begin with this assumption and treat the resource from the viewpoint of a social planner to analyze the optimal extraction of groundwater. Under such a circumstance, each water user extracts cooperatively (or under the guidance of a social planner) in order to maximize the total benefit from utilizing groundwater (Model 1). Thus, he is called a *cooperative user*. Cooperative users are solving the following *quantity-only* problem:

$$\begin{aligned} &\underset{\{w_t\}}{\text{Max}}\ M\sum_{t=0}^{\infty}\beta^t f(w_t,h_t) \\ &\text{s.t.}\quad h_{t+1}=h_t+\phi_1 M w_t-\phi_2 R \end{aligned} \tag{1}$$

In Model 1, M is the number of water users. We assume water users are identical so they have the same net benefit function $f(\cdot)$. The net benefit is a function of the pumping rate w_t. It is also a function of the pumping lift h_t, since the pumping cost is directly associated with the pumping lift.[1] The change in h_t is a function of the difference between the recharge to the aquifer R and the net aggregate withdrawal

[1]Pumping lift is defined as the depth from the ground surface to groundwater.

of all M users Mw_t. The parameters ϕ_1 and ϕ_2 convert changes in the volume of groundwater stock into changes in the lift of pumping. The net benefit function satisfies the following properties: $f_w > 0$, $f_{ww} < 0$, $f_h < 0$, $f_{hh} \leq 0$ and $f_{wh} < 0$. In Eq. (1), we did not include the return flow of irrigation water since the return flow will mostly stay in the shallow aquifer. In this way, we are modeling the actual water consumption.[2]

Model 2

In the HRB, when there is a saline water layer present in the aquifer, the increase in the pumping lift will induce the intrusion of saline water and then leads to an increase in the salinity level of the water resource, denoted as E_t in this contribution. Net benefits are now a function of three variables: w_t, h_t, and E_t. Since the marginal productivity of a given quantity of water decreases with its salinity level ($f_{wE} < 0$), the net benefit decreases in E_t ($f_E < 0, f_{EE} < 0$ and $f_{Eh} = 0$).[3] In regions with salinization problems, water users need to solve a more complicated problem that we henceforth call the *quantity–quality* problem:

$$\begin{aligned} &\underset{\{w_t\}}{Max}\, M\sum_{t=0}^{\infty}\beta^t f(w_t, E_t, h_t) \\ &s.t. \quad h_{t+1} = h_t + \phi_1 M w_t - \phi_2 R \\ &\qquad E_{t+1} = E_t + \delta(h_{t+1} - h_t) \end{aligned} \tag{2}$$

where δ measures the impact of changes in the depth to groundwater on the changes in its level of quality. One unit increase in the pumping lift leads to a δ unit increase in the salinity level of groundwater. The derivation of the equation of motion for E_t is in Appendix 1.

After solving problem (2), cooperative users will follow the optimal allocation rule (Appendix 2):

$$f_{wt} = -M\sum_{l=1}^{\infty}\beta^{\ell}\phi_1 f_{h_{t+\ell}} - M\sum_{l=1}^{\infty}\beta^{\ell}\phi_1\delta f_{E_{t+\ell}} \tag{3}$$

Equation (3) says that along the optimal pumping path, the marginal net benefit of groundwater f_{w_t} is equated with the marginal cost of pumping, which is made of costs that occur in the future. Since pumping one unit groundwater leads to both

---

[2]In practice, what determines the level of the water table is the actual water consumption, not pumping rates. In most uses of groundwater, some of the water pumped is returned to the groundwater system. The only water that does not return to the aquifer is what evapotranspires from crops and soils. The part of evapotranspiration is the actual water consumption. Pumping rates may be irrelevant to the level at which a water table stabilizes. For example, Kendy (2003) shows that pumping decreases in some counties in Hebei province by more than 50%, yet the water table declines at the same rate over years. The modeling in this study also reflects this fact.

[3]The benefit and cost function are separate in the net benefit function. Since E_t only enters the benefit function and h_t only enters the cost function, the cross-derivative, f_{Eh}, is zero.

a ϕ_1 unit increase in the pumping lift and a $\phi_1\delta$ unit increase in the salinity level (both of these changes are in the future), the marginal cost has two components. The first term $-M\sum_{\ell=1}^{\infty}\beta^{\ell}\phi_1 f_{h_{t+\ell}}$ reflects an increase in the pumping cost of all M water users due to a larger pumping lift in all the future periods. The second term $-M\sum_{\ell=1}^{\infty}\beta^{\ell}\phi_1\delta f_{E_{t+\ell}}$ reflects the decrease in the benefit for all M water users due to a higher salinity level in all the future periods. The term on the right-hand side discounts future costs into current ones.

Comparing the decision rules in the quantity-only problem (Model 1) and that in the quantity–quality problem (Model 2), it can be seen that the decisions made by the social planner in seeking the optimal extraction of groundwater resources are different. One fundamental difference is that the optimal steady-state water level differs. In a region in which the pumping of groundwater does not affect the salinity level (i.e., in the case of the quantity-only problem), the second term on the RHS of Eq. (3) vanishes. Therefore, the marginal cost of extraction is higher in a region with a salinization problem. Higher marginal costs results in lower pumping rates, which in turn leads to more water left in the ground at the steady state. Based on this set of ideas, we develop the first proposition:

Proposition 1: *The socially optimal pumping lift is smaller in a region with a salinization problem, compared to that in a region without such a problem* (*Proof in Appendix* 3).

A second difference is that the value of the groundwater resources also differs in the two cases. Specifically, when δ is higher (i.e., when the change in the quantity of groundwater by pumping has a greater impact on the change in quality), the aquifer becomes more saline given the same volume of pumping. Since higher salinity levels reduce benefits from using groundwater in the future, the value of groundwater is lower. Following this logic, we establish a second proposition:

Proposition 2: *In a region with a salinization problem, the value of groundwater (the present value of net benefits from using groundwater in all future periods) decreases in the magnitude of impact that groundwater extraction has on the salinity level of groundwater* (*Proof in Appendix* 4).

2.2 Non-cooperative Extraction Model

Model 3

Unlike the assumption of the social planner model, water users in China are not regulated when withdrawing water from a common aquifer. Without any regulations, water users are not likely to cooperate among themselves. Each individual water user extracts groundwater in order to maximize his own profit independent of that of others and thus is called a *non-cooperative user*. Mathematically, a non-cooperative user i is solving the following *non-cooperative extraction* problem (Model 3):

$$\begin{aligned}
&\operatorname*{Max}_{\{w_{it}\}} \sum_{t=0}^{\infty} \beta^t f(w_{it}, E_t, h_t) \\
&\text{s.t.} \quad h_{t+1} = h_t + \phi_1[w_{it} + \sum_{j\neq i}^{M} w_j^*] - \phi_2 R \\
&\qquad E_{t+1} = E_t + \delta(h_{t+1} - h_t)
\end{aligned} \tag{4}$$

Here, user i is involved in a non-cooperative difference game with other water users. When user i makes his decision, he takes the pumping rates of other users, w_j^* as given. The solution of this model is a feedback Nash equilibrium. In our work, since we will only solve the non-cooperative extraction problem in the case in which there is a saline water layer in the aquifer, we identify Model 3 as the non-cooperative extraction problem (although its complete name would be the non-cooperative extraction problem in the presence of salinization). After solving problem (4), the non-cooperative user i will follow the decision rule that can be expressed as (Appendix 5):

$$\begin{aligned}
f_{w_{it}} = &-\sum_{\ell=1}^{s} \beta^\ell \phi_1 f_{h_{t+\ell}} - \sum_{\ell=1}^{s} \beta^\ell \phi_1 \delta f_{E_{t+\ell}} + \sum_{\ell=1}^{s} \beta^\ell \sum_{j\neq i}^{M} \phi_1 \frac{\partial w_j^*}{\partial h_{t+\ell}} \\
&+ \sum_{\ell=1}^{s} \beta^\ell \sum_{j\neq i}^{M} \phi_1 \delta \frac{\partial w_j^*}{\partial E_{t+\ell}}
\end{aligned} \tag{5}$$

Comparison of the RHSs of Eqs. (3) and (5) shows that, given the same pumping lift and the same level of salinity, non-cooperative users (Model 3) extract more than cooperative users (Model 2, Appendix 6). Similar to the quantity–quality case, the term $-\sum_{\ell=1}^{s} \beta^\ell \phi_1 f_{h_{t+\ell}} - \sum_{\ell=1}^{s} \beta^\ell \phi_1 \delta f_{E_{t+\ell}}$ reflects the marginal cost of pumping due to higher pumping lifts and higher salinity levels in the future. However, when water users are pumping groundwater in an environment characterized by non-cooperative extraction, no single individual accounts for the social cost of his pumping, which is the increased future pumping costs of other users that will accrue due to the drawing down of the water level. Hence, non-cooperative users underestimate the marginal cost by $(M-1)\sum_{\ell=1}^{s} \beta^\ell \phi_1 f_{h_{t+\ell}} + (M-1)\sum_{\ell=1}^{s} \beta^\ell \phi_1 \delta f_{E_{t+\ell}}$. In addition, water users also react to a lower pumping lift or a salinity level by increasing their pumping rates ($\frac{\partial w_j^*}{\partial h_t} < 0$ and $\frac{\partial w_j^*}{\partial E_t} < 0$, Appendix 6). This strategic behavior of water users (a water user may pump more than what he would had there been no other users to discourage the extraction of others) is discussed in detail in Negri (1989) and Provencher and Burt (1993). Knowing this, user i places a lower value on the marginal cost of pumping (by a degree of $\sum_{\ell=1}^{s} \beta^\ell \sum_{j\neq i}^{M} \phi_1 \frac{\partial w_j^*}{\partial h_{t+\ell}} + \sum_{\ell=1}^{s} \beta^\ell \sum_{j\neq i}^{M} \phi_1 \delta \frac{\partial w_j^*}{\partial E_{t+\ell}}$). This lower valuation occurs because any water he conserves for future use (as would occur in the case of the social planner in the quantity–quality problem) will be pumped out by other users. As a result, given the same pumping lift and the same salinity level, non-cooperative users will extract more than what they would extract had water users cooperate in pumping.

Although the over-extraction of non-cooperative users does not necessarily lead to both higher pumping lifts and higher salinity levels at the steady state, it does in this case. The linear relationship between changes in salinity level and changes in pumping lifts makes it that higher pumping lifts are always accompanied by higher salinity levels. Following this logic, we form

Proposition 3 *In regions with salinization problems, compared to cooperative extraction, non-cooperative extraction leads to both a higher pumping lift and a higher salinity level at the steady state* (*Proof in Appendix* 6).

3 Empirical Specification and Parameterization of the Model

Analyses of the theoretical models in the previous models have provided a basic understanding of the way to optimally use groundwater in the specific environment of the HRB where extraction affects both the quantity and quality of groundwater. The next step is to empirically examine the propositions developed from the theoretical models. In this section, before presenting the results of the empirical analysis, we will first specify the functional form of the benefit function and introduce the data sources and information that will be drawn on to parameterize the models.

The specified net benefit function is as follows:

$$f(w_t, h_t, E_t) = \mathrm{e}^{\alpha E_t - \theta (E_t)^2}(a w_t - 0.5 b w_t^2) - c h_t w_t \qquad (6)$$

Following several economics studies that analyze a quantity-only problem (Feinerman and Knapp, 1983; Rubio and Casino, 2001), we use a quadratic benefit function, $a w_t - 0.5 b w_t^2$, where a and b are the intercept and slope of the demand function for irrigation water, respectively. Pumping cost is a function of the volume pumped, w_t, and the pumping lift, h_t. The parameter c is the marginal cost of lifting one unit of groundwater by one unit of pumping lift. Unlike the studies that analyze a quantity-only problem, in the net benefit function there is an exponential function $\mathrm{e}^{\alpha E_t - \theta (E_t)^2}$ that measures the magnitude of the reduction in the crop yield in response to higher salinity level of the irrigation water. This exponential function has been used in several agronomic studies, and the parameters α and θ in the exponent are also estimated in these studies [e.g., Van Genuchten and Hoffman, 1985].

3.1 Parameterization of the Model

Parameter values are obtained from different sources. Values of parameters in the quadratic benefit function and the pumping cost are estimated using the China Water Institutions and Management (CWIM) survey data that we collected in 2004.[4] Net benefit from water use is calculated as revenue from agricultural production minus cost of variable inputs other than water. These inputs include labor, fertilizer, pesticide, herbicide seed, plastic sheeting, and machinery (for most rural households in China, this means the cost of renting machine). The linear parameter a and quadratic parameter b are estimated using a random-coefficient model with net benefit as the dependent variable and water use in linear and quadratic terms as explanatory variables. In the 2004 CWIM, we collected information on the depth of water in the village, characteristics of pumps (size, water per hour, lift, etc.), electricity price, level of water use, and the amount farmers pay for water. Using this set of information, we calculate the average pumping cost to be used in the model.

Rarely will farmers know the level of salinity of the irrigation water they use, so we are unable to estimate parameters in the exponential function $e^{\alpha E_t - \theta (E_t)^2}$ using our survey data. These parameters are estimated using the experimental data on the levels of crop yields and different salinity levels of irrigation water that are reported in Shao et al. (2003).

Even less is known about the exact relationship between changes in salinity level and changes in the depth of water. Hengshui County is among the areas that have the highest degree of salinization problem in Hebei province. The salinity level measured by EC increases by 0.81 dS/m (≈0.5427 g/L) between 1975 and 1995; the water levels dropped by 43.53 m during the same period (Fig. 1).

The value of δ, the parameter that measures the relationship between the changes in the salinity level and the water level, is around 0.012 gL/ m for Hengshui County. Hence, a value of 0.02 can be considered as large. Since the value of δ will differ across places, in the simulations, we choose a range of values of δ. These values range from very small to large, 0.0001, 0.04, 0.1, and 0.2.

---

[4] In the 2004 China Water Institutions and Management Survey, the enumeration team collected data in 24 communities in Hebei province. In order to guarantee an adequate sample of communities in each of several water usage situations, the communities were chosen randomly from three randomly selected counties according to location, which in the Hai River Basin often is correlated with water scarcity levels. Xian County is located along the coastal belt (the most water scarce area of China); Tang County is located along the inland belt (an area with relatively abundant water resources that are next to the hills and mountains that rise in the eastern part of Hebei province); and Ci County is located in the region between the coastal and inland belts. The survey was conducted by interviewing three different types of respondents in each community (or village): the community leader; well manager (typically three randomly selected well managers per community); and households (four randomly selected households). We use separate questionnaires for each type of respondents. Although most of the data in the analysis come from the household questionnaire, we also use some data from the community leader and well manager questionnaires. Two major blocks of data are used from the household survey: data on household production activities and data on household water use.

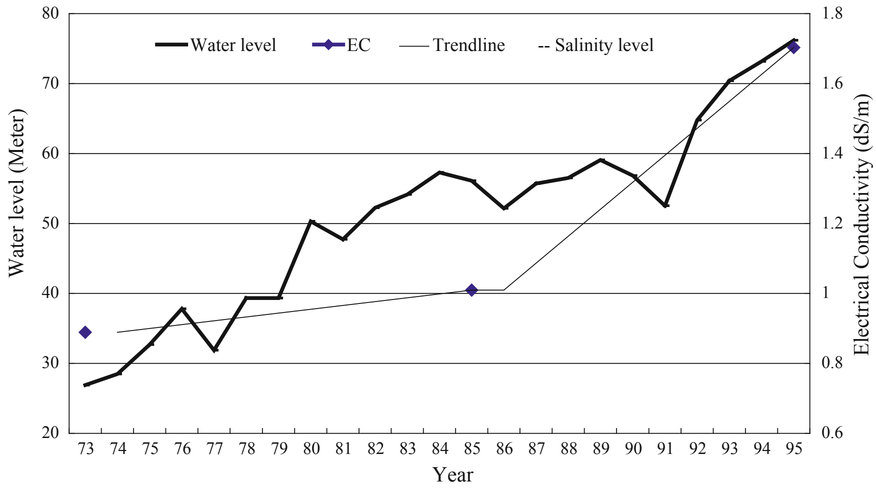

Fig. 1 Changes in the water level and the salinity level 1975–1995, Hengshui County, Hebei Province. *Source* Hebei Bureau of Geology Reconnaissance (2003)

Table 1 Values of parameters in the model

Parameter	Description	Value
β	Discount factor	0.9434
a	Intercept of marginal net benefit	\$0.39/m^3
b	Slope of marginal net benefit	0.007
EC_0	Initial salinity level	0.582 g/L
h_0	Initial depth-to-water	60 m
S	Specific yield	0.00157
c	Marginal pumping cost	\$0.000128/m3/m lift
α	Coefficient on the linear term in salinity-yield function	0.0025627
θ	Coefficient on the quadratic term in salinity-yield function	0.0111101

A discount rate of 6% is used. The initial water depth is set at 60 m. The value of a specific yield is taken from Chen et al. [P167, 1999]. The level of recharge is set at 5 cm expressed in terms of water depth, which is the value used by Shen et al. (2000). In an analysis in Table 2, we also increase it to 1 m. The rest of the parameter values are listed in Table 1.

4 Results of the Empirical Analysis

We solve the dynamic optimization problems numerically (the quantity-only problem, the quantity–quality problem, and the non-cooperative extraction problem) using the general algebraic modeling system (GAMS). When using the dynamic programming technique with the value-iteration algorithm to solve the problems numerically, we use the "collocation" method described in Judd (1998) and Miranda and Fackler (2002). In particular, the Chebychev polynomial is used to approximate the infinite horizon value function of water users. The solutions will provide the pumping path, the level of value function, and the level of the pumping lift and salinity level at the steady states in different problems. Using these results, we will compare the dynamic properties of the pumping path and test the propositions developed in the previous section.

4.1 Quantity-Only Problem Versus Quantity–Quality Problem

In the first part of the empirical analysis, we vary the value of the parameter, δ which measures the impact of groundwater extraction on changes in the salinity level. When δ is zero, there is no salinization problem and users are solving a quantity-only problem (Model 1). When δ takes on nonzero values, there is a salinization problem and users are solving a quantity–quality problem (Model 2). Comparison of the solutions to the two problems will enable us to examine Propositions 1 and 2 in order to answer the question "How does the optimal allocation of groundwater resource differ between the regions with salinization problems and regions without such problems?"

Comparison of the solution to a quantity-only problem and the solution to a quantity–quality problem shows that the way to optimally use groundwater is different between a region with a salinization problem and a region without a salinization problem. The pumping paths differ (Fig. 2a).

When the value of δ is between 0.04 and 0.2, compared to a water user in the region without a salinization problem, in a region with a salinization problem a water user pumps less at all times. The pumping path is more complicated when the value of is 0.0001 (Fig. 3a).

Compared to a water user in the region without a salinization problem and in a region with a salinization problem, a water user pumps more at the beginning of extraction. The rate of fall in his pumping rate is higher. After a certain period of time, his pumping rate drops below that of a water user in a region without a salinization problem. Intuitively, in the region with a salinization problem, pumping will lead to an increase in the salinity level of groundwater and thus reduces the benefit of future groundwater use. Under such a circumstance, if the impact of changes in water depth on the salinity level is small, it is optimal to pump more at the beginning when the

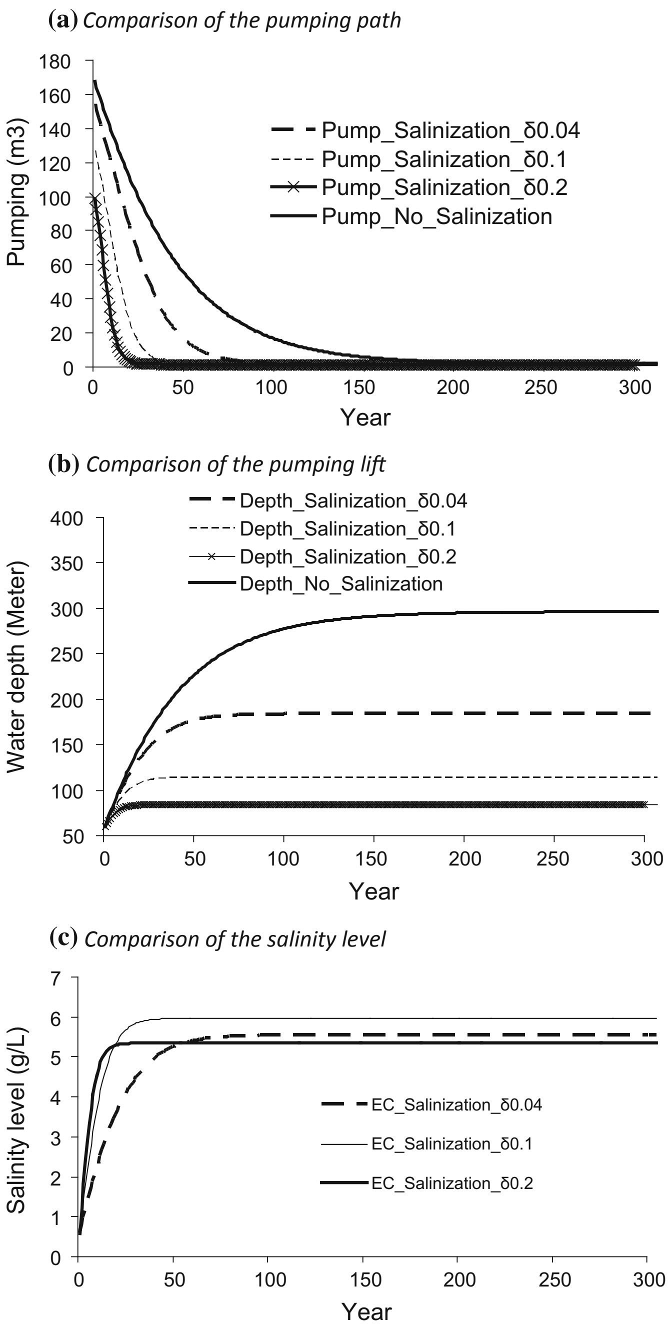

Fig. 2 Comparison of cooperative extraction with and without salinization problem ($\delta = 0.04; 0.1; 0.2$). **a** Comparison of the pumping path. **b** Comparison of the pumping lift. **c** Comparison of the salinity level

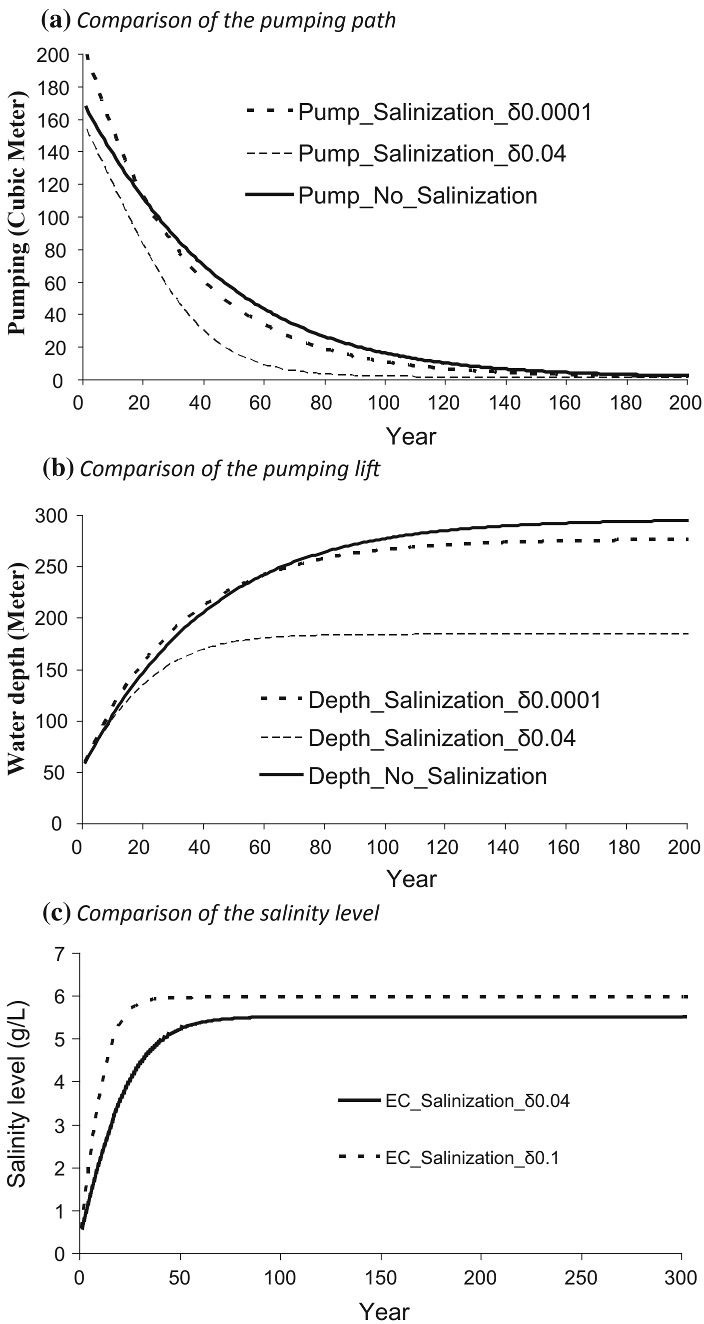

Fig. 3 Comparison of cooperative extraction with and without salinization problem ($\delta = 0.0001$; 0.04). **a** Comparison of the pumping path. **b** Comparison of the pumping lift. **c** Comparison of the salinity level

quality of groundwater is high. It is then optimal to reduce the pumping rate rapidly as the quality of groundwater deteriorates. If the impact of changes in water depth on salinity level is large, pumping will be penalized heavily in terms of benefit reduction even at the beginning, so it is optimal to pump less at all periods.

The comparison also provides evidence for Proposition 1 that the socially optimal pumping lift is smaller in regions with salinization problems (Figs. 3b and 4b). Even in the case of δ is 0.0001, although the heavy pumping of water users in the early stage leads to a more rapid increase in the pumping lift in the region with a salinization problem, the increase slows later when the pumping rate drops rapidly. At the steady state, the pumping lift is smaller when there is a salinization problem compared to when there is no salinization problem. Intuitively, since the pumping of groundwater leads to increases in the salinity level (Figs. 2c and 3c), the marginal net benefit of groundwater decreases and more water is left in the groundwater at the steady state.

The result of comparison also supports Proposition 2 that in regions with salinization problems, the value of groundwater decreases in the magnitude of impact that groundwater extraction has on the salinity level of groundwater. When δ is 0.04 and assume the level of recharge is five centimeter expressed in terms of water depth, the present value of net benefits from using groundwater is reduced by more than 10% compared to the scenario when δ is 0 (Table 2, Column 3).

The present value of net benefits from using groundwater is reduced by almost half when δ is 0.2. It is also consistent with what we have observed in the field. In our pretest and formal interviews with farmers in the HRB, we asked the following question: "Suppose China's government starts a payment for water program. You will be paid to stop cultivation to conserve water, how much is your willingness to accept?" We interviewed farmers in two different counties. In Cang County, there is a serious salinization problem in most places, and in Luancheng County there is no salinization problem. In Cang County, the willingness of farmers to accept ($656/ha/year) is about $300 less than that in Luancheng County ($938/ha/year). In the region with a salinization problem, since the future benefits water users could obtain are lower due to the more saline water, water users value groundwater less.

4.2 *Cooperative Extraction Versus Non-cooperative Extraction*

In the second part of the empirical analysis, we solve the quantity–quality problem numerically using the same nonzero value of δ as in the non-cooperative extraction problem (Model 3). A different game approach is used for the non-cooperative extraction problem. By comparing the solutions to the two problems, we are able to examine Proposition 3 in order to answer the question "In regions with salinization problems, how does the impact of pumping on groundwater quantity and quality differ when water users extract in a cooperative way and when water users extract in a non-cooperative way?"

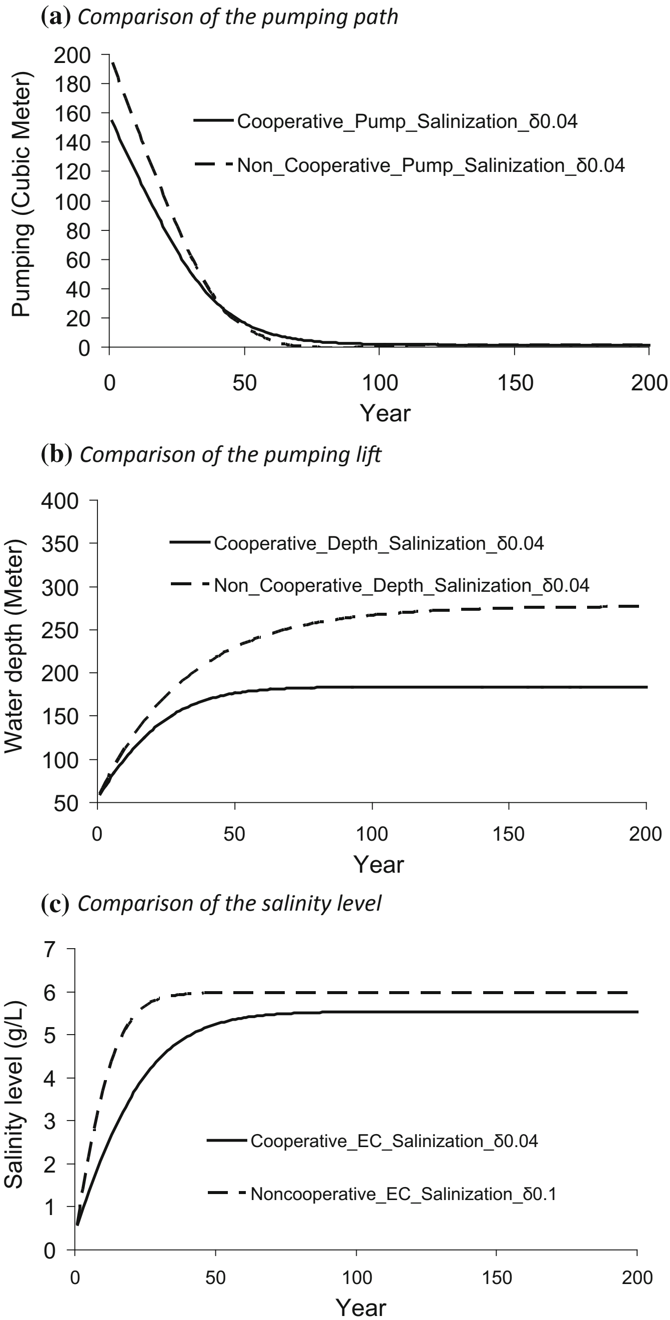

Fig. 4 Comparison of cooperative and non-cooperative extraction with salinization. **a** Comparison of the pumping path. **b** Comparison of the pumping lift. **c** Comparison of the salinity level

Table 2 Comparison of present value and gain from management

(1) Level of recharge	(2) δ	(3) Present value under optimal pumping	(3) Present value under non-optimal behavior		(4) Gain from management	(5) Ratio of gain from management under salinization to that under no salinization
5 cm	0[b]	537,118	Purely myopic	518,135	18,983	
	0.04	479,839	Purely myopic	415,987	63,852	3.4
			Myopic with update	425,131	54,708	2.9
			Only salinity myopic	443,331	36,508	1.9
	0.2	293,390	Purely myopic	197,143	96,247	5.1
			Myopic with update	201,792	91,598	4.8
			Only salinity myopic	239,776	53,614	2.8
1 m	0	588,543	Purely myopic	564,063	24,480	
	0.04	550,350	Purely myopic	451,341	99,008	4
			Myopic with update	474,531	75,818	3.1
			Only salinity myopic	540,501	9848	0.4
	0.2	343,943	Purely myopic	186,802	157,140	6.4
			Myopic with update	232,826	111,116	4.5
			Only salinity myopic	321,709	22,233	0.9

Notes [a]Present value is the total value of benefits from water use of all households within a village in dollars
[b]$\delta = 0$ is the case when there is no aquifer salinization problem

Comparison of the solution to the quantity–quality problem and the non-cooperative extraction problem uncovers the difference in the pumping path of cooperative users and non-cooperative users (Fig. 4a). At the early stage of extraction, non-cooperative users pump more than cooperative users since non-cooperative users ignore the social cost of their pumping. As a result, the pumping lift that non-cooperative users face increases rapidly and they are forced to reduce their pumping rates sooner due to the higher pumping costs. In fact, after a period of time, non-cooperative users are pumping less than cooperative users.

It is also observed that at the steady state, both the pumping lift and the salinity level in the cooperative extraction problem are smaller than that in a non-cooperative problem (Figs. 4b and 5c). This result supports Proposition 3. Thus, over-extraction by non-cooperative users causes both the depletion of the water resource and the deterioration of water quality.

4.3 Cooperative Extraction Versus Different Types of Myopic Extraction

Since currently there is no effective management in the Hai River Basin, we are also interested in the extraction behaviors when there is no regulation. We look at three different types of water users that display different types of myopic behavior: the *purely myopic* water users; the *myopic with update* water users; the *only salinity myopic* water users. The purely myopic water users maximize their own net benefit from the current period. They completely ignore the dynamics of both water stock and water salinity. The myopic with update water users also only maximize one-period net benefit. However, they realize that somehow the benefit they obtain from the same amount of water is less than that from previous years, although they do not realize it is the result of the interaction between water stock and salinity level. So they will update their benefit function based upon observations from previous years.[5] The only salinity myopic water users maximize own net benefit over time, but they do not realize the interaction between water stock and salinity level. The difference between the cooperative water users and the only salinity myopic water users is the latter does not consider the dynamics of salinity when maximizing the present value of net benefit.

The results show that all types of myopic water users lead to higher water depth (Fig. 5b) and higher salinity level (Fig. 5c) at the steady state in comparison with the optimal pumping case. Among the myopic users, the purely myopic users are pumping more than other types of myopic users in all periods (Fig. 5a). As a result, they deplete the water stock and quality more severely than other types of myopic users. Between the only salinity myopic users and the myopic with update water users, the pumping behavior of the only salinity myopic users is closer to that of optimal users. This is because they are maximizing the net benefit over time. Hence,

[5]In simulations, the benefit function is parameterized using values from the year before.

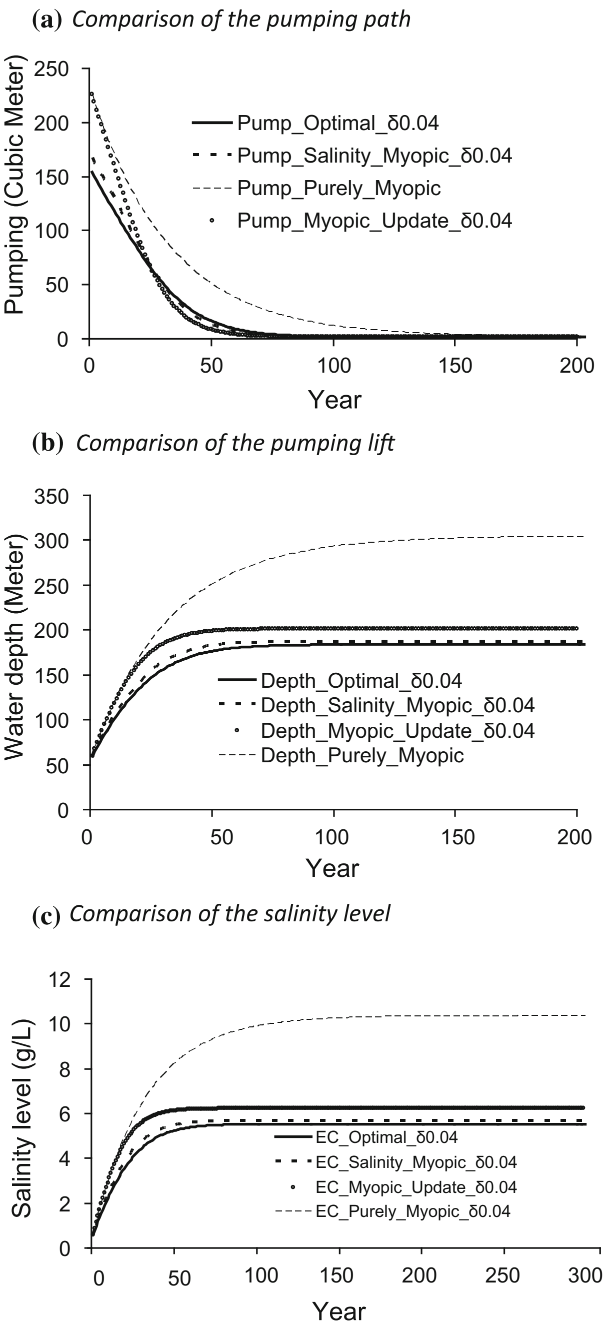

Fig. 5 Comparison of cooperative and myopic extraction with salinization. **a** Comparison of the pumping path. **b** Comparison of the pumping lift. **c** Comparison of the salinity level

depending on the characteristics of farmers and the village, pumping patterns are different when there is no regulation. In villages where the village leader puts efforts in playing the role of the social planner, farmers may be more like the only salinity myopic users since they are guided by the village leader to internalize their externality on other users. In those villages, the pumping of users is quite close to the optimal pumping case. In villages where farmers are more experienced or more motivated, farmers may behave like the myopic with update users, taking time to update their benefit function and revise their input uses. In these villages, farmers will also pump less than purely myopic users.

Our results also show that magnitudes of Gain From Management (GFM) are also different, depending on the degree of salinization types of behavior (Table 2, columns 5 and 6). In most cases, the GFM is larger when there is a salinization problem in comparison with when there is no salinization problem. The GFM can be double or more than five times higher when there is a salinization problem (column 6). The GFM is higher when water users are purely myopic than when water users are myopic with an update or only salinity myopic. However, when the recharge rate is high and farmers are only salinity myopic users, the GFM is much smaller than other cases. It is only 0.4 or 0.9 of the GFM when there is no salinization. Our findings indicate that the management decisions may differ across places. In villages where the aquifer receives high volume of recharge and farmers are only salinity myopic users, the cost of managing groundwater may be higher than the benefit, as pointed out in Gisser and Sanchez (1980). In other villages, however, the benefit of managing groundwater may overweigh the cost.

5 Conclusion

In this contribution, we have discussed the analytical framework for the optimal allocation of groundwater, both when there is no salinization problem and when there is a salinization problem. We also have compared the case when water users cooperate in pumping and the case when water users do not. Results from the numerical computation of the models are used to empirically examine propositions developed from the theoretical models. The lessons learned can help scholars and policymakers understand water use patterns in the HRB and provide some insights into how China should manage water (or whether they should manage it at all.

Results of the empirical analysis show that in a region where there is a salinization problem like in the HRB, the way to manage groundwater depends crucially on local conditions. For example, when extraction leads to small declines in the salinity level, it is actually optimal to pump at high rates in the early stage of extraction. Given this characteristic of the optimal pumping path, the heavy extraction we observe in some areas in the HRB does not necessarily indicate that groundwater resources are being overused. To judge whether we are overusing groundwater resources, we need to know which stage of extraction we are in. In fact, as already pointed out by Howitt and Nuckton (1981), even over-extraction is not necessarily bad during the earlier stages of extraction. If we are still in the early stage of extraction and it is still a long time before the steady state is reached, the pumping lift is still lower than the socially optimal level. Over-extraction is not harmful since it accelerates the convergence to the steady state by increasing the pumping lift rapidly.

Despite the potential danger of losing freshwater stock caused by both aquifer salinization and groundwater extraction, currently, the quality of groundwater and its quantity are managed in isolation of each other in the HRB. In most counties in the HRB, environmental protection bureaus are responsible for maintaining water quality and water resource bureaus are in charge of regulating groundwater extraction (Ministry of Water Resource et al. 2001). Under such a disjoint managing scheme, policies recommended by environmental protection bureaus may be inefficient. For example, massive investments in improving water quality may stimulate more extraction by farmers. Heavy extraction of water, by raising the salinity level of groundwater, makes these investments totally wasted. In fact, if the extraction of groundwater is not regulated, any measures that are intended to maintain the salinity level of freshwater will be in vain. In addition, the target level of optimal pumping lift set by the water resource bureaus will be incorrect. Hence, joint quantity–quality management is required in the HRB.

An equally important aspect that requires consideration before leaders make policies to manage groundwater resources is the cost and benefit of management. Empirical results of our study show that without regulations, the total benefit of all non-cooperative users obtain from extracting groundwater is lower than that obtained by cooperative users who act as if their extraction is managed by a social planner. Hence, there is a gain from managing groundwater and it is measured by the increase in the total benefit from using groundwater. However, the cost of management may easily exceed the benefit due to the fragmented and small-scale nature of China's farmers in hundreds of villages in the HRB. For example, the cost of measuring and enforcing water use on tens of millions of small parcels throughout HRB and collecting fees on a farmer-by-farmer basis may exceed the benefits of volumetric pricing.

If the result of cost-benefit analysis favors the implementation of a certain policy, regions with lower cost of implementation should be given priority. One such policy could be the 'payment-for-water' program. Empirical results of our study indicate that the payment that farmers are willing to accept to retire land is lower in regions with salinization problem. Thus, if China's government is to implement payment-for-water program, regions with salinization problems should be given priority since the cost (the payment to retire land) is lower there.

Appendix 1. Derivative of the Equation of Motion for the Groundwater Salinity Level (E_t)

In this study, we simplify our analysis by only focusing on the case when the saline water moves into the deep aquifer. In Fig. 6, the extraction of deep aquifer water,

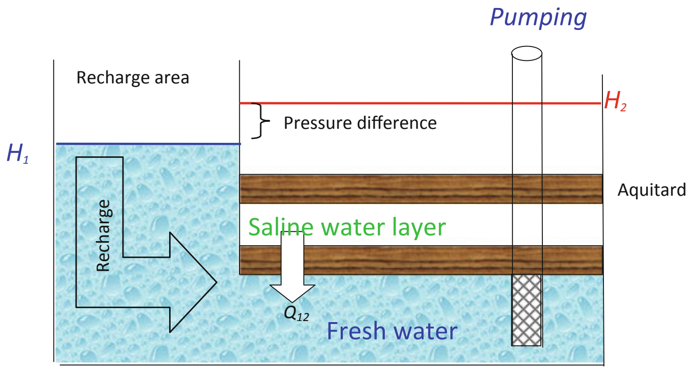

Fig. 6 The process of saline water intrusion due to groundwater extraction

Mw_t, leads to an increase in the depth-to-water in the deep aquifer, h_t. The hydraulic head in the deep aquifer, H_1, keeps declining as a result. When the pressure difference between the head in the saline water layer, H_2, and that in the deep aquifer is large enough, the saline water can move into the deep aquifer through the aquitard. The movement of saline water in response to the change in the head difference, Q_{12}, accounts for the phenomenon of increasing salinity level in the deep aquifer.

In the language of hydrology, Darcy's Law can be employed to formalize the movement of saline water.[6] Suppose the hydraulic head of the deep aquifer is linear in the depth-to-water: $H_{1t} = -c_1 h_t + d_1$ and $H_{2t} = c_2 \frac{Q_t}{As} + d_2$, where Q_t is the stock of the saline water, A is the area of saline water layer, and s is the specific yield, the volume of saline water that moves into the deep aquifer at time t can be expressed as

$$Q_t - Q_{t+1} = Q_{12} = -KA\frac{H_{1t} - H_{2t}}{b} = -\frac{KA}{b}\left[(-c_1 h_t + d_1) - \left(c_2 \frac{Q_t}{As} + d_2\right)\right] \tag{7}$$

where K is the hydraulic conductivity of the aquitard (unit: volume per unit of time), and b is the thickness of the aquitard. From (7), we have $Q_t - Q_{t+1} \propto h_{t+1} - h_t$.[7] We assume that the change in the level of salinity, $E_{t+1} - E_t$, is proportional to the total amount of intruded saline water at time t, Q_{12}. Hence, we have $E_{t+1} - E_t \propto h_{t+1} - h_t$. Suppose $E_{t+1} - E_t = \delta(h_{t+1} - h_t)$, we obtain the equation of motion for E_t:

$E_{t+1} = E_t + (h_{t+1} - h_t)$

[6]Darcy's Law states that the volume discharge rate Q is directly proportional to the head drop $h_1 - h_2$ and to the cross-section area A, but it is inversely proportional to the length difference, L (Wang and Anderson 1995): $Q = -KA\frac{h_1 - h_2}{L}$ where K is the hydraulic conductivity of the medium (e.g., clay or sand). The negative sign signifies that groundwater flows in the direction of head loss.

[7]$Q_t - Q_{t+1} = \frac{KAc_1}{b} h_t + \frac{KAc_2}{Asb} Q_t - \frac{KA}{b}(d_1 - d_2) \Rightarrow Q_{t+1} = (1 - \frac{KAc_2}{Asb}) Q_t - \frac{KAc_1}{b} h_t + \frac{KA}{b}(d_1 - d_2) \Rightarrow Q_{t+2} = (1 - \frac{KAc_2}{Asb}) Q_{t+1} - \frac{KAc_1}{b} h_{t+1} + \frac{KA}{b}(d_1 - d_2) \Rightarrow Q_{t+1} - Q_{t+2} = (1 - \frac{KAc_2}{Asb})(Q_t - Q_{t+1}) + \frac{KAc_1}{b}(h_{t+1} - h_t)$.

Appendix 2. Derivation of the Euler Equation for the Cooperative Extraction Model

We rewrite (2) as

$$\begin{aligned}\underset{\{w[t]\}}{\operatorname{Max}} L = {} & M \sum_{t=0}^{\infty} \beta^t f(w_t, E_t, h_t) \\ & - \sum_{t=0}^{\infty} \beta^{t+1} \lambda_{t+1} [h_{t+1} - (h_t + \phi_1 M w_t - \phi_2 R)] \\ & - \sum_{t=0}^{\infty} \beta^{t+1} \mu_{t+1} [E_{t+1} - (E_t + \delta(\phi_1 M w_t - \phi_2 R)]\end{aligned}$$

The first-order condition for this problem gives:

$$\frac{\partial L}{\partial w_t} = f_{w_t}(w_t, E_t, h_t) + \beta\phi_1(\lambda_{t+1} + \delta\mu_{t+1}) = 0 \tag{8}$$

$$\frac{\partial L}{\partial h_t} = M f_{h_t}(w_t, E_t, h_t) + \beta\lambda_{t+1} - \lambda_t = 0 \tag{9}$$

$$\frac{\partial L}{\partial E_t} = M f_{E_t}(w_t, E_t, h_t) + \beta\mu_{t+1} - \mu_t = 0 \tag{10}$$

$$=> \lambda_{t+1} + \delta\mu_{t+1} = -f_{w_t}(w_t, E_t, h_t)/(\beta\phi_1) \tag{8}$$

Lagging (8) by one period gives:

$$\lambda_t + \delta\mu_t = -f_{w_{t-1}}(w_{t-1}, E_{t-1}, h_{t-1})/(\beta\phi_1) \tag{11}$$

(9) $+\delta^*$ (10)$\Rightarrow$

$$\beta(\lambda_{t+1} + \delta\mu_{t+1}) - (\lambda_t + \delta\mu_t) = -M[f_{h_t}(w_t, E_t, h_t) + \delta f_{E_t}(w_t, E_t, h_t)] \tag{12}$$

Plugging (8) and (11) into (12) gives:

$$\begin{aligned}f_{w_{t-1}}(w_{t-1}, E_{t-1}, h_{t-1}) = {} & \beta f_{w_t}(w_t, E_t, h_t) - \beta\phi_1 M[f_{h_t}(w_t, E_t, h_t) \\ & + \delta f_{E_t}(w_t, E_t, h_t)]\end{aligned} \tag{11}$$

Rolling equation (11) forward one period gives:

$$f_{w_t} = \beta f_{w_{t+1}} - \beta\phi_1 M(f_{h_{t+1}} + \delta f_{E_{t+1}})$$

$$= \beta[\beta f_{w_{t+2}} - \beta\phi_1 M\left(f_{h_{t+1}} + \delta f_{E_{t+1}}\right)] - \beta\phi_1 M\left(f_{h_{t+2}} + \delta f_{E_{t+2}}\right)$$

$$= \cdots$$
$$= \beta^s f_{w_{t+s}} - \sum_{\ell=1}^{s} \beta^\ell \phi_1 M\left(f_{h_{t+\ell}} + \delta f_{E_{t+\ell}}\right) \tag{12}$$

Leading it forward into infinite future, we obtain

$$f_{w_t} = \lim_{s\to\infty}\left\{\beta^s f_{w_{t+s}} - \sum_{\ell=1}^{s} \beta^\ell \phi_1 M(f_{h_{t+\ell}} + \delta f_{E_{t+\ell}})\right\} \tag{13}$$

Since β is the discount factor that is well within the range of 0 and 1, (13) gives

$$f_{w_t} = -M\sum_{\ell=1}^{\infty} \beta^\ell \phi_1 f_{h_{t+\ell}} - M\sum_{\ell=1}^{\infty} \beta^\ell \phi_1 \delta f_{E_{t+\ell}} \tag{14}$$

Appendix 3. Derivation of Proposition 1

At the steady state, $w_{t\text{-}1} = w_t = w*$. We use w^*, E^*, and $h^*$to denote the value at the steady state. Equation (14) now becomes:

$$(1-\beta)f_w(w^*, E^*, h^*) + \beta\phi_1 f_h(w^*, E^*, h^*) + \beta\phi_1\delta f_E(w^*, E^*, h^*) = 0$$

Using the implicit function theorem gives:

$$\frac{\partial h^*}{\partial \delta} = -\frac{\beta\phi_1 f_E(w^*, E^*, h^*)}{(1-\beta) f_{wh}(w^*, E^*, h^*) + \beta\phi_1 f_{hh}(w^*, E^*, h^*)}$$

We have $f_E < 0$, $1- > 0$, $f_{wh} < 0$ and $f_{hh} < 0$. Therefore,

$$\frac{\partial h^*}{\partial \delta} = -\frac{(-)}{(+)(-)+(-)} < 0$$

Appendix 4. Derivation of Proposition 2

Bellman equation for problem s(2) is:

$$\begin{aligned} V(h, E) &= \underset{w}{\text{Max}}\left\{f(w, h, E) + \beta V(h', E')\right\} \\ \text{s.t.}\quad h' &= h + \phi_1 M w - \phi_2 R \\ E' &= E + \delta(h' - h) \end{aligned} \tag{15}$$

Using the Envelope Theorem gives:
$V_\delta = \beta V_{E'}(h',\ E') \cdot (h' - h) \Rightarrow V_\delta < 0$ since $V_{E'} < 0$ and $h' > h$.

Appendix 5. Derivation of the Euler Equation for Non-cooperative Extraction Model

we rewrite (4) as

$$\begin{aligned}\underset{\{w_{it}\}}{\text{Max}}\, L = &\sum_{t=0}^{\infty} \beta^t f^i(w_{it}, E_t, h_t) - \sum_{t=0}^{\infty} \beta^{t+1} \lambda_{t+1}(h_{t+1} - h_t - \phi_1 w_{it} - \phi_1 \sum_{j\neq i}^{M} w_j^* + \phi_2 R)\\ &- \sum_{t=0}^{\infty} \beta^{t+1} \mu_{t+1}(E_{t+1} - E_t - \phi_1 \delta w_{it}\\ &- \phi_1 \delta \sum_{j\neq i}^{M} w_j^* + \phi_2 \delta R)\end{aligned}$$

The first-order condition for this problem gives:

$$\frac{\partial L}{\partial w_t} = f^i_{w_t}(w_t, E_t, h_t) + \beta\phi(\lambda_{t+1} + \delta\mu_{t+1}) = 0 \tag{16}$$

$$\frac{\partial L}{\partial h_t} = f_{h_t}(w_t, E_t, h_t) + \beta\lambda_{t+1} - \lambda_t + \phi_1\beta(\lambda_{t+1} + \delta\mu_{t+1}) \sum_{j\neq i}^{M} \frac{\partial w_j^*}{\partial h_t} = 0 \tag{17}$$

$$\frac{\partial L}{\partial E_t} = f_{E_t}(w_t, E_t, h_t) + \beta\mu_{t+1} - \mu_t + \phi_1\beta(\lambda_{t+1} + \delta\mu_{t+1}) \sum_{j\neq i}^{M} \frac{\partial w_j^*}{\partial E_t} = 0 \tag{18}$$

Using the same manipulations as in the steps to obtain the Euler equation for the cooperative model, we obtain:

$$\begin{aligned}f_{w_{it-1}}(w_{it-1}, E_{t-1}, h_{t-1}) = \beta f_{w_{it}}(w_{it}, E_t, h_t)\left\{1 + \sum_{j\neq i}^{M} \phi_1 \frac{\partial w_j^*}{\partial h_t} + \sum_{j\neq i}^{M} \phi_1 \delta \frac{\partial w_j^*}{\partial E_t}\right\}\\ - \beta\phi_1 f_{h_t}(w_{it}, E_t, h_t) - \beta\phi_1\delta f_{E_t}(w_{it}, E_t, h_t)\end{aligned} \tag{19}$$

Rolling equation (19) forward one period and continuing to substitute for terms in $t + 1$ gives:

$$f_{w_{it}} = \beta^s f_{w_{it}} + \sum_{\ell=1}^{s} \beta^\ell \sum_{j\neq i}^{M} \phi_1 \frac{\partial w_j^*}{\partial h_{t+\ell}} + \sum_{\ell=1}^{s} \beta^\ell \sum_{j\neq i}^{M} \phi_1 \delta \frac{\partial w_j^*}{\partial E_{t+\ell}}$$

$$-\sum_{\ell=1}^{s}\beta^{\ell}\phi_1 f_{h_{t+\ell}}-\sum_{\ell=1}^{s}\beta^{\ell}\phi_1\delta f_{E_{t+\ell}} \tag{20}$$

which as s goes to infinity becomes

$$f_{w_{it}}=\sum_{\ell=1}^{s}\beta^{\ell}\sum_{j\neq i}^{M}\phi_1\frac{\partial w_j^*}{\partial h_{t+\ell}}+\sum_{\ell=1}^{s}\beta^{\ell}\sum_{j\neq i}^{M}\phi_1\delta\frac{\partial w_j^*}{\partial E_{t+\ell}}-\sum_{\ell=1}^{s}\beta^{\ell}\phi_1 f_{h_{t+\ell}}-\sum_{\ell=1}^{s}\beta^{\ell}\phi_1\delta f_{E_{t+\ell}} \tag{21}$$

Appendix 6. Derivation of Proposition 2: Over-Extraction Under Non-cooperative Extraction

Applying the implicit function theorem to (16) gives

$\frac{\partial w_t^*}{\partial h_t}=-\frac{f_{wh}}{f_{ww}}=-\frac{(-)}{(-)}<0$ and $\frac{\partial w_t^*}{\partial E_t}=-\frac{f_{wE}}{f_{ww}}=-\frac{(-)}{(-)}<0.$

Therefore the term, $\phi_1\sum_{j\neq i}^{M}\left(\frac{\partial w_j^*}{\partial h_t}+\frac{\partial w_j^*}{\partial E_t}\right)$, is negative.

Given the same depth-to-water and the same salinity level, the right-hand side of (5) is lower than that of (3). Consequently, $w_{it\text{-}1} > w_{t\text{-}1}$ since $f_{ww} < 0$.

References

Burt, O. R. (1964). Optimal resource use over time with an application to ground water. *Management Science, 11*(1), 80–93.

Chen, W. (1999). *Groundwater in Hebei province*. Beijing: Earthquake Publishing House. [in Chinese].

Dinar, A., et al. (1993). A dynamic model of soil salinity and drainage generation in irrigated agriculture: A framework for policy analysis. *Water Resources Research, 29*(6), 1527–1537.

Feinerman, E., & Knapp, K. C. (1983). Benefits from groundwater management: Magnitude, sensitivity, and distribution. *American Journal of Agricultural Economics, 65*(4), 703–710.

Gisser, M., & Sanchez, D. A. (1980). Competition versus optimal control in groundwater pumping. *Water Resources Research, 16*(4), 638–642.

Hebei Bureau of Geology Reconnaissance. (2003). *Report on the evaluation of groundwater resources in Hebei Plain*. Hebei: Shijiazhuang.

Howitt, R., & Nuckton, C. F. (1981). Is overdrafting groundwater always bad? *California Agriculture, 35*(1), 10–12.

Judd, K. L. (1998). *Numerical methods in Economics*. Cambridge, MA: The MIT Press.

Kan, I., et al. (2002). Microeconomics of irrigation with saline water. *Journal of Agricultural and Resource Economics, 27*(1), 16–39.

Kendy, E. (2003). The false promise of sustainable pumping rates. *Journal of Ground water, 41*(1), 2–10.

Knapp, K. C. (1992a). Irrigation management and investment under saline, limited drainage conditions: 1. Model formulation. *Water Resources Research, 28*(12), 3085–3090.

Knapp, K. C. (1992b). Irrigation management and investment under saline, limited drainage conditions: 2. Characterization of optimal decision rules. *Water Resources Research, 28*(12), 3091–3097.

Knapp, K. C. (1992c). Irrigation management and investment under saline, limited drainage conditions: 3 policy analysis and extensions. *Water Resources Research, 28*(12), 3099–3109.

Maas, E. V., & Hoffman, G. (1977). Crop salt tolerance: Current assessment. *Journal of the Irrigation and Drainage Division, 103*(2), 115–134.

Ministry of Water Resource et al. (2001). *Agenda for water sector strategy for North China*. Ministry of Water Resources (MWR), World Bank and AusAID.

Miranda, M. J., & Fackler, P. L. (2002). *Applied computational economics and finance*. Cambridge, MA: The MIT Press.

Mu, C., & Zhang, J. (2002). The current status of downward movement of saline and freshwater interface and the mechanism of saline water intrusion. *Hebei Hydrology and Water Electricity Technology*(1), 37–39.

Negri, D. H. (1989). The common property aquifer as a differential game. *Water Resources Research, 25*(1), 9–15.

Nickum, J. E. (1988). All is not wells in North China: Irrigation in Yucheng county. In G. T. O'Mara(Ed.), *Efficiency in irrigation: A world bank symposium* (pp. 87–94). Washington, D.C.: World Bank.

Provencher, B., & Burt, O. R. (1993). The externalities associated with the common property exploitation of groundwater. *Journal of Environmental Economics and Management, 24,* 139–158.

Roseta-Palma, C. (2003). Joint quantity/quality management of groundwater. *Environmental & Resource Economics, 26*(1), 89–106.

Roseta-Palma, C. (2002). Groundwater management when water quality is endogenous. *Journal of Environmental Management, 44,* 93–105.

Rubio, S. J., & Casino, B. (2001). Competitive versus efficient extraction of a common property resource: The groundwater case. *Journal of Economic Dynamics and Control, 25*(8), 1117–1137.

Shao, Y., et al. (2003). Technique of brackish water for farmland irrigation. *Tianjin Agricultural Sciences, 9*(4), 25–28.

Shen, Z., et al. (2000). *The evolution of the groundwater environment in North China Plain*. Beijing: Geology Publishing House.

Song, W., & He, J. (1996). Current status of water resources in Cangzhou City. *Hebei Hydrology and Water Electricity Technology* (2).

Van Genuchten, M. T., & Hoffman, G. J. (1985). Analysis of crop salt tolerance data. In I. S. a. J. Shalhevet (Ed.) *Soil salinity under irrigation processes and management*. (pp. 258–271). New York: Springer.

Wang, H. F., & Anderson, M. P. (1995). *Introduction to groundwater modeling: Finite difference and finite element Methods*. New York: Academic Press.

Zhu, J., et al. (2002). Causes for degradation of the environment of the deep aquifers in Hebei Plain and the countermeasures. *Journal of Shijiazhuang Teachers College, 4*(2), 39–42.

Chapter 7
Managing Urban and Agricultural Water Demands in Northern China: The Case of Luancheng County, Hebei Province

Siwa Msangi

Abstract Despite efforts to reform management of water resources, groundwater levels have continued to decline steadily on the North China Plain, leading to serious environmental concerns and impacts. While policy makers have looked to efforts aimed at improving the efficiency of field-level irrigation and strengthening ownership and property rights in local resource management, hydrologists have asserted that more direct control of consumptive use patterns of water is needed. In this contribution, we show how both agricultural and urban demands for water can be managed, so as to ameliorate the depletion of groundwater resources in the North China Plain and promote long-run sustainability of limited water resources.

1 Introduction

Northern China's deepening water crisis has attracted increasing attention from policy makers and water specialists over recent years (Crook and Diao 2000). The threat of a water crisis has always loomed over the region, given the bias in the distribution of water resources toward the South and an increasingly dense urban and rural populace. Although the Huai, Hai, and Huang (Yellow) river basins in the North China Plain account for 24% of GDP (Kahrl et al. 2005) and generate 27% of the national production of grain (Liu 1998), the region has 7.5% of China's water (Varis and Vakkilainen 2001). The decreasing reliability of water available for irrigation has also raised concerns about national food security (Heilig et al. 2000; Moench et al. 2003; Crook 1999) and the potential deepening of poverty and inequality for rural farmers who depend on irrigated agriculture for household income (Huang et al. 2002).

The gradual expansion of irrigated acreage in North China has had a severe effect on groundwater resource levels over time. A tremendous increase in grain production was achieved largely by expanding the acreage under maize and wheat, while that

S. Msangi (✉)
International Food Policy Research Institute, Washington, DC, USA
e-mail: siwamsangi@gmail.com

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_7

of dryland crops (millet and sorghum) decreased (Crook 1999). The expanded cultivation of winter wheat, in particular, requires an increase in irrigation during the dry winter months and is one of the main causes for the drawdown on groundwater levels (Yang et al. 2002). China's falling water table also has created other environmental problems, such as land subsidence (Zhang and Zhang 1995) and seawater intrusion (Lohmar et al. 2003; World Bank 2002).

Urban demand for water has also been increasing rapidly in China, causing greater water stress and competition with the agricultural sector for limited water resources. While agriculture accounts for most of the water use in China (about 65% in 2003), the increasing rural to urban migration and the resulting domestic and industrial water requirements have intensified the competition for water (World Bank 2001b). Migration from rural to urban regions is taking place at an unprecedented rate in China; from 1952 to 2003 the proportionate urban population tripled from 13 to 39% to more than half a billion (Liu and Diamond 2005). This has exerted pressure on water resources for domestic and industrial use, and between 1980 and 1993, urban water consumption increased by 350%, and industrial water consumption doubled at the cost of irrigation use, which declined by about 4%, even though irrigated area increased by about 10% (World Bank 2001b).

Given the increasing stress on water resources on the North China Plain from urban and agricultural uses, policymakers and resource managers have considered several options to expand supply—including that of large-scale surface water transfers. To alleviate the heavy dependence of the North on groundwater, which accounts for 64% of its total water use (Kahrl et al. 2005), a large-scale project to transfer water between the Yangtze River in the south to the North China Plain was conceived. The infrastructure costs of such an undertaking are high, and the Ministry of Water Conservancy undertook a feasibility study to determine which transfer route (Eastern, Western, or Middle) will be chosen.[1] However, given the high unit costs of transferred water, it is challenging for agricultural and rural users to afford it. Given the increasing levels of urbanization and industrialization in Northern China, it would primarily be used to meet growing urban domestic and domestic demands.

Given that transfers may not benefit agricultural users, other policy interventions need to be considered to address the continuing competition between agricultural and urban demands for limited groundwater resources. According to some experts (Zhong et al. 2017; World Bank 2001a), one of the reasons why China's past policy efforts failed to halt the decline of the groundwater table on the North China Plain is because they did not promote any *real water savings*.[2] Although the promotion of improved irrigation technologies and irrigation management practices aid in reducing the total water applied to a farmer's fields, increasing evidence has pointed to the fact that such measures are not enough in promoting sustainability of water usage within

[1]The middle route is the most likely option to be chosen and would supply the north with high-quality water with the help of gravitational force (Liu 1998).

[2]In the rest of the chapter, the term "real water savings" refers to the reduction in non-recoverable water losses that occur through such mechanisms as evapotranspiration or nonessential transportation, rather than through the reduction in seepage losses, which can be recovered further downstream in the water basin (Foster et al. 2003).

a groundwater basin (Zhong et al. 2017). In fact, with the exception of reducing water flows into "sinks," only those modifications to irrigation and cropping patterns that reduce non-beneficial evaporation and evapotranspiration actually represent "real" water savings (Foster 2000). As Kendy (2003) points out, in groundwater villages in most places in Northern China, the real losses in water from the aquifer are represented by that water which does not return to the aquifer after it is pumped onto the field; water accounting exercises mostly show clearly that the main consumption of water from agriculture is mostly that which evapotranspires from the crop and the soils.

Therefore, researchers and policy analysts should consider a range of policy instruments that might achieve real water savings through a reduction in the consumptive use of water on the North China Plain, including the regulation of land usage, so as to limit the total irrigated area of a highly consumptive crop such as winter wheat. Such a policy would run counter to the government's efforts to promote grain security, which necessitates a strong case being made for this type of an instrument, over an alternative price-based one for agricultural or urban water use.

In this contribution, we examine possible strategies to deal with the steady depletion of the groundwater resources on the North China Plain and to compare their efficacy and impact. In order to carry out our analysis, we employ an economic framework to demonstrate the impact of optimally managing the consumptive use of water in agriculture at the aggregate basin level through the adjustment of cropping patterns, so as to create real water savings and promote both the long-term hydrological and economic sustainability of the groundwater aquifer. We also show what other policy instruments might be applied to manage the demand of urban water use and evaluate the overall effect on user benefits and the groundwater basin, in relation to the optimal policy. Embedded within our analytical framework is a dynamic economic model of groundwater usage for a specific area of the North China Plain—Luancheng County—which determines the optimal cropping pattern that will achieve the reduction in total evapotranspiration necessary to maintain sustainable groundwater use within the region's agricultural economy.

2 Materials and Methods

In this section, we describe the essential features of an integrated economic model of agricultural production and groundwater usage that is used to prescribe necessary adjustments to the cropping patterns on the North China Plain, aimed at reducing the consumptive use of water to a sustainable level. The essential features of both the local hydrology and the agricultural economy are combined into an integrated analytical framework in which the usage of groundwater in agricultural production can be studied. The results of this analysis are then compared to the recommendations of other authors in the literature and motivate the institutional analysis that follows in the later sections of the contribution.

The economic model used to study the aggregate-level management of crop consumptive use and groundwater usage in Luancheng County consists of two essential components. The first component of the model is an agricultural production model that represents the cropping behavior of agricultural producers in Luancheng County and their use of groundwater for irrigation. The second component is a simplified hydrological model of the aquifer underlying Luancheng County and the neighboring urban center of Shijiazhuang City. We briefly discuss each of these components and describe how they are integrated into a unified analytical framework.

2.1 Agricultural Production Model

The agricultural model is calibrated to the cropping patterns that are observed within the county and incorporates a measured relationship between yields and the amount of water and fertilizer used as inputs. By embedding an explicit yield function within the model, we allow the model to respond to changes in water availability by either changing the intensity of the inputs used in production or by adjusting the area of land put into cultivation. The specific form of the yield function is given by the following quadratic function

$$y = \begin{bmatrix} \alpha_N & \alpha_W \end{bmatrix} \begin{bmatrix} x_N \\ x_W \end{bmatrix} - \begin{bmatrix} x_N & x_W \end{bmatrix} \begin{bmatrix} \gamma_{NN} & \gamma_{NW} \\ \gamma_{WN} & \gamma_{WW} \end{bmatrix} \begin{bmatrix} x_N \\ x_W \end{bmatrix} = \boldsymbol{\alpha}'\mathbf{x} - \mathbf{x}'\mathbf{0}\mathbf{x} \qquad (2.1.1)$$

where y is the yield a given crop, and x_N are the quantities of water and fertilizer allocated to crop production. The equation below for the full multi-input, multi-output production model shows how the yield function is incorporated into the profit maximization problem of the producer.

$$\begin{aligned} &\max_{\mathbf{x}_k} \quad \sum_k \left[p_k a_k \left(\boldsymbol{\alpha}'_k \mathbf{x}_k - \mathbf{x}'_k \boldsymbol{\Gamma}_k \mathbf{x}_k \right) - \phi_k (a_k)^2 - \boldsymbol{\omega}'_k \mathbf{x}_k \right] \\ &\text{subject to} \quad \sum_k a_k \leq \overline{\mathrm{A}} \quad \text{(land)} \end{aligned} \qquad (2.1.2)$$

In this multi-input and multi-output production model, each crop has a separate yield function, and the crop-specific production decisions are linked solely through the constraint on total land available—in manner similar to that used by Just et al. (1983).

The calibration in crop area is achieved through the employment of the positive mathematical programming principle of Howitt (1995)—referred to, popularly, as PMP. The PMP calibration is embodied in the parameter ϕ_k, which is derived from a shadow value λ_k on the constraints that restrict the non-calibrated model to the base crop areas $\{\overline{a}_1 \ldots \overline{a}_k\}$. The parameter value is calculated from the relationship, $\phi_k = \frac{2\lambda_k}{\overline{a}_k}$, and is added to the model as a coefficient of the nonlinear term for the cultivated area, as shown above, which allows the model to calibrate exactly to the observed acreages. The PMP calibration method has found wide-ranging applications to the construction of policy analysis models, and numerous examples of its use can be found in the work of various authors in the agricultural economics literature (Barkaoui and Boutault 2000; Heckelei et al. 2012; Mérel and Howitt 2014; Merel et al. 2011).

2.2 Hydrological Model of Luancheng County

The second essential component of the economic model used to study the aggregate-level management of consumptive use and groundwater usage is a simplified model of the groundwater aquifer underlying Luancheng County and the neighboring urban center of Shijiazhuang City. We use a single-cell aquifer model to represent the groundwater aquifer and link it to the agricultural production model through groundwater pumping and return flow relationships, and the consumptive use of water implied by a given economically driven cropping pattern. While this model does not match the complexity and detail of the hydrological model used by Kendy et al. (2003) in their study of Luancheng County, it captures the important linkages between the withdrawals of the agricultural and urban users, and the return flow to the aquifer, that are essential to the modeling framework used in this contribution, and which determine the evolution of the groundwater stock over time.

The hydrological model is specified with parameters appropriate for Luancheng County and Hebei Province (Table 1), and the flow relationships in and out of the groundwater aquifer are shown in Fig. 1.

The withdrawals of agricultural producers in Luancheng County and industrial/urban users in Shijiazhuang city from the aquifer are represented by $W_{\mathrm{i=ag,ur}}$,

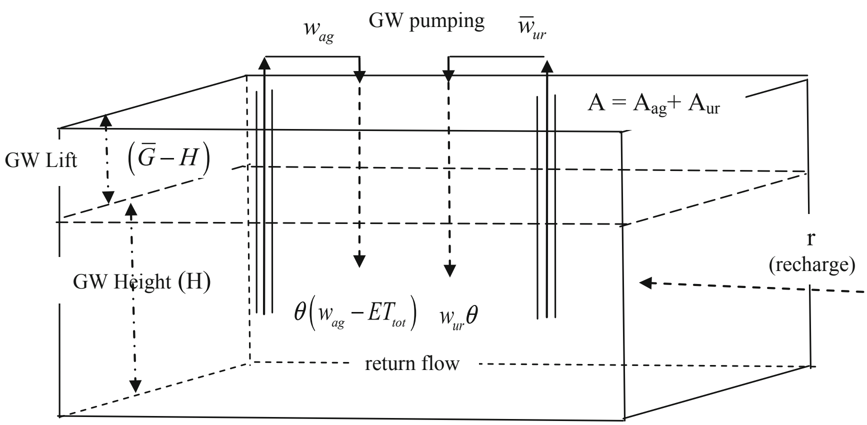

Fig. 1 Single-cell model of groundwater aquifer in Luancheng County

Table 1 Parameters for integrated economic-hydrological model

	Description	Value	Source
A_{ag}	Area overlying aquifer for agricultural sector	30,700 ha	Kendy et al. (2003)
A_{ur}	Area overlying aquifer for urban area	25,000 ha	Kendy et al. (2003)
S_y	Specific yield of aquifer	0.2	Kendy et al. (2003); Zhang et al. (2003)
θ	Deep percolation coefficient	0.2	Own estimate
e	Energy cost per unit pumping lift	15.46 yuan/m-ha per m lift	Hebei 2001 survey data
$\overline{R}$	Reference level for aquifer recharge	370 m	Own calculation
$\overline{G}$	Ground surface level	370 m	Foster et al. (2001)
r	Maximum recharge flow to aquifer	0.057 m/yr	Kendy et al. (2003)
α_R	Proportionality constant for recharge flow	1.907 m^{-1}	Own calculation
$\overline{w}_{ur}$	Fixed urban groundwater pumping	43,900 m-ha	Kendy et al. (2003)
$\overline{p}$	Annual average precipitation	469.2 mm/yr	Zhang et al. (2003)

Note Hebei 2001 survey carried out by Jinxia Wang of the Chinese Center for Agricultural Policy, Beijing, PRC, and referenced in Zhang et al. (2010)

while the return flows to the aquifer are governed by the deep percolation coefficient θ and the total evapotranspiration of crops grown by agricultural producers (ET_{tot}). The difference between the water applied to the irrigated crops (w_{ag}) and the total evapotranspiration of the crops being cultivated (ET_{tot}) will determine the amount of water that returns to the aquifer from agricultural users, according to the relationship $\theta(w_{ag} - ET_{tot})$, while the return flow from urban uses is θw_{ur}. Therefore, as the consumptive use of water increases in the agricultural region (through an increase in total evapotranspiration), then the return flow to the aquifer will decrease, accordingly.

In addition to this return flow from applied irrigation water, there is also subterranean inflow into the aquifer from natural sources of recharge (r). Using a constant of proportionality α_R, we get the amount of recharge in a given period as $\alpha_R\left(\overline{R} - H\right)r$, where the difference between the height (H) and the reference level $\left(\overline{R}\right)$ provides the hydraulic gradient that drives the flow of recharge into (or out of) the aquifer.

In Fig. 1, H represents the groundwater level in the aquifer, such that the distance over which the user must lift the groundwater that is pumped from the aquifer to the surface, in order to make beneficial use of it, is given by the difference $\left(\overline{G} - H\right)$, called the "lift"—where $\overline{G}$ is the height of the ground surface above a common reference level.

The height of the water in the aquifer evolves from one period to the next according to the following equation of motion,

$$H^{+} = H - \frac{\left(w_{\mathrm{ag}} - \theta\left(w_{\mathrm{ag}} - \mathrm{ET}_{\mathrm{tot}}\right) + (1-\theta)w_{\mathrm{ur}}\right)}{A \cdot s_{\mathrm{y}}} + \alpha_{\mathrm{R}}\left(\overline{R} - H\right)r \tag{2.1.3}$$

where H^{+} is the height of the groundwater level in the next period and s_{y} is the specific yield of the aquifer material, and A is the areal extent of the aquifer.

2.2.1 Integrated Dynamic Economic Model of Groundwater Usage

In order to completely specify the formulation of the dynamic economic model of water usage that will be used in this contribution, we combine the economically based objective criterion, which reflects agricultural production behavior and benefits from water usage, with the hydrology of the single-cell groundwater model. By doing so, we are able to link groundwater usage and return flows to the aquifer, explicitly, with the cropping patterns observed from the model behavior. The groundwater usage by urban users is fixed at a constant level ($\overline{w}_{\mathrm{ur}}$) which corresponds to the observed withdrawals reported by Kendy et al. (2003), while the groundwater withdrawals by the agricultural users are driven by the economic criterion embedded in the agricultural production model.

The full dynamic problem of combined agricultural production and groundwater usage can now be written, in terms of the Bellman equation shown below.

$$V(H) = \max_{\substack{w_{\mathrm{ag}} \\ \{a_k\}_{k=1}^{k=K}}} \left\{ \begin{array}{l} \sum_k \left[p_k a_k\left(\boldsymbol{\alpha}'_k \mathbf{x}_k - \mathbf{x}'_k \boldsymbol{\Gamma}_k \mathbf{x}_k\right) - \phi_k (a_k)^2 - \boldsymbol{\omega}'_k \mathbf{x}_k\right] - e\left(\overline{G} - H\right) w_{\mathrm{ag}} \\ \quad + \beta V\left(H^{+}_{\mathrm{ag}}\left(w_{\mathrm{ag}}, \overline{w}_{\mathrm{ur}}, H, \mathrm{ET}_{\mathrm{tot}}\right)\right) \\ \text{s.t.} \quad \mathrm{ET}_{\mathrm{tot}} = \sum_k \left(\mathrm{ET}_k a_k\right),\ w_{\mathrm{ag}} + \overline{p} = \sum_k x^k_W,\quad \sum_k a_k \le \overline{\mathrm{A}} \end{array} \right\}$$

where

$$H^{+} = H - \frac{\left[w_{\mathrm{ag}} - \theta\left(w_{\mathrm{ag}} - \mathrm{ET}_{\mathrm{tot}}\right) + (1-\theta)w_{\mathrm{ur}}\right]}{A \cdot s_{\mathrm{y}}} + \alpha_{\mathrm{R}}\left(\overline{R} - H\right)r \tag{2.3.1}$$

and where the water requirements of the crops can also be met by precipitation ($\overline{p}$). In this model, the economic planner behaves in a manner which is dynamically optimal, by maximizing the combined agricultural and urban net benefits for the current period, while also taking into account the future state of the groundwater table as a result of withdrawals from the aquifer and total evapotranspiration in the current period. In this way, the equation of motion links the decisions made in the current period to the evolution of the groundwater over time. The function $V(H)$ represents the maximized value of the dynamic problem starting with the current groundwater level (H), for a given time period forward, assuming that the actions taken in subsequent periods will be done optimally, with respect to the groundwater stock carried over to later periods. This recursive relationship linking the implied optimality of behavior from period-to-period captures the essence of Bellman's "Principle of Optimality" (Bellman 1957) and holds for the entire time horizon over which groundwater is withdrawn.

In this formulation, we have an explicit linkage between the agricultural production decisions taken in every period of the optimal path and the evolution of the groundwater resource over time. The cropping patterns are linked to the groundwater hydrology through the total evapotranspiration of a given cropping pattern, given by the relationship, $\mathrm{ET}_{\mathrm{tot}} = \sum_k (a_k \cdot \mathrm{ET}_k)$, where ET_k is the consumptive use requirement of crop k, and a_k is its irrigated area. There is also a direct tie between the water pumped from the aquifer cell underlying the agricultural area, and the total water used in agricultural production, through the relationship $w_{\mathrm{ag}} = \sum_k x_W^k$.

3 Results and Discussion

By solving the dynamic economic model (2.3.1), we obtain results (Table 2) which recommend a drastic change of cultivation patterns in Luancheng County and demonstrate a sizable reduction of the total ET requirements in the county, in order to stabilize the groundwater levels and maintain long-run economic sustainability for the groundwater basin. The groundwater level in the aquifer model is stabilized at 47 m depth below the ground surface, which is slightly below the current 40 m of depth that is observed in the area (Foster et al. 2001; Crook 1999). In order to achieve this equilibrium level, the model makes a significant shift from the cropping patterns currently observed in Luancheng County, as given in Table 2.

This table shows that the cultivation of winter wheat is abandoned in favor of maize and that more cotton and millet are grown than before. This shift in cultivation pattern implies an overall reduction of nearly 17% in the total evapotranspiration

Table 2 Cropping pattern changes suggested by economic model (ha)

Irrigated crops	Observed pattern in Luancheng	Pattern suggested by model results	ET requirements (mm/year)[a]
Wheat	24,219	0	490.9
Cotton	427	985	649.8
Summer maize	15,793	34,217	359.5
Spring maize	100	182	431.3
Millet	43	82	341
Implied ET of water (mm/year)	441.1	367.9	

[a] *Source* Zhang et al. (2003)

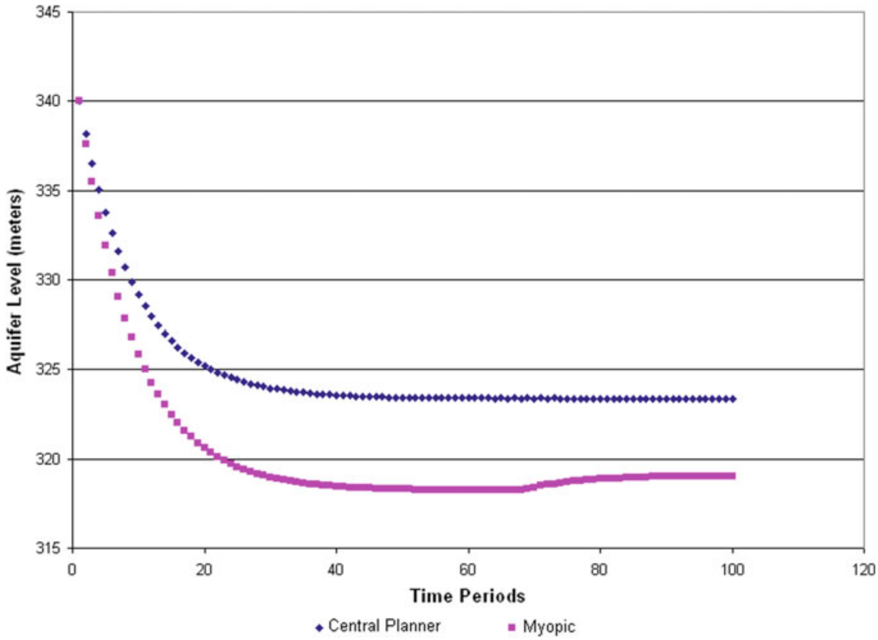

Fig. 2 Comparison of long-run groundwater levels under optimally chosen cropping patterns versus myopic extraction

within the county and is a little more than half of the 30% reduction endorsed by Kendy et al. (2003).

In comparison to this, we observe the groundwater level that is reached when there is sub-optimal and myopic extraction from the aquifer, as shown in Fig. 2.

From Fig. 3, we see that the groundwater stabilizes at a lower level of 315 m, which is achieved solely through the equilibration of pumping benefits and costs, which increase to the level that forces the pumping level to reduce to a level that allows for equilibration with the recharge into the aquifer. This is in contrast to the results of Kendy et al. (2003), which show, for a fixed cropping pattern, that the groundwater levels would drop far below that, which would be economic for extraction, and which are derived purely from considering the physical water balance and the extraction rates implied by the ET requirements of a given fixed cropping pattern. This contrast illustrates the difference between an economically motivated analysis and one which is driven purely by the physical characteristics and underlying hydrological processes in the system.

Alternatively, we could impose a number of possible policy instruments and compare their efficacy vis-à-vis the optimal, dynamic result, which are, namely a volumetric tax on pumping or a per-unit tax on the irrigated area under cultivation of winter wheat. The results from these policies are shown in Fig. 3.

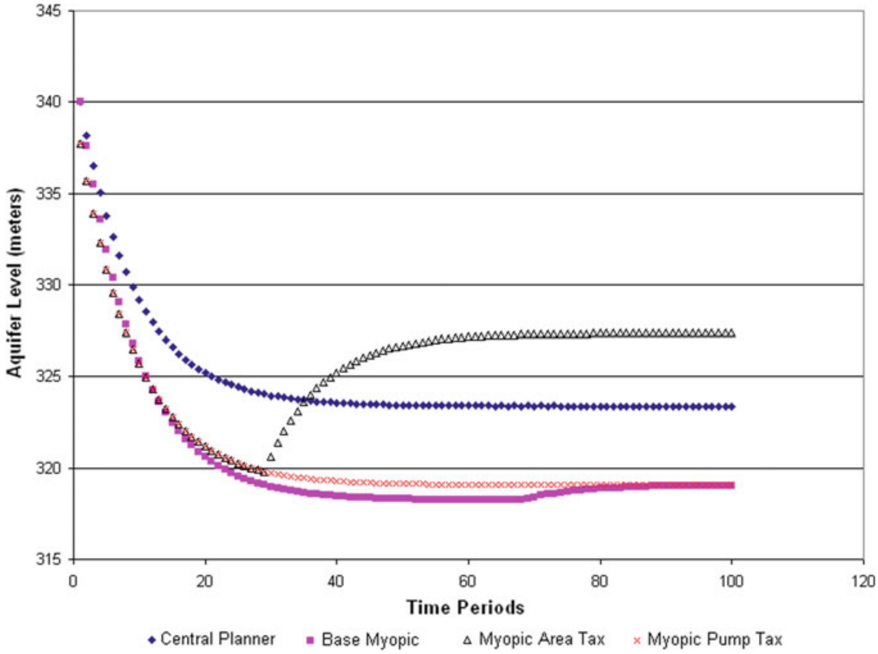

Fig. 3 Comparison of long-run groundwater levels under alternative policies

From this figure, we see that the pump tax policy differs only slightly from the base myopic extraction case in which no policy is applied. The per-unit area tax on the irrigated area under winter wheat, however, has a much more dramatic effect, from the point at which it is applied ($t = 30$). In this case, the tax was set at a level that caused a switch out of winter wheat, but no compensating increase in the area under maize, cotton, or millet, unlike the result in Table 2. Because of this, the overall groundwater table is higher than under the optimal solution, due to the less total irrigated area, even though the total revenues from agriculture are lower than optimal.

In both cases, the tax revenue was assumed to be redistributed as lump sum payments to prevent overall welfare loss under the policies.

4 Conclusions

In this contribution, we address the urgent problem of rapidly declining groundwater tables on the North China Plain through the application of a broad range of analytical, economic methods so as to better understand the behavioral dimensions of an issue that has been treated as purely hydrological in nature. The results of our analysis show that a nearly 17% reduction in total evapotranspiration is possible through affecting a shift in cropping pattern that abandons the cultivation of winter wheat in favor of expanded summer maize production as well as continued cotton and millet production. While this result echoes the recommendations given by many experts who have studied the groundwater resource problem in Northern China, the criteria that were applied in reaching this conclusion were both economic and hydrological in origin and not solely based on the physical water balances performed in other studies.

This contribution has shown that while the physical depletion of the groundwater table is hydrological in nature, the economic behavior underlying the cropping patterns that contribute to it cannot be ignored. This fact should be recognized by any policy that aims to reduce irrigated acreage of winter wheat in favor of alternate cropping patterns. The pump tax has no real effect on groundwater levels, as it does not promote real water savings. The tax on irrigated area, however, does have potential to reduce the overall consumptive use of water, although its effect on the area of other crops may be sub-optimal compared to the central planner's solution. The practical question of how to implement land usage policy is critical, as it will determine the overall welfare effects on farmers. This contribution has shown that policies which directly affect the usage of land can have an important role to play in solving Northern China's deepening groundwater depletion problem.

References

Barkaoui, A., & Boutault, J.-P. (2000). Cereal and oil seeds supply with EU under Agenda 2000: A positive mathematical programming application. *Agricultural Economics Review, 1*(2), 7–17.

Bellman, R. (1957). *Dynamic programming*. Princeton, New Jersey: Princeton University Press.

Crook, F. W. (1999). Water use and crop production in China's Hai river basin. In *Proceedings of WCC-101 Chinese Agriculture and the WTO*, Seattle, Washington, December 2–3, 1999.

Crook, F., & Diao, X. (2000). Water pressure in China: Growth strains resources. In *Agricultural Outlook Economic Research Service, USDA* (pp. 25–29), January–February.

Foster, S. (2000). Sustainable groundwater exploitation for agriculture: Current issues and recent initiatives in the developing world. In *Uso Intensivo de las Aguas Subterraneas: Aspectos Eticos, Tecnologicos y Economicos*, Serie A (no.6). Fondacion Marcelino Botin, Pedrueca, 1 (Santander).

Foster, S., Garduño, H., Evans, R., Olson, D., Tian, Y., Zhang, W., et al. (2001). Quaternary aquifer of the North China Plain: Assessing and achieving groundwater resource sustainability. *Paper for IAH-Hydrogeology Journal, 12*(1), 81–93. (Special Issue).

Foster, S., Tuinhof, A., Kemper, K., Garduño, H., & Nanni, M. (2003). *Groundwater management strategies: Facets of the integrated approach.* GW-MATE Briefing Note Series, Briefing Note 3, The World Bank, Washington, D.C., http://documents.worldbank.org/curated/en/354221468136800362/Groundwater-management-strategies-facets-ofthe-integrated-approach.

Heckelei, T., Britz, W., & Zhang, Y. (2012). Positive mathematical programming approaches – recent developments in literature and applied modelling. *Bio-based and Applied Economics, 1*(1), 109–124.

Heilig, G. K., Fischer, G., & van Velthuizen, H. (2000). Can China feed itself? An analysis of China's food prospects with special reference to water resources. *International Journal of Sustainable Development and World Ecology, 7,* 153–172.

Howitt, R. E. (1995). Positive mathematical programming. *American Journal of Agricultural Economics, 77*(02), 329–342.

Huang, Q. Q., Rozelle, S., Wang, J. X., & Huang, J. K. (2002). *Irrigation, agricultural performance and poverty reduction in China*. Working Paper, May 2002, Department of Agricultural and Resource Economics, University of California, Davis.

Just, R. E., Zilberman, D., & Hochman, E. (1983). Estimation of multicrop production functions. *American Journal of Agricultural Economics, 65*(4), 770–780.

Kahrl F., Roland-Holst, D., & Zilberman, D. (2005). *New horizons for rural reform in China: Resources, property rights, and consumerism, Giannini foundation of agricultural economics*, http://www.agecon.ucdavis.edu/uploads/update_articles/v9n1_4.pdf.

Kendy, E. (2003). The false promise of sustainable pumping rates. *Technical Commentary Ground Water, 41*(1), 2–4.

Kendy, E., Molden, D. J., Steenhuis, T. S., & Liu, C. M. (2003). *Policies drain the North China Plain: Agricultural policy and groundwater depletion in Luancheng County, 1949–2000* (IWMI Research Report 17). Columbo, Sri Lanka: IWMI.

Liu, C. M. (1998). Environmental issues and the South-North Water transfer scheme. *The China Quarterly, 156*, 899–910 (Special Issue: China's Environment).

Liu, J., & Diamond, J. (2005). China's environment in a globalizing world. *Nature, 435,* 1179–1186.

Lohmar, B., Wang, J. X., Rozelle, S., Huang, J. K., & Dawe, D. (2003). China's agricultural water policy reforms: Increasing investment, resolving conflicts and revising incentives. *Agricultural Information Bulletin Number, 782* (Market and Trade Economics Division, Economic Research Service, U.S. Department of Agriculture, Washington, DC).

Moench, M., Burke, J. & Moench, Y. (2003). *Rethinking the approach to groundwater and food security* (Water Reports 24). Rome: FAO.

Mérel, P.R., Simon, L.K., & Yi, F. (2011). A fully calibrated generalized constant-elasticity-of-substitution programming model of agricultural supply. *American Journal of Agricultural Economics, 93*(4), 936–948. https://doi.org/10.1093/ajae/aar029.

Mérel, P., & Howitt, R. (2014). Theory and application of positive mathematical programming in agriculture and the environment. *Annual Review of Resource Economics, 6*(1), 451–470.

Varis, O., & Vakkilainen, P. (2001). China's 8 challenges to water resources management in the first quarter of the 21st century. *Geomorphology, 41*(2–3), 93–104.

World Bank. (2001a). *China: Agenda for water sector strategy for North China* (Report prepared jointly with Sinclair Knight Merz & Egis Consulting Australia, The General Institute of Water Resources & Hydropower Planning and Design (MWR), The Institute of Water and Hydropower Research (Beijing), The Institute of Hydrology and Water Resources (Nanjing) and the Chinese Research Academy for Environmental Sciences (Beijing), Report No. 22040-CHA). Washington, DC: World Bank.

World Bank. (2001b). China: Air, land, and water. Washington, DC: World Bank.

World Bank. (2002). *China: Country water resources assistance strategy*. Strategy paper, East Asia and Pacific Region, Washington, DC: World Bank.

Yang, Y. H., Watanabe, M., Sakura, Y., Changyuan, T., & Hayashi, S. (2002). Groundwater-table and recharge changes in the Piedmont region of Taihang mountain in Gaocheng city and its relation to agricultural water use. *Water SA (Water Resources Commission of South Africa), 28*(2), 171–178.

Zhang, Q., & Zhang, X. (1995). Water issues and sustainable social development in China. *Water International, 20*(3), 122–128.

Zhang, X. Y., Pei, D., & Hu, C. S. (2003). Conserving groundwater for irrigation in the North China Plain. *Irrigation Science, 21*(4), 159–166.

Zhang, L, Wang, J., Huang, J., Huang, Q., & Rozelle, S. (2010). Access to groundwater and agricultural production in China. *Agricultural Water Management, 97*(10), 1609–1616.

Zhong, H., Sun, L., Fischer, G., Zhan, T., van Velthuizen, H., & Liang, Z. (2017). Mission impossible? maintaining regional grain production level and recovering local groundwater table by a cropping system adaptation across the North China Plain. *Agricultural Water Management, 193*(c), 1–12. https://doi.org/10.1016/j.agwat.2017.07.014

Part III
Application of Information-Theoretic Methods

Chapter 8
Using Moment Constraints in GME Estimation

Richard E. Howitt and Siwa Msangi

Abstract In this contribution, we explore the sensitivity of parameter estimates derived through the generalized maximum entropy (GME) approach under alternative specifications of the width of the error term supports. Although many recommend a "three-sigma" rule for setting the width of this term, there can be noticeable differences in the results if it is expanded beyond that, as others in the literature have suggested. We use a Monte Carlo analysis to see how imposing a moment-based condition into the GME problem, as an additional constraint, affects the results. We find that it removes the sensitivity of the parameter estimates to the width of the supports for the error term and that this remains robust even when the data is ill-conditioned. Based on this, we recommend that researchers impose this condition when doing GME-based estimation, to improve the performance of the estimator.

1 Introduction

The use of information theoretic approaches to statistical estimation and inference has become increasingly widespread in the econometric literature and has proved to be a useful alternative to classical estimation techniques. The generalized maximum entropy (GME) estimation approach, introduced by Golan, Judge and Miller (GJM, hereafter) in their seminal monograph "*Maximum Entropy Econometrics: Robust Estimation with Limited Data*" (1996a) is an extension of Jaynes' original maximum entropy procedure for solving inverse problems (1957a, b, 1984). This method of estimation has made several inroads to applied economic research particularly where the data sample is small or ill-conditioned (Fernandez 1997; Ferreira 2013; Fragaso

R. E. Howitt (✉)
Agricultural and Resource Economics Department,
University of California, Davis, CA 95616, USA
e-mail: howitt@primal.ucdavis.edu

S. Msangi
International Food Policy Research Institute, Washington, DC, USA
e-mail: siwamsangi@gmail.com

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_8

and Carvalho 2012; Gohin 2000; Golan et al. 1996b; Howitt and Msangi 2014; Kaplan et al. 2003; Lence and Miller 1998; Wu 2009; Zhang and Fan 2001).

While many researchers have made use of the basic GME framework offered by GJM, we feel that there has been a misunderstanding in the empirical literature as to the implications of using their basic model formulation. GJM offer three alternative GME formulations—the most basic of which uses just the proposed data-generating process ($y = X\beta + e$) as an informative constraint for each observation, and the other two make use of moments of the data. While GJM give some brief discussions as to the choice of data constraint to use (p. 114), the significance of this choice and its implications for the resulting properties of the estimator seem to have been lost on most researchers. Most of the applied research using GME methods makes use of the basic GJM model, regardless of the sample size, and makes no mention of alternative formulations of the estimating equations (Fraser 2000; Zhang and Fan 2001; Lansink et al. 2001; Leon et al. 1999).

Furthermore, most researchers implicitly assume that the property of large sample consistency properties that GJM demonstrate for their normalized cross-entropy model (GCE-NM) carry over to the commonly used basic GME formulation. However, we have found this not to be the case and that the resulting bias can be remedied if you move beyond the basic GME formulation and use sample moments as estimating equations.

While more recent theoretical econometric research has considered the incorporation of moment-based estimation within the generalized entropy framework (van Akkeren et al. 2002), this has been done only within the context of stochastic regressors and instrumental variable estimation. We show that moment-based estimation is appropriate for the basic GME problem and use the primal formulation of the GME problem to demonstrate this.

GJM use the moment-based formulations to prove that GME estimates are consistent in large samples, but don't mention the potential for bias in small samples, when comparing the properties of the resulting estimates with other estimators (least squares, maximum likelihood, and Bayes estimators). In their discussion of finite sample properties, they use the basic GME formulation and don't compare its results to the moment-based one, citing the potential problems of moment-based estimation in small samples. This contribution arose from an analysis of the small sample bias of GME estimates, but it also shows that the estimates from the basic GJM model formulation are biased in both large and small samples. We show that the inherent large sample bias in the basic GME estimate can be avoided by adding a simple moment constraint to the solution approach.

The rest of the chapter will be organized as follows. In the next section, the solution to the primal GME problem is derived, and the effect of the error support specification on the resulting estimates is shown. The moment constrained basic GME estimator (GME-MC) is shown to eliminate the error support bias in the basic GME estimate and also decouple the resulting GME-MC estimate from the error support values. In the following section, a Monte Carlo example is used to compare the estimates from four methods, namely the method of moments, empirical likelihood, and the GME formulation both with (GME-MC) and without the moment constraints. These initial results suggest that the GME-MC estimator is robust and scalable in that it is unbiased for standard sample sizes, but also substantially outperforms traditional estimators and is equivalent to the GME estimator when the sample is ill-conditioned or ill-posed.

2 Primal GME Estimation

Let us consider a simple data-generating process described by the following equation:

$$y_j = x_j \beta + \varepsilon_j \tag{1}$$

where there are n observations on y and x, which are indexed over the set j. For simplicity, the parameter of interest is a scalar and is estimated by solving the standard primal GME optimization problem. The problem defined in Eq. (2) is exactly equivalent to that defined by GJM in their Eqs. (6.3.1)–(6.3.4).

$$\begin{aligned}
&\underset{\vec{p}^{\,b},\vec{p}^{\,e}}{\text{Max}} - \sum_i p_i^b \ln p_i^b - \sum_j \sum_i p_{i,j}^e \ln p_{i,j}^e \\
&\text{subject to} \\
&y_j = \left(\sum_i p_i^b z_i^b\right) x_j + \sum_i p_{i,j}^e z_i^e \quad \forall j \\
&\sum_i p_i^b = 1, \quad \sum_i p_{i,j}^e = 1 \quad \forall j
\end{aligned} \tag{2}$$

where the β parameter and ε error terms are expressed as expected values and defined over a discretized support space defined by z_i^b and z_i^e and indexed over the set i. The support values for β are restricted to the positive orthant, whereas those for the error term ε can take on both positive and negative values and are symmetric around zero. The range of the support values for e typically satisfies the three-sigma rule suggested by GJM, when applied in the empirical literature (Pukelsheim 1994; Fraser 2000). The Lagrangian associated with this optimization problem can be written as:

$$\begin{aligned}
L = &- \sum_i p_i^b \ln p_i^b - \sum_i \sum_j p_{i,j}^e \ln p_{i,j}^e \\
&+ \sum_j \lambda_j \left[y_j - \left(\sum_i p_i^b z_i^b\right) x_j - \sum_i p_{i,j}^e z_i^e \right] + \mu^b \left[\sum_i p_i^b - 1 \right] \\
&+ \sum_j \mu_j^e \left[\sum_i p_{i,j}^e - 1 \right]
\end{aligned} \tag{3}$$

For simplicity, however, we will consider a simple case where there are only three supports for the parameter and error terms, so that we can rewrite the Lagrangian as:

$$
\begin{aligned}
L = & -p_1^b \ln p_1^b - \left(1 - p_1^b - p_3^b\right) \ln\left(1 - p_1^b - p_3^b\right) - p_3^b \ln p_3^b \\
& - \sum_j \left[p_{1,j}^e \ln p_{1,j}^e + \left(1 - p_{1,j}^e - p_{3,j}^e\right) \ln\left(1 - p_{1,j}^e - p_{3,j}^e\right) + p_{3,j}^e \ln p_{3,j}^e\right] \\
& + \sum_j \lambda_j \left[y_j - \left(p_1^b z_1^b + \left(1 - p_1^b - p_3^b\right) z_2^b + p_3^b z_3^b\right) x_j \right. \\
& \left. - \left(p_{1,j}^e z_1^e + \left(1 - p_{1,j}^e - p_{3,j}^e\right) z_2^e + p_{3,j}^e z_3^e\right)\right]
\end{aligned}
\tag{3a}
$$

The first-order conditions for (3a) are:

$$
\begin{aligned}
\partial L / \partial p_1^b &= -(1 + \ln p_1^b) + \ln\left(1 - p_1^b - p_3^b\right) + 1 - \sum_j \lambda_j x_j \left(z_1^b - z_2^b\right) = 0 \\
\partial L / \partial p_3^b &= -(1 + \ln p_3^b) + \ln\left(1 - p_1^b - p_3^b\right) + 1 - \sum_j \lambda_j x_j \left(z_3^b - z_2^b\right) = 0 \\
\partial L / \partial p_{1,j}^e &= -(1 + \ln p_{1,j}^e) + \ln\left(1 - p_{1,j}^e - p_{3,j}^e\right) + 1 - \lambda_j \left(z_1^e - z_2^e\right) = 0 \quad \forall j \\
\partial L / \partial p_{3,j}^e &= -(1 + \ln p_{3,j}^e) + \ln\left(1 - p_{1,j}^e - p_{3,j}^e\right) + 1 - \lambda_j \left(z_3^e - z_2^e\right) = 0 \quad \forall j
\end{aligned}
\tag{4}
$$

By re-arranging the FOCs for the error terms, we obtain:

$$
\ln\left(1 - p_{1,j}^e - p_{3,j}^e\right) - \ln p_{1,j}^e = \lambda_j \left(z_1^e - z_2^e\right) \quad \forall j \tag{4a}
$$

$$
\ln\left(1 - p_{1,j}^e - p_{3,j}^e\right) - \ln p_{3,j}^e = \lambda_j \left(z_3^e - z_2^e\right) \quad \forall j \tag{4b}
$$

And similarly, the FOCs for the β supports can be rewritten as

$$
\ln\left(1 - p_1^b - p_3^b\right) - \ln p_1^b = \sum_j \lambda_j x_j \left(z_1^b - z_2^b\right) \tag{4c}
$$

$$
\ln\left(1 - p_1^b - p_3^b\right) - \ln p_3^b = \sum_j \lambda_j x_j \left(z_3^b - z_2^b\right) \tag{4d}
$$

By subtracting Eq. (4b) from (4a), we obtain Eq. (5a) as

$$
\begin{aligned}
& \ln p_{3,j}^e - \ln p_{1,j}^e = \lambda_j \left(z_1^e - z_3^e\right) \quad \forall j \\
\text{and} \quad & \lambda_j = \frac{\ln\left(p_{3,j}^e / p_{1,j}^e\right)}{\left(z_1^e - z_3^e\right)} \qquad \forall j
\end{aligned}
\tag{5a}
$$

Similarly, by subtracting Eq. (4d) from (4c), we obtain Eq. (5b) as

$$\ln p_3^b - \ln p_1^b = \ln\left(p_3^b \big/ p_1^b\right) = \left(z_1^b - z_3^b\right) \sum_j \lambda_j x_j \tag{5b}$$

Substituting the expression for λ_j into 5b, we obtain:

$$\ln\left(p_3^b \big/ p_1^b\right) = \left(z_1^b - z_3^b\right) \sum_j \left[x_j \frac{\ln\left(p_{3,j}^e \big/ p_{1,j}^e\right)}{\left(z_1^e - z_3^e\right)} \right]$$

$$\ln\left(p_3^b \big/ p_1^b\right) = \frac{\left(z_1^b - z_3^b\right)}{\left(z_1^e - z_3^e\right)} \sum_j \left[x_j \ln\left(p_{3,j}^e \big/ p_{1,j}^e\right) \right] \tag{6}$$

which further simplifies to

$$\begin{aligned}
\ln\left(p_3^b \big/ p_1^b\right) &= \frac{\left(z_1^b - z_3^b\right)}{\left(z_1^e - z_3^e\right)} \sum_j \ln\left[\left(p_{3,j}^e \big/ p_{1,j}^e\right)^{x_j} \right] \\
&= \frac{\left(z_1^b - z_3^b\right)}{\left(z_1^e - z_3^e\right)} \ln\left[\prod_j \left(p_{3,j}^e \big/ p_{1,j}^e\right)^{x_j} \right] \\
&= \ln\left[\left\{ \prod_j \left(p_{3,j}^e \big/ p_{1,j}^e\right)^{x_j} \right\}^{\frac{\left(z_1^b - z_3^b\right)}{\left(z_1^e - z_3^e\right)}} \right]
\end{aligned}$$

So by taking the anti-log of both sides, we obtain

$$\left(p_3^b \big/ p_1^b\right) = \left[\prod_j \left(p_{3,j}^e \big/ p_{1,j}^e\right)^{x_j} \right]^{\frac{\left(z_1^b - z_3^b\right)}{\left(z_1^e - z_3^e\right)}} \tag{7}$$

Given the supports specified, the ratio of parameter probabilities in Eq. (7) determines the resulting estimation value. Equation (7) shows that the parameter probabilities are a function of the data, the error probabilities, and the ratio of support values for the parameter and the error term. Two papers that examine the effect on GME parameter estimates of changing the error support values are [Preckel (2001) and Paris and Caputo (2001)]. Both papers use Monte Carlo experiments, but come to differing conclusions. Paris and Caputo conclude that GME estimates are sensitive to changes in error support bounds, while Preckel concludes that wide, symmetric error support values result in identical estimates for OLS and GME. It can be seen from Eq. (7) that as the numerical value of the error supports z_1^e and z_3^e increase (note that they have opposite signs) the exponent will go to zero, and the expression for the ratio of the parameter probabilities will go to one. This unit value for the ratio will result in the GME estimate for beta converging on the mean of the parameter support values as the error support bounds are widened. This analytical conclusion is consistent with both the Preckel and Paris and Caputo Monte Carlo results. The Preckel results converge on the unbiased OLS parameter values because the parameter supports are centered on these values, not because of any property of the GME estimator apart from those shown in Eq. (7). An empirical counter example where the parameter supports are centered far from the unbiased coefficient values is shown in the next section. Preckel's results seem to reinforce the general folk wisdom that using wider error supports is "safer" (Paris and Caputo 2001). In contrast, Lence and Miller (1998) conclude that "GME results are not sensitive to changes in the width of the error supports." The results of this contribution support the Paris and Caputo conclusion that GME estimates are influenced by error support bounds, but also show that wider error support values are not "safer" as they introduce greater bias in the resulting GME estimators.

2.1 Moment Constrained GME (GME-MC)

The bias from the error support values and the resulting sensitivity to their bounds can be avoided by solving a simplified GME estimation problem with a moment condition on the errors. The resulting estimator can be termed the GME-MC estimator.

The moment constraint is specified in terms of the previous problem as

$$\sum_{j=1}^{n}\sum_{i=1}^{3} x_j \left(z_i^e p_{i,j}^e\right) = 0, \text{ where} \sum_{i=1}^{3}\left(z_i^e p_{i,j}^e\right) = y_j - x_j \sum_{i=1}^{3} z_i^b p_i^b, \quad \forall j,$$

Substituting for the error terms in the moment condition, the Lagrangian for the GME-MC primal problem can be written as

$$\begin{aligned}\max_{\vec{p}^{\,b}} L = & -p_1^b \ln p_1^b - (1 - p_1^b - p_3^b)\ln(1 - p_1^b - p_3^b) - p_3^b \ln p_3^b \\ & + \lambda \sum_j x_j[y_j - (p_1^b z_1^b + (1 - p_1^b - p_3^b)z_2^b + p_3^b z_3^b)x_j]\end{aligned} \tag{8}$$

where we only solve for the parameter probabilities that maximize the Shannon entropy measure. The FOCs for this reduced problem, simplify to

$$\ln\left(1 - p_1^b - p_3^b\right) - \ln p_1^b = \lambda \sum_j x_j x_j \left(z_1^b - z_2^b\right) \tag{9a}$$

$$\ln\left(1 - p_1^b - p_3^b\right) - \ln p_3^b = \lambda \sum_j x_j x_j \left(z_3^b - z_2^b\right) \tag{9b}$$

After further simplification and taking the anti-log of both sides, we can obtain

$$\frac{(1-p_1^b-p_3^b)}{p_1^b}\Big/\frac{(1-p_1^b-p_3^b)}{p_3^b}=\frac{p_3^b}{p_1^b}=\frac{\exp\left[\lambda\sum_j x_jx_j\left(z_1^b-z_2^b\right)\right]}{\exp\left[\lambda\sum_j x_jx_j\left(z_3^b-z_2^b\right)\right]}$$

$$=\exp\left[\lambda\left(z_1^b-z_2^b\right)\sum_j x_jx_j-\lambda\left(z_3^b-z_2^b\right)\sum_j x_jx_j\right]$$

$$\frac{p_3^b}{p_1^b}=\exp\left[\lambda\left(z_1^b-z_3^b\right)\sum_j x_jx_j\right] \tag{10}$$

A comparison of Eqs. (7) and (10) shows that the GME-MC problem with the embedded moment restriction on the estimated errors gives a solution for the parameter probabilities that only depend on the parameter supports and the data. The shadow value λ is also a function of the parameter supports and the data and is a scalar value. So by imposing moment conditions on the GME errors, we have freed the parameter probabilities—and their resulting estimates—from any dependence on the error supports. Paris (2001) and Marsh and Mittelhammer (2001) have proposed other alternatives to traditional GME procedures in order to address the sensitivity of the parameter estimates to support specification.

2.2 *Bias in GME Estimates*

The explicit dependence of traditional GME estimates on the support specification of the error terms is not the only issue that raises concern when considering the properties of the GME estimator. There is also a potential for bias in the estimate, in both large and small samples, which is best illustrated by considering the result from Eq. (5a) and expressing it in anti-log form as

$$\frac{p_{3,j}^e}{p_{1,j}^e}=\exp\left[\lambda_j\left(z_1^e-z_3^e\right)\right]\quad\forall j\quad\text{or}\quad\frac{p_{1,j}^e}{p_{3,j}^e}=\exp\left[\lambda_j\left(z_3^e-z_1^e\right)\right]\quad\forall j \tag{11}$$

This can be rewritten as the following pair of equations, after multiplying on both sides by the appropriate error support term

$$z_1^e p_{1,j}^e=z_1^e p_{3,j}^e\exp\left[\lambda_j\left(z_3^e-z_1^e\right)\right]\quad\forall j \tag{12a}$$

$$z_3^e p_{3,j}^e=z_3^e p_{1,j}^e\exp\left[\lambda_j\left(z_1^e-z_3^e\right)\right]\quad\forall j \tag{12b}$$

By adding (12a) and (12b) together, we obtain an expression for the expected error $\hat{e}_j$ for the jth observation as:

$$\widehat{e}_j = z_1^e p_{1,j}^e + z_3^e p_{3,j}^e = z_1^e p_{3,j}^e \mathrm{e}^{\lambda_j (z_3^e - z_1^e)} + z_3^e p_{1,j}^e \mathrm{e}^{\lambda_j (z_1^e - z_3^e)} \quad \forall j \tag{13}$$

Note that the symmetry of the error supports implies that $z_2^e = 0$, $\left|z_1^e\right| = \left|z_3^e\right| = \bar{z}^e$ and $z_1^e - z_3^e = 2\bar{z}^e$

So that we can rewrite the expected error expression and sum over all j to obtain

$$\sum_{j=1}^{n} \widehat{e}_j = z_1^e \sum_{j=1}^{n} p_{3,j}^e \mathrm{e}^{-2\lambda_j \bar{z}^e} + z_3^e \sum_{j=1}^{n} p_{1,j}^e \mathrm{e}^{2\lambda_j \bar{z}^e} = \bar{z}^e \left[\sum_{j=1}^{n} p_{3,j}^e \mathrm{e}^{-2\lambda_j \bar{z}^e} - \sum_{j=1}^{n} p_{1,j}^e \mathrm{e}^{2\lambda_j \bar{z}^e} \right] \tag{14}$$

So for the zero bias condition $\sum_{j=1}^{n} x_j \widehat{e}_j = 0$ to hold, it requires that the GME errors satisfy the condition

$$\sum_{j=1}^{n} p_{3,j}^e x_j \mathrm{e}^{-2\lambda_j \bar{z}^e} = \sum_{j=1}^{n} p_{1,j}^e x_j \mathrm{e}^{2\lambda_j \bar{z}^e} \tag{15}$$

Recalling the moment condition that yields unbiased estimates, we imposed on the modified GME problem

$\sum_{j=1}^{n} \sum_{i=1}^{3} x_j \left(z_i^e p_{i,j}^e \right) = 0$ we realize that it requires the following condition

$$\sum_{j=1}^{n} \left[x_j \left(\underbrace{z_1^e}_{=\bar{z}^e} p_{1,j}^e \right) + x_j \left(\underbrace{z_2^e p_{2,j}^e}_{=0} \right) + x_j \left(\underbrace{z_3^e}_{=-\bar{z}^e} p_{3,j}^e \right) \right]$$

$$= \sum_{j=1}^{n} \left[x_j \left(\bar{z}^e p_{1,j}^e \right) - x_j \left(\bar{z}^e p_{3,j}^e \right) \right] = \bar{z}^e \sum_{j=1}^{n} \left[x_j p_{1,j}^e - x_j p_{3,j}^e \right] = 0$$

which in turn implies that the following condition must hold

$$\sum_{j=1}^{n} x_j p_{1,j}^e = \sum_{j=1}^{n} x_j p_{3,j}^e \tag{16}$$

However, the only situation in which (15) and (16) can both hold in the basic GME specification is if the following is true

$$\mathrm{e}^{-2\lambda_j \bar{z}^e} = \mathrm{e}^{2\lambda_j \bar{z}^e} = 1 \quad \forall j \quad \Rightarrow \quad \lambda_j = 0 \quad \forall j$$

But this is an implausible condition, since it implies that the marginal contribution of each observation to the objective function in terms of expected information is zero. We would only expect this to be the case if the data is completely uniformative, and the optimal GME solution is the equiprobable solution that maximizes the entropy and satisfies the adding up conditions on the probabilities.

So we can conclude, from this, that the GME estimation will result in biased parameter estimates—*despite* the symmetry of the error supports—unless we impose an explicit moment condition on the error terms. Notice that this bias is inherent in the basic GME specification and is not changed by an increase in the sample size or by an increase in the width of the support space. Also, this result is unchanged if there are more than three support values for the error, say five, as long as the error support values are distributed around zero. Differences between the number of support values for the errors and parameters likewise do not change the fundamental effect of the support structure on the parameter bias; see Appendix.

By adding the moment constraint to the GME primal problem (2), we obtain a new primal GME-MC problem (17), which satisfies the moment condition $\sum_{j=1}^{n} x_j\left(y_j - x_j\hat{\beta}\right) = 0$ for the resulting estimate. This moment condition arises naturally out of the first-order necessary conditions for maximum likelihood estimation and from the normal equations of ordinary least-squares estimation.

The new GME-MC primal problem is given as

$$\underset{\bar{p}^b,\bar{p}^e}{\text{Max}} - \sum_i p_i^b \ln p_i^b - \sum_j \sum_i p_{i,j}^e \ln p_{i,j}^e$$

subject to

$$y_j = \left(\sum_i p_i^b z_i^b\right) x_j + \sum_i p_{i,j}^e z_i^e \quad \forall j$$

$$\sum_{j=1}^{n} \sum_i x_j \left(p_{i,j}^e z_i^e\right) = 0$$

$$\sum_i p_i^b = 1, \sum_i p_{i,j}^e = 1 \quad \forall j \tag{17}$$

The GME-MC problem bears a strong resemblance to maximum entropy empirical likelihood (MEEL) formulation put forward by Mittelhammer et al. (2000), which serves to extend the empirical likelihood estimation framework (Owen 1988, 1991; Qin and Lawless 1994). The MEEL primal problem can be stated as

$$\underset{\bar{p},\beta}{\text{Max}} - \sum_j p_j \ln p_j$$

subject to

$$\sum_{j=1}^{n} p_j h_m\left(y_j, x_j, \beta\right) = 0 \quad \forall m$$

$$\sum_j p_j = 1 \tag{18}$$

where there are m possible moment conditions $h_m(y_j, x_j, \beta) = 0$, that serve as estimating equations. The Shannon entropy measure serves, in effect, as a criterion function for optimally choosing the weights for the estimating equations, over all observation, in similar fashion to empirical likelihood. Notice, however, that the parameter of interest, β, is not expressed in terms of discretized support values and associated probability weights, as in GME, thereby freeing it of any dependence on defined support values supplied by the researcher. However, this also prevents the researcher from incorporating any prior information into the estimation of β, which is one of the appealing aspects of entropy-based estimation methods. So the modified GME model that we propose in (17) extends the MEEL framework to include an added Shannon entropy measure for choosing the optimal probability weights that define the parameter β in terms of expected value over defined supports. While the resulting estimate of β will ultimately depend on the defined parameter support values, within our estimation framework, they do not depend on the support values for the error terms.

In Chap. 6 of their book, GJM use a dual specification of the more general cross-entropy specification to prove that the resulting estimates are asymptotically consistent. This result has been used to increase the comfort level of practitioners who have applied GME to obtain small and medium sample estimates. While it is theoretically possible to have an estimator with a large sample bias that converges in probability to the true parameters in the limit, the ability to have an estimator that reflects the mean of the data sample and is invariant to the selection of the error support values seems much more comforting for those small sample problems where the GME method has a comparative advantage.

3 An Empirical Example

An empirical test of the analytic development was performed using a Monte Carlo generation of a range of 100–3 observations for a five explanatory variable linear equation. Four different methods were used to recover the distribution from the generated sample. *Maximum likelihood,* empirical likelihood (Owen 1988, 1991 (EL); Qin and Lawless 1994), the method of moments (MM), standard GME, and moment constrained GME-MC.

The observations were generated from a normal distribution with a true mean value of 10 and standard deviation of 1. The support values for the two GME estimations had five values for both the parameter and error distributions. The values for the mean were [−6.0, −3.0, 0.0, 3.0, 6.0] while the supports for the error terms were based on the three-sigma rule [−6.0, −3.0, 0.0, 3.0, 6.0]. The true values for the five coefficients are [1.0, 0.75, 0.5, 0.25, 0.1] note that the mean of the parameter support values (0) is not in the set of the true coefficient values to minimize any "prior" influence on the resulting estimates and to test the assertion that the GME parameter values will tend toward the value zero as the error bounds are increased. The model was run using GAMS (Brooke et al. 1988).

Table 1 Comparison of alternative parameter estimation methods (50 observations)

		Beta 1	Beta 2	Beta 3	Beta 4	Beta 5
True	Betas	1	0.75	0.5	0.25	0.1
OLS	Betas	0.9718	0.68	0.6333	0.1458	0.1491
MM	Betas	0.9718	0.68	0.6333	0.1458	0.1491
EL	Betas	0.9718	0.68	0.6333	0.1458	0.1491
GME-MC	Betas	0.9718	0.68	0.6333	0.1458	0.1491
GME	Betas	0.8834	0.6948	0.6389	0.1507	0.1518

A comparison of the parameter results using different estimators and a sample size of 50 is shown in Table 1.

The GME-MC estimator results in an unbiased estimate of all coefficients and is exactly the same as the benchmark methods of OLS, the method of moments, and empirical likelihood. In contrast, all coefficients from the standard GME estimator differ from the unbiased OLS estimates. Comparison of the average mean squared error of the beta estimates from the generating values shows less difference in the estimators. The parameter mean squared error for OLS and GME-MC is 0.0073. The basic GME is slightly larger at 0.0097.

The selection of error bounds and their influence on the parameter estimates is examined in Table 2. From Eqs. (7) and (10), we would expect that the GME estimate values converge toward the mean of the parameter supports (0) as the error bounds are increased, but that the GME-MC estimates remain unchanged by the error bounds. Table 2 empirically shows this to be the case. The first five increases in error bounds from 3 sigma to 40 sigma are those used by Preckel (2001) who concludes that:

> as the width of the support increases, the entropy-based values approach the least squares-based values

In Preckel's paper, the parameter support space[1] of the GME problem is centered on the OLS estimates of the true parameters. His example shows that his GME estimates converge to the OLS estimates as the width of the support space for the GME errors increases, and seems to suggest that by doing so one reduces the bias in the resulting parameter estimates, that would otherwise cause it to deviate from OLS results. However, the results in Table 3 show that as the error bounds increase, the parameter estimates systematically depart from the OLS values and tend to the

[1] Preckel refers to a "reference distribution" when describing the support space, since he is using the cross-entropy formulation to motivate the similarity between GME and the least-squares estimation procedure, in terms of minimizing deviations. However, the "reference" distribution actually used in his discussion is uniform, which makes the Generalized Cross-Entropy (GCE) minimization problem equivalent to the primal GME maximization problem that we use for our discussion, here. So we will avoid confusion of terminology by solely referring to the support space of the parameter—which Preckel suggests should be chosen by placing the OLS estimate at its center and placing values symmetrically on either side that are equal to multiples of the standard error of the OLS estimate (pp. 370 & 373).

Table 2 Effect of changing error support bounds on estimates (50 observations)

	Beta1	Beta2	Beta3	Beta4	Beta5
True value	1	0.75	0.5	0.25	0.1
OLS	0.9718	0.68	0.6333	0.1458	0.1491
GME					
3 sigma	0.8834	0.6948	0.6389	0.1507	0.1518
5 sigma	0.1303	0.6447	0.6165	0.2364	0.2344
10 sigma	0.0916	0.5297	0.5212	0.2924	0.2896
20 sigma	0.0584	0.3101	0.3267	0.2602	0.2601
40 sigma	0.0282	0.1435	0.1554	0.1345	0.1348
80 sigma	0.0152	0.0766	0.0836	0.0736	0.0738
160 sigma	0.0093	0.0464	0.0508	0.045	0.0451
320 sigma	0.0062	0.0308	0.0337	0.03	0.0301
GME-MC					
3 sigma	0.9718	0.68	0.6333	0.1458	0.1491
5 sigma	0.9718	0.68	0.6333	0.1458	0.1491
10 sigma	0.9718	0.68	0.6333	0.1458	0.1491
20 sigma	0.9718	0.68	0.6333	0.1458	0.1491
40 sigma	0.9718	0.68	0.6333	0.1458	0.1491
80 sigma	0.9718	0.68	0.6333	0.1458	0.1491
160 sigma	0.9718	0.68	0.6333	0.1458	0.1491
320 sigma	0.9718	0.68	0.6333	0.1458	0.1491

mean value of the parameter supports. This shows that Preckel's conclusion results from his selection of the center of his support space and is not a result of the width of the support space.

The results in Table 3 also show that the GME-MC estimator is equal to the OLS under well posed and well-conditioned data. Under these data sets, it performs better than the basic GME estimator. However, when the data is ill-conditioned, but well posed, the GME-MC estimator outperforms OLS and is equal to the GME results.

It seems that the GME-MC moment constraint dominates the resulting estimates when the data set has a sufficient structure to uniquely determine the parameters. However, if the data is collinear the coefficient support values provide the necessary structure and the GME-MC estimates mirror the standard GME results. The performance of our numeric OLS solution using the ill-posed data set (three observations) in Table 3 is currently unexplained.

Table 3 MSE of beta estimates using well- and Ill-conditioned data (symmetric supports, sigma = 1)

Number of Observations	No collinearity			
	OLS	GME-MC	GME	
100	0.0884	0.0884	0.0926	
50	0.0018	0.0018	0.0019	
10	0.0354	0.0354	0.0554	
3	0.1905	0.1903	0.19	
	Collinearity			
	OLS	GME-MC	GME	Condition number
100	4.611E+07	0.0605	0.0621	3.4172E+07
50	1.136E+07	0.055	0.0514	4.1607E+07
10	1.402E+08	0.0542	0.0467	8.5139E+06
3	0.2288	0.2285	0.2281	2.2972E+06

4 Conclusions

We have demonstrated that estimates using the standard GME approach (GJM) can be very different from any consistent estimate of the true distribution parameters if the error bounds are increased beyond a reasonable bound which, in the current empirical literature, remains the three-sigma bound originating from Pukelsheim (1994) and advocated by Golan et al. (1996a, b). On the other hand, adopting a moment-based approach to GME estimation removes this sensitivity to the error support bounds and still allows the GME-MC estimates to remain robust under conditions of high collinearity and sparse observations. In addition, within the GME-MC formulation, the only prior information required is that which defines the supports for the parameter estimates, and this can be based on economic theory and prior empirical studies. In contrast, the GME formulation requires additional prior information in order to specify the error supports. The only cited theoretical basis for specifying these priors is the three-sigma rule, which in turn relies on estimates of sigma that may be unreliable when there is limited data. In addition, this chapter shows that the results in Preckel's paper showing the convergence of standard GME estimates to those from OLS results from his choice of reference distribution for his experiments and is not a general property of standard GME estimators.

We conclude that when making a choice between GME and GME-MC estimators that researchers use the quality of prior information to choose whether to use the GME formulation over the GME-MC moment-based model. The choice should be based on the assessed precision of the researcher's prior information, which is defined in terms of the range of the parameter and error support bounds. Despite the warning offered by Golan et al. (1996a, b) in adopting moment-based GME formulations in empirical applications, we recommend it over expanding the error support bounds that appears to be suggested by Preckel. If remaining consistent with OLS under "close-to-ideal" conditions is of concern to the researcher, then incorporating the moment constraints into the GME-MC problem seems reasonable, as they are in essence the first-order conditions of MLE and the normal equations of OLS. By doing so, one not only achieves equivalence with classical estimation results under "close-to-ideal" conditions, but one is assured of robust estimation results under adverse conditions—which is really where the strength of the GME approach lies. To use a historical analogy, it seems that the GME-MC specification is an "estimator for all seasons."

Appendix

General Derivation for Basic GME Problem

We can write the GME problem as

$$\max -\sum_{i=1}^{2I+1} p_i^b \ln p_i^b - \sum_{j=1}^{N} \sum_{k=1}^{2K+1} p_{jk}^e \ln p_{jk}^e$$

$$\text{s.t.} \quad y_j = x_j \left(\sum_{i=1}^{2I+1} p_i^b z_i^b \right) + \sum_{k=1}^{2K+1} p_{jk}^e z_k^e \quad \forall j$$

$$\text{where} \quad z_k^e \in \left\{ z_1^e, \ldots z_{K+1}^e, \ldots, z_{2K+1}^e \right\}, z_{K+1}^e = 0 \quad \text{and} \quad -z_k^e = z_{2(K+1)-k}^e \quad \forall k \neq K+1$$

which can be rewritten as

$$\begin{aligned} \max - &\left[\sum_{i=1}^{I} p_i^b \ln p_i^b + (1 - \sum_{i \neq I+1} p_i^b) \ln(1 - \sum_{i \neq I+1} p_i^b) + \sum_{i=I+2}^{2I+1} p_i^b \ln p_i^b \right] \\ - \sum_{j=1}^{N} &\left[\sum_{k=1}^{K} p_{jk}^e \ln p_{jk}^e + (1 - \sum_{k \neq K+1} p_{jk}^e) \ln(1 - \sum_{k \neq K+1} p_{jk}^e) + \sum_{k=K+2}^{2K+1} p_{jk}^e \ln p_{jk}^e \right] \end{aligned}$$

s.t.

$$\begin{aligned} y_j = x_j &\left(\sum_{i=1}^{I} p_i^b z_i^b + (1 - \sum_{i \neq I+1} p_i^b) z_{I+1}^b + \sum_{i=I+2}^{2I+1} p_i^b z_i^b \right) \\ &+ \left(\sum_{k=1}^{K} p_{jk}^e z_k^e + (1 - \sum_{k \neq K+1} p_{jk}^e) z_{K+1}^e + \sum_{k=K+2}^{2K+1} p_{jk}^e z_k^e \right) \quad \forall j \end{aligned}$$

Which give the first-order conditions

$$\frac{\partial L}{\partial p_i^b} = -\left(1 + \ln p_i^b\right) + \ln(1 - \sum_{i' \neq I+1} p_{i'}^b) + 1 - \sum_{j=1}^{N} \lambda_j x_j \left(z_i^b - z_{I+1}^b\right) = 0 \quad \forall i \neq I+1$$

$$\frac{\partial L}{\partial p_{jk}^e} = -\left(1 + \ln p_{jk}^e\right) + \ln(1 - \sum_{k' \neq K+1} p_{jk'}^e) + 1 - \lambda_j \left(z_k^e - z_{K+1}^e\right) = 0 \quad \forall j, k \neq K+1$$

taking the $k = 1$ and $k = 2\ K + 1$ cases, we can derive the result $\lambda_j = \frac{1}{z_1^e - z_{2K+1}^e} \ln\left(\frac{p_{j,2K+1}^e}{p_{j,1}^e}\right)$ and substitute it into the following expression derived from the FOCs for p_i^b

$$\ln p_{i'}^b - \ln p_i^b = \sum_{j=1}^{N} \lambda_j x_j \left(z_i^b - z_{i'}^b\right) \quad \forall i, i' \in \mathbf{I} = \{1, \ldots, I, I+2, \ldots, 2I+1\},$$

$$I + 1 \notin \mathbf{I}$$

And we obtain $\frac{p_{i'}^b}{p_i^b} = \left[\prod_{j=1}^{N} \left(\frac{p_{j,2K+1}^e}{p_{j,1}^e}\right)^{x_j}\right]^{\frac{z_i^b - z_{i'}^b}{z_1^e - z_{2K+1}^e}} \quad \forall i, i' \in \mathbf{I} \quad \text{and} \quad I + 1 \notin \mathbf{I}.$

References

Brooke, A., Kendrick, D., & Meeraus, A. (1988). *GAMS: A users's guide*. The Scientific Press.

Fernandez, L. (1997). Estimation of wastewater treatment objectives through maximum entropy. *Journal of Environmental Economics and Management, 32,* 293–308.

Ferreira, P. (2013). An application of general maximum entropy to utility. *International Journal of Applied Decision Sciences, 6*(3), 228–244.

Fragaso, R. M., & Carvalho, M. L. (2012). Estimation of joint costs allocation coefficients using the maximum entropy: A case of mediterranean farms. *Journal of Quantitative Economics, 10*(2), 91–111.

Fraser, I. (2000). An application of maximum entropy estimation: The demand for meat in the United Kingdom. *Applied Economics, 32*(1), 45–59.

Gohin, A. (2000). Positive mathematical programming and maximum entropy: Economic tools for applied production analysis. In *INRA-ESR-Rennes Economics*, Paris.

Golan, A., Judge, G., & Miller, D. (1996b). *Maximum entropy econometrics: Robust estimation with limited data*. New York: Wiley.

Golan, A., Judge, G. & Karp, L. (1996b). A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea using maximum entropy. *Journal of Economic Dynamics and Control* 20(4): 559–582.

Howitt, R. E., & Msangi, S. (2014). Entropy estimation of disaggregate production functions: An application to Northern Mexico. *Entropy, 16,* 1349–1364. https://doi.org/10.3390/e16031349.

Kaplan, J. D., Howitt, R. E., & Farzin, Y. H. (2003). An information theoretic analysis of budget-constrained nonpoint source pollution. *Journal of Environmental Economics and Management, 46*(1), 106–130.

Lansink, A. O., Silva, E., & Stefanou, S. (2001). Inter-firm and intra-firm efficiency measures. *Journal of Productivity Analysis, 15,* 185–199.

Lence, S. H., & Miller, D. J. (1998). Estimation of multi-output production functions with incomplete data: A generalized maximum entropy approach. *European Review of Agricultural Economics, 25,* 188–209.

Leon, Y., Peeters, L., Quinqu, M., & Surry, Y. (1999). The use of maximum entropy to estimate input-output coefficients from regional farm accounting data. *Journal of Agricultural Economics, 50*(3), 425–439.

Marsh, T. L., Mittlehammer, R., & Cardell, N. S. (2001). Generalized maximum entropy analysis of the linear simultaneous equations model. *Entropy, 16*(2), 825–853.

Mittelhammer, R. C., Judge, G. G., & Miller, D. J. (2000). *Econometric foundations*. New York: Cambridge University Press.

Owen, A. (1988). Empirical likelihood confidence ratio confidence intervals for a single functional. *Biometrika, 75*(2), 237–249.

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics, 19,* 1725–1747.

Paris, Q. (2001, April). *MELE: Maximum entropy leuven estimators*. Working Paper 01-003. UC Davis: Department of Agricultural and Resource Economics, UCD. Retrieved from https://escholarship.org/uc/item/66q143ht.

Paris, Q., & Caputo, M. R. (2001, August). *Sensitivity of the GME estimates to support bounds*. Working Paper 01–008. Davis: Dept of Agricultural & Resource Economics, University of California.

Preckel, P. V. (2001). Least squares and entropy: A penalty function perspective. *American Journal of Agricultural Economics, 83*(2), 366–377.

Pukelsheim, F. (1994). The 3-sigma rule. *American statistician, 48*(2), 88–91.

Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics, 22*(1), 300–325.

Van Akkeren, Judge, M. G., & Mittelhammer, R. (2002). Generalized moment based estimation and inference. *Journal of Econometrics* 107 (1–2): 127–148.

Wu, X. (2009). A weighted generalized maximum entropy estimator with a data-driven weight. *Entropy, 11*. https://doi.org/10.3390/e11040917.

Zhang, X., & Fan, S. (2001). Estimating crop-specific production technologies in chinese agriculture: A generalized maximum entropy approach. *American Journal of Agricultural Economics, 83*(2), 378–388.

Chapter 9
Estimating Field-Level Rotations as Dynamic Cycles

Duncan MacEwan and Richard E. Howitt

Abstract Crop rotation systems are an important part of agricultural production for managing pests, diseases, and soil fertility. Recent interest in sustainable agriculture focuses on low input-use practices which require knowledge of the underlying dynamics of production and rotation systems. Policies to limit chemical application depending on proximity to waterways and flood management require field-level data and analysis. Additionally, many supply estimates of crop production omit the dynamic effects of crop rotations. We estimate a dynamic programming model of crop rotation which incorporates yield and cost intertemporal effects in addition to field-specific factors including salinity and soil quality. Using an Optimal Matching algorithm from the Bioinformatics literature, we determine empirically observed rotations using a geo-referenced panel dataset of 14,000 fields over 13 years. We estimate the production parameters which satisfy the Euler equations of the field-level rotation problem and solve an empirically observed four-crop rotation.

1 Introduction

The history of agricultural production over the past 100 years has been one of a steady increase in input intensification and a simplification of crop rotations. An example of this long-term shift toward intensification of inputs can be shown in the nitrogen use statistics in Fig. 1.

Most economic prescriptions for a more sustainable agriculture devolve to input reduction by moving back down the static intensive and extensive margins of adjustment with consequent reductions in productivity. This contribution characterizes the

D. MacEwan · R. E. Howitt (✉)
ERA Economics, Davis, USA
e-mail: howitt@primal.ucdavis.edu

D. MacEwan
e-mail: duncan@eraeconomics.com

R. E. Howitt
Department of Agricultural Economics, University of California at Davis, Davis, USA

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_9

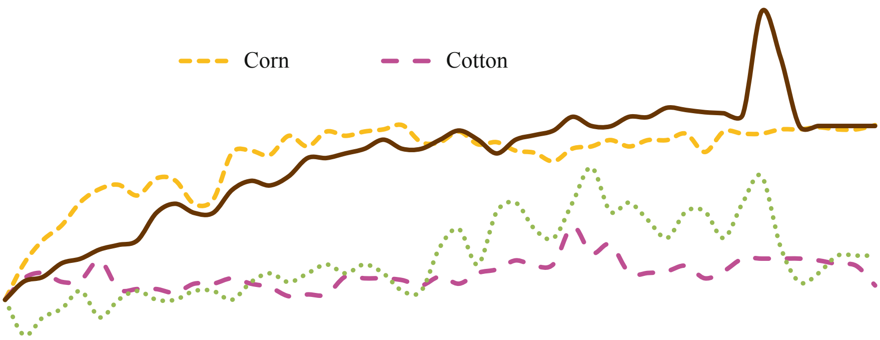

Fig. 1 Average nitrogen application index for US crops, base 1964. *Note* Data compiled from the USDA's fertilizer consumption and use index

potential productivity gains from crop rotations as a third potential adjustment along the dynamic margin and shows how an economic model of rotations can be specified and estimated.

The present interest in sustainable systems is stimulated by recognition that intensified use of agricultural inputs can impose serious external effects on other areas of agricultural production, or environmental effects such as soil erosion, groundwater contamination, or the eutrophication of rivers. Recognition of these impacts makes it clear that the pricing of many agricultural inputs is incomplete and does not reflect the external costs that result from their use.

The reduction of agricultural externalities by reducing the intensification of input use is clearly a movement back along the intensive margin of production, with consequent negative effects on agricultural productivity. In the long term, this is particularly difficult when faced with the increasing needs of agricultural production in many areas to address changing diets and food demands as well as the basic food requirements in developing nations. In addition, the technical change that has simultaneously boosted much of agricultural production is predicted by some studies to be slowing down due to a lack of investment (Alston et al. 2010).

Crop rotations have been an integral part of agricultural production for the past 300 years, but have received scant attention from agricultural economists due to their decline in importance of agricultural production. Practical policies for sustainable agriculture should be based on existing farmer's behavior and response. Thus if rotations are a promising agronomic avenue to a more sustainable system, we need to understand the current behavior and motivation behind input allocation if we are to propose sustainable policies that utilize rotations to a greater extent. It follows that with increasing interest in the full costs of agricultural intensification, there may be a renewal of interest in the economics of agricultural rotations. Accordingly, we are attempting to analyze rotations from an economic perspective and measure the switching costs of rotations as reflected by farmer's decisions. If we can measure and understand switching costs, we will be able to predict whether the adoption of more rotational practices is an effective tool in moving toward a more sustainable agricultural production system.

At this point, we want to draw the distinction between agricultural diversity and agricultural rotations. The positive externalities of rotations in terms of fertility, pest control, and weed control are agronomically dependent on rotating a given field through different crops. A farm may have diverse production without using rotations, and from a single cross section, one cannot identify which fields are being rotated and which are in single-crop diversified production. It is only by using field-level panel data, that rotations can be identified and estimated. This significant data problem is very likely one of the reasons why economists have made so few empirical studies of field-level rotations. A second reason may be that the theoretical structure of the rotation problem is a complex control problem, with a combination of continuous states driven by agronomic and physical processes combined with discreet switch points driven by economic considerations. In addition, the resulting problem must demonstrate the ability to reflect a steady-state cycle of crop rotations, but also be responsive to exogenous shifts in prices, costs, or technology. Clearly, the decision of whether and how many crops to rotate is a joint economic and agronomic question which we term the dynamic margin of agricultural production. Another reason for the lack of research on rotations is that while this interdependence between multi-output production is ubiquitous in agriculture, in most other multi-output production systems it is much less common.

Historically, decisions at the field level of detail have been difficult to observe consistently across time in anything other than experimental plots. We employ a unique geo-referenced panel dataset of field-level production covering over 14,000 fields (over 1 million acres) and 13 years. Using these data, we estimate the observed rotations using an Optimal Matching algorithm from the Bioinformatics literature originally developed for determining common genetic sequences (DNA base pairs). We specify a stochastic-dynamic economic field-level rotation problem and solve the model using dynamic programming. We estimate the parameters of the rotation problem, including yield and cost carry-over effects as well as soil and salinity effects for a four-crop, seven-year, alfalfa–cotton–grain–fallow rotation. We estimate the ill-posed stochastic-dynamic rotation parameters using Generalized Maximum Entropy (GME). We show that the estimated crop rotations are an endogenous result of relative stochastic prices and field physical capital and result in field-level dynamic cycles. We apply the model to estimate the dynamic margin response to changes in groundwater salinity in California's Central Valley and discuss policy implications for moving toward a more salinity-sustainable agriculture.

We want to emphasize four main contributions to the literature on sustainability and crop rotations. In this contribution, we (i) apply a unique remote sensing panel dataset of land and water use, (ii) specify the rotation decision as a stochastic-dynamic programming problem with switching costs, (iii) estimate the economic rotation parameters using Generalized Maximum Entropy, and (iv) quantify the dynamic margin of agricultural supply response. We conclude with an empirical example and estimate grower's response to changes in groundwater salinity in California's San Joaquin Valley.

2 Existing Literature on Crop Rotations

Heady (1948) first formalized the crop rotation problem with a static analysis of the hay–grain rotations observed in the US Corn Belt. Heady (1948) followed work by Johnson (1933), who should be credited as the first to consider the rotation problem. Burt and Allison (1963) formalized the dynamics of rotations in the context of a wheat–fallow rotation. They considered a dynamic programming approach to crop planting decisions, wheat or fallow, in every year depending on the underlying state of the field (soil moisture). Contemporary research on the economics of crop rotations stems from these seminal works and falls into four main areas: (i) linear programming models of production with fixed proportion rotation constraints, (ii) models that lend themselves to econometric analysis and control for lagged crop choice, (iii) dynamic analysis which models crop rotation as a control variable consisting of the proportion of total land use, and (iv) multiple-phase optimal control dynamic models which estimate the switch point between two successive crops.

Linear programming models of production with fixed rotation constraints were introduced to the literature by Hildreth and Reiter (1951). They analyzed a corn–oats–hay rotation in the Corn Belt of the USA and treated specific rotations as individual production processes. Linear programming models impose rotation constraints, in essence fixed proportions, on production activities (Swanson 1956; Peterson 1955; Beneke and Winterboer 1973). El-Nazer and McCarl (1986) built on the previous methodology and specified a set of rotation constraints that made the optimal rotation endogenous. They specified a static model that they hypothesized would satisfy steady-state conditions and, consequently, represent a dynamic solution. The tendency for overspecialization limits linear programming methods and typically requires significant constraints in order to reproduce observed rotation decisions.

Hennessey (2006) formalized the theory behind models that lend themselves to econometric specification for a two-crop rotation. He considered rotation effects through changes in yield or changes in input use in a framework that allows for positive effects without excluding the possibility of negative effects. Other studies that employ a reduced-form econometric specification include Wu et al. (2004) and Wu and Babcock (1998). Both authors specified a reduced-form multinomial logit model to analyze land-use decisions. Tanaka and Wu (2004) evaluated the Conservation Reserve Program in the USA in a similar framework to investigate the effect of taxes on fertilizer, payments for land retirement, and payments on rotations. Other researches analyzed the effects of rotations in the context of environmental concerns including Langpap and Wu (2008), Langpap et al. (2008), Antle and Stoorvogel (2006), and Antle and Valdivia (2006).

The field-level rotation is the solution to a complicated discrete-switching stochastic-dynamic control problem. To recast the problem in a continuous framework, one can view the control variable as proportion of total land use in a region or farm. A continuous control variable, as well as the underlying state equations, makes it such that traditional control theory applies. Jaenicke (2000) and Orazem and Miranowski (1994) apply this framework in the context of dynamic data envelop analysis and fertility carry-over rotation effects, respectively. The central vein of this literature focuses on the optimal fertilizer application rate and dynamic carry-over effect, following the work of Kennedy et al. (1973), Kennedy (1981, 1986), and Taylor (1983). Thomas (2002) developed a structural dynamic model to account for this trade-off and estimated a restricted version of the model using generalized method of moments (GMM).

Dynamic multi-phase optimal control represents a relatively new approach to modeling crop rotations and focuses on the field-level decisions and switch points between crops. The method specifies a set of controls which determine a set of optimal switching times between regimes. The number of regimes (stages) must be exogenously specified in this approach since large state spaces render these models intractable. Doole (2009) provided an algorithm for solving these problems with transition costs, and Doole (2008) provided an application of the algorithm.

Livingston et al. (2012) adopt a hybrid estimation approach that uses field observations and other empirical data to estimate a corn–soybean rotation model. The fundamental framework of their model is similar to our approach, in that the sequence of crops affects profitability and farmers act as dynamically optimizing agents. However, our estimation approaches are very different due, in large part, to differences in available data.

We want to design a framework to estimate the soil quality, salinity, and rotation parameters that the grower faces. In order to do this, we need to specify a structural dynamic model. Structure includes addressing unobservable variables, such as management effects and weather shocks, in addition to specifying functional forms and assumptions about grower actions. In the next section, we review the application region, Kern County, California, and geo-referenced panel dataset. In the following section, we develop the GME estimation framework, add structure to the model, and specify error terms and unobservable variables.

3 An Estimable Model of Field Rotations

In contrast to previous approaches which consider the dynamics of rotations in terms of aggregate land-use proportions, fixed rotation constraints, or lagged crop choice, we explicitly model the discrete switching of field-level decisions subject to continuous underlying agronomic and economic states. We specify fertility and physical capital at the field level and represent rotations as an endogenous cycle resulting from observed behavior of dynamically optimizing farmers. We allow for one-year rotation carry-over effects in both yield and cost.

Consider a farmer managing a specific field within the farm, which can be planted to annual or multi-year crops on a seasonal (annual) basis. The field has a fixed (unit) size and is not sub-divided in any given year. The farmer seeks to maximize the present discounted value of a future stream of profits by choosing the sequence of crops planted every season, i.e., the crop rotation. We allow for one-year carry-over rotation effects and treat the physical capital endowment of the field as fixed and exogenous. We characterize the state of a given field in terms of its aggregate "fertility" that depends on both rotation- and field-specific effects. Rotation effects include pest and disease management and soil fertility, as reflected in the sequence of crops. Field-specific effects include physical capital such as soil quality and shallow groundwater salinity, both of which are known to affect yield. Furthermore, the optimal rotation decision may be affected by farmer expectations for stochastic crop prices and exogenous weather shocks.

Formally, consider the management of a single field, of unit size, in a multi-crop rotation system. We define the two sets of crops as the current crop i and k in the previous time period that includes activities such as fallow and perennial crops. We denote the crop-specific prices as p_{ti} for crop i at time t. We define yields as y_i and variable costs of production as F_i where y_i^* and F_i^* represent the average yield and average variable cost. Note that this model allows for prices to change over time but assumes stationary yields and production costs.

Yield variation may stem from weather shocks, changes in water supply, management effects, and other factors which we define as ν_i. Prices and yields are typically correlated; however, we specify these as independent processes in order to keep an estimable dynamic model. To account for yield shocks, define the yield variance as σ_i^2 and let $\nu_i \sim N(0, \sigma_i^2)$. We allow for stochastic yields in the estimation framework; the dynamic programming model only allows for stochastic crop prices as a first-order Markov process.

Crop rotation affects both crop yields and production costs. Yield effects result from changes in soil fertility and management of pests and disease cycles. Cost effects stem from soil fertility management which may reduce fertilizer and other chemical costs. We introduce parameters $\Gamma_{i|k}$ and $\Psi_{i|k}$ to represent the yield and cost carry-over effects, respectively. In general, these parameters may be functions that represent the effect of planting crop i today, given that crop k was planted in the previous period. Furthermore, this framework can accommodate multiple-year carry-over effects if we define the set k to include the relevant crop planting sequence history. Accounting for yield and cost rotation effects and stochastic crop yields, we can define field profits at any time t as

$$\pi_{ti} = p_{ti}\big((y_i + \nu_i) - \Gamma_{i|k}\big) - \big(F_i - \Psi_{i|k}\big). \tag{1}$$

Farmers form expectations about future prices in order to make current production decisions in addition to considerations of rotation effects and field-specific physical capital. We model farmer's price expectations as a stationary first-order Markov process. Thus, current prices and price transition probabilities completely describe the future period price expectations. Formally,

$$\Pr\big(p_{t+1} = p' | p_t = p''\big) = \Pr\big(p_t = p' | p_{t-1} = p''\big). \tag{2}$$

Intertemporal cost and yield rotation effects are, in general, crop-specific functions which depend on the relevant crop history of the field. To write down an estimable model, we consider rotation effects as deviations from the average yield and cost with a one-year memory. As such, $\Gamma_{i|k}$ and $\Psi_{i|k}$ represent i x k matrices of rotation effects. On any given field, yield and cost equal the mean within a region, plus or minus a rotation adjustment effect. Rotation effects may be positive or negative, depending on the relationship between the crop planted in the current and previous period. For example, cotton and grain extract a relatively large amount of nutrients from the soil whereas alfalfa replaces soil nutrients. As such, a cotton–grain rotation may increase production costs and decrease mean yield, or both, whereas an alfalfa–cotton (or grain) rotation may have the opposite effect.

Yields and production costs may vary with changes in physical capital. For example, poor soil quality requires more intensive input use and management and, therefore, higher production costs. We allow physical capital to affect crop yields but not crop prices in order to keep the model tractable. We introduce two coefficients to capture the effects of salt and soil on yields, β_i^1 and β_i^2, which represent the marginal effects of salt (ec) and soil (sl) on mean crop yield, respectively. We will assume these effects are stationary and unaffected (directly) by crop rotation such that we can write crop yield as

$$y_i = \left(y_i^* + \nu_i\right) - \Gamma_{i|k} - \beta_i^1 \cdot \text{ec} - \beta_i^2 \cdot \text{sl}. \tag{3}$$

Combined we can write the current period profits on any given field of unit size as

$$\pi_{ti} = p_{ti}\left((y_i + \nu_i) - \Gamma_{i|k}\right) - \left(F_i - \Psi_{i|k}\right) - \beta_i^1 \cdot \text{ec} - \beta_i^2 \cdot \text{sl}. \tag{4}$$

This equation fundamentally describes the rotation decision between two crops at any point in time, given field physical capital. We use this equation to estimate the parameters of the model based on observed rotation (switching) decisions using a panel dataset of land-use and production data, discussed in the following section.

3.1 Stochastic-Dynamic Programming Model of Field Rotations

To transform the rotation problem into a dynamic programming framework, we need to specify the evolution of state variables and the nature of rotation effects. We slightly redefine the i, k notation in order to clearly define the dynamic programming problem. Let c_t denote planting crop c in period t for $t = 1, 2, \ldots, T$ on a field of unit size. Consider discrete time and assume an infinite time horizon. Let s_t be the state variable that represents the underlying rotation state (fertility) of the field, which depends solely on a function of the crop planted in the previous period, $s_t = g(c_{t-1})$. Finally, let δ signify the discount factor of the farmer and allow average crop yields to be deterministic. We can write the farmer dynamic programming problem as follows.

$$\max_{c_t} \sum_{t=0}^{T} \delta^t \big\{ p_t(c_t)\big(y^*(c_t) - \Gamma(c_t, s_t) - \beta^1(c_t)\cdot \mathrm{ec} - \beta^2(c_t)\cdot \mathrm{sl}\big) - \big(F^*(c_t) - \Psi(c_t, s_t)\big)\big\} \tag{5}$$

subject to Eq. (2) and

$$s_{t+1} = g(c_t). \tag{6}$$

Under this formulation, we can derive the Bellman equation which becomes a vector fixed-point equation in the value function. Bertsekas (1976) showed that if the discount factor equals less than one, then the mapping underlying the Bellman equation is a strong contraction on Euclidean space. Consequently, the Contraction Mapping Theorem guarantees the existence and uniqueness of the solution value function. In other words, the dynamic programming formulation of the rotation problem has a theoretically guaranteed unique solution which we can find by solving for the fixed point of the Bellman equation.

4 Kern County Geo-referenced Panel Data

Kern County, California, is located at the southern end of the San Joaquin Valley and produced over $5.3 billion in gross value of agriculture in 2011. The top grossing commodities include milk, almonds, grapes, citrus, carrots, alfalfa, cotton, and tomatoes. The top crops by acreage include cotton, alfalfa, wheat, and almonds. Crops are primarily irrigated with water coming from state and federal surface water projects in addition to local surface supplies and groundwater. The data we have compiled include all irrigated agricultural land in Kern County between 1997 and 2009. On each field and year, we observe the crop grown, field size in acres, farm owner, and farm manager of the field. We are able to uniquely identify and track fields across time using a geo-referenced dataset provided by the Kern County Agricultural Commissioner's Office.

We observe physical characteristics of each field from data that we aggregate up to the field level. Soil data are from the United States Department of Agriculture (USDA) Soil Survey Geographic Database (SSURGO). Since soil type is unchanged from year to year over the time horizon of the data, we take a cross section from 2002. The data are geo-referenced, provided in polygon layers, and include seven classifications for agricultural uses, developed by USDA. We use the Soil Capability Class Index polygon layer and the tools in ArcGIS to estimate the dominant soil class for each field in the sample. Shallow groundwater salinity data are from a 2002 survey analysis completed by the California Department of Water Resources (DWR). DWR surveyed salinity levels, measured in electrical conductivity (mS/cm), over a sample grid in California's Central Valley and created a polygon layer map of salinity levels. The average depth to groundwater varies across the Central Valley;

the sample average reported by DWR was 3 m. We use the DWR survey polygon layers to determine the salinity level at each field in the sample.

We additionally observe actual crop evapotranspiration (ET) and dry biomass production on a 30-by-30-m scale for 2002, 2005, 2008, and 2009. These data are provided as monthly and seasonal raster layers by SEBAL North America. The Surface Energy Balance Algorithm for Land (SEBAL) is an algorithm which uses LANDSAT thermal images and a series of energy balance equations to estimate actual crop ET and dry biomass on a 30-by-30-m pixel scale (Thoreson et al. 2009). We use the standard tools in ArcGIS to smooth the seasonal ET and biomass raster layers and remove pixel observations near field boundaries. We average the remaining pixels within each field to estimate average ET and biomass for each field in the sample. We augment the geo-referenced field data with economic data from the Kern County Agricultural Commissioner's Office. The Kern Agricultural Commissioner's Office conducts farmer surveys and consults with county experts and extension agents in order to estimate county average prices, yields, and input costs. We include Kern County average crop price, yield, input use, and input costs for 1997–2009.

GIS field data in Kern County enable us to identify crop rotations that are directly observable as crops change on the same field across years and consist of grains, cotton, corn, and processing tomatoes.

4.1 Rotation Identification—Optimal Matching

In order to specify and estimate a dynamic programming model of rotations, we need to know common rotation practices. Kern County produces vegetables, grains, cotton, forage crops, grapes, citrus, and nuts, in addition to a range of other crops. For rotation systems with two crops, such as corn and soybeans, even with a multiple-year lag effect there are a limited number of potential rotation systems. Rotation systems in Kern County and other diverse regions will likely include multiple crops and depend on a number of factors such as field characteristics, farmer knowledge, micro-climate, pests and disease, and inputs. We need to identify some common rotation sequences in order to specify and estimate the rotation problem.

We can think of a rotation as a subsequence of crops on a given field, and we want to identify common subsequences across all fields. This is a parallel problem to identification of common DNA subsequences, such as single-nucleotide polymorphisms, in the genetic sciences. To apply this approach, we aggregate crops into 20 groups according to DWR Crop Group Classification used for land use and planning in California. After excluding perennial crops and groups not produced in Kern County, we are left with ten groups. We use a Sequence Analysis algorithm to determine common subsequences across fields in Kern County and define the identified sequences as base crop rotations.

Sequence Analysis (SA) is a branch of research within the field of Bioinformatics that identifies sequences of amino acids and DNA base pairs. Needleman and Wunsch (1970) were the first to consider the problem and developed an algorithm based on the principles of dynamic programming to estimate common subsequences between two sequences. A sequence can be transformed using insertion (additional elements), deletion (removing elements), or gap (adding/subtracting breaks). Each of these transformations has an associated cost, similar sequences or subsequences have a low (or zero) cost of transformation. Given a range of possible transformations, it becomes a dynamic programming problem to estimate the minimum distances and identify similar subsequences. Subsequent to Needleman and Wunsch's work, the method has been expanded to include Multiple Sequence Analysis where similarities are identified across and within multiple sequences. Other fields that use similar algorithms include finance, string editing, and language processing.

We employ a version of a SA algorithm called Optimal Matching in order to empirically identify crop rotations. We use the package, SQ-Ados, developed in Stata by Brzinsky-Fay et al. (2006) with the default substitution and insertion/deletion costs (2 and 1, respectively). We use a sequence "suppression" option that condenses multiple sequential crops, of the same type, into a single observation and identifies commonalities across reduced-form sequences. We take this approach because rotations are a dynamic process, subject to external shocks, and we model the underlying process that results in multiple years in the same crop. Economic and agronomic considerations such as price expectations and heterogeneity in land characteristics will affect rotation decisions. Table 1 summarizes the aggregate data, and Table 2 summarizes the results. We report the five most common sequences in Table 2.

We will focus on the base rotation of alfalfa–cotton–grain–fallow in the rest of this chapter.

Since alfalfa is a perennial crop, we use satellite data to estimate the mean yield in any given year and treat different years of alfalfa as different crops. Specifically, we allow for four years of alfalfa and estimate the mean yield of a field at any point in the four-year sequence. We use SEBAL satellite data to identify the mean alfalfa

Table 1 Summary of rotations, by field, in Kern County

Fields in top 20 rotation	4500
Total fields in annual crops (plus Alfalfa)	7939
Total fields in perennials	6290
Total fields	14,229

Table 2 Summary of top five common rotation systems in Kern County

Rotation system	Number of fields	Percent of total
Alfalfa–cotton–grain–fallow	963	12.14
Alfalfa–corn–grain	479	6.04
Vegetable–grain	357	4.50
Alfalfa–grain	356	4.49
Alfalfa–corn–cotton–grain	353	4.45

Table 3 Conditional rotation statistics for Kern County 2000–2009

Base crop	Rotation crops	# of fields	% of total	Base crop	Rotation crops	# of fields	% of total
Cotton		4424		Alfalfa		3459	
	Monoculture	131	2.96		Monoculture	16	0.46
	Alfalfa, grain	534	12.07		Cotton, grain	534	15.44
	Corn, alfalfa, grain	353	7.98		Corn, grain	479	13.85
	Veg, grain	328	7.41		Grain	356	10.29
	Alfalfa	297	6.71		Corn, cotton, grain	353	10.21
	Alfalfa, veg, grain	208	4.70		Cotton	297	8.59
	Grain	194	4.39		Veg, cotton, grain	208	6.01
Fallow		2425		Grain		5240	
	Monoculture	226	9.32		Monoculture	60	1.11
	Veg, grain	357	14.72		Alfalfa, cotton	534	10.19
	Veg	286	11.79		Alfalfa, corn	479	9.14
	Alfalfa	226	9.32		Veg	357	6.81
	Grain	215	8.87		Alfalfa	356	6.79
	Cotton, grain	194	8.00		Alfalfa, corn, cotton	353	6.74
	Veg, grain	130	5.36		Veg, cotton	328	6.26

yield by field for 2002 and the Kern geo-referenced land-use data to determine the age of the stand.

The SA algorithm does not differentiate between the relative numbers of sequences; only the most common sequences are reported. For example, cotton, alfalfa, and grain are the most commonly observed crops in Kern County. However, vegetables are likely part of an important rotation even though they are only observed on a small portion of fields. SA on the subsets of fields that are observed to be in a given crop at any point in the data illustrates what we term “conditional” rotation statistics. In other words, the statistics show other crops that are grown on a field, conditional on a specific crop being produced. Table 3 summarizes the results.

For example, we observe cotton at least once on 4242 fields in Kern County between 2000 and 2009. Cotton is grown in monoculture on 131 fields and rotated with alfalfa on 297 fields over the 2000–2009 data.

5 Parameter Estimation

We observe marginal rotation decisions in terms of switching between crops on a set of n fields in addition to average (county-wide) data on prices, yields, and costs. Our estimation strategy is to observe the field-level rotation, as a sequence of discrete switches, from which we infer the parameter values that farmers are responding to, given they behave according to the model specified above. We use the data to model farmer rotation decisions in terms of deviations from the mean yield and costs, depending on the dynamic sequence of crops planted.

We estimate yield variance and Markov price transition probabilities from empirical data. We use the County Agricultural Commissioner time series data, from 1980 to 2009, to estimate the yield variance for alfalfa, cotton, and grain. We additionally use these data to estimate price state transition probabilities and allow for eight price states for each crop. We select the eight states as the high and low observed between 2000 and 2009 (the range of our field observations) and the evenly distributed percentiles for the other six states. The soil state of the field is allowed to take six values, corresponding to the USDA index described in the previous section. Finally, the salinity state is allowed to take eight discrete values of 0, 3, 5, 7, 11, 13, 16, and 30, corresponding to min, max, and percentile values.

Farmers typically plant alfalfa for four years, thus the base rotation sequence that we identified with the SA algorithm is alfalfa1–alfalfa2–alfalfa3–alfalfa4–cotton–grain–fallow. It follows that the model we have outlined has 106 parameters. After imposing restrictions on second-, third-, and fourth-year alfalfa, the dynamic model reduces to 54 parameters. These include 4 by 1 vectors, β^1 and β^2, of the crop-specific soil and salinity yield effects. We expect β^1 and β^2 to have positive signs because decreasing soil quality and increasing shallow groundwater salinity decrease yields. The 7 by 7 parameter matrices Ψ and Γ represent cost and yield carry-over effects due to crop rotation, respectively. The i, k entry of each matrix represents the yield or cost effects from planting crop i today given that the farmer planted crop k in the previous year. We anticipate that these parameters can take any sign, representing both positive and negative agronomic effects from rotating crops.

The model we specified in the previous section yields a of 42 Euler equations that define the base alfalfa–cotton–grain–fallow rotation. Importantly, note that multiple years in a single crop or other variations on this rotation system can occur. Variations on the base rotation system result from economic factors including changes in relative prices, costs, resource constraints, or changes in field-specific physical capital, which we will simulate in the following section.

With 54 parameters and 42 equations, the estimation problem is underdetermined and standard econometric techniques do not apply. We could impose restrictions to reduce the number of unknown parameters; however, this would lead to a miss-specified model because we need all of the cross-crop rotation effects to satisfy the Euler equations. Cross-crop rotation effects become important when we allow relative crop prices to change in the dynamic programming model policy simulations.

Generalized Maximum Entropy (GME) provides an estimation procedure that can handle underdetermined problems (Mittelhammer et al. 2003). Given that we have incomplete observations about a statistical process, an information-theoretic consistent method to recover parameters for inference is to impose probabilistic structure on the model in such a way that it is consistent with observed data and imposes as little additional information as possible. This concept represents an extension of Laplace's "principal of insufficient reason." Subsequent to Shannon (1948) various researchers have expanded and extended the entropy estimation procedure (Kullback 1959; Levine 1980; Csiszar 1991; Skilling 1989). We introduce estimation error terms and write the estimation equations as

$$\pi_{n,c|a4} \geq \pi_{n,ia4} + \varepsilon_j \quad \text{for all} \quad i \neq c \tag{7}$$

$$\pi_{n,g|c} \geq \pi_{n,i|c} + \varepsilon_j \text{ for} \quad \text{all} \quad i \neq g \tag{8}$$

$$\pi_{n,\,f|g} \geq \pi_{n,\,i|g} + \varepsilon_j \quad \text{for all} \quad i \neq f \tag{9}$$

$$\pi_{n,a1|f} \geq \pi_{n,i|f} + \varepsilon_j \quad \text{for all} \quad i \neq a1 \tag{10}$$

$$\pi_{n,a2|a1} \geq \pi_{n,i|a1} + \varepsilon_j \quad \text{for all} \quad i \neq a2 \tag{11}$$

$$\pi_{n,a3|a2} \geq \pi_{n,i|a2} + \varepsilon_j \quad \text{for all} \quad i \neq a3 \tag{12}$$

$$\pi_{n,a4|a3} \geq \pi_{n,i|a3} + \varepsilon_j \quad \text{for all} \quad i \neq a4 \tag{13}$$

In order to estimate the parameters using GME, we need to transform the basic problem from one of finding specific parameters, as defined above, to one of finding probability weights over parameter-specific support values. Let the convex sets described by Eqs. 7–13 define the set of relevant possible parameter solutions.

$$\vartheta^{\gamma} = \left\{ \tilde{\gamma}_{k,l} : \tilde{\gamma}_{k,l} = \sum_{i=1}^{s} \pi_i^{\gamma_{k,l}} \gamma_{k,l}^{i}, \pi_i^{\gamma_{k,l}} \geq 0, \sum_{i=1}^{s} \pi_i^{\gamma_{k,l}} = 1 \right\} \quad \text{for all} \quad k, l, \tag{14}$$

$$\vartheta^{\psi} = \left\{ \tilde{\psi}_{k,l} : \tilde{\psi}_{k,l} = \sum_{i=1}^{s} \pi_i^{\psi_{k,l}} \psi_{k,l}^{i}, \pi_i^{\psi_{k,l}} \geq 0, \sum_{i=1}^{s} \pi_i^{\psi_{k,l}} = 1 \right\} \quad \text{for all} \quad k, l, \tag{15}$$

$$\vartheta^{\beta^1} = \left\{ \tilde{\beta}_{k,l}^{1} : \tilde{\beta}_{k,l}^{1} = \sum_{i=1}^{s} \pi_i^{\beta_k^1} \beta_{k,l}^{1,i}\,, \ \pi_i^{\beta_k^1} \geq 0\,, \ \sum_{i=1}^{s} \pi_i^{\beta_k^1} = 1 \right\} \quad \text{for all } k, \tag{16}$$

$$\vartheta^{\beta^2} = \left\{ \tilde{\beta}_{k,l}^{2} : \tilde{\beta}_{k,l}^{2} = \sum_{i=1}^{s} \pi_i^{\beta_k^2} \beta_{k,l}^{2,i}, \pi_i^{\beta_k^2} \geq 0, \sum_{i=1}^{s} \pi_i^{\beta_k^2} = 1 \right\} \quad \text{for all } k, \tag{17}$$

Thus, given that the solution to the underdetermined system is restricted by Eqs. 14–17 as the convex hull of the respective support values, where s represents the number of discrete support values for each parameter. We define the unknown parameters as

$$\sum_{i=1}^{s} \pi_i^{\gamma_{j,k}} \gamma_{j,k}^{i}, \tag{18}$$

$$\sum_{i=1}^{s} \pi_i^{\psi_{j,k}} \psi_{j,k}^{i}, \tag{19}$$

$$\sum_{i=1}^{s} \pi_i^{\beta_k^1} \beta_k^{1,i}, \tag{20}$$

and

$$\sum_{i=1}^{s} \pi_i^{\beta_k^2} \beta_k^{2,i}. \tag{21}$$

Similarly, let the error terms lie in a convex set,

$$\vartheta^{\varepsilon_j} = \left\{ \tilde{\varepsilon}_j : \tilde{\varepsilon}_j = \sum_{i=1}^{s} \pi_i^{\varepsilon_j} \varepsilon_j^i, \pi_i^{\varepsilon_j} \geq 0, \sum_{i=1}^{s} \pi_i^{\varepsilon_j} = 1 \right\} \text{ for all } j, k, \tag{22}$$

and we write the unknown error terms as

$$\sum_{i=1}^{s} \pi_i^{\varepsilon_j} \varepsilon_j^i. \tag{23}$$

We have transformed the problem from finding the parameters into finding the solutions for the parameter-specific probability weights $\pi_1, \ldots, \pi_s$ which define the convex combinations of the parameter-specific support values. Restricting solutions to probability weights over a finite support space may seem like a strong assumption; however, the parameters of the rotation problem naturally conform to a reasonable support range because of agronomic conditions. The transformed problem is to choose the parameter-specific probability weights (π) over the parameter-specific support values (s). However, the problem is still underdetermined as there are an infinite number of probability distributions which satisfy the Euler equations and probability distribution requirements. GME offers a solution procedure which chooses the maximally uninformative probability weights over each parameter's support values subject to the known data constraints (Euler equations). Mittelhammer et al. (2003) (pg. 23) showed that a unique solution to the GME problem exists and that one can solve for it numerically. Intuitively, GME maximizes a concave function subject to a compact constraint set, which guarantees a unique solution.

Table 4 Soil, salinity, cost ($/ac), yield (tons/ac), and error support values for GME program

Crop	Avg. yield	Soil (%)	Salinity (%)	Yield (%)	Avg. variable cost	Cost (%)	Avg. revenue (Total)	Error (%)
Alfalfa 1	8.16	±35	±35	±50	152	±50	1016	±100
Alfalfa 2	8.30	±35	±35	±50	152	±50	1016	±100
Alfalfa 3	7.40	±35	±35	±50	152	±50	1016	±100
Alfalfa 4	6.90	±35	±35	±50	152	±50	1016	±100
Cotton	0.67	±35	±35	±50	441	±50	1016	±100
Grain	2.72	±35	±35	±50	257	±50	1016	±100

The GME program maximizes the cumulative entropy measure over all the parameter distributions

$$\max\left\{\begin{array}{l} -\sum_i\sum_k\sum_l \pi_i^{\gamma_{k,l}}\ln(\pi_i^{\gamma_{k,l}}) - \sum_i\sum_k\sum_l \pi_i^{\psi_{k,l}}\ln(\pi_i^{\psi_{k,l}}) - \sum_i\sum_k \pi_i^{\beta_k^2}\ln(\pi_i^{\beta_k^2}) \\ -\sum_i\sum_k \pi_i^{\beta_k^1}\ln(\pi_i^{\beta_k^1}) - \sum_i\sum_j \pi_i^{\varepsilon_j}\ln(\pi_i^{\varepsilon_j}) \end{array}\right\} \tag{24}$$

by choosing $\pi_i^{\gamma_{j,k}}$, $\pi_i^{\psi_{j,k}}$, $\pi_i^{\beta_k^1}$, $\pi_i^{\beta_k^2}$, and $\pi_i^{\varepsilon_j}$ subject to the Euler conditions, 7–13, with Eqs. 18–21 and 23 substituted in for parameter values. Additionally, we have to add two more sets of constraints that ensure that the probabilities sum to one, and the probability weights have a positive signs.

The objective function is concave and the constraint set is compact, thus a global maximum to the problem exists. The probability weights over the support values for each of the parameters that minimize additional information subject to the data constraints form the solution.

5.1 Generalized Maximum Entropy Estimation Results

For the rotation problem, we set $s = 5$ and use a truncated uniform support space for each of the parameters. The parameters we estimate include the rotation yield and cost effects, salinity effects, soil effects, and the first-order condition error terms. Thus, each parameter has a set of five support values for which the GME program will choose the optimal respective probability weights. The support space is based on agronomic priors and summarized in Table 4.

Soil, salinity, and yield rotation effects have a support space (in percentage terms) around the average yield as specified in columns 3, 4, and 5, respectively. Cost rotation effects have a support space of plus or minus 50% of average variable cost.

Table 5 Salt and soil parameter estimates (standard errors in parentheses)

Salinity yield parameters						
ALF1	ALF2	ALF3	ALF4	COT	GRN	FAL
0.6750	0.6750	0.6750	0.6750	0.0576	0.2305	n/a
(0.000)	(0.00)	(0.00)	(0.000)	(0.030)	(0.110)	.
Soil yield parameters						
ALF1	ALF2	ALF3	ALF4	COT	GRN	FAL
0.1197	0.1197	0.1197	0.1197	0.0120	0.0386	n/a
(0.000)	(0.000)	(0.000)	(0.000)	(0.011)	(0.002)	.

We solve the GME program in the General Algebraic Modeling Software (GAMS) using the CONOPT3 nonlinear solver and bootstrap standard errors. Table 5 summarizes the marginal effect of salinity and soil on average yields.

Parameters are interpreted as the marginal adjustment in tons per acre to mean yield due to a one unit change in salinity or soil quality. Salinity is measured in dS/m and soil is by SSURGO definitions, as discussed previously. The estimated marginal effect of salinity on crop yield is consistent with the literature. Namely, alfalfa is relatively salt-intolerant and cotton and grain are more salt-tolerant. The estimated marginal effects, in percentage terms, reflect this agronomic information with alfalfa realizing the largest yield decrease.

Parameter estimates for the rotation adjustment effects for costs and yield are reported in Tables 6 and 7.

An entry in the matrix is interpreted as given that crop (column) was planted last period the marginal change in costs/yield relative to the average if crop (row) is planted this period. Parameter estimates are based on farmer behavior and, as such, should be interpreted as the implied yield and cost rotational adjustments based on observed farmer behavior. Entries denoted with "n/a" represent imposed restrictions.

The elements below the main diagonal (and in the top right corner) in Table 6 represent the key cost carry-over effects in the rotation problem. Alfalfa year 1 through alfalfa year 4, cotton, grain, and then fallow represents the base sequence of crops in the rotation is. Then the cycle repeats. When grain follows cotton, it translates into an average cost savings of \$51.40 per acre. For example, the UC Davis Integrated Pest Management Web site recommends rotating cotton with grains to control nematodes and seedling diseases. This, in turn, translates into reduced chemical applications and cost savings in the following season. When cotton follows cotton, our estimates imply an increase in variable production costs of \$88.20 per acre. Part of this cost increase may stem from additional chemical costs to control nematodes and seedling diseases. We also know that grains replenish soil organic content back into the soil when they follow cotton in a rotation. Thus, some of the cost increase from planting cotton after cotton likely comes from reduced soil fertility, which requires additional fertilizer application.

Table 6 Estimated effect of rotation on costs (standard errors in parentheses)

Cost adjustment parameters							
	ALF1	ALF2	ALF3	ALF4	COT	GRN	FAL
ALF1	14.73	15.59	16.32	14.26	22.83	14.03	−17.45
SE	(0.113)	(0.024)	(0.011)	(0.000)	(0.393)	(0.020)	(0.010)
ALF2	−17.60	n/a	n/a	n/a	n/a	n/a	n/a
SE	(0.005)	.	.	.	.	.	.
ALF3	n/a	−21.85	n/a	n/a	n/a	n/a	n/a
SE	.	(0.711)	.	.	.	.	.
ALF4	n/a	n/a	−26.83	n/a	n/a	n/a	n/a
SE	.	.	(0.006)	.	.	.	.
COT	51.64	61.06	77.62	−58.30	88.20	94.50	64.54
SE	(0.000)	(0.130)	(0.045)	(0.056)	(0.000)	(0.002)	(0.004)
GRN	29.50	29.55	29.55	29.50	−51.40	55.20	30.72
SE	(0.000)	(0.000)	(0.003)	(0.001)	(0.002)	(0.000)	(0.000)
FAL	23.00	23.00	23.00	23.00	23.00	−200.00	23.00
SE	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Note This table is read as crop (row) follows crop (column)

Table 7 Estimated effect of rotation on yields (standard errors in parentheses)

Yield adjustment parameters							
	ALF1	ALF2	ALF3	ALF4	COT	GRN	FAL
ALF1	−0.90	−1.10	−1.11	−0.72	−1.20	−0.61	1.57
SE	(0.003)	(0.004)	(0.007)	(0.000)	(0.000)	(0.000)	(0.009)
ALF2	1.02	n/a	n/a	n/a	n/a	n/a	n/a
SE	(0.004)	.	.	.	.	.	.
ALF3	n/a	1.19	n/a	n/a	n/a	n/a	n/a
SE	.	(0.001)	.	.	.	.	.
ALF4	n/a	n/a	1.19	n/a	n/a	n/a	n/a
SE	.	.	(0.004)	.	.	.	.
COT	−0.07	−0.12	−0.13	0.07	−0.10	−0.33	−0.12
SE	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
GRN	−0.22	−0.33	−0.22	−0.22	0.58	−1.09	−0.24
SE	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.053)
FAL	n/a	n/a	n/a	n/a	n/a	n/a	n/a
SE	.	.	.	.	.	.	.

Note This table is read as crop (row) follows crop (column)

We want to emphasize and important point for interpreting our parameter estimates. Parameter estimates represent implied cost and yield effects based on observed farmer decisions and the dynamic program specified in the previous section. It is tempting to interpret results, such as yield effects, in terms of physical units and compare to observed yields at the field. These numbers should only be used as a general guide, not for direct comparison of magnitudes.

6 Model Simulation

Given the estimated parameters of the model, we formulate and solve the dynamic programming model defined in the previous section. Figure 2 demonstrates the optimal sequence of crops for the model solution, which, as expected, reproduces the observed base rotation.

As shown, the base alfalfa–cotton–grain–fallow rotation represents the optimal crop rotation, with alfalfa as a four-year crop. In the absence of external shocks, the field stays in an infinite cycle of alfalfa–cotton–grain–fallow. In years 1–4, the farmer plants alfalfa, then rotates into cotton in year 5, grain in year 6, and fallow in year 7 before the cycle repeats again.

Variations on the base rotation from changes in relative prices, soil quality, salinity, or resource constraints constitute the main results. We expect that rotations vary with differences in field-specific physical capital in addition to changes in relative crop prices. Our model shows that the optimal rotation shifts as soil quality and field salinity levels change, even with average relative prices. For example, with average prices over low salinity the optimal rotation is alfalfa (four years) followed by one year of cotton. Thus, the model satisfies one of the proposed criteria, the flexibility to represent spatial variation in rotation systems. For brevity, we omit policy simulations

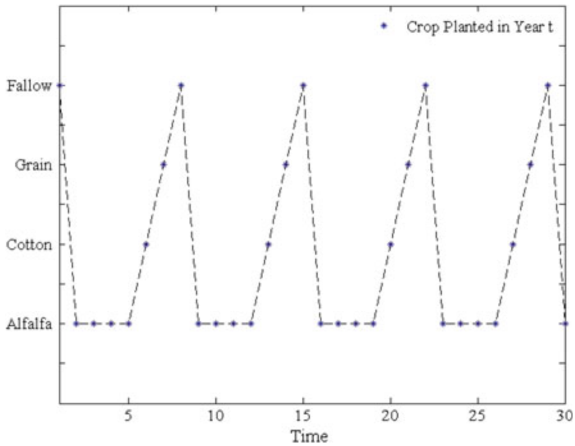

Fig. 2 Simulation of base crop solution—average conditions over 30 years

of the model with changes in relative prices and focus on simulation of a salinity policy.

6.1 Policy Application: The Dynamic Margin Response to Salinity

In California's Central Valley, direct agricultural salinity costs are estimated between \$450 and \$750 million per year. Including damages to urban users, the environment, and industry increase total cost estimates to over \$1 billion per year. Most of the costs to agriculture are due to declining crop yields as salinity increases. However, not all crops are affected by salinity in the same way and, as such, we expect that farmers may be able to mitigate some of the losses though increased rotation management. The cost estimates cited above treat production as a static process and only allow for adjustments along the intensive and extensive margin. It follows that the costs of increasing salinity may change if we additionally consider the dynamic margin.

Reduced crop yields account for the largest direct cost of salinity to agriculture. The relationship between root-zone salinity and crop yield varies by crop and field-specific conditions. Farmers faced with high salinity have a number of management options which may change the nature of this relationship. For example, with sufficient depth to groundwater, applying water in excess of crop consumption will allow leaching of salts below the root zone. With saline groundwater intrusion into the root zone, switching to micro-irrigation or facilitating drainage may mitigate the impacts of salinity. In addition to adjusting irrigation and technology, other salinity management strategies include crop rotation, field flushing, adjusting fertilizer application, field drainage, establishing native salt-tolerant vegetation, and land fallowing. The extent to which farmers use these different options depends on the type of salinity and field-specific factors including micro-climate, soil characteristics, and the quality of the available irrigation water.

We expect that the farmer considers the relative costs of the management alternatives, the marginal crop-specific yield effects of management strategies, and relative crop prices when making rotation system management decisions. This suggests that adjustment to higher salinity is not an immediate extensive margin shift, as static models would predict, but rather a gradual intensive and dynamic margin adjustment. Figure 3 shows the effect of salinity levels on the optimal rotation when salinity changes from 5, 13, 16, and 30 dS/m.

These levels represent the upper range of the salinity state space estimated in the dynamic programming model. With moderate salinity of 5 dS/m, the optimal rotation is two years of alfalfa followed by one year of cotton. Cotton is relatively salt-tolerant whereas alfalfa is relatively intolerant but valuable, thus the farmer manages higher salinity levels by rotating between these crops. Additionally, alfalfa is known to fix nitrogen, which benefits cotton production in the following year. As salinity increases, it is no longer profitable to include cotton and the optimal rotation

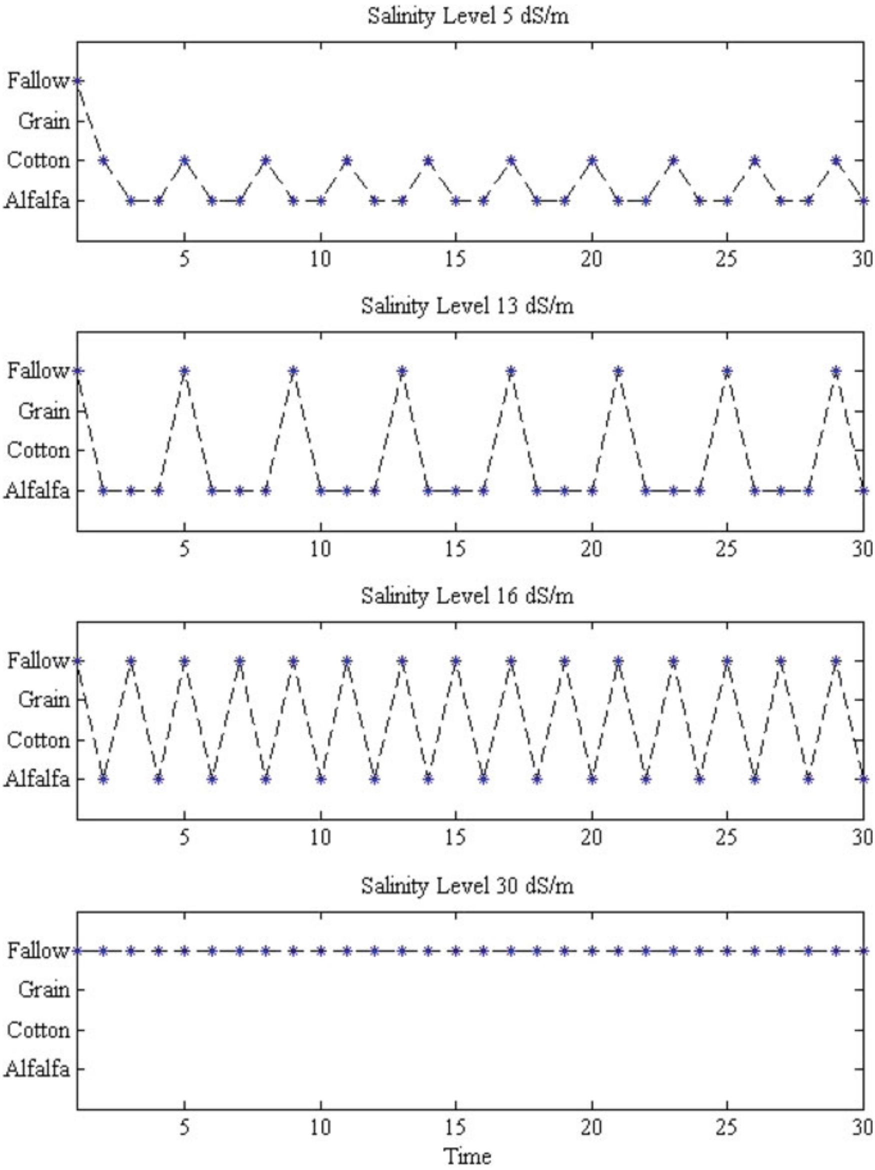

Fig. 3 Crop rotation with average prices and good soil (USDA SCCI = 2) under varying salinity levels

shifts to an alfalfa–fallow rotation. At some point between 15 and 30 dS/m, the field is removed from production. Note that this analysis assumes constant relative crop prices, at average observed levels, and that a change in relative prices would shift the optimal rotation.

If we estimate the model under a range of relative prices and physical capital, the optimal rotation shifts as expected. For example, with a higher price, cotton remains in the optimal rotation system longer with higher salinity levels. Similarly, high grain prices outweigh rotation adjustment costs and salinity effects and grain is included in the optimal rotation. Thus, the model satisfies one important criterion, the ability to reproduce a range of rotation systems which respond, endogenously, to changing relative process and field conditions. Furthermore, the model reproduces shifts out of a base rotation into nested rotation cycles in response to changing relative prices or salinity levels.

To demonstrate the rotation response to changes in relative prices, we simulate an alfalfa price increase from the sample average of $132.60 per ton to $257 per ton. We simulate the spike in years 7 through 12, hold soil quality constant, and allow salinity levels to increase. Figure 4 illustrates the results.

The relevant comparison is to Fig. 3. First, note that under relatively low salinity levels, of 5 dS/m, the optimal rotation shifts from 2 years of alfalfa and one year of cotton to alfalfa monoculture for the duration of the spike. Note that this actually corresponds to two four-year alfalfa stands planted in succession. The increase in the price of alfalfa, combined with elevated field salinity levels, outweighs the negative rotation costs of sequential alfalfa and the farmer switches to alfalfa monoculture.

Figure 4 also illustrates an interesting response to the alfalfa price spike under moderate salinity levels (13 dS/m). With higher alfalfa prices, the model estimates that the dynamically optimizing grower will switch to a two-year, one-year fallow rotation cycle instead of three-year alfalfa under the lower prices. This counterintuitive result is explained by noting that first- and second-year alfalfa stands have higher yields than third-year alfalfa. As such, the farmer can increase profits on the field during the price spike by shortening the rotation, incurring higher switching costs, and realizing higher yields during the price spike. Under higher salinity levels, Fig. 4 shows that the optimal rotation cycle remains unchanged.

Our model demonstrates that the endogenous dynamic cycle of a field rotation responds as expected to changes in field capital and relative prices. The dynamic margin of adjustment leads to a more gradual shift in rotation systems in response to changes in salinity than that which static models may predict. If we think about the continuous nature of field capital, then we can visualize how fields in a region like Kern County are in a continuous range of dynamic rotation cycles. Importantly, fields are at different points in their respective rotation cycles at any point in time (i.e., different years of the alfalfa stand, cotton, or grain) and this will additionally affect the response of individual fields to changes in salinity. Policies aimed at more sustainable agriculture, such as a salt balance, need to be aware of the intensive, extensive, and dynamic margins of response to salinity.

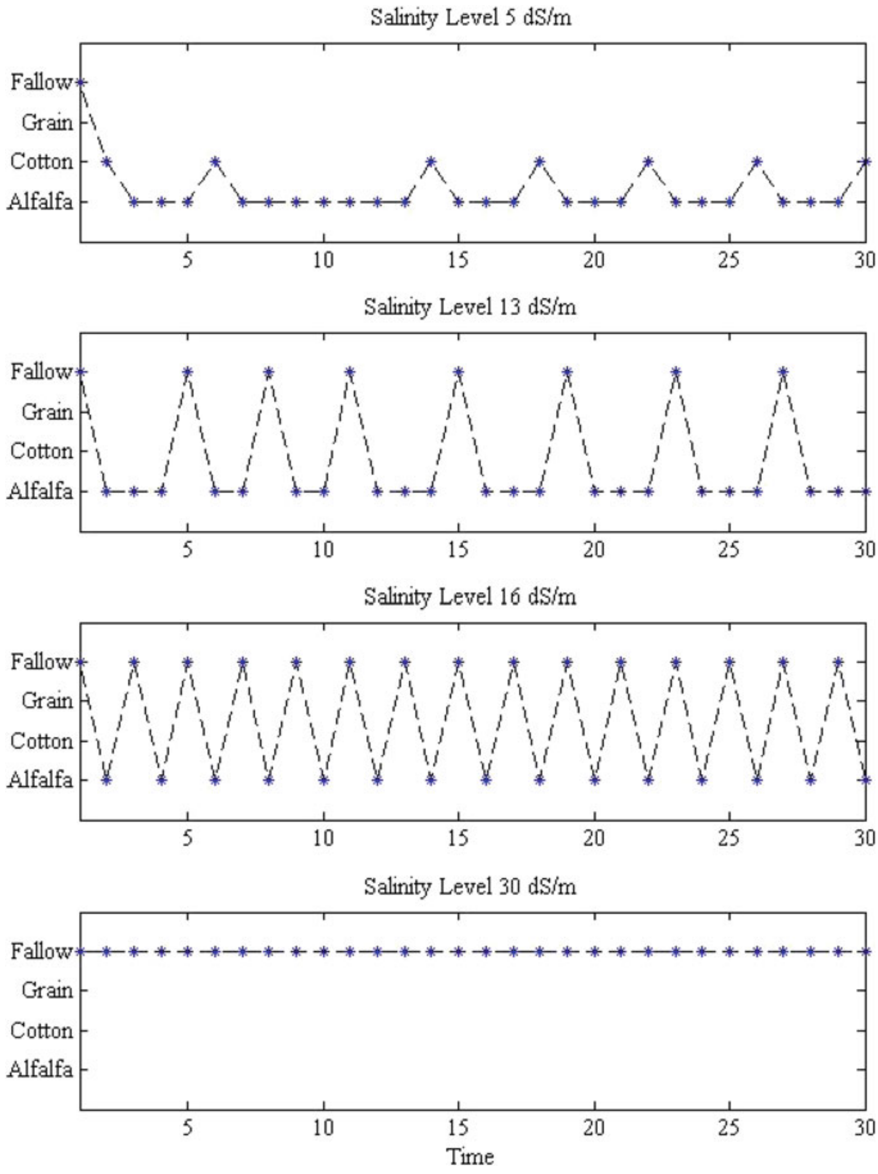

Fig. 4 Crop rotation with average prices and good soil (USDA SCCI = 4) under varying salinity levels with alfalfa price spike in years 7 through 12

7 Conclusion

Cropping decisions are made on a field and farm basis. In many cases, it is the relevant level of disaggregation for agricultural policy analysis. Farm-wide considerations such as risk and input smoothing are certainly important for aggregate planting decisions but, fundamentally, we hypothesize farmers understand variation in land characteristics and manage rotations on individual fields. Within a region, we observe that agricultural production exhibits significant spatial specialization, and it follows that *where* may be as important as *what* for many agricultural–environmental policies. For example, environmental effects of nitrogen runoff depend on the spatial location (e.g., proximity to waterways) of the field(s) producing a specific set of crops. As another example, salinity levels vary across fields and our model shows that the rotation response will vary accordingly. Treating agricultural production as part of a dynamic cycle at the field level offers valuable insights into these and related policy questions.

In addition to sustainable agriculture and spatially dependent agricultural–environmental policies, field-level analysis is relevant for incorporating the dynamic margin into supply response. Researchers have established, in various contexts, that there is a difference between dynamic and static supply elasticities (Orazem and Miranowski 1994; Tegene et al. 1988). An interesting extension, which our framework encompasses, is how production cycles, or rotation systems, respond to price shocks. For example, two different sets of fields may be growing cotton in a given season but one may be part of a one-year cotton–vegetable rotation, to control seedling diseases, whereas the other may be in a multi-year cotton–alfalfa–vegetable rotation, perhaps to control nematodes. Both changes in relative prices and the dynamic cost of breaking the rotation are important.

In this contribution, we propose that understanding how farmers determine rotations and mono-cropping is important for the implementation of effective sustainable agricultural policies. We hypothesized that an economic analysis of sustainable agriculture should account for the dynamic margin of adjustment in addition to the static intensive and extensive margins. The dynamic programming model we formulated explicitly accounts for the switching costs of rotations as reflected by farmer decisions. To measure and understand switching costs is one step toward predicting the adoption of more rotational practices and represents an effective tool in evaluation of policies that seek to move toward a more sustainable agricultural production system.

References

Alston, J. M., Andersen, M. A., James, J. S., & Pardey, P. G. (2010). *Persistence pays: U.S. agricultural productivity growth and the benefits from public R&D spending*. New York: Springer.

Antle, J., & Stoorvogel, J. J. (2006). Predicting the supply of ecosystem services from agriculture. *American Journal of Agricultural Economics, 88,* 1174–1180.

Antle, J., & Valdivia, R. O. (2006). Modeling the supply of ecosystem services from agriculture: A minimum data approach. *Australian Journal of Agricultural and Resource Economics, 50,* 1–15.

Beneke, R., & Winterboer, R. (1973). *Linear programming applications to agriculture*. Ames: Iowa State University Press.

Bertsekas, D. P. (1976). Dynamic programming and stochastic control. *Mathematics in Science and Engineering, 125*, 222–293.

Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal, 6,* 435–460.

Burt, O., & Allison, J. (1963). Farm management decisions with dynamic programming. *Journal of Farm Economics, 45*(1), 121–136.

Csiszar, I. (1991). Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Annals of Statistics, 19*(4), 2032–2066.

Doole, G. J. (2008). Optimal management of annual ryegrass (*Lolium rigidum* Guad.) in phase rotations in the Western Australian Wheatbelt. *The Australian Journal of Agricultural and Resource Economics, 52*(3), 339–362.

Doole, G. J. (2009). A practical algorithm for multiple-phase control systems in agricultural and natural resource economics. *Journal of Agricultural and Resource Economics, 34*(1), 1–21.

El-Nazer, T., & McCarl, B. (1986). The choice of crop rotation: A modeling approach and case study. *American Journal of Agricultural Economics, 68*(1), 127–136.

Heady, E. O. (1948). The economics of rotations with farm and production policy applications. *Journal of Farm Economics, 30*(4), 645–664.

Hennessy, D. A. (2006). On monoculture and the structure of crop rotations. *American Journal of Agricultural Economics, 88*(4), 900–914.

Hildreth, C., & Reiter, S. (1951). On the choice of crop rotation. In T. Koopmans (Ed.), *Analysis of production and allocation*. New York: Wiley.

Jaenicke, E. C. (2000). Testing for intermediate outputs in dynamic DEA Models: Accounting for soil capital in rotational crop production and productivity measures. *Journal of Productivity Analysis, 14*(3), 247–266.

Johnson, S. E. (1933). The theory of combination enterprises on individual farms. *Journal of Farm Economics, 15*(4), 656–667.

Kennedy, J. (1981). An alternative method for deriving optimal fertilizer rates: Comment and extension. *Review of Marketing and Agricultural Economics, 49*(3), 203–209.

Kennedy, J. (1986). Rules for optimal fertilizer carryover: An alternative explanation. *Review of Marketing and Agricultural Economics, 54*(2), 3–10.

Kennedy, J., Whan, I., Jackson, R., & Dillon, J. (1973). Optimal fertilizer carryover and crop recycling policies for a tropical grain crop. *Australian Journal of Agricultural Economics, 17*(2), 104–113.

Kullback, J. (1959). *Information theory and statistics*. New York: Wiley.

Langpap, C., Hascic, I., & Wu, J. (2008). Protecting watershed ecosystems through targeted local land use policies. *American Journal of Agricultural Economics, 90*(3), 684–700.

Langpap, C., & Wu, J. (2008). Predicting the effect of land-use policies on wildlife habitat abundance. *Canadian Journal of Agricultural Economics, 56*(2), 195–217.

Levine, R. D. (1980). An information theoretical approach to inversion problems. *Journal of Physics A: Mathematical General, 13,* 91–108.

Livingston, R., Roberts, M., & Zhang, Y. (2012, Nov). Optimal sequential plantings of corn and soybeans under price uncertainty. Working Paper.

Mittelhammer, R. C., Judge, G. G., & Miller, D. J. (2003). *Econometric foundations*. New York: Cambridge University Press.

Needleman, S., & Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48*(3), 443–453.

Orazem, P., & Miranowski, J. (1994). A dynamic model of acreage allocation with general and crop-specific soil capital. *American Journal of Agricultural Economics, 76*(3), 385–395.

Peterson, G. (1955). Selection of maximum profit combinations of livestock enterprises and crop rotation. *Journal of Farm Economics, 37*(3), 546–554.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423.

Skilling, J. (1989). The axioms of maximum entropy. In J. Skilling (Ed.), *Maximum entropy and Bayesian methods in science and engineering* (pp. 173–187). Dordrecht: Kluwer.

Swanson, E. R. (1956). Application of programming analysis to Corn Belt Farms. *Journal of Farm Economics, 38*(2), 408–419.

Tanaka, K., & Wu, J. (2004). Evaluating the effect of conservation policies on agricultural land use: A site specific modeling approach. *Canadian Journal of Agricultural Economics, 52*(3), 217–235.

Taylor, C. (1983). Certainty equivalence for determination of optimal fertilizer application rates with carry-over. *Western Journal of Agricultural Economics, 8*(1), 64–67.

Tegene, A., Huffman, W., & Miranowski, J. (1988). Dynamic corn supply functions: A model with explicit optimization. *American Journal of Agricultural Economics, 70*(1), 103–111.

Thomas, A. (2002). A dynamic model of on-farm integrated nitrogen management. *European Review of Agricultural Economics, 30*(4), 439–460.

Thoreson, B., Clark, B., Soppe, R., Keller, A., Bastiaanssen, W., & Eckhardt, J. (2009). Comparison of evapotranspiration estimates from remote sensing (SEBAL), water balance, and crop coefficient approaches. In *World Environmental and Water Resource Congress*, May 17–21 2009, Kansas City, MO, USA, https://doi.org/10.1061/41036(342)437.

Wu, J., Adams, R. M., Kling, C. L., & Tanaka, K. (2004). From microlevel decisions to landscape changes: An assessment of agricultural conservation policies. *American Journal of Agricultural Economics, 86*(1), 26–41.

Wu, J., & Babcock, B. (1998). The choice of tillage, rotation, and soil testing practices: Economic and environmental implications. *American Journal of Agricultural Economics, 80*(3), 494–511.

Part IV
Using Quantitative Methods to Inform Decision-Making in Agricultural and Resource Policy

Chapter 10
Water into Wine and Cheese: Implications of Substitution and Trade for California's Perennial Water Woes

Daniel A. Sumner and Qianyao Pan

Abstract Water woes are growing globally as farmers and others struggle to develop infrastructure and institutions that allow the agricultural economy to thrive in the face of competing uses for water. While not new, these struggles are deeply important, and nowhere more so than to agriculture within arid regions. This chapter uses the California water context to trace through the simple economics of how irrigation water availability and price affect prices and quantities of tradeable food products. We highlight a few key relationships within the supply chains for wine and cheese using the simplest framework possible—fixed proportions and elastic input supplies at each stage of a multi-market chain. First, we consider irrigation water used to produce grapes that are transformed into wine and highlight the role of cost shares and final product demand elasticities. We show that irrigation water is a far more important driver of prices and quantities in the low-cost San Joaquin Valley region, which faces a more elastic demand for wine than in the high-cost (and price) North Coast region, which faces a less elastic demand. When we consider the irrigation water used to produce feed crops for dairy cows—which, in turn, produce milk that is transformed into cheese, we find that water has only a moderate cost share in forage production and that forage has a moderate share in milk output. Nevertheless, because California cheese faces an elastic demand in the global market, a rise in water costs could reduce California cheese production significantly.

1 Introduction

Like many arid agricultural regions, California faces perennial irrigation water concerns including (a) periodic droughts, (b) competing demands from environmental and urban uses, (c) limited storage capacity, (d) distribution costs, and (e) complexities of property rights in surface water and groundwater

D. A. Sumner (✉) · Q. Pan
Department of Agricultural and Resource Economics, Univerisity of California Agricultural Issues Center, University of California at Davis, Davis, USA
e-mail: dasumner@ucdavis.edu

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_10

(Hanak et al. 2018a, b). Market signals struggle to be heard fully within this cacophony of concerns.

Market signals have not traditionally been used to guide allocation and management of water resources. Surface water is often allocated for irrigation under horrendously complicated rules, with prices charged to growers that seem to be sometimes only loosely connected to underlying supply and demand fundamentals. As with other storable commodities, long-term contract prices deviate from spot-market prices, while temporal and locational variations in prices are based on physical and regulatory costs of moving water over time and space. At the same time, a legacy of government mandates and regulations limit market transactions. Groundwater pumping is often expensive, and overuse of groundwater causes external impacts and the potential for rising costs over time (Hanak and Jezdimirovic 2016).

The implementation of the Sustainable Groundwater Management Act (SGMA) passed in 2014 may help address the groundwater overdraft issue, and the further use of water markets, and incentives for groundwater recharge and groundwater storage (Hanak and Stryjewski 2012), could also help in aligning water use with costs and benefits (Hanak et al. 2018a, b).

Farming and processing of farm production are vulnerable to water availability and price. This contribution takes the simplest possible approach to trace through the impacts from changes in irrigation water costs to the supply of processed farm products that face alternative demand conditions. In California, wine and cheese are important tradeable products derived from local irrigation water, through the production of grapes used for wine and through the production of hay and silage used to produce the milk that is used for cheese. We illustrate how cost shares and demand elasticities are key parameters influencing the market-level implications of changes in water availability and price.

2 Some Economic Implications of Higher Irrigation Water Costs

Economic trends augmented by new regulations and climate change are likely to increase irrigation water scarcity and price in the San Joaquin Valley. Population growth in California and shifts toward water-intensive crops raise additional water scarcity concerns. Regulations also shift irrigation costs. More demand for environmental uses of water has been the most prominent of the regulation-driven impacts on water scarcity. Limits on groundwater pumping may further reduce availability of irrigation water. Additional or more streamlined tradability of water rights raises explicit costs of water in some districts while lowering the water prices elsewhere. Climate trends toward higher growing season temperatures raise the demand for irrigation, while higher winter temperatures reduce snowpack available for irrigation supply. Less surface water available for irrigation reduces groundwater recharge and raises subsequent costs of groundwater access. Thus, long-term prospects point

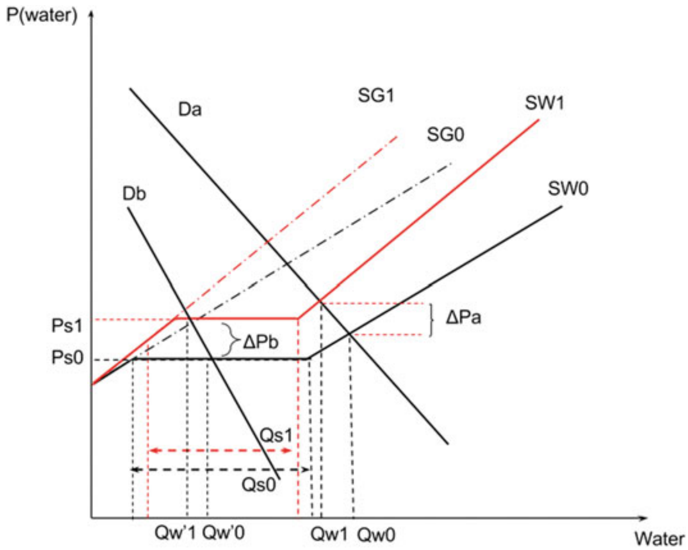

Fig. 1 Heterogeneous farm-level water demand for surface and groundwater supply

toward higher average water costs and for periodic droughts to cause higher spikes in water prices.

Before considering some simple simulations, let us review the impacts of more scarcity in water supply for irrigation in a simple conceptual framework. Farms use both groundwater and surface water. Groundwater incurs pumping costs that rise as more is pumped and the vertical distance to the aquifer (i.e., the 'lift') increases. Surface water scarcity is reflected in lower allocation quantities and higher prices for the quantity that is made available. Movement (sales or other transactions) of water between farms is limited, so no full equilibrium of equal prices or marginal value products is established. For illustration, Fig. 1 shows the case of two farms with distinct demand functions that share access to groundwater and have separate surface water allocations, but the same prices.

In Fig. 1, SG_0 represents the upward sloping groundwater supply function that represents marginal pumping costs as a function of collective use of the two farms. SW_0 is the combined supply of groundwater and surface water available. We assume a constant price P_{s0} for surface water, which is higher than the cost to pump the first unit of groundwater. The amount of surface water is initially limited to Q_{s0} for each farm. For simplicity, we allow each farm to be allocated equal shares of surface water. With N farms, the total amount of surface water available for irrigation would be $Q_{s0} * N$.

The demand curve D_b represents the water demand from farm B, which is small relative to farm A, represented by D_a. Each farm uses both groundwater and surface water for irrigation. In Fig. 1, D_b intersects with SW_0 at $Q_{w'0}$, which means farm *B* uses $Q_{w'0}$ amount of water in total. Farm B uses an initial amount of groundwater that costs less than surface water and uses only about half of its allocation of surface water.

The demand curve for farm A, D_a, intersects SW_0 at Q_{w0}, which means farm A uses Q_{w0} amount of water in total. Farm A utilizes its full allocation of surface water available (Q_{s0}) and pumped the additional groundwater ($Q_{w0} - Q_{s0}$). Notice that the initial amount of groundwater costs less per unit than does surface water and the second amount of groundwater costs more per unit than does surface water. We assume that the two farms cannot trade surface water with each other.

Consider application of three shocks to water supply in the context of this simple framework. First, the amount of surface water allocated to each farm is reduced. In Fig. 1, the amount of surface water declines from Q_{s0} to Q_{s1}. Second, the price of surface water rises.

In Fig. 1 the price of surface water increases from P_{s0} to P_{s1}. Third, the marginal cost of groundwater rotates up, say because of an area-wide decline in the water table or some other added cost imposed on the use of groundwater. In Fig. 1, groundwater marginal cost rotates from SG_0 to SG_1.

Three outcomes follow from the three changes in the water supply system. For farm A, the marginal cost of water is still that of the supplementary groundwater that is used. Farm A uses less water in total and specifically less surface water. As drawn in Fig. 1, groundwater use rises for farm A, but the amount is ambiguous in theory. For farm B, the marginal price of water is that for surface water which increases by ΔP_b. Farm B uses less surface water and more groundwater. Total water for farm B falls from $Q_{w'0}$ to $Q_{w'1}$. Note that the changes in the effective marginal price of water differ between the two farms and the responses differ as well.

The analysis illustrated in Fig. 1 shows how an observed change in water prices may differ across farms and water districts when there are shocks in water supply. Therefore, it may be important to include the heterogeneity in the changes in water prices when evaluating the impact of water price change. We do not consider the potential heterogeneity in water price change across farms in this chapter when we simulate the impact of water price change on hay production. We assume the same percentage change in water prices for both farms growing alfalfa hay and corn silage. This simplification may introduce bias in the projected impacts of water price changes, especially when considering different water demands from alfalfa hay production and silage production.

A rising cost of irrigation results in an increase in the cost of production for irrigated crops. The magnitude of the increase in cost of production depends on the derived demand elasticity and the supply elasticity for irrigation water. Farms adapt to the higher water cost in several ways. Farms may switch to less water-intensive crops—crops with higher marginal value product for water utilized. In equilibrium, the expected marginal value of water will equalize across crops. Farms also use less water for each given crop by being more water-efficient, such as by substituting capital or labor for irrigation water. Reduced irrigation per area is a form of substituting land for irrigation water.

3 Simulating Water Price Impacts Using the SWAP Model

One alternative for simulating responses to water scarcity is the Statewide Agricultural Production (SWAP) model–a regional optimization model of irrigated agricultural production tailored to California (Howitt et al. 2012; CH2MHill 2012). The SWAP model has 27 base regions in the Central Valley in addition to agriculture in the Central Coast, South Coast, South Lahontan, and Colorado River regions—giving a total of 37 regions in the current model. The SWAP model is a mathematical programming-based model that maximizes the objective of producer profits, and is calibrated to observed crop production and area patterns using the positive mathematical programming method (Howitt 1995a, b). An early application to Sacramento Valley Water was Lee et al. (1999, 2001).

SWAP incorporates surface water supplies and access to groundwater into its representation of inputs that go into agricultural production. When the quantity of surface water or the cost of groundwater changes, the model optimizes crop choice by adjusting the production mix, water input, and other inputs (Howitt et al. 2012, 2015). SWAP generates the agricultural water demand curves that are used in a larger engineering model of California's water infrastructure—CALVIN (Draper et al. 2003). Howitt et al. (2015) estimate the drought impacts on California agriculture using SWAP model—where the water allocation to the agricultural sector coming from the larger CALVIN model determines the amounts of surface and groundwater that the SWAP model has available.

In this contribution, we examine multi-market dynamics that link the market for feed to that for raw milk and processed dairy products. We model the equilibrium displacements within the input and product markets and trace the impacts through the supply chain. We consider first impacts from supply shifts of irrigation water on markets for grapes and wine under alternative wine market conditions. We then consider irrigation water price impacts on prices and quantities of dairy feed, raw milk, and California cheese that is sold on global markets.

4 From Water to Wine in California's San Joaquin Valley

Almost all water used in wine production is applied to vineyards to grow grapes. Wine output and price are affected (over a relatively long horizon) as a change in the cost of water passes through the vertical chain from water to grapes to wine. In this section, we develop a detailed example to demonstrate the impact of high water costs on the wine market. We use a single-product, partial-equilibrium model with two inputs at each stage (water and other inputs to produce grapes and grapes and other inputs to produce wine) and assume fixed proportions. The supply conditions of inputs and final output are discussed first. We then proceed with derived demand for grapes and water given wine demand conditions. We include numerical examples to illustrate the impacts of high water costs on quantities and prices of wine in two representative regions in California.

With fixed proportions in grapes production, the supply of grapes is defined by the supply conditions for the underlying inputs. Let us assume water supply facing grape production is perfectly elastic, which is a reasonable approximation because water use for grapes alone is a relatively small share of total water use in the regions. This means that the price of irrigation water, P_{w}, is exogenous and set to P'_{w}. Writing these conditions more specifically, grape production and the supply function for water to grapes are given by,

$$G = \min(\lambda W, \ \beta Z), \text{ and } P_{\mathrm{w}} = P'_{\mathrm{w}},$$

where G is the quantity of grape production, and W is the amount of water going into grapes. Other inputs to grape production are represented by the composite, Z, and the supply of Z is an upward-sloping function of quantity used,

$$P_{\mathrm{z}} = m^{\mathrm{s}}(Z).$$

These considerations yield an inverse supply function for grapes, G, written as

$$P_{\mathrm{g}} = d[(P'_{\mathrm{w}})(G/\lambda) + m^{\mathrm{s}}(G/\beta)(G/\beta)]dG = g^{\mathrm{s}}(G).$$

Because the inputs are used in fixed proportions, the marginal cost of grape production is the sum of marginal cost from water and marginal cost from non-water inputs. If the supply function for non-water inputs is upward sloping, the supply function for grapes is also upward sloping, even with the perfectly elastic supply of water to the wine grape industry.

Now, move downstream to wine production using grapes and non-grape inputs. Again, we assume fixed proportions and composite non-grape inputs. The marginal cost of wine is again the sum of the marginal cost of underlying inputs. The derived demand function for grapes is derived using the wine demand and non-grape input supply. Similarly, the derived demand function for non-grape inputs is given by the wine demand and the grape supply. The production function, where Q is the quantity of wine, is

$$Q = \min(\alpha G, \gamma O).$$

The inverse supply function for non-grape inputs, O, may be written simply as

$$P_{\mathrm{o}} = h^{\mathrm{s}}(O).$$

Putting these together means that the inverse supply of wine expressed as the price of wine quantity, P_{q}, as a function of the parameters is

$$P_{\mathrm{q}} = d(g^{\mathrm{s}}(Q/\alpha)(Q/\alpha) + h^{\mathrm{s}}(Q/\gamma\lambda)(Q/\gamma\lambda))dQ = f^{\mathrm{s}}(Q).$$

Turning to the demand side, the inverse demand for wine is

$$P_{\mathrm{q}} = f^{\mathrm{d}}(Q), \text{ and}$$

the implied derived demand for the two inputs, grapes and other inputs, is as follows:

$$P_{\mathrm{g}} = f^{\mathrm{d}}(G\alpha) - (G\alpha/\gamma)\cdot h^{\mathrm{s}}(G\alpha/\gamma) = g^{\mathrm{d}}(G), \text{ and}$$
$$P_{\mathrm{g}} = f^{\mathrm{d}}(O\gamma) - (O\gamma/\alpha)\cdot h^{\mathrm{s}}(O\gamma/\alpha) = h^{\mathrm{d}}(O).$$

Going one link further up the supply chain we can now express the derived demand for water and non-water inputs. The inverse demand functions for water and non-water inputs are expressed as follows:

$$P_{\mathrm{w}} = g^{\mathrm{d}}(W\lambda) - (W\lambda/\beta)\cdot m^{\mathrm{s}}(W\lambda/\beta) = w^{\mathrm{d}}(W), \text{ and}$$
$$P_{\mathrm{z}} = g^{\mathrm{d}}(Z\beta) - (Z\beta/\lambda)\cdot P'_{\mathrm{w}} = m^{\mathrm{d}}(Z).$$

4.1 Decrease in the Supply of Water for Wine Grapes

Now, consider implications of an increase in the price of water. The water price change affects the equilibrium in all markets involved in the multi-stage wine production, including the quantity of grapes supplied, the quantity of wine supplied, the quantity demanded for non-grape inputs in wine production, and the quantity demanded of non-water inputs in grape production. Suppose the marginal cost for wine increases by k percent due to an increase in water costs. This would generate a decrease in quantity demanded for wine and thus the quantity demanded for grapes and water. In log terms, change in the supply price of wine is

$$\mathrm{d}\ln P^{\mathrm{s}} = (1/\eta)\mathrm{d}\ln\ Q^{\mathrm{s}} + k$$

where the supply elasticity for wine is η. The log change in the quantity of wine demanded is

$$\mathrm{d}\ln Q^{\mathrm{d}} = \epsilon\ \mathrm{d}\ln\ P^{\mathrm{d}}$$

where the demand elasticity is ϵ. Market clearing in price and quantities require that

$$\mathrm{d}\ln P^{\mathrm{s}} = \mathrm{d}\ln P^{\mathrm{d}} \text{ and}$$
$$\mathrm{d}\ln Q^{\mathrm{s}} = \mathrm{d}\ln Q^{\mathrm{d}}.$$

Therefore, the equilibrium change in quantity of wine is

$$\mathrm{d}\ln \mathrm{Q}* = (\eta\epsilon/(\eta - \epsilon))k.$$

4.2 Numerical Simulations of the Effects of Water Cost Increases on the Markets for California Grapes and Wine

This section provides an example of the impacts of high water prices on the quantity demanded for wine and grapes in two representative regions in California: the North Coast and the San Joaquin Valley. The wine grapes grown in the North Coast face very different demand from wine grapes grown in the San Joaquin Valley due to their distinct market and production conditions.

4.2.1 Data for Simulations of Effects of Water Cost Increases on the Wine Market

Based on Lapsley and Sumner (2016), wine grapes in the San Joaquin Valley use about 7616 m^3/ha of irrigation water. Wine grapes yield about 32 metric tons per hectare. These assumptions indicate that 232 m^3 of water produces one metric ton of grapes. One metric ton of grapes produces about 730 liters (l) of wine. Finally, about 4–6 liters of water is used directly used within a winery to produce one liter of wine (Lapsley and Sumner 2016).

In the North Coast of California, less water is used for irrigation due to the cooler and wetter climate. We assume 1523 m^3 of irrigation water per hectare of grapes and 8.9 metric tons of grapes per hectare. This implies that 169 m^3 of irrigation water produces one metric ton of grapes in the North Coast district (Lapsley and Sumner 2016). Within a winery, one metric ton of North Coast grapes produces 667 l of wine. The water directly used within the winery ranges from 4 to 6 l of water/l of wine (Lapsley and Sumner 2016).

For prices, we assume that 440 US \$/metric ton of grapes in the San Joaquin Valley and wine using San Joaquin Valley grapes has a wholesale price of 2.2 US \$/l of wine. Grapes and wine produced in the North Coast are much more costly. We assume grapes produced in the North Coast cost 4409 US \$/metric ton and the wholesale price of wine using the North Coast grapes is 27.8 US \$/l of wine.

Wine market demand parameters also differ by region. In the wine market, the demand for the wine produced in the San Joaquin Valley is more elastic than the wine produced in the North Coast. The San Joaquin Valley wine competes with imported bulk wine in the USA and exported wine from all over the world outside the USA (Gabrielyan 2018). Wine produced in the North Coast faces a less elastic demand thanks to the long established unique characteristics of the market. Here, we assume the short run (one year or less) demand elasticity is −0.5 for both San Joaquin Valley

Table 1 Statistics for grape and wine productions in the San Joaquin Valley and the North Coast

Simplified estimates	Unit	San Joaquin Valley	Sonoma/Napa
Grape price	\$/metric ton	440	4409
Water usage	m^3/ha	7616	1523
Grape yield	metric ton/hectare	32	8.9
Wine from grapes	l/metric ton	730	667
Water use in processing wine	l/l	4–6	4–6
Wholesale price for wine	\$/l	2.2	27.8
Demand elasticity short run/long run	–	−0.5/−3.0	−0.5/−1.0

Source Lapsley and Sumner (2016), Smith et al. (2010), CDFA (2013), and authors' estimates

wine and North Coast wine. In the long run, the demand elasticity is −3 for the San Joaquin Valley wine and −1 for the North Coast wine.

Table 1 lists data and parameters and allows a comparison of grapes and wine grown in the San Joaquin Valley and in the North Coast.

4.2.2 Simulations of the Impacts of Water Cost Increases on Grapes and Wine

Now, we simulate the impact of a higher water cost, by increasing water price from 0.16 US \$/m^3 to 0.32 US \$/m^3. The increase in the cost of production in grapes increases the price of wine. For simplicity, we assume the marginal cost of non-water inputs and non-grape inputs are constant in the vertical chain of wine production. In that case, the marginal cost of grapes and wine production increases as follows, where Δ is a discrete change,

$$\begin{aligned} \Delta P_{\mathrm{g}} &= (1/\lambda)\Delta P_{\mathrm{w}} \\ \Delta P_{\mathrm{q}} &= (1/\alpha)\Delta P_{\mathrm{g}} = (1/\lambda)(1/\alpha)\Delta P_{\mathrm{w}} \end{aligned}.$$

The percentage changes in prices and equilibrium quantities of wine are just,

$$\begin{aligned} \mathrm{d}\ln P_{\mathrm{q}} &= \Delta P_{\mathrm{q}}/P_{\mathrm{q}}*, \text{ and} \\ \mathrm{d}\ln Q* &= \epsilon\, \mathrm{d}\ln P_{\mathrm{q}}* \end{aligned}.$$

Using the data in Table 1, the price of grapes increases by 9% in the San Joaquin Valley and only 1% in the North Coast. The price of wine increases by 2% in the San Joaquin Valley and only 0.15% in the North Coast. In the short run, the quantity demanded for wine produced in San Joaquin Valley decreases 1%, and the quantity demanded for wine produced in North Coast decreases 0.08%. The long run impacts

Fig. 2 Estimates of the impact of higher water prices on grape costs, wine prices, and quantities demanded. *Source* Author estimates and calculations

are substantially higher for the San Joaquin Valley. For the long run, we consider a shock that is expected to be permanent. The quantity demanded for wine produced in the San Joaquin Valley decreases 7%, while the change in quantity demanded for wine produced in the North Coast remains very small.

Figure 2 summarizes the estimates of the changes in prices and quantities for wine and allows comparisons of impacts for the San Joaquin Valley and in the North Coast.

4.3 Implications for Wine Grapes and Wine

Based on simulation results presented in Fig. 2, the cost of water has a much larger impact on grapes and wine where the demand is more elastic. Wine produced in the San Joaquin Valley has close substitutes in the US market from bulk imports and faces similar substitutes in the world market (Gabrielyan 2018). North Coast wine faces less potential substitution and thus demand for wine from that region is less elastic, even in the long run.

Given the inelastic demand for water in grape production, the significant cost share of irrigation water in that region, and fixed proportions in wine production, the increase in cost of irrigation water significantly increases the cost of production for wine produced in San Joaquin Valley. North Coast grapes use less irrigation water, and the cost share of water in grape production is lower. That, plus the less elastic demand for North Coast wine, means that the increase in water cost in grape production will change the quantity demanded for North Coast wine only slightly.

5 Water into Cheese in the San Joaquin Valley

Next, we use a similar framework to investigate a more complex vertical chain. The linkage from forage for dairy cows to raw milk to cheese is a multi-stage production process that takes place in the San Joaquin Valley, where almost all California dairy cows reside and almost all milk and cheese is produced. Moreover, the two main forage crops, alfalfa hay and corn silage, are important in terms of acreage and value and use substantial amounts and shares of irrigation water.

We investigate how higher water costs in California affect the multi-stage supply chain where the end product is cheese. We could have examined a similar chain for non-fat dry milk and butter, which also uses both milk fat and milk solids other than fat. The vertical chain of markets that affects cheese supply includes markets for water, forage crops, raw milk, and cheese. Again, we assume a single final product in partial equilibrium, where each stage has two inputs used in fixed proportion. We include a realistic numerical illustration to show the potential impacts of higher water costs on the California dairy sector.

Two forage crops are fed to the dairy cows in the San Joaquin Valley—wet roughage (mostly corn silage) and dry roughage (mostly alfalfa hay). Alfalfa and silage face different water prices and non-water input costs because the silage market is very local, due to the high transport costs that cause silage to be cultivated very near to dairy farms. Hay is grown both in the San Joaquin Valley and transported from other places. We assume both crops have two-input, fixed-proportion production functions using water and non-water inputs.

According to Long et al. (2015), alfalfa hay production uses 10,668 m^3 of irrigation water per hectare and yields 15.7 metric tons per hectare. This implies that 680 m^3 of water produces one metric ton of alfalfa hay. Based on Klonsky et al. (2015), corn silage uses 9144 m^3 of water per hectare and yields 71.7 metric tons per hectare. This implies that 127.5 m^3 of water produces one metric ton of corn silage. The prices for hay and silage are weighted feed costs based on 2015 Holstein Feed Summary reported by California Department of Food and Agriculture (CDFA 2015b) and the authors' calculation. We use a representative alfalfa hay price of 262 US $/metric ton and an estimated silage price of 80 US $/metric ton.

Feed use is based on the assumed life cycle of a milk cow (Anderson and Sumner 2016), milk yields from California Dairy Statistics reported by CDFA, the feed rations reported in the Holstein Feed Summary (CDFA 2015a, b), and authors' estimates. We find that a dairy cow on average consumes 1.67 metric tons of hay and 5.89 metric tons of silage per year, with annual milk production per milking cow of 10.6 metric tons within the standard 305 milking days per year. Thus, we estimate 0.22 metric tons of hay and 0.79 metric tons of silage are used to produce 1 metric ton of milk for a dairy cow, on average across its life cycle. We use a price of milk used in

Table 2 Estimated statistics for forage, milk, and cheese productions in California

	Unit	Value
Water in hay	m^3/metric ton	680
Water in silage	m^3/metric ton	128
Hay in milk	metric ton/metric ton	0.22
Silage in milk	metric ton/metric ton	0.79
Milk in cheese	kg/kg	9.80
Hay price	$/metric ton	262
Silage price	$/metric ton	80
Milk price	$/kg	0.34
Non-milk cost in cheese	$/kg	0.52
Estimated cheese price	$/kg	3.81
Demand elasticity for cheese	–	−5

Source Long et al. (2015), Klonsky et al. (2015), Anderson and Sumner (2016), CDFA (2015a, b) and author estimates

cheese production based on the minimum milk price for California class 4b, which is 0.34 US $/kg (CDFA 2016).

Milk price regulations imply that the price of milk components used for cheese may differ from the price of identical components in the same region when that milk is used for other products. Farmers receive regulated prices for milk that are a blend of the component prices by end use based on the market-wide use of milk in California. California pricing rules changed in November 2018, but we use milk prices, from 2015 to be consistent with the other data.

According to California Dairy Statistics, about 9.8 kg of milk produces one kilogram of cheese. Based on CDFA manufacturing cost reports, (CDFA 2014), processors spend 0.52 US dollars on non-milk inputs to produce one kilogram of cheese. The cheese price is estimated based on the sum of the milk cost and non-milk cost, given input prices and production ratios, which is 3.81 US $/kg.

We assume the demand facing California cheese is very elastic because California cheese is mostly generic, competes in global markets, and has a high degree of substitution with cheese produced outside California. California's dairy processing plants account for almost 40% of US dairy exports (Matthews et al. 2016). Here, we use −5 as the demand elasticity for California cheese in the long run. Table 2 summarizes the parameters used to estimate impacts of water cost increases in the multi-stage cheese production.

Because much alfalfa hay is produced in the North Sacramento Valley or other places and corn silage is produced in the San Joaquin Valley, we assume different water prices for the two crops. Based on the 2015 data in Long et al. (2015) and Klonsky et al. (2015), the base prices of water are 0.04 US $/$m^3$ in alfalfa hay production, and 0.08 US $/$m^3$ in silage production. We simulate the impacts of higher water prices by doubling water prices from 0.04 to 0.08 US $\$/\text{m}^3$ for alfalfa hay production, and from 0.08 to 0.16 US $\$/\text{m}^3$ for silage production.

The derivation of model equations is similar to what was presented for grapes and wine except that here we have two inputs affected by the water price change and we have an addition stage (animal product production) from crop production through to the final farm product sold to consumers. Here, we show the key equations used in the numerical simulations. The subscript h represents hay, and the subscript s represents silage. The subscript m stands for milk, and c stands for cheese. As before, P is price and Q is the quantity of the final output. Parameters are defined analogously as in the wine modeling.

$$
\begin{aligned}
\Delta P_{\mathrm{h}} &= (1/\lambda_1)\,\Delta P_{\mathrm{w}} \\
\Delta P_{\mathrm{s}} &= (1/\lambda_2)\Delta P_{\mathrm{w}} \\
\Delta P_{\mathrm{m}} &= (1/\alpha_1)\,\Delta P_{\mathrm{h}} + (1/\alpha_2)\Delta P_{\mathrm{s}} \\
\Delta P_{\mathrm{c}} &= (1/\omega)\Delta P_{\mathrm{m.}}
\end{aligned}
$$

We consider proportional changes as,

$$\mathrm{d}\ln P_{\mathrm{c}} = \Delta P_{\mathrm{c}}/P_{\mathrm{c}}*.$$

The equilibrium cheese market impact in quantity is

$$\mathrm{d}\ln Q_{\mathrm{c}}* = \epsilon\mathrm{d}\ln P_{\mathrm{c}}*.$$

A higher water cost causes an increase in the cost of production of hay and silage. As before, we assume the marginal cost of non-water inputs in crop production does not change and the non-forage inputs per unit of milk production do not change. We also assume that the non-milk inputs per unit of cheese production do not change. Hence, moving through the vertical chain of cheese production, the marginal cost of cheese production increases by a fixed amount, which is determined by the increase in the cost of irrigation water.

Before the water cost shock, the water represents about 10% of total hay production cost and about 13% of total silage production costs. Hay represents about 17% and silage about 19% of average milk production costs. The cost share of milk in cheese production is about 86.5%. When water price doubles, the water cost share rises to 18% in hay production and 23% in silage production. As a consequence, the hay cost share in milk production rises to 18%, and the silage cost share rises to 21%. The cost share of milk in cheese rises to 87%. For simplicity, these calculations assume that the prices and usage of other inputs involved in the multi-stage production remain constant.

Table 3 shows the impacts of an irrigation water cost increase (a doubling of cost of water for each crop) that is expected to be long-lasting on the price and quantity of California cheese.

Table 3 Estimated impacts of higher irrigation water prices on forage, raw milk, and cheese prices and quantities in California

Estimates	Unit	Before water price change	After water price change
Water price for hay/silage	$\$/m^3$	0.04/0.08	0.08/0.16
Hay price	$/metric ton	262	293
Silage price	$/metric ton	80	92.4
Milk price	$/kg	0.34	0.36
Cheese price	$/kg	3.81	4.01
Cheese price change (%)	–	–	4.2
Cheese quantity change (%)	–	–	−21

Source Author calculations

The higher cost of water implies a higher cost of alfalfa hay by 12% and a higher cost of silage by about 16%. As a result of these feed cost increases, the cost of producing milk and therefore the farm price of milk rise. A higher price of milk translates into as higher cost of production of California cheese. The price of California cheese increases by about 4%, and because demand facing California cheese is elastic, the quantity of California cheese falls substantially.

The challenge for the California dairy industry is that the own-price elasticity of demand for California cheese is very high. We assume an own-price demand elasticity of −5. The implication is that the quantity demanded of California cheese falls by more than 20% in the long run. A similar result would follow from a similar simulation non-fat dry milk (or skim milk powder) and butter. Together cheese and the butter/powder production comprise about 80% of the use of California farm milk. That means a substantial irrigation water cost increase would reduce competitiveness of the California dairy industry, which faces a highly elastic demand.

6 Conclusions

This chapter has highlighted the vulnerability to irrigation water increases of San Joaquin Valley farm industries that face elastic demands for processed products to which they contribute. We used fixed proportion frameworks to link water costs to farm outputs and farm outputs to processed products. This simple approach shows clearly how changes in the water price translate directly into changes in the market prices for grapes, wine, milk, and cheese. These price increases mean less quantity demanded and less production in California.

The more that farms can use capital and innovations to substitute away from increasingly costly water, the more these major impacts can be mitigated. For example, if higher water costs cause more sub-surface drip irrigation use to reduce water per ton of alfalfa hay, the smaller the impact on the price of hay and therefore a smaller resultant cost increase for of milk and cheese. Similarly, if grapevine root innovations can reduce the amount of water used per ton of grapes—and, therefore, reduce the water used per liter of wine—then the smaller will be the impact of higher water prices on that of wine. As has been true for more than a century, innovation is the key for sustainability of California agriculture.

References

Anderson, N. M. & Sumner, D. A. (2016). Which California Foods You Consume Makes Little Impact on Drought-Relevant Water Usage. *ARE Update, 19*(3), 5–8, University of California Giannini Foundation of Agricultural Economics. https://s.giannini.ucop.edu/uploads/giannini_public/f3/91/f39155ce-574f-43f8-9a23-6068128e4180/v19n3_2.pdf

California Department of Food and Agriculture. (2013). *Grape crush report final 2012*. https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Grapes/Crush/Final/2012/201203gcbtb00.pdf

California Department of Food and Agriculture. (2014). *Manufacturing cost annual California 2014 data. 22 July 2016*. https://www.cdfa.ca.gov/dairy/pdf/Annual/2015/ManufacturingCostAnnual2014Data.pdf

California Department of Food and Agriculture. (2015a). *California Dairy Statistics Annual 2015. 8 August 2016*. https://www.cdfa.ca.gov/dairy/pdf/Annual/2015/2015_Statistics_Annual.pdf

California Department of Food and Agriculture. (2015b). *2013–2015 holstein feed summary excel file*. http://www.cdfa.ca.gov/dairy/uploader/postings/feedsummarydata/Default.aspx

California Department of Food and Agriculture. (2016). Minimum prices for class 2, 3, 4a, and 4b market milk—F.O.B. In *Processing Plant with Commodity Prices Used to Calculate These Minimum Prices Pursuant to the Stabilization and Marketing Plans for Market Milk (Plans)*. Released Oct 3 2016. https://www.cdfa.ca.gov/dairy/uploader/docs/October%20&%20November%202016%20Class%202,%203%20and%20September%202016%20Class%204a,%204b.pdf

CH2MHill. (2012). Draft Agricultural Economics Technical Appendix. A draft technical memorandum prepared for the California Department of Water Resources by CH2MHill. Aug 1, 2012. https://water.ca.gov/LegacyFiles/economics/downloads/Models/SWAP_TechAppendix_080612_Draft.pdf

Draper, A. J., Jenkins, M. W., Kirby, K. W., Lund, J. R., & Howitt, R. E. (2003). Economic-engineering optimization for California water management. *Journal of water resources planning and management, 129*(3), 155–164.

Gabrielyan, G. T. (2018). *Wine trade and economics of import duty and excise tax drawbacks*. Ph.D. dissertation. Davis: Department of Agricultural and Resource Economics, University of California.

Hanak, E., & Jezdimirovic, J. (2016). *California's water market: A report of the PPIC WATER POLICY Center*. San Francisco: Public Policy Institute of California (PPIC). https://www.ppic.org/publication/californias-water-market/

Hanak, E., & Stryjewski, E. (2012). *California's water market, by the numbers: Update 2012*. A report of the PPIC Water Policy Center. Public Policy Institute of California (PPIC). San Francisco. https://www.ppic.org/content/pubs/report/R_1112EHR.pdf

Hanak, E., Mount, J., Lund, J., Ajami, N., Allaire, M., Baerenklau, K. et al. (2018a). *California's water: A report of the PPIC water policy center*. Public Policy Institute of California (PPIC). San Francisco. https://www.ppic.org/wp-content/uploads/californias-water-november-2018.pdf

Hanak, E., Escriva-Bou, A., Jezdimirovic, J., Green, S., Harter, T., & Lund, J. et al. (2018b). *California's water: Water for farms. A report of the PPIC water policy center*. San Francisco: Public Policy Institute of California (PPIC). https://www.ppic.org/publication//californias-water-water-for-farms/

Howitt, R. (1995a). Positive mathematical programming. *American Journal of Agricultural Economics, 77*(02), 329–342.

Howitt, R. E. (1995b). A calibration method for agricultural economic production models. *Journal of Agricultural Economics, 46*(2), 147–159.

Howitt, R. E., Medellín-Azuara, J., MacEwan, D., & Lund, J. R. (2012). Calibrating disaggregate economic models of agricultural production and water management. *Environmental Modelling and Software, 38,* 244–258. https://doi.org/10.1016/j.envsoft.2012.06.013.

Howitt, R. E., MacEwan, D., Medellín-Azuara, J., Lund, J. R., & Sumner, D. A. (2015). Economic Analysis of the 2015 Drought for California Agriculture. Center for Watershed Sciences, University of California Davis, Davis, CA. https://watershed.ucdavis.edu/files/biblio/Final_Drought%20Report_08182015_Full_Report_WithAppendices.pdf

Klonsky, K. M., Mitchell, J., & Stewart, D. (2015). Sample costs to produce silage corn conservation tillage in the northern san joaquin valley. UC Cooperative Extension. https://coststudyfiles.ucdavis.edu/uploads/cs_public/a9/c5/a9c55e16-83cf-46f3-86b8-436ae2deb41f/15ctsilagecornsanjoaquinvalleyfinaldraftjuly20.pdf

Lapsley, J., & Sumner, D. A. (2016). *Water into wine: Myths and realities regarding California grapes and the drought*. Working Paper. Agricultural Issues Center, University of California.

Lee, H., Sumner, D. A., & Howitt, R. E. (1999). *Economic impacts of irrigation water cuts in the Sacramento Valley*. Agricultural Issues Center: University of California.

Lee, H., Sumner, D. A., & Howitt, R. (2001). Potential economic impacts of irrigation-water reductions estimated for Sacramento Valley. *California Agriculture, 55*(2), 33–40.

Long, R., Leinfelder-Miles, M., Putnam, D. H., Klonsky, K. & Stewart, D. (2015). Sample Costs to Establish and Produce Alfalfa Hay in the Sacramento Valley and Northern San Joaquin Valley Flood Irrigation. UC Cooperative Extension. https://coststudyfiles.ucdavis.edu/uploads/cs_public/39/f2/39f29aa5-b991-4a13-816e-c695ed243249/alfalfa-flood-sv-2015.pdf

Matthews, W. A., Gabrielyan, G. T., Putnam, D. H., & Sumner, D. A. (2016). The role of California and Western US dairy and forage crop industries in Asian Dairy Markets. *International Food and Agribusiness Management Review: Special Issue, 19B,* 147–162.

Smith, R. J., Klonsky, K. M. & De Moura, R. L. (2010). Sample costs to establish a vineyard and produce winegrapes cabernet Sauvignon in North Coast Region Sonoma County. UC Cooperative Extension. https://coststudyfiles.ucdavis.edu/uploads/cs_public/33/41/3341f086-690a-4d70-a113-c864e26614de/2016winegrapessonomafinaldraft111816.pdf

Chapter 11
Climate Policies as Water Policies

Kazim Konyar and George Frisvold

Abstract This study uses an updated version of the U.S. Agricultural Resource Model (USARM)—a multi-region U.S. agricultural sector programming model—to examine effects of climate change mitigation policies on U.S. water resources. One scenario considers effects of increasing prices of energy and energy-intensive inputs (primarily fertilizers) through a carbon tax or cap-and-trade program. A second scenario combines the first scenario with an agricultural offset program where farmers are paid to retire cropland for carbon sequestration. The consequences of climate mitigation policies for agricultural water use and pollution control have received relatively little attention in part because—unlike USARM—many national agricultural sector models do not explicitly include water as an input. USARM also allows for input substitution among seven inputs in a CES framework, while accounting for all major crops as well most specialty crops, federal commodity programs, and crop exports. Major results are as follows. First, climate mitigation policies have scope to significantly reduce agricultural water use. Whether domestic offsets are included has little effect on the total amount of water conserved, but has a large effect on which parts of the country the conservation takes place. Second, either carbon taxes or cap-and-trade combined with domestic offsets combines two policies often modeled as potential solutions to the hypoxic "dead zone" in the Gulf of Mexico—increased fertilizer prices and land retirement. Climate policies may have unanticipated, near-term, environmental benefits by addressing the hypoxia problem. Third, while domestic offsets reduce total fertilizer and agricultural chemical use, they increase their use per acre. Particularly in watersheds with significant land retirement, there could be unintended intensive margin effects where fertilizer and chemical use are increased. Despite this last, cautionary finding, a key insight into decision makers is that climate policies can have unanticipated, near-term benefits of

K. Konyar (✉)
Department of Economics, California State University at San Bernardino, San Bernardino, USA
e-mail: KKonyar@csusb.edu

G. Frisvold
Department of Agricultural and Resource Economics, The University of Arizona, Tucson, USA
e-mail: frisvold@ag.arizona.edu

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_11

water pollution control and water conservation that could be included in benefit-cost analyses of climate policy proposals.

1 Introduction

In 2009, the U.S. House of Representatives passed the American Clean Energy and Security Act (ACESA) of 2009 (H.R. 2454, known as the Waxman-Markey bill). H.R. 2524 would impose a “cap-and-trade” system on carbon dioxide and other greenhouse gas (GHG) emissions. While the effects of H.R. 2454 on agricultural markets have received considerable attention, there has been little consideration of effects on agricultural water use and water quality. This contribution seeks to fill this gap in assessment of policies to mitigate climate change. H.R. 2524 has three features with important implications for agriculture and water resources. First, agricultural practices or GHG emissions are not regulated directly. Second, however, it would increase the costs of fossil fuels and energy-intensive products and production practices. It would significantly increase costs of fertilizers and irrigation water pumping and affect water demand (for irrigation) and quality (via fertilizer and pesticide use). Third, farmers would be able to earn offset credits by reducing carbon emissions or sequestering carbon, either in soils or through tree planting. The most likely means of earning offset credits would be for farmers to plant trees on cropland (afforestation) to sequester carbon (Brown et al. 2010). Offsets would become a new source of income for agricultural landowners. By taking cropland out of production, it would also act as a supply-control program, raising the price of agricultural commodities. Even growers who did not participate in an offset program would receive higher output prices as an indirect result. In sum, farmers’ practices would not be directly regulated, but farmers would face higher input costs. Offset sales would be a new source of income. Afforestation under the offset program would limit supply and increase agricultural prices.

Under H.R. 2454’s cap-and-trade policies, industrial manufacturers and energy utilities would need a permit for every unit of GHGs they emit. Regulations would cap the total number of permits issued and reduce the number of permits issued over time. Generators of greenhouse gases would have to limit their emissions consistent with the permits they own or purchase additional permits from willing sellers. Anyone selling permits would need to cut emissions accordingly. Firms that can cut emissions at low costs can sell permits to firms with high costs of cutting emissions. Tradable permit systems can substantially reduce costs of cutting emissions compared to command-and-control methods that require industries to adopt specific technologies or that set emission limits at the industry or plant level (Field and Field 2009). The USA has implemented tradable permit systems for other air pollutants (e.g., sulfur dioxide and nitrous oxides) (Schmalansee and Stavins 2013; Goulder 2012) and water pollutants (Fisher-Vanden and Olmstead 2013). Trading systems for carbon have been initiated in California, the Northeast United States, the EU, Quebec, Australia, New Zealand, and South Korea (Newell et al. 2013). A critical issue in cap-and-trade policy design is how emission permits are distributed. As the government distributes fewer permits over time, permits become increasingly valuable. If the government auctions off all permits, the market-clearing price would be equivalent to a carbon tax rate required to cut emissions to the permitted level. Cap-and-trade policies provide greater incentives for reducing emissions by equating marginal abatement costs, which reduces overall compliance costs. It also creates greater incentives to innovate to reduce emissions. In these two respects, cap-and-trade policies are superior to command-and-control approaches of uniform emission mandates or reliance on technology standards and similar to the effects of a carbon tax (Field and Field 2009; Goulder 2013). A significant distinction between cap-and-trade and carbon taxes is uncertainty about quantities and prices. Under a carbon tax, the price of carbon (the tax rate) is known, but the actual level of emission reduction is uncertain ex ante. Under cap-and-trade, the level of emissions is set, but the price of carbon (price of emission allowances) is uncertain. In tradable permit markets, permit prices can be highly volatile. This volatility could present challenges for agricultural offset markets that would rely on long-lived decisions about tree planting. A full discussion of the scope and limits of a cap-and-trade system is beyond the scope of this contribution. We direct interested readers to papers published in a symposium on Tradable Pollution Permits published in the *Journal of Economic Perspectives* (Fisher-Vanden and Olmstead 2013; Goulder 2013; Newell et al. 2013; Schmalansee and Stavins 2013).

While other climate change bills have been introduced in the USA, H.R. 2454 is the first passed by any house of Congress. Consequently, potential impacts H.R. 2454 might have on U.S. agriculture have received considerable attention (USDA, OCE 2009a, b; FAPRI 2010; Brown et al. 2010; Baker et al. 2009; De la Torre Ugarte et al. 2009; Pan et al. 2011; Gramig 2012; Golden et al., 2009 survey several earlier studies). These studies have examined impacts on agricultural production, prices, exports, and farm income, but have not addressed effects on agricultural water use or potential impacts on water quality. H.R. 2454 has never come up for a full vote in the U.S. Senate and is not scheduled to do so in the near future. However, given continuing policy debates over federal deficits and the national debt, the auctioning of carbon permits or carbon taxes has been continually discussed as a future source of revenue and as an alternative to other forms of taxation (Metcalf 2010; Gale et al. 2013; Rausch and Reilly 2012).

This study uses the U.S. Agricultural Resource Model (USARM) a multi-region programming model of the U.S. agricultural sector to examine effects of H.R. 2454 on U.S. agriculture. It differs from previous research in that we explicitly examine how H.R. 2454 might affect national and regional water use by agriculture. We also consider the implications of changes in fertilizer, pesticide, and land use for water quality. Konyar (2001) illustrated how U.S. adherence to the Kyoto Protocol on climate change would have significantly altered U.S. fertilizer, pesticide, and water use. Though not addressing H.R. 2454 specifically, Pattanayak et al. (2002) and Feng et al. (2003) examined how carbon sequestration may provide environmental "co-benefits" of reduced fertilizer and sediment pollution of water bodies.

This study also differs from previous research in the scenarios used for future energy prices. Previous agricultural studies have relied on price estimates from U.S. Environmental Protection Agency (EPA) analysis (U.S. EPA 2009) or from the "basic" scenario developed by Energy Information Administration of the U.S. Department of Energy (U.S., DOE-EIA 2009). The EIA basic scenario projects higher energy price increases than does the EPA scenario, but also assumes "key low-emissions technologies, including nuclear, fossil with carbon capture and storage (CCS), and various renewables, are developed and deployed on a large scale … without encountering any major obstacles" (p. viii). The EIA report acknowledges, "There is great uncertainty about the costs of these technologies, as well as the feasibility of introducing them rapidly on a large scale" (p. 6). However, safety concerns raised by the Fukushima nuclear disaster in 2011 may be just such a "major obstacle" to deploying nuclear power "rapidly" in the USA. Rapid deployment of large-scale solar plants has also met with political opposition (Schwartz 2011). Given uncertainties about non-fossil fuel technologies, we assume energy prices will change according to the EIA "high-cost" scenario, where costs of nuclear, coal with CCS, and biomass technologies are assumed 50% higher than in the basic scenario. The effects of H.R. 2454 on energy markets would be phased in slowly over time. Some previous analyses suggest that near-term impacts on agricultural input markets would be relatively modest (USDA, OCE 2009a, b). In particular, there are additional provisions that would limit increases in fertilizer costs until 2025. Our interest is in impacts of H.R. 2454 once provisions have been fully phased in, so the analysis focuses on impacts as of 2030.

2 USARM Model

USARM is a 32-commodity, 12-region U.S. agricultural sector model. Earlier versions have been applied to examine impacts of U.S. commodity and conservation programs, agricultural biotechnology adoption, water shortages in the Western U.S., and U.S. adherence to the Kyoto Protocol (Howitt 1991; Ribaudo et al. 1994; Konyar and Howitt 2000; Konyar 2001; Kim et al. 2002; Frisvold and Konyar 2012). USARM is a nonlinear, mathematical programming model designed to simulate farmer behavior under external market and policy shocks. It includes 10 major field and 22 fruit and vegetable crops (Table 1).

Nested constant elasticity of substitution (NCES) production functions determine each production activity, with separate functions for irrigated and dryland crops. Use of seven inputs—land, water, labor, capital, fertilizer, agricultural chemicals, and energy/other inputs—is determined endogenously for each activity. The NCES functions allow for different substitution possibilities between inputs, allowing acres planted and per-acre input use to vary. Deficit irrigation is possible where less water is applied per acre. This reduces costs, but lowers yields. Each production activity has a cost function that is quadratic in the land input, but linear in all the others. The quadratic function captures the fact that as more land is allocated to a specific

Table 1 Commodity and regional groupings in USARM model

Field crops		
Barley	Rice	Sugar cane
Corn	Sorghum	Soybeans
Cotton	Sugar beets	Wheat
Alfalfa hay		
Fruit, vegetable, and nut crops		
Almond	Raisin Grapes	Pears
Apple	Green Beans	Peppers
Asparagus	Green Peas	Potatoes
Broccoli	Lettuce	Strawberry
Cauliflower	Melons	Tomato, Fresh
Citrus	Onions	Tomato, Processing
Cucumber	Peaches	Walnuts
Grapes		

Production region	States
Appalachian	Kentucky, North Carolina, Tennessee, Virginia, West Virginia
California	California
Corn Belt	Illinois, Indiana, Iowa, Missouri, Ohio
Delta States	Arkansas, Louisiana, Mississippi
Lake States	Michigan, Minnesota, Wisconsin
Northeast	Connecticut, Delaware, Maine, Maryland, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
Northern Mountain	Idaho, Montana, Wyoming
Northern Plains	Kansas, Nebraska, North Dakota, South Dakota
Pacific Northwest	Oregon, Washington
Southeast	Alabama, Georgia, Florida, South Carolina
Southern Mountain	Arizona, Colorado, Nevada, New Mexico, Utah
Southern Plains	Oklahoma, Texas

crop, the marginal cost increases as marginal lands with lower yield potential come into production. It also allows for the exact calibration of the model solutions to base year crop acreage following Howitt's (1995a, b) positive mathematical programming (PMP) method. More details of USARM equation specification and data sources are provided in Frisvold and Konyar (2012).

USARM divides the USA into 12 production regions (Table 1). Regions are modeled as aggregate farm units producing crops in their respective areas. Regional agriculture adjusts to shocks by changes at the extensive margin (change in total acres planted) and the intensive margin (choices of crops to grown, whether or not to irrigate, how much to irrigate, and application rates for other inputs).

The model includes domestic and export demand equations for each crop that (along with production behavior) endogenously determine crop prices. USARM accounts for federal commodity programs. Loan deficiency payments (LDPs), when in effect, are coupled payments that encourage greater production at the margin. Countercyclical payments (CCPs) are based on the difference between the market price and a target price, but payments are based on historic, rather than current, production. CCPs do not affect marginal production decisions, but aggregate production alters the level of CCPs via its impact on market prices.

Variable inputs (fertilizer, chemicals, labor, capital, and energy/other) are supplied elastically at fixed national prices. Initially, total dryland and total irrigated land uses are calibrated to match actual total regional land use patterns in each region in the base period. For each region, an upward-sloping supply curve for agricultural land allows land rent to be determined endogenously for each region. Total water applications to agriculture in each region are subject to a regional water constraint in the base scenario, consistent with empirical findings that irrigation water use is frequently quantity-constrained (e.g., Moore and Dinar 1995). For each region, the cost of water is a weighted average, capturing the purchase price of water and the cost of water pumping and accounting for regional differences in surface and groundwater use.

3 Market and Policy Shocks

3.1 2030 Baseline

USARM was originally calibrated to match acreage and price data for field crops in 2002 and specialty crops in 2000. Because our interest is the effects of climate legislation in 2030, we develop a new baseline to reflect projected market conditions in that year. This was done in two steps. First, crop demand curves and variable input prices were shifted to replicate relative output prices, input prices, and production levels in 2012. Next, 2012 values were adjusted based on data from various projections. These included USDA's *Agricultural Baseline Projections to 2022*, the Food and Agricultural Policy Research Institute's (FAPRI) *U.S. and World Agricultural Outlook (with projections to 2022),* a FAPRI (2010) analysis of H.R. 2454, which included projections to 2030, and energy price projections to 2030 of the Energy Information Administration (EIA) of the U.S. Department of Energy. The changes in relative agricultural prices from 2002 to 2012 were greater than projected changes to 2022. For our analysis, we assumed minimal changes in relative prices from 2022 to 2030.

There are some significant differences between the 2002 baseline and the updated 2030 baseline. The first is the increase in feed grain and oilseed prices along with higher costs of energy and energy-intensive inputs, such as fertilizer. The second difference is that U.S. agriculture uses more land, but less water. Water use in the

Southeast, Southern Plains, Southern Mountain States, Pacific Northwest, and California all fall below 2002 levels, while regional water constraints remain binding elsewhere in the USA. More land comes into production, primarily in the Corn Belt, Lake States, and Northern Plains. Dryland acreage increases while irrigated acreage declines nationally.

These projections are consistent with some projections from some studies, but not others. Between the 2002 and 2007 *Census of Agriculture*, harvested cropland increased 2%. Since 2002, there are nearly 7 million fewer acres of cropland retired under the Conservation Reserve Program as higher commodity prices have induced farmers to place more land under cultivation. The USARM simulation projects a 0.35 million acre-foot (MAF) reduction in agricultural water demand in the Southern Mountain Region for 2030. This is reasonably close to the 0.3–0.5 MAF range reductions projected for 2035 for the same area in the Bureau of Reclamation's *Colorado River Basin Water Supply and Demand Study*. USARM projects irrigation water demand for the Southern Plains (Texas and Oklahoma) falling by 14% from 2002 to 2030. The Texas Water Development Board has projected an 8% reduction from 2010 to 2030 (TWDB 2012). The TWDB projections have been criticized, however, for overstating future irrigation water demands (Ball and Kelly 2003). Brown (1999) projected reductions in the Arkansas-White-Red River Basin and Texas Gulf Coast Basins of 9% from 2005 to 2030 and a 12% reduction in the Rio Grande Basin for the same period. These basins roughly correspond to the Southern Plains. Brown's (1999) estimates for the basins most comparable to the Southern Mountain Region project declines of 3% from 2005 to 2030 compared to the USARM projections of 4% from 2002 to 2030. Brown projected reductions of 11% in the Pacific Northwest compared to USARM's 13%. Houston et al. (2003) projected a more modest decrease for the Pacific Northwest, 4% from 2000 to 2030. Most projections for California suggest greater reductions than the USARM projections of 2%. For California, Houston et al. project a decline of 7%, while Brown projects a decline of 4%. The Pacific Institute (Gleick et al. 2005) and the California Department of Water Resources (Groves et al. 2005) project reductions of 5–10%. For all the regions that USARM projects the regional water supply constraint to be binding, Brown projects increases in demand. Our projections treat those demands as supply-constrained.

Our projections diverge from some other projections for the Southeast. The USARM projections suggest a decline in agricultural water demand of 11% from 2002 to 2030. The Florida Department of Environmental Protection (Florida DEP 2010) projected that irrigation demand would rise 6% from 2005 to 2030, but revised this estimate to a <1% increase the next year (Florida DEP 2011). Hutson et al. (2004) projected a 37% increase in water use in the Tennessee Valley from 2000 to 2030. However, according to the two most recent USDA Farm and Ranch Irrigation Surveys, irrigation water applications fell by 19% between 2003 and 2008 in the Southeast. According to the two most recent USGS *Water Use in the United States* reports, water withdrawals in the Southeast fell 27% from between 2000 and 2005 (Hutson et al. 2005; Kenny et al. 2009). These are sizeable reductions, so even with significant growth in demand between 2005–2008 and 2030, Southeast water demands could quite conceivably remain 11% below 2002 levels. Revised projections of Tennessee Valley Authority irrigation water use show an actual drop of 51% from 2000 to 2010, but project a 35% increase from 2010 to 2035, for a net reduction of 33% from 2000 to 2010 (Bohuc and Bowen 2012).

3.2 Cap-and-Trade Without Carbon Offsets

This scenario considers the impacts of the cap-and-trade portion of H.R. 2454 without the possibility that agricultural producers can earn carbon offsets by planting trees on cropland (afforestation) to sequester carbon and selling the offsets for a return. Here, the effect of cap-and-trade is simply to increase input prices. Energy price increases were based on the "high-cost" scenario used by the EIA to evaluate effects of H.R. 2454 (EIA 2009). Industrial natural gas prices increase by 39.9%. Using estimates from Huang (2007) the elasticity of fertilizer price with respect to natural gas price is about 0.8, so it is assumed that cap-and-trade increases average fertilizer price by 31%. Following procedures used in Konyar (2001) to map energy price changes to agricultural chemical price changes, agricultural chemical prices are assumed to increase by 4%. Increases in diesel fuel, gasoline, propane, and electricity affect the price of the energy/other input categories in two ways. First, there are increases in direct costs for purchasing energy inputs. Second, there are costs of hauling and other custom services that increase with energy costs and fall under the category of other inputs. It is assumed that the price of energy/other increases by 45%. The costs of irrigation pumping are treated separately from other energy-related costs and are specific to each of the 12 regions in the model. Pumping costs per acre-foot depend on depth to water, relative reliance on surface versus groundwater, and relative reliance on different fuel sources to pump water. The costs of energy price shocks were computed based on data from the 2008 Farm and Ranch Irrigation Survey. Implicitly, this assumes that fuel sources for pumps in each region are used in fixed proportions and does not allow for fuel substitution for irrigation pumping. This lack of substitution may overstate effects on irrigation pumping costs. Another limitation of this approach is that pumping costs are not tied to changes in the groundwater table through a hydrological model. The increase in pumping costs per acre-foot ranges from 21% in the Southern Plains to more than 41% in the Northeast (Table 2).

3.3 Cap-and-Trade with Carbon Offsets

In this scenario, input prices increase as in the previous scenario, but farms can now earn and sell carbon offset credits. Producers may obtain payments for taking land out of crop production to sequester carbon through afforestation. Cropland is taken out of production to varying degrees in different regions. Different areas have different potentials to sequester given amounts of carbon. If industrial emitters pay

Table 2 Simulated input price increases (%) from cap-and-trade and acres of cropland converted to forest (million acres) from carbon offset program

Input/region	Percent change price	Million acres of cropland converted
Agricultural chemicals	4	
Fertilizer	31	
Capital	0	
Labor	0	
Energy/other	45	
Irrigation pumping costs (per acre-foot)		
Appalachia	41	0
California	32	0
Corn Belt	39	25
Delta	41	16
Lake States	35	2
Northeast	41	0
Northern Mountain	30	0
Northern Plains	38	0
Pacific Northwest	30	1
Southeast	39	2
Southern Mountain	28	0
Southern Plains	21	5
U.S. total		51

farmers to sequester carbon, they will have an incentive to pay for such sequestration where it will achieve the most carbon reduction per dollar spent. Brown et al. (2010) developed regional estimates of acreage afforested by 2023 based on initial analysis by the EPA (2005), which projected 50 million acres converted from cropland to forests. Areas identified with the most scope for afforestation correspond closely with those identified in separate analysis by McNulty et al. (2011). Brown et al. (2010) estimated that the Delta and Corn Belt states had the largest absolute potential for sequestering carbon. Agricultural landowners may currently receive payments for planting trees on cropland under USDA's Conservation Reserve Program. Currently, about 153,000 acres are enrolled for bottomland timber establishment. Of these, 67% are in the Delta and 14% are in the Corn Belt. About 250,000 acres are enrolled for longleaf pine establishment, 71% of which is in the Southeast and 10% is in the Delta (USDA, FSA 2013). We do not explicitly model receipt of offset payments and acreage reallocation by region. Rather, the simulation removes land from crop production by the same proportion as the Brown et al. (2010) analysis. The percentage of total cropland retired for carbon sequestration is 89% in the Delta, 26% for the Corn Belt, 11% in the Southern Plains, 4% for the Southeast and Pacific Northwest, 2% in the Lake States, and zero elsewhere. Our simulation initially removes 51 million acres from crop production. The ultimate number of acres reduced in the simulation differs slightly from this initial shock. This occurs because taking 51 million acres out of production reduces agricultural production and raises agricultural prices. Higher production prices induce some producers to put more land back into production.

4 Results

4.1 Output, Prices, and Revenue

Production of all crops declines under the cap-and-trade policy (Table 3).

Including carbon offsets leads to further reductions in output for every crop. For many specialty crops, the differences are less than 0.1%, however. Rice is the crop most affected by the climate policies, with production declining as much as 70% under offsets. Brown et al. (2010) and USDA-OCE (2009b) also found that rice was the crop whose output was most affected by H.R. 2454. Our analysis assumes larger fuel, and fertilizer price increases than those studies. Compared to most crops, rice production has relatively high fertilizer costs per acre and per-acre fuel costs 2–4 times other crops (USDA-OCE 2009b). Although rice faces the greatest output reduction, the USA sells most of its rice in the world market and faces a relatively elastic demand curve. The price of rice only rises 2–3% under the climate policies. The sharp input price increases, and limited scope for passing those costs on to consumers causes a sharp contraction in U.S. production in general and Delta production in particular. Prices of corn, soybeans, and sugarcane rise by the highest percentages. For most crops, the percent increase in price is greater than the percent reduction in output, so that gross revenues increase. Important exceptions are barley, cotton, rice, sorghum, wheat, and citrus. For most specialty crops, adding carbon offsets to the cap-and-trade policy has little additional effect on production, prices, or revenue.

4.2 Water Use Trends, 2002–2030

Table 4 shows changes in regional water use compared to the original 2002 model baseline.

With no climate policy, irrigation demand is projected to decline 3% nationally by 2030, with demands falling in five of twelve regions. The cap-and-trade provisions by themselves accelerate this process, so that by 2030, national agricultural water demand falls by 15%, compared to 2002. In half the regions, irrigation water demand falls by 19% or more below 2002 levels. Nationally, water use is slightly higher under offsets, so that water use falls 13% from 2002 levels. This occurs because although acreage is taken out of production, higher output prices encourage greater water use per acre on remaining cropland.

Table 3 Effects of climate change policy on agricultural output, prices, and gross revenues (% change from 2030 baseline)

Crop	Output		Price		Gross revenues	
	Cap-and-trade (%)	Cap-and-trade with offsets (%)	Cap-and-trade (%)	Cap-and-trade with offsets (%)	Cap-and-trade (%)	Cap-and-trade with offsets (%)
Alfalfa	−4.5	−10.3	5.2	11.8	0.4	0.3
Barley	−14.8	−18.1	7.3	9.0	−8.5	−10.8
Corn	−5.5	−14.2	15.6	40.2	9.2	20.3
Cotton	−4.0	−16.9	2.6	11.1	−1.5	−7.7
Rice	−45.2	−70.1	2.1	3.3	−44.0	−69.1
Sorghum	−24.5	−33.1	4.5	6.1	−21.1	−29.0
Soybeans	−2.0	−14.9	4.9	36.5	2.8	16.2
Sugarcane	−4.7	−10.1	21.2	45.1	15.5	30.5
Sugar beets	−2.3	−3.1	11.6	15.7	9.0	12.0
Wheat	−18.6	−24.5	6.8	9.0	−13.0	−17.7
Almonds	−3.9	−3.9	3.1	3.1	−0.9	−0.9
Apples	−2.6	−3.0	4.1	4.8	1.4	1.6
Asparagus	−2.3	−2.6	3.1	3.5	0.7	0.8
Broccoli	−0.6	−0.6	13.4	13.4	12.7	12.7
Cauliflower	−5.7	−5.7	13.4	13.5	7.0	7.0
Citrus	−5.1	−5.1	4.7	4.7	−0.6	−0.6
Cucumbers	−2.6	−3.1	6.9	8.1	4.1	4.8
Grapes	−3.3	−3.4	3.5	3.6	0.1	0.1
Grapes, raisin	−4.6	−4.6	5.2	5.2	0.4	0.4
Green beans	−5.7	−9.7	4.7	7.9	−1.3	−2.5
Lettuce	−3.3	−3.3	8.4	8.4	4.8	4.8
Melons	−0.4	−0.5	7.8	8.7	7.4	8.2
Onions	−1.8	−1.8	8.7	8.9	6.8	6.8
Peaches	−1.9	−2.1	1.8	2.0	−0.1	−0.1
Pears	−0.9	−0.9	2.8	2.9	1.9	1.9
Peas	−14.3	−19.6	6.3	8.7	−8.9	−12.6
Peppers	−3.0	−3.0	10.1	10.2	6.8	6.9
Potatoes	−3.5	−3.7	11.8	12.5	7.9	8.4
Strawberries	−5.2	−5.2	10.1	10.2	4.4	4.4
Tomatoes, fresh	−2.9	−3.0	10.2	10.4	7.0	7.1
Tomatoes, processed	−2.7	−2.7	9.0	9.2	6.1	6.2
Walnuts	−2.8	−2.8	8.9	8.9	5.9	5.9

Table 4 Percentage change in water use from 2002 baseline

Region	2030 baseline (no policy) (%)	2030 cap-and-trade (%)	2030 cap-and-trade with offsets (%)
Delta	0	−11	−48
California	−2	−19	−13
Southeast	−11	−27	−21
Southern Mountain	−4	−24	−8
Southern Plains	−14	−24	−7
Pacific Northwest	−13	−28	−18
Remaining Regions	0	0	0
U.S. total	−3	−15	−13

Including offsets affects regions differently. Water use falls dramatically in the Delta, where a huge share of acreage goes out of crop production. With no climate policies, Southern Plains water use is projected to decline 14% from 2002 levels. With cap-and-trade and afforestation, the decline is only 7%. The higher crop prices induce Southern Plains growers to demand more water than they would otherwise. For other regions where the regional water constraint is no longer binding, afforestation causes water use to fall less than under cap-and-trade alone. For six regions, the regional water constraint remains binding under all scenarios and water use remains unchanged. For California, the Pacific Northwest, the Southern Mountain States, and the Southeast, declines in water use from 2002 are greatest with cap-and-trade alone, followed by cap-and-trade with afforestation, then no climate policies.

4.3 *Agricultural Inputs*

Table 5 compares national input use under climate policies to a “no-policy” 2030 baseline.

Cap-and-trade alone has no effect on total cropland acres as reductions in irrigated cropland are balanced by increases in dryland acres. When effects of afforestation under an offset program are included, total cropland falls by >18% with larger percentage reductions in dryland than irrigated acreage. Final crop acreage falls by 49.4 million acres even though 51 million were originally converted to forestland. Higher farm prices encourage other acreages to come into production. Compared to the 2030 baseline, water use falls 11.5% under cap-and-trade and by 9.8% with afforestation. With afforestation, water use per irrigated acre increases by 5.4%. As cropland declines with afforestation, a higher share of acreage is irrigated and that share is irrigated more intensively.

Fertilizer and energy use decline significantly under both climate policy scenarios, again with slightly larger effects under cap-and-trade only. Fertilizer use falls

Table 5 Effects of climate policies on national use of agricultural inputs (percentage change from 2030 baseline)

Input	Cap-and-trade (%)	Cap-and-trade with offsets (%)
Cropland	0.0	−18.4
Irrigated cropland	−1.3	−14.4
Non-irrigated cropland	0.2	−19.1
Water	−11.5	−9.8
Water per acre	−11.5	10.5
Water per irrigated acre	−10.4	5.4
Fertilizer	−18.7	−16.3
Fertilizer per acre	−18.7	2.6
Agricultural chemicals	−0.5	−2.8
Agricultural chemicals per acre	−0.7	20.2
Capital	−0.3	−3.0
Labor	1.4	0.2
Energy/other	−20.4	−19.0

16–19%, as it is an energy-intensive input and thus sensitive to rising energy costs. Fertilizer use per acre increases slightly under afforestation as acreage falls more than fertilizer use. Reductions in agricultural chemicals are far more modest, but are larger under cap-and-trade with afforestation. In addition, under afforestation agricultural chemical use per acre increases significantly (by >20%). Konyar (2001) estimated that U.S. adherence to the Kyoto Protocol would reduce pesticide use more than would H.R. 2454, although both raise energy prices. That analysis was based on mid-1990 prices. Since then, agricultural chemical prices have fallen relative to average input prices by 23%, while fertilizer and fuel prices have risen by 34 and 51%. Thus, there is less incentive to switch away from agricultural chemicals and more incentive to switch from fertilizers and fuel.

4.4 Farm Income and Economic Surplus

Cap-and-trade by itself reduced grower returns per acre about $1/acre nationally, although effects ranged from losses of $20/acre in California to $4/acre gains in the Lake States and in the Pacific Northwest (Table 6).

Gains were <$0.50/acre in the Corn Belt and Northern Plains, but these two regions accounted for 57% of cropland. The effect of cap-and-trade with afforestation on per-acre returns was positive in all regions except California, with losses of $1/acre. Nationally, average returns rose by $75/acre. In the Delta, per-acre returns rose by $352/acre. However, recall that 89% of Delta cropland is converted to forestland in this scenario. The only cropland that remains under cultivation earns high returns.

Table 6 Effect of climate policies on net returns per acre (change in \$/acre from 2030 no-policy baseline)

Region	Cap-and-trade	Cap-and-trade with afforestation
Appalachian	−\$7	\$56
Corn Belt	\$0[a]	\$131
Delta	−\$12	\$352
California	−\$20	−\$1
Lake	\$4	\$123
Northeast	−\$4	\$50
Northern Mountain	−\$0[b]	\$9
Northern Plains	\$0[a]	\$53
Southeast	−\$17	\$19
Southern Mountain	\$2	\$36
Southern Plains	−\$2	\$21
Pacific Northwest	\$4	\$24
USA	−\$1	\$75

[a]gains < \$0.50/acre; [b]losses < \$0.50/acre

Table 7 Change in economic surplus in \$U.S. billions from climate policies

Change in	Cap-and-trade	Cap-and-trade with afforestation
Consumer surplus (CS)	−\$9.2	−\$27.0
Producer surplus (PS)	−\$0.3	\$10.0
Government payments (GP)	\$0.2	−\$1.7
Total surplus (CS + PS − GP)	−\$9.7	−\$15.3

While per-acre returns rose dramatically, total income from crop production fell by 63% in the region. Table 7 reports changes in economic welfare to producers and consumers nationally.

Under cap-and-trade alone, producer surplus (crop net income) falls by \$0.3 billion, while consumer surplus falls by \$9.2 billion. These latter are losses from having fewer agricultural commodities to buy and having to pay a higher price for them. Foreign importers of U.S. exports feel some of this loss, but most are felt by U.S. consumers. Consumers here may be best thought of as first purchasers of agricultural commodities. These are often agricultural producers themselves, specifically livestock, poultry, and dairy producers. Prices for feed grains (corn, sorghum, barley), soybeans, and alfalfa all increase substantially (Table 3). Most soybeans are processed for oil and protein for animal feed. Alfalfa hay is also a major expense in U.S. dairy production.

When the land retirement effects of carbon offsets on acreage and output prices are included, producer surplus increases by \$10 billion, while consumer surplus falls by \$27 billion. By taking land out of production, afforestation acts as a supply control program, raising agricultural prices and incomes. The positive effect of cap-and-trade with afforestation on crop producer income may seem counterintuitive to some readers. However, other studies have similarly found that policies that take agricultural land out of production (e.g., the CRP and wetland protection) have a similarly positive effect (Ribaudo et al. 1994: Claassen et al. 1998). Purchasers of agricultural commodities suffer from higher prices. Within the agricultural sector, livestock, poultry, and dairy producers will face losses from higher feed prices. The offset program also reduces government price support payments, by raising market prices. This would reduce the U.S. budget deficit by $1.7 billion. These payments are already accounted for in producer surplus but funded by taxpayers. Economic welfare measures net out this transfer of income.

The payments for carbon offsets to agricultural producers are not included in our welfare calculations. Although important to producers, they represent transfers from industries wishing to emit more GHGs. Thus, receipts by farmers will equal payments by emitters and be netted out of a social welfare calculation. Under the high-cost scenario, EIA (2009) projects the price of carbon offsets to be $72/ton. Brown et al. (2010) estimate that Delta farmers could sequester 6.3 ton of carbon per acre by converting cropland to forestland, while Corn Belt farmers could sequester 3.43 ton per acre. This would represent offset income of $250–$450 per acre per year. Including other regions, national agricultural offset income could exceed $20 billion. The one-time, up-front costs of planting trees would also be a consideration. Average Forest Service costs of establishing forest vegetation are >$500 per acre, but can be as low as $200 per acre (Gorte 2009).

However, total offset payments per acre and total absolute payments could be considerably less than the $20 billion figure above. Brown et al.'s (2010) simulation model results suggested that the target land retirement could be reached with offset payments of $103/acre. At this payment level, sequestering carbon through tree planting would yield higher per-acre returns that farm on over 25% of the region's cropland. Brown et al. (2010) compared this simulation result with regional cash rent values for cropland in the Corn Belt. They found that, "an offset value of $117 per acre, 25.3% of cropland in the Corn Belt would have an annual value in an afforestation program greater than its current value as estimated by cash rents" (p. 17). In USARM, afforestation payments needed to exceed per-acre returns on 25% of Corn Belt cropland would have to reach $131/acre at the margin. In the Delta, payment rates would need to reach $165/acre at the margin. Marginal rates in the Lake States would be comparable to the Corn Belt. Elsewhere, rates could be much lower (roughly half) to hit the regional acreage targets. Crop-specific, regional net returns in USARM suggest that offset payments of roughly $6 billion could be sufficient (assuming a single, regional offset price) to retire the 51 million acres for carbon sequestration.

The net cost of the climate policies (from impacts on agricultural producers, consumers, and taxpayers) ranges from $9.7 billion to $15.3 billion. Some of this cost will be passed on to foreign importers of U.S. exports. These cost estimates do not include the benefits of limiting effects of climate change, nor do they include benefits to industrial emitters of GHGs who are purchasing carbon offsets from farmers and reducing their regulation compliance costs. Such emitters will only purchase agricultural offsets if this is cheaper than reducing their own emissions or buying emission permits from other industrial emitters. Thus, our cost estimates do not include these “cost-savings” as benefits.

Also missing from these estimates are calculations of the costs of implementation, monitoring, enforcement, offset market establishment and participation and other transaction costs associated with the legislation. The U.S. Congressional Budget Office (CBO)-estimated administrative costs for federal agencies implementing H.R. 2454 would be $8.2 billion from 2010 to 2019 (CBO 2009). Their estimates relied on information provided by EPA and other agencies responsible for implementing the law. It is not clear from the CBO report the extent to which costs associated with a domestic offsets program were included. The Congressional Research Service (CRS) noted that transaction costs to farmers and others participating in offset markets could be substantial, especially in the first years of implementation (Johnson et al. 2010). CRS also noted that EPA and USDA analyses did not account for such transaction costs and criticized earlier analyses of H.R. 2454 for this omission.

Existing USDA land retirement programs may serve as a guide about the order of magnitude of these costs. The Conservation Reserve Program has retired between 30 and 36 million of acres of farmland. From 1996 to 2002, a USDA agency transaction costs averaged $93.3 million per year with cumulative acres enrolled averaging 32.4 million acres (Heimlich 2005). This is approximately $2.88 (1996 constant) per cumulative acres enrolled. Heimlich (2005) notes that per-acre costs fall over time because there are many up-front initiating costs, while agency staff “learn-by-doing” over the longer term, reducing implementation costs. At $3 per cumulative acre retired, a 50-million acre offset program could have $150 million per year in administrative costs. This may be a lower bound estimate of transaction costs, however. First, it does not account for farmer time spent in program participation. Second, a program of afforestation for long-term carbon sequestration may require longer-term easements to prevent land conversion. USDA administers the Wetlands Reserve Program (WRP) and the Healthy Forests Reserve Program (HFRP) that includes such easements. Heimlich (2005) notes the process of arranging for easements greatly increases the cost of administering the WRP relative to the CRP. The HFRP is at present a small-scale pilot program with only some states participating. To date, <250,000 acres are enrolled in the program. Payments for forest restoration are combined with payments for 10 years, 30 years, or permanent easements. Information about this program’s performance could be an important indicator of the nature and size of transaction costs associated with an offset program. Finally, the federal government administers these programs centrally although multiple agencies may be involved. If an offset program were more decentralized, then there will be transaction costs (particularly during initiation) as markets form so that willing buyers and sellers can negotiate terms. Federal agencies such as USDA may serve the role of monitoring and enforcing offset contracts.

5 Implications for Water Use and Pollution

5.1 Water Use

Climate policies have potentially large, underappreciated implications for water use, both positive and negative. With no climate policy, our projections suggest a decline in agricultural water use of 3% between 2002 and 2030. Under cap-and-trade alone, the decline would be 15%. With both cap-and-trade and the offsets, the decline is 13%. Reducing irrigation demand may reduce threats to aquatic species. Moore et al. (1996) note that in the 17 Western contiguous U.S. states, agriculture is reported as a "factor in decline" in federal decisions to list 50 fish as threatened or endangered under the Endangered Species Act (ESA). Irrigated agriculture in 235 counties relies on rivers with ESA-listed fish. Moore et al. (1996) also found a positive correlation between agricultural reliance on surface water and the number of ESA-listed species in a county. Reducing agricultural water demand may ease groundwater management in those areas where agriculture is a major source of aquifer depletion. Baker and Murray (2009) have modeled the issue from both directions, considering GHG reduction effects on groundwater management and how optimal groundwater management may reduce GHG emissions.

While climate policies may lower national irrigation water use and reduce competition for water in some regions, they may increase competition for water in other regions. Our analysis suggests significant regional differences in water conservation potential. In 6 of 12 U.S. regions, we found no water conservation effect of the climate policies. In the Southern Plains, the combination of cap-and-trade with offsets *increased* water use, relative to a no-policy baseline. Instead of irrigation water demands declining by 14% from 2002 to 2030, they decline by only 7% under cap-and-trade with offsets. By bidding up farm output prices, the offset program encourages greater production and resource use in the Southern Plains.

Another important area of research would be the water consumption implications of converting 51 million acres of cropland to forestland. Especially compared to dryland crops, forests can use more water and reduce surface runoff and aquifer recharge (Jackson et al. 2005; van Dijk and Keenan 2007; Dymond et al. 2012). Forest water use may come at the expense of agricultural users or wetland downstream (Dymond et al. 2012; Nordblom et al. 2012). This may present less of a problem in the Delta where cap-and-trade with afforestation reduced agricultural water use by 48%. However, our simulations assumed that landowners did not necessarily relinquish water rights with land conversion from cropland to forestland. Irrigation water use in the Corn Belt was initially assumed supply-constrained and continued to be so even with the climate policies. Under afforestation, irrigation water use remained constant absolutely, but water use per irrigated acre increased by >28%. To put this differently, under afforestation, the Corn Belt uses as much water for irrigated crops as it did before, but there is additional water consumption demand from 26 million acres of forestland. This suggests two extreme scenarios. If agriculture maintains all its pre-policy water use, afforestation would place significant (and likely unsustainable) pressure on regional water resources. If water rights were acquired from farmers to support all the new consumptive uses of forestland, this would decrease agricultural production in the Corn Belt significantly. Scenarios between these two extremes are possible, but the capacity to assess trade-offs between the two is limited by the state of knowledge of forest water use. This in turn is limited by the sheer number of factors affecting forest water use and limitations in the state of the art in regression-based and model-based methods of estimating forest–water relationships (van Dijk and Keenan 2007).

Water use considerations might become a factor in tree planting for carbon sequestration. The choice of tree species planted and the spatial configuration of tree plantings can influence impacts on streamflows (van Dijk and Keenan 2007). Dense forests of fast-growing tree species will use more water. For a given level of carbon sequestration, species and planting designs that reduce streamflows less may be preferable. At a regional level, water carbon trade-offs may become important considerations.

5.2 Water Pollution

By reducing national agricultural fertilizer use by roughly 6–19%, the climate policies may provide various water quality benefits. Nutrients (from fertilizers and manure) were top sources of impairment of 7% of U.S. river and stream miles, 12% of lake and reservoir acres, and 4% of square miles of bays and estuaries that were assessed in the EPA Water Quality Inventory (EPA 2004). Reducing fertilizer use can reduce pollution avoidance and water treatment costs. Consumers spend >$800 million each year on bottled water because of taste and odor problems associated with nutrients (Dodds et al. 2009). Ribaudo et al. (2011) estimated that agriculture's share of costs to water treatment plants of removing nitrate from U.S. drinking water supplies was about $1.7 billion per year. They further suggest that reducing nitrate concentrations in source waters by 1% would reduce U.S. water treatment costs by over $120 million per year.

H.R. 2454 would produce "co-benefits" of addressing problems of the hypoxic "dead zone" in the Gulf of Mexico. Fertilizer loadings (both nitrogen and phosphorus) have been identified as major contributors to hypoxia problems there. These loadings come from agricultural production throughout the Mississippi Basin, with much of it originating in the Corn Belt (Ribaudo et al. 2001). Studies considering ways to address this problem have focused on reductions in fertilizer use and land retirement and reductions in fertilizer use, applied to a large regional basis as 31 states drain the Mississippi Basin (Ribaudo et al. 2001; Pattanayak et al. 2002; Rabotyagov et al. 2010, 2012). By raising fertilizer prices, the cap-and-trade provisions of H.R. 2454 act as a fertilizer tax. Retiring land from crop production would further reduce fertilizer loadings. Planted forests contribute to reductions in nutrient runoff (Jackson et al. 2005; van Dijk and Keenan 2007), and "tree belts" have the potential to act as filter strips that can capture runoff and prevent nutrients and sediments from reaching water bodies (Ellis et al. 2006). Pattanayak et al. (2002) suggest that co-benefits of water pollution control from an offset program would be largest in Corn Belt, Delta, and Southern Plains states.

Converting cropland to forestland also has potential to reduce soil erosion. USDA's CRP already provides incentives to convert cropland to trees for soil erosion control and other environmental benefits. Tree planting accounts for only 6% of CRP acreage, where planting of legumes and grasses as cover crops dominates (USDA-FAS 2013). Estimates of water-related benefits of erosion control range from \$1.46 to \$7.12 per ton of sediment (Hansen and Ribaudo 2008). Feng et al. (2003), however, have illustrated that cropland targeted for carbon sequestration may provide quite different (and lower) erosion control benefits than acreage currently in the CRP or acreage targeted specifically for erosion control. The *types* of benefits under carbon offsets may be similar to those obtained under the CRP. However, the dollar value of per-acre benefits could be quite different.

The environmental implications of the climate policies on pesticide use are more problematic. Cap-and-trade alone produces only slight reductions in agricultural chemicals (herbicides, insecticides, fungicides). Cap-and-trade with offsets reduces agricultural chemical use less than 3%, but increases applications per acre >22%. As more acres are taken out of production, more pesticides are used on the remaining acres. Moreover, both pesticide and water use increase in remaining cropland. This may suggest increased runoff of pesticides on those acres that remained cropped.

6 Conclusions

This study used a mathematical programming model of the U.S. crop sector to examine how the proposed American Clean Energy and Security Act (H.R. 2454) would affect U.S. agriculture and water resources. Unlike many previous studies, we consider a more pessimistic scenario for deployment of non-fossil fuels, which leads to higher regulatory costs. By 2030, the bill's cap-and-trade provisions for greenhouse gases reduce U.S. irrigation water use by >11% and fertilizer use by >18% with positive implications for water conservation and quality. Carbon offset provisions reduce agricultural production and raise prices. Including the offset program causes the legislation to increase total U.S. crop producer income. However, not all crop producers benefit and feed grain purchasers (livestock, poultry, and dairy producers) would suffer from higher costs. The offset program has complex implications for water use and pollution that vary by region. In some areas, it reduces agricultural demand for water via land retirement. By raising agricultural prices, however, it creates incentives to apply more water per acre.

The simulation results also highlight several interesting questions that merit future research. First, with large-scale afforestation (on a scale of 50 million acres), how much will regional consumptive water demands increase? What is the scope and nature of trade-offs between carbon emission reduction and water conservation? Second, what is the value of co-benefits from water pollution reduction provided by policies to mitigate climate change? Our simulations suggest that these are potentially quite large. Yet, more research is needed to quantify the water pollution control benefits of climate policies. Third, how will climate policies interact with other agricultural conservation programs? Design of programs will provide different (and sometimes competing) incentives to limit carbon, reduce erosion, or reduce other water pollutants. Optimal design of water conservation or pollution control policies will depend on carbon abatement impacts, while optimal climate policies will be affected by water use and quality impacts.

Acknowledgements This work was supported by the National Oceanic and Atmospheric Administration's Climate Program Office through grant NA07OAR4310382 with the Climate Assessment for the Southwest Program at the University of Arizona.

References

Baker, J. S., McCarl, B. A., Murray, B. C., Rose, S. K., Alig, R. J., Adams, D., et al. (2009). *The effects of low-carbon policies on net farm income*. Working Paper 09–04, Nicholas Institute for Environmental Policy Solutions, Duke University.
Baker, J. S., & Murray, B. C. (2009). Groundwater management in the presence of greenhouse gas mitigation: Incentives for agriculture. In *Agricultural and Applied Economics Association, AAEA and ACCI Joint Annual Meeting*, Milwaukee, Wisconsin, July 26–29, 2009.
Ball, L., & Kelly, M. (2003). *Irrigation demand in Texas: An analysis of methodologies to predict irrigation trends*. Austin, TX: Environmental Defense.
Bohac, C. E., & Bowen, A. K. (2012). *Water use in the Tennessee Valley for 2010 and projected use in 2035*. Knoxville, TN: Tennessee Valley Authority, River Operations and Renewables.
Brown, T. C. (1999). *Past and future freshwater use in the United States: A technical document supporting the 2000 USDA forest service RPA assessment* (General Technical Report RMRS-GTR-39). U.S. Department of Agriculture. Fort Collins, CO: Forest Service, Rocky Mountain Research Station.
Brown, T., Elobeid, A., Dumortier, J., & Hayes, D. J. (2010). *Market impact of domestic offset programs*. CARD Working Paper 10-WP 502, Center for Agricultural and Rural Development, Iowa State University.
Claassen, R., Heimlich, R. E., House, R. M., & Wiebe, K. D. (1998). Estimating the effects of relaxing agricultural land use restrictions: Wetland delineation in the Swampbuster program. *Review of Agricultural Economics, 20*(2), 390–405.
Congressional Budget Office (CBO). (2009). *American clean energy and security act of 2009: Cost estimate*. June 5, 2009, As ordered reported by the House Committee on Energy and Commerce on May 21, 2009.
De la Torre Ugarte, D., English, B. C., Hellwinckel, C., West, T. O., Jensen, K. L., Clark, C. D., et al. (2009). *Implications of climate change and energy legislation to the agricultural sector*. Knoxville: University of Tennessee.
Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., et al. (2009). Eutrophication of U.S. freshwaters: Analysis of potential economic damages. *Environmental Science and Technology, 43*(1), 12–19.
Dymond, J. R., Ausseil, A.-G., Ekanayake, J., & Kirschbaum, M. U. F. (2012). Tradeoffs between soil, water, and carbon: A nationalscale analysis from New Zealand. *Journal of Environmental Management, 95*(1), 124–131.
Ellis, T. W., Leguėdois, S., Hairsine, P., & Tongway, D. (2006). Capture of overland flow by a tree belt on a Pastured Hillslope in South-Eastern Australia. *Australian Journal of Soil Research, 44*(2), 117–125.

Feng, H., Kling, C. L., & Gassman, P. W. (2003). Carbon sequestration, co-benefits, and conservation programs. In *Choices 3rd Quarter* (pp. 19–23).
Field, B. G., & Field, M. K. (2009). *Environmental economics: An introduction* (5th ed.). New York, NY: McGraw-Hill/Irwin.
Fisher-Vanden, K., & Olmstead, S. (2013). Moving pollution trading from air to water: Potential, problems, and prognosis. *Journal of Economic Perspectives, 27*(1), 147–171.
Florida (State of) Department of Environmental Protection (DEP). (2010). *Sustaining our water resources: Annual report on regional water supply planning 2010*. Tallahassee, FL: Department of Environmental Protection.
Florida (State of) Department of Environmental Protection (DEP). (2011). *Regional water supply planning annual status report 2011*. Tallahassee, FL: Department of Environmental Protection.
Food and Agricultural Policy Institute (FAPRI). (2010). *Impacts of climate change legislation on US agricultural markets: Sources of uncertainty FAPRI-MU report #06-10*. Columbia, MO: FAPRI, University of Missouri.
Frisvold, G. B., & Konyar, K. (2012). Less water: How will agriculture in southern mountain states adapt? *Water Resources Research, 48,* W05534. https://doi.org/10.1029/2011wr011057.
Gale, W., Brown, S., & Saltiel, F. (2013). *Carbon taxes as part of the Fiscal solution*. Washington, DC: Urban-Brookings Tax Policy Center.
Gleick, P. H., Cooley, H., & Groves, D. (2005). *California water 2030: An efficient future*. Oakland, CA: The Pacific Institute.
Golden, B., Bergtold, J., Boland, M., Dhuyvetter, K., Kastens, T., Peterson, J., & Staggenborg, S. (2009, December). A comparison of select cost-benefit studies on the impacts of H.R. 2454 on the agriculture sector of the economy. In *AgManager.info*. Manhattan, KS: Department of Agriculture Economics, Kansas State University.
Gorte, R. W. (2009, May). *U.S. tree planting for carbon sequestration* (CRS Report for Congress R40562). Washington, DC: Congressional Research Service.
Goulder, L. H. (2013). Markets for pollution allowances: What are the (new) lessons? *Journal of Economic Perspectives, 27*(1), 87–102.
Gramig, B. M. (2012). Some unaddressed issues in proposed cap-and-trade legislation involving agricultural soil carbon sequestration. *American Journal of Agricultural Economics, 94*(2), 360–367.
Groves, D. G., Matyac, S., & Hawkins, T. (2005). *Quantified scenarios of 2030 California water demand* (California Water Plan Update 2005). Sacramento, CA: California Department of Water Resources.
Hansen, L., & Ribaudo, M. (2008). Economic measures of soil conservation benefits: Regional values for policy assessment, TB-1922. In *USDA, Economic Research Service*, USA.
Heimlich, R. (2005). *The policy-related transactions costs of land conservation in the United States: Evolution and comparison between programs*. Paper presented at the OECD Workshop on Policy-Related Transaction Costs, Paris, January 20–21, 2005. Paris: Organisation for Economic Co-operation and Development.
Houston, L. L., Kline, J. D., & Alig, R. J. (2003). Economics research supporting water resource stewardship. In *General Technical Report PNW-GTR-550*. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station.
Howitt, R. E. (1991). Water policy effects on crop production and vice versa: An empirical approach. In R. E. Just & N. Bockstael (Eds.), *Commodity and resource policies in agricultural systems* (pp. 234–253). New York: Springer.
Howitt, R. E. (1995a). Positive mathematical programming. *American Journal of Agricultural Economics, 77*(2), 329–342.
Howitt, R. E. (1995b). A calibration method for agricultural economic production models. *Journal of Agricultural Economics, 46*(2), 147–159.
Huang, W. (2007). *Impact of rising natural gas prices on U.S. ammonia supply* (Outlook Report No. WRS-0702). Washington, DC: USDA Economic Research Service.

Hutson, S., Barber, N., Kenny, J., Linsey, K., Lumia, D., & Maupin, M. (2005). Estimated use of water in the United States in 2000. In *USGS Circular 1268*. Reston, VA: U.S. Geological Survey.

Hutson, S. S., Koroa, M. C., & Murphree, C. M. (2004). *Estimated use of water in the Tennessee River watershed in 2000 and projections of water use in 2030* (U.S. Geological Survey, Water-Resources Investigations Report 03–4302). Nashville, Tennessee.

Jackson, R. B., Jobbagy, E. G., Avissar, R., Roy, S. B., Barrett, D. J., et al. (2005). Trading water for carbon with biological sequestration. *Science, 310,* 1944–1947.

Johnson, R., Ramseur, J. L., Gorte, R.W., & Stubbs, M. (2010). *Potential implications of a carbon offset program to farmers and landowners* (CRS Report R41086). Washington, DC: Congressional Research Service.

Kenny, J. F., Barber, N. L., Hutson, S. S., Linsey, K. S., Lovelace, J. K., & Maupin, M. A. (2009). *Estimated use of water in the United States in 2005* (U.S. Geological Survey Circular 1344). Reston, VA: U.S. Geological Survey.

Kim, H., Konyar, K., & Sargent, K. (2002). *Economic viability of Bt-corn in the U.S.* Selected paper, presented at the American Agricultural Economics Association Annual Meeting, Long Beach, CA, July 28–31, 2002.

Konyar, K. (2001). Assessing the role of US agriculture in reducing greenhouse gas emissions and generating additional environmental benefits. *Ecological Economics, 38*(1), 85–103.

Konyar, K., & Howitt, R. E. (2000). The cost of the Kyoto protocol to U.S. crop production: Measuring crop price, regional acreage, welfare, and input substitution effects. *Journal of Agricultural and Resource Economics, 25*(2), 347–367.

McNulty, S., Sun, G., Myers, J., Cohen, E., & Caldwell, P. (2011). Robbing Peter to Pay Paul: Tradeoffs between ecosystem carbon sequestration and water yield. In K. W. Potter & D. K. Frevert (Eds.), *Watershed management 2010: Innovations in watershed management under land use and climate change* (pp. 103–114). Reston, VA: American Society of Civil Engineers.

Metcalf, G. E. (2010). *Submission on the use of carbon fees to achieve fiscal sustainability in the federal budget*. Boston, MA: Tufts University. http://works.bepress.com/gilbert_metcalf/86.

Moore, M. R., & Dinar, A. (1995). Water and land as quantity rationed inputs in California agriculture: Empirical tests and water policy implications. *Land Economics, 74,* 445–461.

Moore, M. R., Mulville, A., & Weinberg, M. (1996). Water allocation in the American West: Endangered fish versus irrigated agriculture. *Natural Resources Journal, 36,* 319–357.

Newell, R. G., Pizer, W. A., & Raimi, D. (2013). Carbon markets 15 Years after Kyoto: Lessons learned, new challenges. *Journal of Economic Perspectives, 27*(1), 123–146.

Nordblom, T. L., Finlayson, J. D., & Hume, I. H. (2012). Upstream demand for water use by new tree plantations imposes externalities on downstream irrigated agriculture and wetlands. *Australian Journal of Agricultural and Resource Economics, 56*(4), 455–474.

Pan, S., Hudson, D., & Mutuc, M. (2011). *The effects of domestic offset programs on the cotton market*. Selected paper, Southern Agricultural Economics Association Annual Meeting, Corpus Christi, TX, February 5–8, 2011.

Pattanayak, S. K., Sommer, A., Murray, B. C., Bondelid, T., McCarl, B. A., & Gillig, D. (2002). *Water quality co-benefits of Greenhouse gas reduction incentives in U.S. agriculture* (Final Report). Washington, DC: Environmental Protection Agency.

Rabotyagov, S. S., Campbell, T., Jha, M., Gassman, P. W., Arnold, J., Kurkalova, L., et al. (2010). Least cost control of agricultural nutrient contributions to the Gulf of Mexico Hypoxic zone. *Ecological Applications, 20*(6), 1542–1555.

Rabotyagov, S. S., Kling, C. L., Gassman, P. W., Rabalais, N. N., & Turner, R. E. (2012). *The economics of dead zones: Linking externalities from the land to their consequences in the sea*. Working Paper 12-WP 534. Center for Agricultural and Rural Development, Iowa State University, Ames, IA.

Rausch, S., & Reilly, J. (2012). *Carbon tax revenue and the budget deficit: A Win-Win-Win solution?* (MIT Joint Program on the Science and Policy of Global Change, Report No. 228). Cambridge, MA: Massachusetts Institute of Technology.

Ribaudo, M., Delgado, J., Hansen, L., Livingston, M., Mosheim, R., & Williamson, J. (2011). *Nitrogen in agricultural systems: Implications for conservation policy* (Economic Research Report 127). Washington, DC: USDA, Economic Research Service.

Ribaudo, M., Osborn, C., & Konyar, K. (1994). Land retirement as a tool for reducing agricultural nonpoint source pollution. *Land Economics, 70*(1), 77–87.

Ribaudo, M. O., Heimlich, R., Claassen, R., & Peters, M. (2001). Least-cost management of nonpoint source pollution: Source reduction versus interception strategies for controlling nitrogen loss in the Mississippi basin. *Ecological Economics, 37*(2), 183–197.

Schmalensee, R., & Stavins, R. (2013). The SO_2 allowance trading system: The ironic history of a grand policy experiment. *Journal of Economic Perspectives, 27*(1), 103–121.

Schwartz, C. (2011). Concentrated solar thermal power and the value of water for electricity. In D. S. Kenney & R. Wilkinson (Eds.), *The water energy Nexus in the American West* (pp. 71–83). Cheltenham, UK: Edward Elgar.

Texas Water Development Board. (2012). *Water for Texas: 2012 state water plan*. Austin, TX: TWDB.

U.S. Dept. of Agriculture, Farm Services Agency (USDA-FSA). (2013, February). *Conservation reserve program: Monthly summary*.

U.S. Dept. of Agriculture, Office of the Chief Economist (USDA, OCE). (2009a). *A Preliminary analysis of the effects of HR 2454 on U.S. agriculture*. Washington, DC: USDA-OCE. July 22, 2009.

U.S. Dept. of Agriculture, Office of the Chief Economist (USDA, OCE). (2009b). *The impacts of the American clean energy and security act of 2009 on U.S. agriculture*. Washington, DC: USDA-OCE. December 18, 2009.

U.S. Dept. of Energy, Energy Information Agency (EIA). (2009). *Energy market and economic impacts of H.R. 2454, the American clean energy and security act of 2009*. Washington, DC: EIA.

U.S. Environmental Protection Agency (EPA). (2004). *National water quality inventory: Report to Congress, 2004 reporting cycle*. Washington, DC: USEPA.

U.S. Environmental Protection Agency (EPA). (2009, June). Analysis of H.R. 2454 in the 111th Congress: The American clean energy and security act of 2009. Washington, DC.

U.S. Environmental Protection Agency (EPA). (2005). Greenhouse gas mitigation potential in U.S. forestry and agriculture. EPA 430-R-05-006. Washington, DC: USEPA.

van Dijk, A. I. J. M., & Keenan, R. J. (2007). Planted forests and water in perspective. *Forest Ecology and Management, 251*(1–2), 1–9.

Chapter 12
Enhancing Productivity and Market Access for Key Staples in the EAC Region: An Economic Analysis of Biophysical and Market Potential

Siwa Msangi, Kennedy Were, Bernard Musana, Joseph Mudiope, Leonidas Dusengemungu, Lucas Tanui, Jean-Claude Muhutu, George Ayaga, Geophrey Kajiru and Birungi Korutaro

Abstract In this chapter, we show how the current crop areas under three key staples—rice, maize, and beans—could be better aligned with the crop suitabilities that are inherent in the East African Community (EAC) region, through some key policy interventions. We take a multi-market model that was constructed for the 5 main countries in the EAC and use it to demonstrate how reducing transport costs, and increasing crop productivities can lead to market-level welfare improvements, as well as a closer alignment between the areas where the crops are cultivated, and the areas with the best agronomic suitability for those crops. At present, a signif-

S. Msangi (✉)
International Food Policy Research Institute, Washington, DC, USA
e-mail: siwamsangi@gmail.com

K. Were · L. Tanui · G. Ayaga
Kenya Agricultural and Livestock Research Organisation (KALRO), Food Crops Research Institute, Kaptagat Rd, Nairobi 00100, Kenya
e-mail: ayagag@yahoo.com

B. Musana
College of Agriculture, Animal Sciences and Veterinary Medicine, University of Rwanda, P.O. Box 210, 3971 Kigali, Musanze, Rwanda
e-mail: bernard.musana@rab.gov.rw

J. Mudiope · B. Korutaro
Kilimo Trust, Plot 42 Princess Anne Drive, Bugolobi, Kampala, Uganda
e-mail: joseph.mudiope@millenniumpromise.org

B. Korutaro
e-mail: bkorutaro@kilimotrust.org

L. Dusengemungu · J.-C. Muhutu
Rwanda Agriculture Board, KK 18 Ave, P.O. Box 5016, Rubilizi, Kigali, Kicukiro District, Rwanda

G. Kajiru
Lake Zone Agricultural Research Development Institute (LZARDI), Mwanza, Tanzania
e-mail: gjkajiru@gmail.com

S. Msangi and D. MacEwan (eds.), *Applied Methods for Agriculture and Natural Resource Management*, Natural Resource Management and Policy 50,
https://doi.org/10.1007/978-3-030-13487-7_12

icant proportion of those staples are grown in areas with limited growth potential, but opening up markets in combination with productivity-focused investments can allow countries to make better use of the crop potential they already have, and take advantage of regional market opportunities.

1 Introduction

Like in the rest of the world, the growing demand for food, feed, and fiber products in East Africa continues to place increasing pressure on the region's agricultural production base and represents both an opportunity and concrete challenges for its agricultural economy. As in much of the rest of sub-Saharan Africa, there remains tremendous biophysical potential for agricultural production, across a range of products (cereal, root and tuber, pulses, fruit and vegetables) within the East African region—although much of that potential still remains under-exploited. Within the geographical region covering the East Africa Community (EAC),[1] a large portion of agricultural area is under semiarid conditions, and with a history of low-yielding, subsistence agriculture and under-exploited irrigation potential (Salami et al. 2010; Waithaka et al. 2013). Much donor interest and nationally driven efforts are focused on improving this history of under-performance, so as to take advantage of the available land and water resources, and to create a more favorable policy environment for encouraging farmers to increase their on-farm productivity of key agricultural products (EAC 2011; USAID 2017).

Given that roughly three-quarters of the region's agricultural production is produced by smallholders (Salami et al. 2010), a good deal of the un-exploited potential lies in the intensification of production to use more yield-enhancing inputs and making greater use of labor-saving technologies that can increase on-farm productivity. Given that the opportunities for area expansion are limited, the emphasis on enhancing productivity to meet regional food needs is urgent (Waithaka et al. 2013). In order to avoid an excessive application of chemical inputs in order to grow crops where the suitability of soils is unfavorable toward its cultivation, the alignment of crop production with the native suitability of the soils and land would be an ideal strategy for intensification. Given constraints on land availability, historical patterns of human settlement and land use, and other socio-economic factors that run contrary to creating a natural alignment of production with potential, we find a good deal of the EAC's agricultural production occurs on soils that are of limited suitabilities for the key staple crops that are grown. In order to meet the agricultural development goals of the EAC regional policy body, however, some efforts will need to be made to re-align crop production to take better advantages of the regions suitabilities.

[1]The East African Community consists of five key countries—Burundi, Kenya, Rwanda, Tanzania, and Uganda. South Sudan is now also an official part of the EAC; however, a lack of reliable data and links to research experts in that area has led to its omission from this study.

In this chapter, we look at the agricultural markets of the EAC region, with respect to the three key staple crops that grow within it, and show how policy-focused investments in transportation and marketing infrastructure as well as productivity-enhancing technologies can shift the supply, demand and trade of the region in favor of the agricultural areas with better suitabilities in certain crops. We employ an agricultural multi-market model to simulate these policy-driven changes and show the resulting impacts on demand, supply, and trade—as well as on the distribution of harvested area in these crops—across the EAC countries. Through these empirical illustrations, and evidence drawn from the literature, we argue for a 'market-focused' approach to the EAC's agricultural policy—given that the best chance for meeting growing food demands occurs when productivity gains are maximized and the market access for the agricultural sector is secure.

The technical novelty of this chapter lies not only in the economic model used—which embeds positive mathematical programming (PMP) methods to calibrate the production side of the multi-market model—but also in the way that information-theoretic methods were used to make use of important biophysical data provided by agronomy experts based in the EAC region. As is the case with many modeling-based exercises, the creation of a robust and reliable database is usually more than half the effort needed and might even exceed the technical demands represented by the main analytical engine that runs the final simulation results. This is certainly the case in this chapter, where the aggregate-level information on crop production, market supply, trade, and demand might not align well with the biophysical information that is available from agronomic assessments. Given the increasing volume of high-quality information on crop suitability, soil characteristics, weather conditions, and other biophysical data for Africa (at various scales of spatial resolution), and the relatively coarse resolution of publicly available socio-economic information that describe the agricultural sector, the need for tools to carry out data reconciliation and processing to generate a complete database that is suitable for policy-relevant economic analysis is clear. As efforts to generate better socio-economic data on agricultural producers (and consumers) continue, through international organizations, NGOs, and private foundations committed to generating public goods,[2] then we might hope to achieve a better match in resolution and detail between the biophysical and socio-economic dimensions of the agricultural landscape, in the future.

The rest of the chapter is designed as follows. The following section describes the agroecological characteristics of the EAC region, with respect to a few key staple crops of interest, and the way in which this data was prepared for our market-focused study. The subsequent section gives a description of the multi-market model for the EAC region that will be used for the empirical analysis, with key references to a detailed technical annex. The subsequent session introduces the scenarios that are used to illustrate how the production patterns of the focus staple crops shift in relation

[2]Like the Bill and Melinda Gates Foundation's efforts to support the World Bank in adding details relevant to agricultural in their Living-Standards and Measurement Surveys (LSMS)—resulting in the LSMS-ISA (i.e., 'Integrated Surveys on Agriculture') project.

to their agro-ecological potential, as well as other market-level impacts within the EAC region. The final section concludes with recommendations for policy and future research.

2 Key Agro-Ecological Characteristics for Staples in the EAC

In this section, we describe the characterization of biophysical suitability for important staples within the EAC region, and how they are captured within customized agro-ecological classifications. We will use these classifications of crop suitability to define the yield potential that exists for three staple crops of interest for this study—rice, maize, and beans. Other crops could have been chosen for this kind of exercise, but we chose to start with these three since their value chains are well-defined, and since there is already a good deal of knowledge about their agronomical potential and growth characteristics, across the five East African countries. Within the five EAC countries, a number of key growing areas were identified as being representative agro-ecological zones that capture the crop growth potential for these three major staples within the EAC region. These were the Lake Victoria Basin in Kenya, the Kyoga plains, and south-western highlands of Uganda, the lake zone and southern highlands area of Tanzania, and essentially all of Rwanda and Burundi.

Carrying out a process of expert consultation with a team of agronomists drawn from the EAC region, a set of criteria for crop suitability were derived for the three key staple crops of interest—although it should be noted that a key distinction was made between upland and lowland rice, and between climbing beans and bush beans. For each of these crops, a set of suitability criteria drawn from those used by Kaaya et al. (1994) were evaluated for key growing regions within the EAC, to determine the level of suitability of the soils with respect to the requirements for growth for each crop. There were five broad land qualities that were evaluated for each crop, namely the moisture availability (in terms of total rainfall within the growing period); the temperature regime (in terms of mean temperature during the growing period); the nutrient availability (measured in terms of soil reaction and topsoil organic carbon); the rooting conditions (in terms of effective soil depth); and the erosion hazard (in terms of the slope angle). Using digitized maps of soil and land characteristics for the target growing areas of the EAC region, the existing growing conditions, in terms of these criteria, can be observed and overlaid with the distribution of crop areas in the region that come from digitized crop maps, such as those produced by *HarvestChoice*[3] and made publically available for analysis. The cut-off values of alternative suitability classes had to be determined for each of the crop suitability criteria defined above, in order to arrive at a determination of whether a given area is highly suitable, moderately suitable, marginally suitable or not suitable for a particular crop.

[3]See the data products available at: https://harvestchoice.org/.

Using this methodology, a map of suitability such as the one shown for rainfed maize (Fig. 1) could be produced.

A similar map for bush beans can be generated, applying a different set of cutoffs for the key criteria of suitability described above, and applying them to the target agro-ecological zones for the EAC region (Fig. 2).

Based on these maps—as well as those for climbing beans, lowland and upland rice—we are able to create a database of crop area and yield under the various zones of suitability that is used to define a model of agricultural production, demand, and trade for the EAC region that we describe in greater detail in the following section.

3 Quantifying Market Impacts to Key Staples of the EAC Region

In this section, we summarize the basic features of a national, agricultural multi-market model for the EAC region and describe how it links the supply and demand of agricultural commodities within the various countries within the EAC, to regional and global agricultural markets. This will provide a useful understanding of how the information on crop suitability, described in the previous section, is applied in the construction of a regional model of production, consumption, and trade in the three key staple crops of interest. The resulting model is then used to evaluate the effect of different market-relevant scenarios on the distribution of production across the suitability zones and the resulting implications for prices, consumption, and trade within the EAC region.

3.1 The Basic Agricultural Multi-market Model

In this study, we have used an analytical approach that can address the issue of trade directly—both within the EAC, as well as the trade between the EAC region and the rest of the world. We use an economic modeling framework that accounts for the supply and demand response of the three commodities of interest (maize, rice, and beans) to prices within the EAC region and the trade flows that result from this interaction. The five EAC countries (Burundi, Kenya, Rwanda, Tanzania, and Uganda) are included within the modeling framework as separate units with sub-national definitions that define both administrative boundaries, as well as different zones of agro-ecological potential. Since we have access to data that can describe the distribution of maize, beans, and rice production across various administrative and agro-ecological boundaries, we have a much more disaggregated view of crop production and market supply, compared to consumption and market demand. Since data on trade is collected at the national level, we have to model the price formation within the market at that level of aggregation. This allows us to calibrate the model to

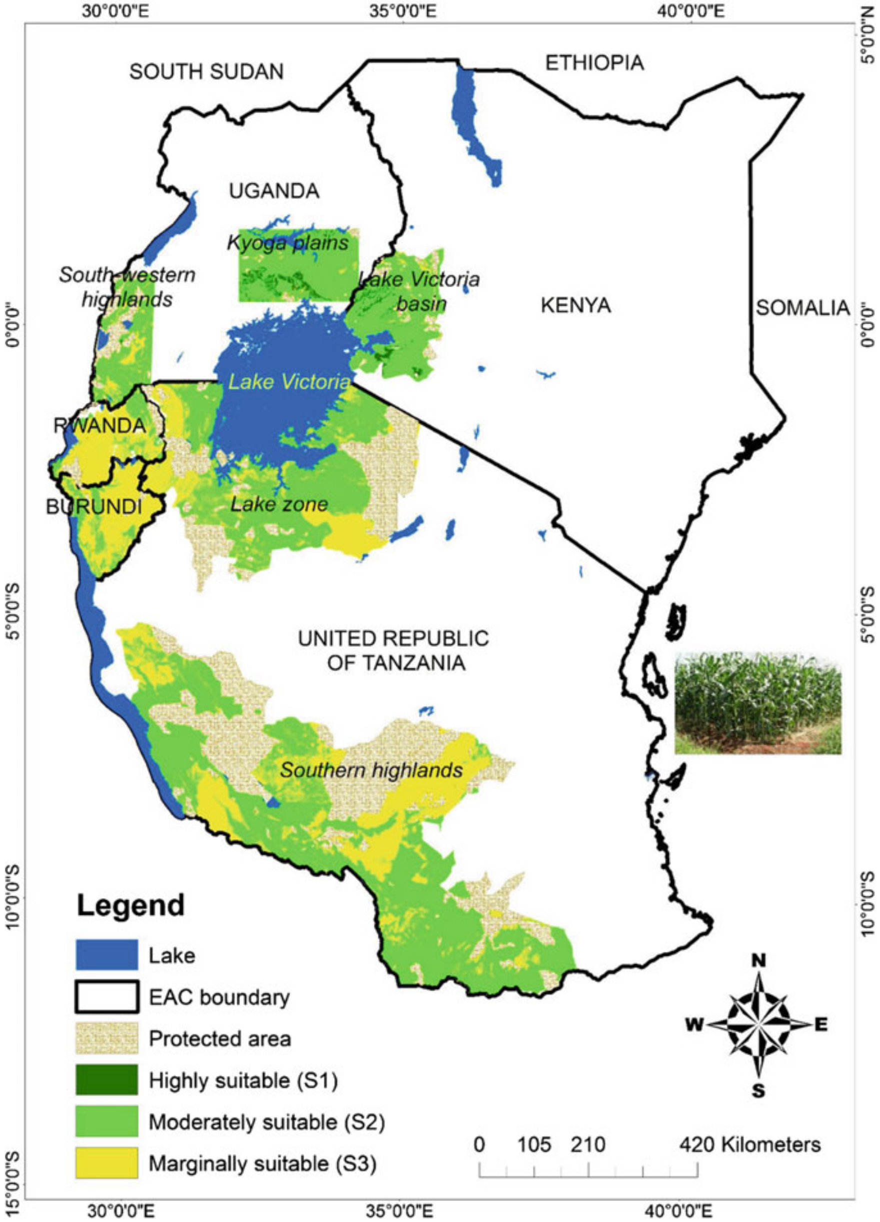

Fig. 1 Agro-ecosystems suitability for rainfed maize production in the EAC region. *Source* Were et al. (2016)

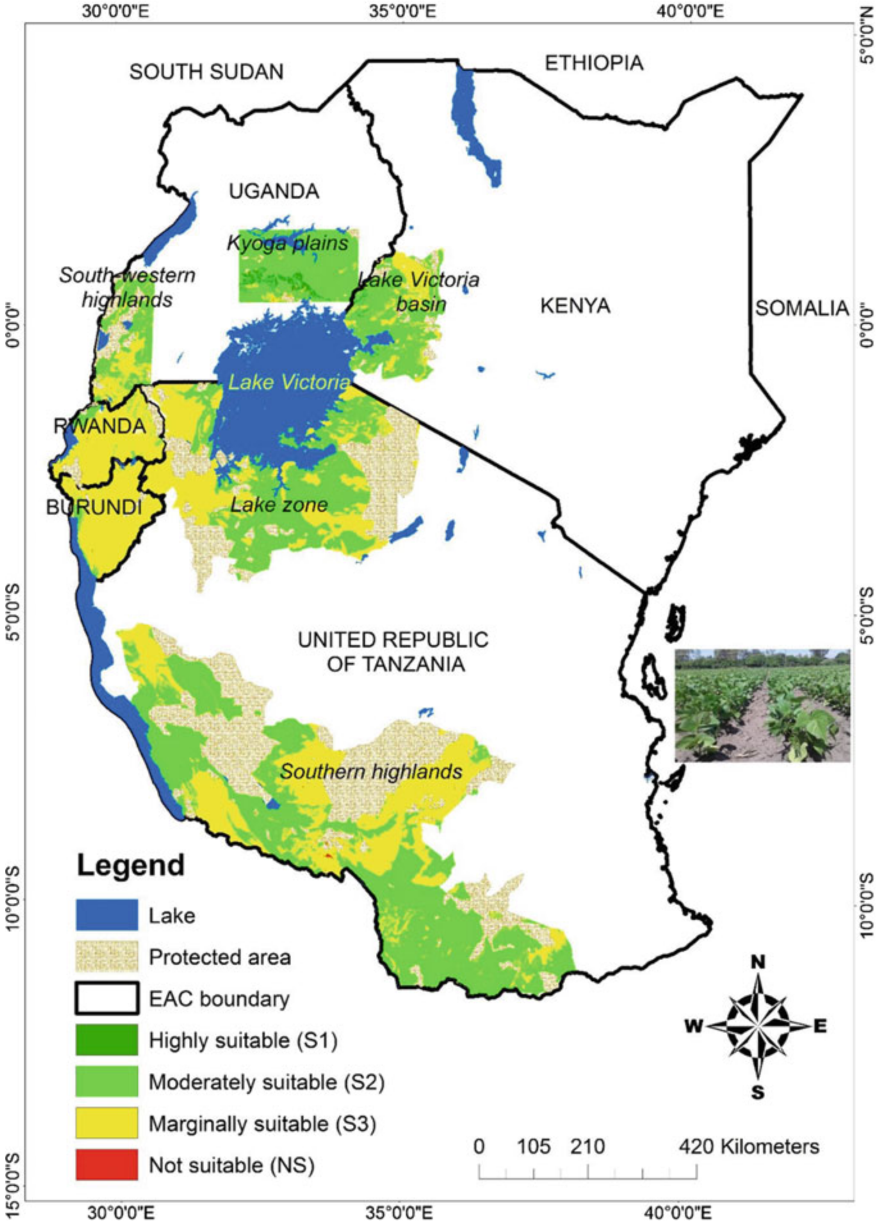

Fig. 2 Agro-ecosystems suitability for rainfed bush beans production in the EAC region. *Source* Were et al. (2016)

replicate the observed patterns of production, consumption, and trade, which gives us confidence in being able to carry out policy-based experiments.

Our model is 'partial-equilibrium' in nature because it focuses on the particular parts of the economy that we're interested in (i.e., the three commodity markets) and leaves out other parts of the economy that are beyond the scope of our interest, or ability to capture at this stage—such as the market for non-agricultural goods, the market for land and labor. While other sectors of the economy do, in reality, interact with the decisions of farms and affect the overall economic environment in which they leave, we omit them from our analysis, so that we can focus on certain aspects of the three commodity markets in more detail. For example, we are able to model agricultural production in more detail than the 'economy-wide' multi-market model of Omamo et al. (2006), which uses highly simplified supply functions in order to allow for more computational detail to be given to the key economy-wide linkages with the agricultural sector that they are interested in. In our case, we separate yield from area response, so that the effects of crop-specific technological improvements that boost productivity can be differentiated from land-use changes that affect the expansion potential of harvested area—allowing us to differentiate between the responses on the intensive and extensive margins of production, when modeling supply changes.

In order to make the best use of the detailed statistical and biophysical data that is available on crop areas and yields for maize, beans, and rice, we subdivide the regional production of the three key crops into spatial units that represent the production potential within the defined agro-ecological zones. Since most of the statistics are collected at the level of administrative units (especially for consumption), we will have to simultaneously make use of provincial or district-level supply data and intersect it with the boundaries of defined agro-ecological zones, so that we can reconcile the supply, demand, and trade relationships that have to be respected.

In constructing the multi-market model, we have built upon a framework that allows for straightforward expansion in the future, should additional data or research questions need to be explored. Rather than using an existing model of agricultural markets, like the global IMPACT model of IFPRI (Rosegrant et al. 2001, 2002, 2012), we build a customized model that can provide greater flexibility with the sub-national data that we have on hand and focus on the issues of interest. Since we focus on the EAC region, we have no need for a global market model and can hold the conditions in the 'rest-of-the-world' constant and exogenous to the model solution. The overall schematic for the model is shown in Fig. 3.

3.1.1 Structure of the Economic Multi-market Model

Here, we give a summary explanation of the structure of the market model, to make the methodological approach clear to the reader. As in any representation of a market which has to 'clear,' the heart of the market model lies in the satisfaction of a fundamental balance between demand, supply, and the exchange of goods between

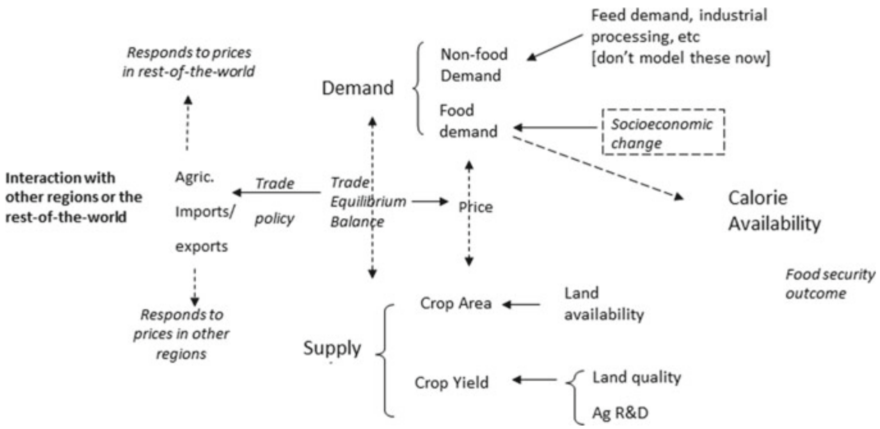

Fig. 3 Schematic of model structure. *Source* Authors

regions. For any of the countries within the EAC region, the following general balance is satisfied:

$$\text{Supply} + \left(\sum_{\text{other EACRegs}} (\text{inflows}) + \text{iMport}_{\text{RoW}}\right) = \text{Demand}$$
$$+ \left(\text{eXprt}_{\text{RoW}} + \sum_{\text{other EACRegs}} (\text{outflows})\right) + \text{StockChange} \quad (1)$$

where supply is comprised of the production of the three key crops of interest in each EAC region (aggregated up from the sub-national level), demand is the national-level food and non-food consumption of each crop, and stock change is the amount of good that is subtracted from the available market supply and put into inventory (i.e., storage) from period to period. In our current model, we have not endogenized stock change, due to lack of information about the actual stocks of goods being held publically or privately in the five EAC countries We have also chosen to keep non-food demand[4] as exogenous and non-price responsive, as it relates to livestock, processing, and other activities which are outside the scope of our model.

In the balance shown in Eq. (1), we have divided the imports into the region and the exports out of the particular EAC region between those trade flows going to other EAC countries, and those going to the rest of the world. This is a useful distinction which we will refer to in the policy analysis that will follow later.

[4]Non-food demand also contains categories such as 'seed use' and 'waste' which are tracked in the FAO food balances, but which we do not have sufficient information for to model explicitly.

Both supply and demand are described by analytical equations which relate the price levels in each region to the level of production that is realized. In the case of demand, the total demand within an EAC country is described by the relationship

$$\bar{Q}_D^{\text{non-food}}\text{Demand} = f_D(P, y) \cdot \text{popn} + \bar{Q}_D^{\text{non-food}} \tag{2}$$

where the non-food demand ($\bar{Q}_D^{\text{non-food}}$) is a fixed and exogenous parameter, and where the per-capita food demand responds to both price (P) and per-capita hold income (y) levels, according to the function $f_D(P, y)$—which is multiplied by the population, in order to give the total food demand in the country. The per-capita food demand function, itself, takes on the form

$$f_D(P, y) = c_{\text{dmd}} \cdot P^{\varepsilon} \cdot y^{\eta} \tag{3}$$

where the parameters ε and η denote the food preferences of the consumer, in terms of their tendency to change income with response to price changes or the level of per-capita income, respectively. The parameter c_{dmd} is a constant and denotes the 'intercept' of the demand equation—which is calculated in a way that allows per-capita food demand to calibrate to the observed data, at a given price and income level.

On the supply side, we take into account both area and yield when calculating the production of each of the three key crops represented in the model, according to the following relationship.

$$\text{Supply} = \text{Area} \cdot \text{Yield} = \sum_R \sum_s \left(A_{R,s} \cdot Y_{R,s}\right) \tag{4}$$

where each country is divided into sub-regions (provinces or districts) denoted by R which, in turn, are subdivided into zones of crop suitability that were defined by the GIS-based analysis of regional experts.[5] The multiplication of harvested area and yield and their addition over the sub-regions and crop suitability zones of each country are what constitute the national production of each crop.

The yield of the crop, in each suitability zone, has been informed by the inputs of agronomic experts from each of the EAC regions covered in the study. The information gained from these experts—which indicated the share of maximum potential yield of each crop that would be attainable within that zone—was used to allocate the base areas and yields of the model, for all the EAC countries, such that the national-level supply was matched, and could be balanced to demand and trade. The process of calibrating base area and yields is described briefly in Sect. 3 of this report—and in more detail in the technical appendix.

The harvested area of each crop responds to prices, according to the net revenues that can be realized per hectare of each crop across the available cropland area.

[5]The GIS-based analysis of crop suitability is described in a separate technical document of Were et al. (2016).

The key essence of this decision process is captured in the following optimization problem

$$\begin{array}{ll} \max & \sum_j \left[p_{j,R} \cdot Y_{j,R,s} \cdot A_{j,R,s} - c\left(A_{j,R,s}\right)\right] \\ \text{s.t.} & \sum_j A_{j,R,s} \leq \bar{A}_{R,s} \end{array} \tag{5}$$

where the regional prices of each crop (j) are $p_{j,R}$, and $c\left(A_{j,R,s}\right)$ represents a nonlinear relationship between the total area harvested and the per-hectare cost of cultivation. This nonlinear relationship is a necessary component of calibrating the model to observed crop areas and is part of the 'positive' mathematical programming (PMP) approach. The PMP approach to calibrating agricultural models is described in further detail, in the technical appendix.

This method for allocating crop area is in contrast to the 'reduced-form' approach used in an earlier version of this work, in which an iso-elastic area response function was hypothesized ($A_{R,r} = f_A(P) = c_A^{R,r} \cdot P^{\gamma}$[6]). In this function, the reaction of area to price changes reflects the behavior that is embodied in the optimization problem that we now model explicitly.

In the model, the prices play a key role in the response of supply and demand—but are also fundamental in determining the direction of trade flows into and out of the region. The difference between prices in two adjacent regions essentially determines the direction of the trade flow. The following relationship shows the no-arbitrage condition that is hypothesized to hold between two regions, where the transportation cost (TC) provides the 'wedge' between the two prices and accounts for the cost of delivering each unit of good from its origin to its destination

$$P_{\text{Region}(R)} + \text{TC}_{R,R'} \geq P_{\text{Other Region}(R')} \tag{6}$$

In this relationship, the price that one receives in the destination region (R') should be no less than the purchase price in the region of origin (R) and the cost of transporting it between the two ($\text{TC}_{R,R'}$). Where the purchase + transport cost is greater, then it is not optimal to export from R to neighboring R'(or, conversely, to import into R' from R). Where this relationship holds with equality (i.e., $P_R + \text{TC}_{R,R'} = P_{R'}$), then one could expect to have nonzero levels of trade. This complementary relationship between the levels of trade and the equality or non-equality of the no-arbitrage constraint can be written as

$$\begin{array}{lll} P_R + \text{TC}_{R,R'} > P_{R'} & \text{and} & \text{Trade}_{R,R'} = 0 \\ P_R + \text{TC}_{R,R'} = P_{R'} & \text{and} & \text{Trade}_{R,R'} \geq 0 \end{array} \tag{7}$$

And it can, in turn, be summarized by the following equality statement

[6]In this function, the parameter $c_A^{R,r}$ is a calibrating constant, and the 'elasticity' γ gives the response of area to a change in price.

$$\left[P_R + \mathrm{TC}_{R,R'} = P_{R'}\right] \cdot \mathrm{Trade}_{R,R'} = 0 \tag{8}$$

which must hold at all times. This type of relationship determines the trade levels between the EAC countries, as well as between each EAC country and the rest of the world.

In addition to this, there could be other taxes on imports or exports, such that the unit price of each traded good is affected, or there could be a quantitative limit on total exports and imports such that we have either a quota on imports

$$\mathrm{Quota}^R_{\mathrm{Import}} \geq \sum_r M^R_r \tag{9}$$

Or a quota on exports

$$\mathrm{Quota}^R_{\mathrm{Export}} \geq \sum_r X^R_r \tag{10}$$

where the quota (and tax) levels are decided by policy, and the levels of import $\left(M^R_r\right)$ and export $\left(X^R_r\right)$ for each country R are summed over the quantities going from each sub-region, r.[7] In our model, we only apply import and export taxes (and quotas) on non-EAC trade, leaving the EAC region a 'free-trade zone,' as it was intended. But this could, of course, be relaxed if we wanted to examine a counter-factual case.

This brief exposition summarizes the basic structure of the multi-market model which we use in this study. Further details are in the Technical Annex.

3.2 Base Data and Model Calibration

In order to carry out the quantitative analysis that is required for this project, we draw upon some key empirical methods for the analysis of agricultural production and trade. The key sources of data that were used in this work can be summarized as follows:

- Biophysical data encompasses the areal extent and productivity of cropping activities, with reference to their rainfed or irrigated nature. The information on crop suitability classes also falls into this class, which is a key to the determination of overall crop productivity.
- Economic data captures the traded volumes of commodities, prices, and marketing costs. Where possible, the physical levels of stockholding can also be used to measure the total supply potential of a commodity (beyond that which is being harvested from the land within a particular time period).

[7] In our implementation, we did not model trade between the sub-regions (r) of each country, as that would have imposed an enormous computational burden on the model, and require detailed data beyond what we possess. So, we only model trade at the national level, in this model.

Table 1 Key data used in the model

	Data source	Notes
Area and yields (irrigated, rainfed) according to administrative boundaries in the five EAC countries	HarvestChoice dataset	Available at www.harvestchoice.com
Crop suitability classifications for cropland area in the EAC countries	GIS-based model analysis of spatial data for EAC	This work was done as a separate component of the WaLETS program of work by Dr. Kennedy Were and colleagues
Food/non-food demand for commodities	FAOSTAT data	Downloaded through the Knoema Web site (knoema.com)
Import and exports for commodities	FAOSTAT data	Downloaded through the Knoema Web site (knoema.com)
Supply response elasticities and demand response elasticities	Database of IFPRI ASARECA regional multi-market model[a]	Technical appendix available from www.ifpri.org
Transportation costs between EAC regions and between regions and the rest of the world	Initial estimates from WB study[b] and various sources	Final calculation of costs was done as part of the model calibration process described in the report
Regional prices for the three commodities in ECA region	Knoema price database	International ‘world’ prices were calibrated in a way consistent with the region-specific data, observed transportation costs, and net trade
Potential and attainable yields across the classifications of crop suitability	Country agronomy experts from within the WaLETS project team	These were used in the determination of the yield distribution across the suitability classes

[a]See technical appendix of Omamo et al. (2006)
[b]See Teravaninthorn and Raballand (2009)

- Other socio-economic data particularly on population (urban and rural), income levels, and the indicators of food security or poverty that help define the overall welfare of the population. These kinds of data are critical to defining the ‘demand’ side of the market and the consumers that depend upon these key staple crops.

Drawing from these broad categories of data, we have a basis upon which to construct a multi-market model of the agricultural sector that focuses on the three commodities of particular interest. The particular sources of data that we drew upon are summarized in Table 1.

3.2.1 Calibration of the Model to Data

Socio-economic parameters

One of the key challenges we face in specifying the model correctly is finding the correct magnitude of transportation costs and border prices that allow the model to find a solution that is consistent with observed data.

Essentially, we have three major unknowns:

- The exact magnitude of the transportation costs that exist between the various countries (or sub-national regions) of the EAC region, both internally, and with the rest of the World
- The exact magnitude of the inter-EAC trade that exists between the countries
- The exact border prices which would define the fixed 'rest-of-the-World' price that defines the export and import parity values that each country faces when engaging in an exchange with the countries outside of the EAC.

Based upon the price relationships given in Eq. (8), that determine the levels of trade, this means that we had to find the value of $\mathrm{TC}_{R,R'}$ that would make the total imports and exports $\left(\mathrm{Trade}_{R,\mathrm{RoW}} + \sum_{R'_{\mathrm{EAC}}} \mathrm{Trade}_{R,R'_{\mathrm{EAC}}}\right)$ match the values observed from the FAO trade balance data, while also solving for the prices in each region (P_R) as well as the trade between each region (R) in the EAC with each neighboring EAC country $\left(\mathrm{Trade}_{R,R'_{\mathrm{EAC}}}\right)$. The fixed border prices that determine the level of imports or exports from the rest of the world and each region—i.e., the import $\left(\bar{P}_M^{\mathrm{RoW}}\right)$ and export $\left(\bar{P}_X^{\mathrm{RoW}}\right)$ parity prices—also had to be found in this calibration process, as we do not observe it directly in the data.[8]

The calibration of the demand side of the model was relatively straightforward, as it only entails calculating the 'intercept' of the demand function (i.e., the constant c_{dmd} in the equation $f_D(P, y) = c_{\mathrm{dmd}} \cdot P^{\varepsilon} \cdot y^{\eta}$) for given values of the demand elasticities (ε, η), per-capita income (y), and prices (P).

Biophysical aspects

The other component of calibrating the supply side to the observed data is obtaining the distribution of areas and yields across the suitability zones that are consistent with the GIS-based modeling outputs, as well as with the distribution of crop area and production (as in Figs. 1 and 2) that are obtained from statistical data and other data sources, like the *HarvestChoice* spatial dataset. Some key information from agronomy experts within our research team also provided a means of calibrating the

[8] In our formulation, we treat the 'rest of the World' as a homogeneous entity, knowing that it represents something different to each EAC country, in reality. For the inland, landlocked countries, for example, the rest of the World, would be the bordering Congo or the Sudan, whereas the coastal countries receive goods from the 'rest of the World' over ocean-based routes at Mombasa, Dar-es-Salaam, Tanga, Mtwara, or from neighboring, Ethiopia, Malawi, Zambia, or Mozambique. By simplifying the representation of RoW to one entity, we cannot (therefore) tie a particular landed price at a given point of entry or exit as the border price—but hypothesize a composite world price which we have to solve for in the calibration process.

yields within the model, such that they conformed to the attainable yield levels that are consistent with the biophysical conditions captured in the suitability classes.

The calibration procedure made use of the following pieces of information:

- The area of cropland that falls into a given suitability classification within each EAC country
- The amount of area under each of the three key staples (maize, rice, and beans) which falls into each administrative sub-region of the five EAC countries.

By using a data processing procedure that makes us of maximum entropy estimation methods,[9] we were able to obtain the following distribution of crop areas, according to suitability classifications in the five EAC countries (Table 2).

This distribution of areas and yields provides the model with the starting point it needs to calibrate the base-year supply, demand, and trade levels to the observed levels. This base-year calibration is described further, both in the following section (briefly)—and in more detail in the technical appendix.

Based on the specified structure of the model that we've described in the previous section, we carried out a calibration of the model to match the model results to the observed supply, demand, and trade for the five EAC countries and the three key commodities. This calibration gives us confidence in the proper functioning of the model and also verifies that the calibration of transportation costs and border prices (with the rest of the World) was done correctly. The calibration results of the model are shown in Tables 13 and 14, in the technical appendix.

We will now consider how the base distribution of areas under the three staple crops changes, according to the scenarios that are discussed in the following section.

4 Key Policy Scenarios for the EAC Region

Once the calibration of the model has been checked, we can then proceed to carry out some illustrative policy simulations to see how the base case will shift with the imposition of key policy-relevant changes that are introduced into the model. We focus on the policies relating to openness of trade as well as to productivity-enhancing agricultural technologies, as we will now discuss in further detail.

4.1 Scenario #1—Reducing Regional Transportation Costs

One of the key policy scenarios that we carry out is that of reducing the transportation costs across the EAC region to see the effect on regional trade and the implications for prices, consumption, and welfare.

[9]The application of maximum entropy inference methods is explained in further detail in the Technical Annex (2).

Table 2 Calibration of base area and yield levels across suitability classes for each EAC country

		Area ('000 ha)			Yield (mt/ha)		
		Rice	Maize	Beans	Rice	Maize	Beans
Kenya	Maize/high suitability		85.2			2.1	
Kenya	Maize/moderate suitability		1520.1			1.8	
Kenya	Maize/low suitability		45.9			1.3	
Kenya	Rice/high suitability	3.7			3.1		
Kenya	Rice/moderate suitability	4.6			2.7		
Kenya	Rice/low suitability	4.7			2.1		
Kenya	Rice/unsuitable	4.5			1.3		
Kenya	Beans/high suitability			279.7			0.5
Kenya	Beans/moderate suitability			377.3			0.4
Kenya	Beans/low suitability			248.5			0.3
Tanzania	Maize/moderate suitability		1477.1			1.5	
Tanzania	Maize/low suitability		1373.1			1.1	
Tanzania	Rice/high suitability	13.1			1.74		
Tanzania	Rice/moderate suitability	201.0			1.69		
Tanzania	Rice/low suitability	203.4			1.3		
Tanzania	Rice/unsuitable	178.2			0.8		
Tanzania	Beans/high suitability			38.2			1.1
Tanzania	Beans/moderate suitability			392.3			0.8
Tanzania	Beans/low suitability			395.1			0.7
Tanzania	Beans/unsuitable			1.1			0.4
Uganda	Maize/high suitability		177.5			2.3	
Uganda	Maize/moderate suitability		434.9			1.5	
Uganda	Maize/low suitability		163.6			0.7	
Uganda	Rice/moderate suitability	33.9			1.5		
Uganda	Rice/low suitability	34.1			0.9		
Uganda	Rice/unsuitable	33.7			0.5		
Uganda	Beans/high suitability			238.3			0.6
Uganda	Beans/moderate suitability			330.4			0.56
Uganda	Beans/low suitability			255.7			0.45

(continued)

Table 2 (continued)

		Area ('000 ha)			Yield (mt/ha)		
		Rice	Maize	Beans	Rice	Maize	Beans
Rwanda	Maize/moderate suitability		54.2			1.1	
Rwanda	Maize/low suitability		53.4			0.6	
Rwanda	Rice/moderate suitability	5.1			6.3		
Rwanda	Rice/low suitability	1.7			2.3		
Rwanda	Rice/unsuitable	1.4			1.4		
Rwanda	Beans/high suitability			32.8			1.2
Rwanda	Beans/moderate suitability			135.1			0.9
Rwanda	Beans/low suitability			142.7			0.5
Burundi	Maize/moderate suitability		56.8			1.4	
Burundi	Maize/low suitability		56.6			0.8	
Burundi	Rice/moderate suitability	8.7			4.1		
Burundi	Rice/low suitability	4.0			1.5		
Burundi	Rice/unsuitable	3.4			0.9		
Burundi	Beans/moderate suitability			119.8			1.2
Burundi	Beans/low suitability			125.2			0.5

In carrying out this scenario, we impose a 50% reduction in all transportation costs (TC), which affects the prices relationships that are determined by the no-arbitrage conditions given in Eq. (8). What we would expect a priori is that the quantities traded in the model (both between the EAC regions and with the rest of the World) would generally increase with a reduction in transportation costs.

To see more clearly what actually happens in the model, as a result of these changes, we divide the total trade for each country between the net trade of that country with the rest of the World—i.e., the quantity $\text{eXprt}_{\text{RoW}} - \text{iMport}_{\text{RoW}}$ in Eq. (1)—and the net trade between that country and other EAC countries [i.e., $\sum_{\text{other EACRegs}} (\text{outflows}) - \sum_{\text{other EACRegs}} (\text{inflows})$].

Table 3 shows the results for net trade[10] with the rest of the world.

These results show that the net imports for the landlocked countries (like Uganda, Rwanda, and Burundi) go to zero, whereas there is an increase in net imports from the rest of the world from the countries with sea access (Tanzania and Kenya) for most of the commodities. In total, the sum of net trade with the rest of the world,

[10]Here, we interpret net trade as net exports (exports minus imports)—which means a positive quantity makes the country a net exporter and a net importer of the good if negative.

Table 3 Changes in net trade with the rest of the World under reduced transportation costs ('000 mt)

		Base case	Reduced transport costs
Rice	Kenya	−241	−209
Rice	Tanzania	−113	−229
Rice	Uganda	−48	
Rice	Rwanda	−12	
Rice	Burundi	−7	
Maize	Kenya	−160	−220
Maize	Tanzania	−134	−213
Maize	Rwanda	−31	
Maize	Burundi	−66	
Beans	Kenya	−6	−8
Beans	Rwanda	−2	
Beans	Burundi	−0.3	
	Total	−820	−881

across the EAC region, increases the level of net imports from 820 thousand metric tons to 881 thousand, with all of the increases in occurring in the coastal countries. In total, this means that the landlocked countries become less reliant on the rest of the world for its imports, since it is now easier to trade with their neighbors who have better sea access. In order to understand what happens to the increased net imports of rice, maize, and beans, we have to examine the internal patterns of trade within the EAC region.

When we look at the volume of trade within the EAC region, between its constituent countries, we see an increase in trade activity, as is shown in Table 4.

Here, we see that the intra-regional trade in rice increases from zero to a point where Tanzania is exporting rice that supplies Uganda, Rwanda, and Burundi. In the case of maize, Tanzania exports the same volume of maize that is now imported by Rwanda. In the case of beans, both Tanzania and Uganda become a net exporters of beans to Rwanda and Burundi, with Burundi now receiving most of the regional imports. Looking at Tables 3 and 4 together, we see that the EAC countries that are big net importers from the rest of the world (i.e., Tanzania and Kenya) remain big net importers—with Tanzania and Kenya increasing net imports in order to supply their EAC neighbors with the goods that they would have otherwise obtained from outside the region.

Given that Kenya and Tanzania are the coastal countries with direct sea access, it makes sense that they would remain the large-volume importers from the rest of the world. But given infrastructure improvements that would lower transportation costs to the interior—including railway and road improvements, besides improvements to the capacity and handling efficiency at the ports of Mombasa, Dar-es-Salaam, Tanga, and Mtwara—the other EAC countries can source their imports directly through the transportation network within their EAC neighbors, rather than trying to face

Table 4 Changes in net trade within the EAC region under reduced transportation costs ('000 mt)

		Base case	Reduced transport costs
Rice	Tanzania		117
Rice	Uganda		−77
Rice	Rwanda		−19
Rice	Burundi		−20
Maize	Kenya	−2	
Maize	Tanzania		69
Maize	Uganda	2	
Maize	Rwanda		−69
Beans	Kenya	−3	
Beans	Tanzania	3	60
Beans	Uganda	9	51
Beans	Rwanda		−4
Beans	Burundi	−9	−107

higher costs of trading directly with regions outside the EAC. So, there is a logical re-alignment of trade, with infrastructure improvements (and lowering of transport costs). The coastal EAC countries—in particular, Tanzania—become the new 'conduits' for sourcing external imported goods into the region. This makes sense, since Tanzania provides the best connection for Rwanda and Burundi, and has three major ports, as opposed to the single port of Mombasa in Kenya.

In terms of other impacts from this scenario, we can consider the effects on commodity prices within the EAC countries, as is shown in Table 5.

Here, we see that there are some notable changes in prices, as a result of transportation cost decreases. In the case of rice, the prices in Uganda, Rwanda, and Burundi—the landlocked EAC countries—go down between 2 and 5% compared to the base case. This reflects the fact that they are now importing goods from their closer EAC neighbors, rather than facing higher import costs from other regions in the rest of the world. We see the same effect for the case of maize and beans in Rwanda and Burundi, where consumers are now receiving goods more cheaply from their neighbors. In contrast to the price decreases that occur in countries receiving more net imports, we see that the prices within the countries exporting maize and beans goes up, due to there being less supply within their national markets, compared to the baseline. The higher prices are a benefit for the producers in those countries, who are encouraged to produce more and supply their neighbors with needed goods. The price decreases, on the other hand, are a clear benefit to consumers, such as those within the landlocked countries of the EAC.

A priori, we'd expect a decrease in price to boost demand in the countries that showed negative price changes in Table 5—which is exactly the case that we observe in Table 6.

Table 5 Changes in regional EAC prices under reduced transportation costs (USD/mt)

		Base case	Reduced transport costs	% difference under scenario
Rice	Kenya	383	383	0
Rice	Tanzania	367	367	0
Rice	Uganda	417	397	−5
Rice	Rwanda	410	402	−2
Rice	Burundi	426	407	−5
Maize	Kenya	319	319	0
Maize	Tanzania	303	303	0
Maize	Uganda	286	333	17
Maize	Rwanda	346	338	−2
Maize	Burundi	363	343	−5
Beans	Kenya	656	656	0
Beans	Tanzania	619	639	3
Beans	Uganda	622	641	3
Beans	Rwanda	683	671	−2
Beans	Burundi	699	679	−3

Table 6 Changes in food demand within the EAC region under reduced transportation costs (‘000 mt)

		Base case	Reduced transport costs	% difference under scenario
Rice	Kenya	276	276	0
Rice	Tanzania	792	792	0
Rice	Uganda	130	137	5
Rice	Rwanda	45	46	2
Rice	Burundi	42	44	5
Maize	Kenya	2942	2942	0
Maize	Tanzania	2348	2348	0
Maize	Uganda	751	692	−8
Maize	Rwanda	110	111	1
Maize	Burundi	163	168	3
Beans	Kenya	344	344	0
Beans	Tanzania	498	484	−3
Beans	Uganda	377	367	−3
Beans	Rwanda	237	240	2
Beans	Burundi	202	207	3
	Total	9258	9199	−0.6

Table 7 Changes in production within the EAC region under reduced transportation costs ('000 mt)

		Base case	Reduced transport costs	% difference under scenario
Rice	Kenya	39	71	81
Rice	Tanzania	763	763	0
Rice	Uganda	95	72	−25
Rice	Rwanda	38	31	−18
Rice	Burundi	45	34	−25
Maize	Kenya	2920	2861	−2
Maize	Tanzania	3735	3725	−0.3
Maize	Uganda	1169	1109	−5
Maize	Rwanda	92	56	−40
Maize	Burundi	125	196	56
Beans	Kenya	397	398	0.1
Beans	Tanzania	626	668	7
Beans	Uganda	452	484	7
Beans	Rwanda	227	230	1
Beans	Burundi	226	134	−41
	Total	10,952	10,832	−1.1

The rice importers increase their demand by the same percentage change as was seen for the price reductions in those countries. The strong increase in prices for maize in Uganda, seen in Table 5, causes the demand for maize to go down there, as it does for beans in Uganda and Tanzania. By contrast, the regions in which we saw a decrease in price in Table 5 are showing a decrease in production in Table 7.

The changes in production that we see in Table 7 reflect a mix of changes within the EAC region that result from the scenario which reduces regional transportation costs. On the one hand, the reduction in net imports from the rest of the world that we observed in Table 3 is reflected here, such as the increase in rice production in Kenya which matches the decrease in imported rice from the rest of the world that was noted in Table 3. In other cases, the changes in production reflect the increased importation of goods from within the region, which require less production domestically—such as the reduction in rice production for the three landlocked countries, which now source it from within the EAC region (as shown in Table 4). In other cases, the changes in production reflect the price changes as shown in Table 5, such as the strong increase in beans prices in Tanzania and Uganda, which now export to neighboring Rwanda and Burundi (as shown in Table 4).

In total, the sum of overall production of the three key staples goes down across the region, as a result of reducing regional transportation costs. This, despite the increase of production in some regions, reflects the fact that freeing up the movement of goods allows some countries to release land from production, so that they can import

Table 8 Changes in harvested area across the various crop suitability classes within the EAC region, under reduced transportation costs ('000 ha)

		Base case	Reduced transport Costs	%diff
Rice	Rice/high suitability	17	17	0
Rice	Rice/moderate suitability	253	251	−1
Rice	Rice/low suitability	248	238	−4
Rice	Rice/unsuitable	221	208	−6
Maize	Maize/high suitability	263	255	−3
Maize	Maize/moderate suitability	3543	3458	−2
Maize	Maize/low suitability	1693	1746	3
Beans	Beans/high suitability	589	597	1
Beans	Beans/moderate suitability	1355	1376	2
Beans	Beans/low suitability	1167	1139	−2
Beans	Beans/unsuitable	1.1		−100
	Total	9350	9284	−0.7

the good they want, and put the agricultural land they have to better use growing commodities that are better suited to their local conditions. This will be reflected in the patterns of the harvested area that we will examine next.

4.1.1 Implications for Cropland Area

Finally, we can see what the impact of this scenario is on the physical landscape of the region, by looking at changes in harvested area of each crop. For this case, we can take advantage of the disaggregation of harvested area and yield to the sub-national boundaries of each country—and aggregate the effects up to the crop suitability classifications that cut across the EAC region. Table 8 shows how the changes in harvested areas across the different suitability classes appear for the region.

Here, we see that there is (overall) a small decrease in total harvested area for these three crops (−0.7%) across the entire EAC region. For rice, there is a reduction in total area, especially in those suitability zones that are moderate or lower. The area for maize in the zones with low suitability increases somewhat, whereas those in zones of moderate or high suitability go down. For beans, the zones which have high or moderately suitability increase their area slightly, whereas those with lower suitability decrease. The area of beans that falls into unsuitable zones goes to zero (albeit from a very small number, in the base case).

Table 9 Changes in harvested area across the various crop suitability classes within the EAC region, under the scenario with increased productivity (‘000 ha)

		Base case	Increased productivity	%diff
Rice	Rice/high suitability	17	17	2
Rice	Rice/moderate suitability	253	242	−5
Rice	Rice/low suitability	248	225	−9
Rice	Rice/unsuitable	221	289	30
Maize	Maize/high suitability	263	272	4
Maize	Maize/moderate suitability	3543	3535	−0.2
Maize	Maize/low suitability	1693	1260	−26
Beans	Beans/high suitability	589	612	4
Beans	Beans/moderate suitability	1355	1389	3
Beans	Beans/low suitability	1167	950	−19
Beans	Beans/unsuitable	1.1		−100
	Total	9350	8791	− 6.0

4.2 Scenario #2—Increasing Crop Productivity

The scenario in which transportation costs (and trade barriers) are lowered, leads to favorable trade patterns, lower overall prices and increased production. But this, by itself, does not necessarily align the production of the 3 staple crops into zones of higher suitability. If we impose, instead of this scenario, an alternative case in which the productivity (i.e., yield) of the key staple crops[11] is increased by 10%, then we see a much stronger re-alignment of crop areas with more favorable suitabilities, as is seen in Table 9.

From these results, we see that the productivity increase has a clear ‘land-saving’ effect that reduces the overall harvested area more strongly than the scenario with reduced transportation costs. We also see that there is a sharp reduction in the areas under low levels of suitability and an increase in the area under higher levels of crop suitability, for each of the staples. This illustrates the effect that on-farm technological improvements can have in helping re-align the production patterns to make better use of the agro-ecological potential that exists on the ground.

In terms of the impact that this particular scenario has on trade patterns, where the original transportation costs between regions are the same as the base case, we show the changes to trade patterns within the EAC region in Table 10.

In contrast to Table 4, where the lowered transportation costs made the coastal countries the largest regional exporters to their neighbors, we see here that the effects are limited to the landlocked countries, which still face high costs of importing goods from either the rest of the world, or their neighbors. With an increase in yield

[11] In this scenario, this yield increase is implemented for those crops falling into zones where less than 60% of the maximum potential is currently being realized.

Table 10 Changes in net trade within the EAC region under increased staples productivity ('000 mt)

		Base case	Increased productivity
Rice	Uganda		−86
Rice	Rwanda		116
Rice	Burundi		−30
Maize	Kenya	−2	
Maize	Uganda	2	149
Maize	Rwanda		−53
Maize	Burundi		−96
Beans	Kenya	−3	
Beans	Tanzania	3	
Beans	Uganda	9	174
Beans	Rwanda		−77
Beans	Burundi	−9	−97

productivity, the local comparative advantage across the landlocked countries in the various key staples shifts, allowing some to become exporters to their neighbors. In terms of rice, Rwanda starts exporting to Uganda and Burundi, whereas in terms of maize and beans, Uganda becomes the major exporter of those commodities to its landlocked neighbors, Rwanda and Burundi. From Table 2, we saw that Uganda has a sizable area of maize and beans under zones of moderate or low suitability—so the effect of this scenarios seems to enable it to meet more of its productive potential, such that it is able to start exporting and supplying the needs of its neighbors, who face even more severe constraints on land availability.

4.2.1 Combined Scenario—Increased Crop Productivity and Reduced Transport Costs

If we now combine the two policy scenarios—such that we cut the costs of transportation within the EAC region by 50% at the same time that we raise the crop productivity levels in the less-favored zones by 10%—we can see the combined effects on trade and production patterns within the EAC region.

The effect on harvested crop area within the EAC region (divided into the various suitability zones) is shown in Table 11.

The table shows an even greater reduction in crop harvested area compared to the case with only yield increases (Table 9) and by a bigger percentage than the simple sum of the area reductions under that scenario and the scenario with only reduced transportation costs (Table 8). The combined scenario leads to a complete shift out of rice production in areas that are unsuitable for that crop, in addition to the shift out of unsuitable bean area (as was the case in the previous two scenarios). Even

Table 11 Changes in harvested area across the various crop suitability classes within the EAC region, under increased productivity and lower transport costs ('000 ha)

		Base case	Increase yield and reduce transport cost	%diff
Rice	Rice/high suitability	17	17	2
Rice	Rice/moderate suitability	253	255	0.5
Rice	Rice/low suitability	248	247	−0.5
Rice	Rice/unsuitable	221		−100
Maize	Maize/high suitability	263	245	−7
Maize	Maize/moderate suitability	3543	3402	−4
Maize	Maize/low suitability	1693	1506	−11
Beans	Beans/high suitability	589	613	4
Beans	Beans/moderate suitability	1355	1380	2
Beans	Beans/low suitability	1167	967	−17
Beans	Beans/unsuitable	1.1		−100
	Total	9350	8631	−7.7

though the area of maize in high suitability areas does not increase as it did in the scenario with just increased yield (Table 9), there is still a strong shift out of areas with low maize suitability, as is also the case with rice and beans. So there is still a strong re-alignment effect of production away from low suitability areas, when the two policy interventions are combined. In order to see a more complete picture of how harvested area changes under the alternative policy scenarios, for each country, we can refer to Table 13 in the annex.

In terms of trade within the EAC region, we see the emergence of strong regional exporters like Tanzania, as is shown in Table 12.

In this table, we see that the landlocked countries become strong importers of products supplied by their EAC neighbors, such as Tanzania, with Uganda remaining a big regional exporter of beans, similar to the case with just yield increases (Table 10). The combination of lowering the barriers to regional trade and increasing crop productivities increases region-wide exchange of key staple goods, while encouraging countries with enhanced comparative production advantage to benefit from greater trade opportunities.

These results illustrate the potential that policy interventions focused on trade can have, in comparison to those coming from technology-focused interventions that are directly aimed at on-farm improvement. The combined effect enhances the environmental benefits of avoided land expansion by allowing the region to reduce overall crop area in favor of higher productivity growth and to reduce areas under crops that can be more productively grown elsewhere and sourced through an enhanced trade network.

Table 12 Changes in net trade within the EAC region under increased staples productivity and lower transport costs (‘000 mt)

		Base case	Increase yield and reduce transport cost
Rice	Tanzania		97
Rice	Uganda		−58
Rice	Rwanda		−19
Rice	Burundi		−20
Maize	Kenya	−2	
Maize	Tanzania		179
Maize	Uganda	2	
Maize	Rwanda		−63
Maize	Burundi		−116
Beans	Kenya	−3	
Beans	Tanzania	3	9
Beans	Uganda	9	194
Beans	Rwanda		−90
Beans	Burundi	−9	−112

Source model simulations

5 Implications for Policy

Based on these results, we can conclude that lowering the costs of transportation within the EAC region, through improvements in infrastructure, can play a big role in promoting trade of staples within the EAC region. For a country like Tanzania which is the ‘doorway’ to the ocean for six landlocked countries in Africa, the infrastructure improvements that it can make to its roads, railways, and harbors can have a tremendous effect on the patterns of production and trade within the region. This has been noted by other studies such as Foster and Briceño-Garmendia (2010) and Anyango (1997). These same patterns of trade and production are also affected strongly by the increases in crop productivity that are introduced in the second policy scenario—and by the combined effect of both of them. The patterns of trade are moved by two primary forces that are in operation in these scenarios: (1) The changes in transport cost open up new possibilities for exchange across the EAC region and allow for goods to flow in ways that take the best advantage of relative regional differences in price, and (2) from the changing levels of supply and demand within each region, that are affected by the scenario-induced changes in price, and which require a shift in trade to bring the individual countries (and the entire EAC region) into a new equilibrium and trade balance.

The lowering of prices that we observe in the EAC countries that can now import goods more easily from their neighbors has a clear benefit to consumers in those countries, who can now source their consumption at a lower price. While this provides a disincentive to the pure producers in those countries, the farm households that both consume and produce these commodities will experience an offsetting effect from the consumption side. In those countries that now have the opportunity to export to their neighbors, the prices will tend to go up as a way of encouraging increased output, and which provide benefits to the producers in those countries (at some disadvantage to the consumers). So, the distribution of welfare benefits is differentiated between the producers and consumers, and the kind of price changes that they face, with some regions experiencing greater consumer benefits for some goods, while others have greater benefits going to the producers of those goods which experience price increases.

Any benefits that go to either producers or consumers, under these scenarios, should be balanced against other consumer- or producer-focused policy measures (such as subsidy or voucher programs) to evaluate the cost-effectiveness of this policy. Since we're only looking at three commodities, we are not even capturing the widespread benefits that improving infrastructure can have on the rest of the agricultural sector and the rest of the economy. The high costs of transport are often cited as some of the biggest barriers to growth and development that many of Africa's landlocked countries face. So we are certainly understating the overall economy-wide benefits to improving infrastructure, which might make this scenario look even more attractive, when compared with other policy alternatives. This kind of comprehensive assessment can (and should) be taken up in further work.

The scenario on productivity gains illustrates that there is an important role for improving on-farm performance potential in re-aligning East Africa's agricultural production such that it takes better advantage of its agro-ecological potential, while also unlocking its latent trade potential. A policy instrument purely focused on trade cannot achieve this re-alignment, by itself, but must be coupled with other measures, in order to be effective in improving both producer and consumer welfare, as well as moving production patterns toward better exploitation of existing suitabilities that exist on the agricultural landscape. The combined scenario in which the reduction of transportation costs was coupled with the increase in crop productivity clearly illustrated this and gave the greatest reduction in overall crop harvested area across the various crop suitability zones of the EAC region.

Given that the goal of sustainable intensification is now becoming more prominent in the agricultural policy objectives of many countries, any combination of instruments which leads to less land conversion and better utilization of existing crop growth potential will contribute strongly toward this goal. The re-alignment of crop production away from areas of low suitability (where more chemical inputs would be required to achieve acceptable growth levels) would also enhance the suitability of agricultural production within the EAC region, and we see that the combined scenario makes the most progress in this direction. This opens up some promising possibilities for further development of the agricultural sector that can go beyond the three key staples that we have focused on for the purposes of this study.

As was noted earlier in this report, we do not have sufficient information on the actual quantities of stocks being held of the three key staple commodities to model the way in which inventories might be accumulated or released, in response to prices or other policy-relevant factors. This may be leaving out an important component of food security-focused policy that governments might be actively taking to manage the quantities of important food grains such as maize on the markets (and their prices). If better information on this were to become available—such that we could model stockholding behavior (on the part of private or parastatal agencies) with some confidence—then the effect of policy-induced changes in stock management on market-level availability and prices of the commodity of interest could be explored further.

6 Limitations of the Study and Extensions

As with all studies, there are limitations to the analysis that we have provided here. Nevertheless, the results from the study, so far, provide intuitive and interesting results that we will continue to build upon in further improvements and extensions to the existing modeling framework. The map of agro-ecological potential that was developed by experts within the study might be updated, in the future—and the disaggregation of area and yield that was used by the model can be adjusted to this data, using the same techniques described here. In addition to this, other data improvements (on prices, base transport costs, the disaggregation of urban and rural demand, etc.) can also be used to refine the analysis in future.

In carrying out this work, we recognize that there are a number of methodological and empirical challenges that need to be met to bring out further details on the functioning of the key staple markets of interest that we examined—such as the issue of stock management that was mentioned in the previous section. Due to time and resource constraints, we were not be able to fully address some of these—but will aim to have a better understanding of how these can be addressed in follow-up activities, as a basis of further research.

An important challenge that we face in our modeling work is being able to capture the process by which market access to the key commodities we're interested in can be improved or change, as a function of infrastructure improvements or policy changes. We provide the 'entry points' in our modeling to capture transaction (i.e., transportation and/or marketing) costs and policies that affect the openness of trade and commercial exchange—but the relationship between these levels of market 'friction' and the degrees of investment in better roads, market storage, infrastructure, etc., cannot be precisely defined or measured. Given this uncertainty, we have limited ourselves to the examination of alternative cases in which these costs can be 'low' or 'high' and describe the change in market outcomes and welfare in relation to specific policy measures or interventions that might be able to bring these changes about. In effect, we have done sensitivity analysis, without being able to ascribe a precise degree of causality between specific interventions and the mechanisms by which they affect the market. A more comprehensive study that could engage experts in thinking more systematically through these linkages could help to unpack these important policy dimensions.

7 Conclusions

In this chapter, we have explored a number of illustrative results for the agricultural sector of the EAC region that point to ways in which policy can intervene to promote greater productivity, food security, and more favorable market conditions for trade within the region. Although we have just focused on three staple crops, the key components of this analysis could be extended to other commodities in a straightforward way, in order to illustrate the implications of lower transport costs and improved productivity for the wider agricultural sector.

Our analysis makes a clear case for coupling investments focused on boosting on-farm productivity with those that improve the marketability of the products produced on-farm. This is not a new message, in agricultural policy circles, but is one which comes out strongly from our results. The way in which we have combined the expert assessment and analysis of agronomists in the EAC region with an economic analysis of production/supply, demand and trade points towards a methodology that can actively engage economic and agronomic experts in the policy analysis process. Taking this approach, useful information from both disciplines is combined and utilized within an analytical engine that can evaluate scenario-based policy alternatives. Oftentimes, this kind of linkage between agronomic and economic assessments is not done in a way that is useful for policy-making and has left many potentially useful analyses of agricultural investment and policy only telling half of the story that decision-makers need to consider. We hope that this chapter helps to point toward a way in which the important biophysical and socio-economic constraints and realities that underlie developing agricultural economies can be captured in policy-relevant, model-based assessments.

Technical Annex: Details of the EAC Multi-market Model

Annex 1: Structure of the Partial Equilibrium, Multi-market Model

The general schematic of a multi-market model is shown in Fig. 4 and encompasses a wide variety of models that are in common use in empirical economic work.

Figure 4 shows the basic conceptual outline of the class of economic model that we are building, following the theoretical treatment described in Sadoulet and De Janvry (1994). A model of agricultural markets has to take into account the markets for the products themselves—which we will mostly focus upon—as well as the markets for key factors that are necessary for production—chiefly labor and marketed inputs like fertilizer. In the case where agriculture is relatively low in input use and mostly uses un-paid household labor—which is common in the African context—then much of what is supplied to crop production may not be captured in this kind of framework and may not have observable data (without a detailed household-level production survey, which is beyond the scope of this study). In the case of high-value and highly commercialized crops, this may be possible—and agricultural may compete with other sectors for these inputs, in which case a multi-sector representation of input and output markets might be needed (as in many applied general equilibrium models). Our modeling work falls into the 'partial equilibrium' category—which signifies the fact that not all inputs and outputs within the economy are modeled, and the circular flow of revenue from production activities, to the consumer and the returns to the use of key inputs like capital are not accounted for fully. We have built upon the basic framework of a partial-equilibrium, multi-market model of agricultural trade described by Minot (2009). There are models of this class that have been built for many of the EAC countries—but they do not have the detail on agriculture that is needed for this project—so we have adopted this approach.

The key elements of Fig. 4 that we will take into account are:

- The assumption of profit-maximization on the part of agricultural producers.
- The assumption of utility maximization on the part of consumers of agricultural products.
- The influence of income growth on demand, although the growth in income will not be fully endogenized, as we do not capture the full picture of payments to households (from ag and non-ag activities) and their expenditure on food and non-food goods.
- We will consider the per-capita demand for products, and use exogenous projections of population growth to extrapolate it to the national-level demand.

The key components of the multi-market model are as follows:

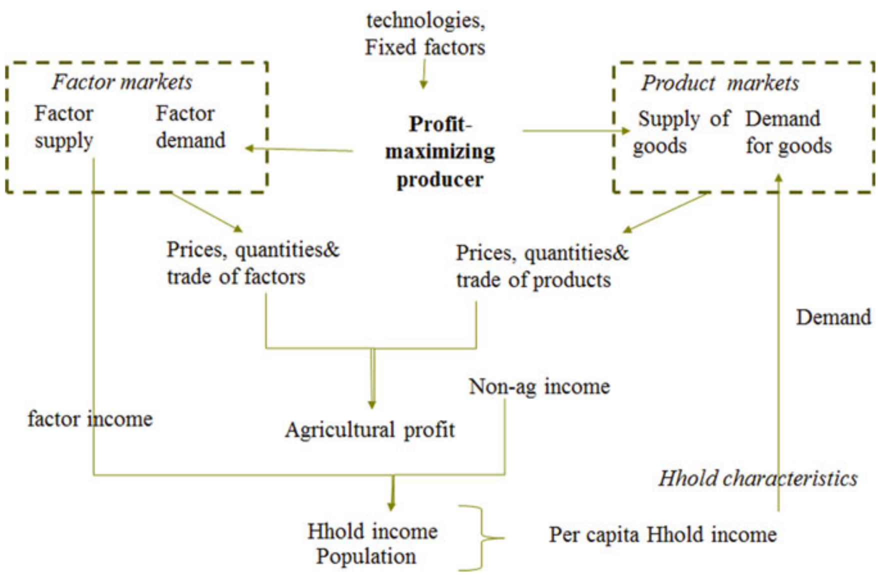

Fig. 4 Key relationships within a generic multi-market model

- The supply, demand, and trade balance that occurs at the national level for each country in the EAC region
- The relationship that describes the response of crop production to price changes—in terms of irrigated and rainfed harvested area and yield
- The relationship that describes the commodity food demand to price and income changes
- The key relationships that link prices across regions and relate them to the transportation costs and the volume of trade between those regions.

These relationships comprise the essential components of the EAC multi-market model that we use in this study.

The overall balance between supply, demand, and trade is captured in this equation

$$\begin{aligned}\text{Supply} + \left(\sum_{\text{other EACRegs}} (\text{inflows}) + \text{iMport}_{\text{RoW}}\right) &= \text{Demand} \\ &+ \left(\text{eXprt}_{\text{RoW}} + \sum_{\text{other EACRegs}} (\text{outflows})\right) + \text{StockChange}\end{aligned}$$

where ‘Supply’ is the aggregate production coming from the harvested area and yields across the sub-regions of each of the EAC countries, as is given here

$$\text{Supply} = \text{Area} \cdot \text{Yield} = \sum_R \sum_s \left(A_{R,s} \cdot Y_{R,s}\right)$$

where each country is divided into sub-regions (provinces or districts) denoted by R which, in turn, are subdivided into zones of crop suitability that were defined by the GIS-based analysis of other WaLETS team members.[12] The multiplication of harvested area and yield and their addition over the sub-regions and crop suitability zones of each country is what constitutes the national production of each crop.

Calibrating the supply side of the model

The yield of the crop, in each suitability zone, has been informed by the inputs of agronomic experts from each of the EAC regions covered in the WaLETS project. The information gained from these experts—which indicated the share of maximum potential yield of each crop that would be attainable within that zone—was used to allocate the base areas and yields of the model, for all the EAC countries, such that the national-level supply was matched, and could be balanced to demand and trade. This process is described in more detail, in the next subsection of this technical appendix.

---

[12]The GIS-based analysis of crop suitability is described in separate technical documentation of the WaLETS project and has been led by Kennedy Were of KALRO and his colleagues. This can be accessed at: https://www.kilimotrust.org/documents/reports/2017/walets/WaLETS_Final_Reports/WaLETS_GIS_TechnicalReport.pdf.

The harvested area of each crop responds to prices, according to the net revenues that can be realized per hectare of each crop across the available cropland area. The key essence of this decision process is captured in the following optimization problem

$$\begin{aligned} \max \quad & \sum_j \left[p_{j,R} \cdot Y_{j,R,s} \cdot A_{j,R,s} - c\left(A_{j,R,s}\right) \right] \\ \text{s.t.} \quad & \sum_j A_{j,R,s} \leq \bar{A}_s \end{aligned}$$

where the regional prices of each crop (j) are $p_{j,R}$, and $c\left(A_{j,R,s}\right)$ represents a nonlinear relationship between the total area harvested and the per-hectare cost of cultivation. This nonlinear relationship is a necessary component of calibrating the model to observed crop areas and is part of the 'positive' mathematical programming (PMP) approach (explained in the next subsection).

This method for allocating crop area is in contrast to the 'reduced-form' approach used in an earlier version of this work, in which an iso-elastic area response function was hypothesized ($A_{R,r} = f_A(P) = c_A^{R,r} \cdot P^{\gamma}$[13]). In this function, the reaction of area to price changes reflects the behavior that is embodied in the optimization problem that we now model explicitly.

The Positive Mathematical Programming Approach

The positive mathematical programming principle of Howitt (1995a, b)—referred to, popularly, as PMP—is a method of calibrating economic models of agricultural production, that exploits the mathematical principles of duality that is embedding in all mathematical programming models. In essence, the PMP approach imposes a degree of curvature upon the objective function of the mathematical programming model, such that it causes the model solution to exactly equate the implied marginal costs of land allocation that are reflected, implicitly, in the observed behavior of the decision-maker. Since we, typically, do not observe the marginal costs that the decision-maker faces—but, rather, the average costs that are reflected in the collected data—we often have trouble calibrating economic models to replicate the land allocation behavior that we observe from farmers.

The PMP approach takes the valuation of land (and other) resources that is implicit in the observed allocation of crop area and uses the 'shadow values' derived from a constrained 'stage 1' problem, in order to reconstruct the nonlinear cost function that allows the model to calibrate exactly to the observed data in 'stage 2'.

To illustrate, let us suppose that we start with a mathematical programming problem of land allocation among alternative crops that is linear in both the revenue and cost terms—such as the following problem.

[13]In this function, the parameter $c_A^{R,r}$ is a calibrating constant, and the 'elasticity' γ gives the response of area to a change in price.

$$\begin{aligned} \max \quad & \sum_j \left[p_{j,R} \cdot Y_{j,R,s} \cdot A_{j,R,s} - c_{j,R,s} \cdot A_{j,R,s} \right] \\ \text{s.t.} \quad & \sum_j A_{j,R,s} \leq \bar{A}_{R,s} \end{aligned}$$

In this case, the parameter $c_{j,R,s}$ reflects the per-hectare cost of cultivation (specific to crop, region, and soil class) which remains constant at all scales of production activity. This is a classic linear programming type of problem which would tend to over-specialize in one activity—such that all of the available cropland would be allocated to the most profitable crop (i.e., the crop with the largest gross margin per hectare—$p_{j,R} \cdot Y_{j,R,s} - c_{j,R,s}$). This type of solution does not typically reflect the kind of behavior that one usually observes among farmers, where there is often a mix of crops in the farming portfolio. Since we only observe the average costs of production, per hectare ($c_{j,R,s}$), we seek to obtain a better measure of the actual marginal costs of production that an economically optimizing farmer would equate with marginal revenue when reaching the mixed allocation of cropland that we observe in data ($\{\bar{a}_1, \ldots, \bar{a}_k\}$ for k crop types).

Therefore, if we posit the existence of a nonlinear cost function in crop area that would allow the model to replicate these implicit (but un-observed) marginal costs of the quadratic form: $TC\left(A_{j,R,s}\right) = \phi^0_{j,R,s} \cdot A_{j,R,s} + \frac{1}{2}\phi^1_{j,R,s} \cdot \left(A_{j,R,s}\right)^2$.[14]

Then, we can use the PMP procedure to obtain the values of the parameters of this cost function in three separate stages.

In the first stage, we modify the linear programming problem, shown above, to include constraints on crop area, such that the model is forced to replicate the solution that we observe in the data

$$\begin{aligned} \max \quad & \sum_j \left[p_{j,R} \cdot Y_{j,R,s} \cdot A_{j,R,s} - c_{j,R,s} \cdot A_{j,R,s} \right] \\ \text{s.t.} \quad & \sum_j A_{j,R,s} \leq \bar{A}_{R,s} \\ & A_{j,R,s} \leq A^{\text{obs}}_{j,R,s} \cdot (1 + \varepsilon) \end{aligned}$$

In the solution of the model, we expect (according to economic principle) for the marginal revenue of production to be equated to the marginal cost of production for each of the crops in the optimal allocation.

The 'shadow value' that comes from the calibration constraints of the constrained programming problem, above, at the optimal solution represents the difference between the marginal costs that the decision-maker is hypothesized to equate at the observed data point and the average costs which we observe in the data. In other words, $\mathrm{MC}_{j,R,s} = \mathrm{AC}_{j,R,s} + \lambda_{j,R,s}$, where MC and AC are the marginal and average costs per hectare of crop activity (j) in region R and on suitability class s. If we take our total cost function and manipulate it to obtain the functional form for the marginal and average cost, i.e.,

[14]The function does not have to be quadratic—but must be convex in curvature. We have chosen the quadratic form simply for analytical convenience (in implementation and exposition to the reader).

$$
\begin{aligned}
&\mathrm{TC}\big(A_{j,R,s}\big) = \phi^0_{j,R,s} \cdot A_{j,R,s} + \tfrac{1}{2}\phi^1_{j,R,s} \cdot \big(A_{j,R,s}\big)^2 \\
&\mathrm{MC}\big(A_{j,R,s}\big) = \frac{\partial \mathrm{TC}\big(A_{j,R,s}\big)}{\partial A_{j,R,s}} = \phi^0_{j,R,s} + \phi^1_{j,R,s} \cdot A_{j,R,s} \\
&\mathrm{AC}\big(A_{j,R,s}\big) = \frac{\mathrm{TC}\big(A_{j,R,s}\big)}{A_{j,R,s}} = \phi^0_{j,R,s} + \tfrac{1}{2}\phi^1_{j,R,s} \cdot A_{j,R,s} \\
&\text{so that} \\
&\mathrm{MC} - \mathrm{AC} = \lambda_{j,R,s} = \big(\phi^0_{j,R,s} + \phi^1_{j,R,s} \cdot A_{j,R,s}\big) - \big(\phi^0_{j,R,s} + \tfrac{1}{2}\phi^1_{j,R,s} \cdot A_{j,R,s}\big) \\
&\qquad\quad = \tfrac{1}{2}\phi^1_{j,R,s} \cdot A_{j,R,s}
\end{aligned}
$$

So we are able to recover the 'slope' of our cost function $\left(\phi^1_{j,R,s}\right)$ from the shadow values derived in 'stage 1', such that $\hat{\phi}^1_{j,R,s} = \frac{2\lambda_{j,R,s}}{A^{obs}_{j,R,s}}$. The 'intercept' of the quadratic cost function $\left(\phi^0_{j,R,s}\right)$ can then be obtained by using the average costs obtained from data $\big(c_{j,R,s}\big)$, and equating it to the functional form we derived earlier, such that:

$$
\begin{aligned}
&c^{\mathrm{obs}}_{j,R,s} = \mathrm{AC}\big(A_{j,R,s}\big) = \phi^0_{j,R,s} + \tfrac{1}{2}\hat{\phi}^1_{j,R,s} \cdot A^{\mathrm{obs}}_{j,R,s} \\
&\text{so that} \\
&\hat{\phi}^0_{j,R,s} = c^{\mathrm{obs}}_{j,R,s} - \tfrac{1}{2}\hat{\phi}^1_{j,R,s} \cdot A^{\mathrm{obs}}_{j,R,s}
\end{aligned}
$$

And we are able to obtain the values that define our cost function $\left(\hat{\phi}^0_{j,R,s}, \hat{\phi}^1_{j,R,s}\right)$.

We can now insert this calibrated cost function into the mathematical programming problem, in place of the linear cost term, such that we obtain the following nonlinear programming problem

$$
\begin{aligned}
&\max \sum_j \left[p_{j,R} \cdot Y_{j,R,s} \cdot A_{j,R,s} - \left(\hat{\phi}^0_{j,R,s} \cdot A_{j,R,s} + \tfrac{1}{2}\hat{\phi}^1_{j,R,s} \cdot \big(A_{j,R,s}\big)^2 \right) \right] \\
&\text{s.t.} \sum_j A_{j,R,s} \le \bar{A}_{R,s}
\end{aligned}
$$

And solving this un-constraint, mathematical programming problem, with the nonlinear term for cultivated area embedded in it, will now allow the model to find an optimal solution that matches exactly to the observed crop areas observed in the data. This is the essence of the PMP methodology, as we have implemented it in our model.

The PMP calibration method has found wide-ranging applications to the construction of policy analysis models, and numerous examples of its use can be found in the work of various authors in the agricultural economics literature (Gohin and Chantreuil 1999; Barkaoui and Boutault 2000; Heckelei and Britz 2000, 2005; Heckelei and Wolff 2003; Heckelei et al. 2012; Judez et al. 2001; Röhm and Dabbert 2003; Henry de Frahan et al. 2007; Kanellopoulos et al. 2010; Howitt et al. 2012; Mérel and

Bucaram 2010; Mérel et al. 2011, 2014; Mérel and Howitt 2014; Doole and Marsh 2014).

Calibrating the demand side of the model with trade

On the demand side, we divide this between food and non-food demand—keeping non-food demand as an exogenous parameter $\left(\bar{Q}_D^{\text{non-food}}\right)$.

$$\text{Demand} = f_D(P, y) \cdot \text{popn} + \bar{Q}_D^{\text{non-food}}$$

Like harvested area, food demand also responds to prices—albeit negatively, as we'd expect demand to go down with increasing prices. The per-capita demand function $f_D(P, y) = c_{\text{dmd}} \cdot P^{\varepsilon} \cdot y^{\eta}$ also has a response to per-capita income, which is generally positive for the three key crops we consider, but could be negative in other cases.[15]

The stock change at the national level is also kept fixed and exogenous, in our model. In principle, this can represent a component of supply response, as public and private managers of cereal stocks could decide to release stocks in response to price and augment national supply—or withdraw supply by adding to stocks. We do not have sufficient data on public and private stockholding, at the moment, to operationalize a behavioral model of this.

The price relationships lie at the heart of the way in which the model adjusts both demand and supply within each region—as well as the trade flows that happen across regions. The difference between prices in two adjacent regions determines the direction of the trade flow. The following relationship shows the 'no-arbitrage'[16] condition that is hypothesized to hold between two regions, where the transportation cost (TC) provides the 'wedge' between the two prices and accounts for the cost of delivering each unit of good from its origin to its destination

$$P_{\text{Region}(R)} + \text{TC}_{R,R'} \geq P_{\text{Other Region}(R')}$$

In this relationship, the price that one receives in the destination region $\left(R'\right)$ should be no less than the purchase price in the region of origin (R) and the cost of transporting it between the two $\left(\text{TC}_{R,R'}\right)$. Where the purchase + transport cost is greater, then it is not optimal to export from R to neighboring R'(or, conversely, to import into R' from R). Where this relationship holds with equality (i.e., $P_R + \text{TC}_{R,R'} = P_{R'}$), then one could expect to have nonzero levels of trade. This complementary relation-

[15]In cases where a good is 'inferior' to other preferred goods, the per-capita consumption could go down with income. An example could be the declining demand for coarse grains (millet/sorghum) as household income increases, in favor of rice- and wheat-based products.

[16]The 'no-arbitrage' condition describes a competitive market equilibrium, where any opportunity for selling a good for a higher price than the original purchase price + transport cost is exhausted. So, at best, an agent can break even by selling a good for exactly the cost at which it was purchased plus the cost of delivering it to the destination, but no more. This assumption could be relaxed in a less competitive market environment—but that is beyond the scope of this study.

ship[17] between the levels of trade and the equality or non-equality of the no-arbitrage constraint, which can be written as

$$\begin{aligned} P_R + \mathrm{TC}_{R,R'} &> P_{R'} \quad \text{and} \quad \mathrm{Trade}_{R,R'} = 0 \\ P_R + \mathrm{TC}_{R,R'} &= P_{R'} \quad \text{and} \quad \mathrm{Trade}_{R,R'} \geq 0 \end{aligned}$$

And it can, in turn, be summarized by the following equality statement

$$\left[P_R + \mathrm{TC}_{R,R'} = P_{R'}\right] \cdot \mathrm{Trade}_{R,R'} = 0$$

which must hold at all times. This type of relationship determines the trade levels between the EAC countries, as well as between each EAC country and the rest of the world.

In addition to this, there could be other taxes on imports or exports, such that the unit price of each traded good is affected—or there could be a quantitative limit on total exports and imports such that we have either a quota on imports

$$\mathrm{Quota}^R_{\mathrm{Import}} \geq \sum_r M^R_r$$

Or a quota on exports

$$\mathrm{Quota}^R_{\mathrm{Export}} \geq \sum_r X^R_r$$

where the quota (and tax) levels are decided by policy, and the levels of import $\left(M^R_r\right)$ and export $\left(X^R_r\right)$ for each country R are summed over the quantities going from each sub-region, r.[18] In our model, we only apply import and export taxes (and quotas) on non-EAC trade, leaving the EAC region a 'free-trade zone', as it was intended.

In terms of defining the prices that exist on the 'border' and which are relevant to determining the levels of exports (P_X) to and imports (P_M) from the rest of the world, these can be adjusted from the fixed and exogenous 'world price' (P_{world}) to determine the import and export parity prices.

We can account for export taxes (t_X) vis-à-vis FOB price that is relevant for exporters as

$$P_X = \mathrm{NER} \times P_{\mathrm{world}} \times (1 - t_X)$$

[17] This describes the kind of 'mixed complementarity' formulation that is commonly applied to solve trade equilibrium problems and is found in other types of mathematical programming problems. In such a problem, one does not need to maximize or minimize an economic objective function, since these complementary relationships summarize the first-order necessary conditions required to solve the implicit optimization problem. See Paris (2010) for more details.

[18] In our implementation, we did not model trade between the sub-regions (r) of each country, as that would have imposed an enormous computational burden on the model and require detailed data beyond what we possess. So, we only model trade at the national level, in this model.

whereas the import taxes (t_M) can be applied to the world price to define the relevant CIF price that an importer would care about

$$P_M = \text{NER} \times P_{\text{world}} \times (1 + t_M)$$

In each case, NER denotes the exchange rate between local currency and the US dollar (which the world price is denominated in).

In summary, the key endogenous variables that are solved by the model are:

$$X, M, P_{\text{region}}, \text{inflows, outflows, Demand, Supply}$$

whereas the exogenous and fixed parameters of the model are

$$P_M, P_X, P_{\text{world}}, \text{Quota}^R_{\text{export}}, \text{Quota}^R_{\text{import}}, \text{TC}, t_M, t_X, \text{NER}$$

The model is solved as a mixed complementarity problem—in which the following relationships must be satisfied by the optimal solution (i.e., the market equilibrium):

Export price relationship

$$\left[P_R + \text{TC} + \text{ImTax}_{\text{export}} - P_X\right] \cdot X = 0$$

Import price relationship

$$\left[P_M + \text{TC} + \text{ImTax}_{\text{import}} - P_R\right] \cdot M = 0$$

Domestic price relationships

$$\left[P_R + \text{TC}_{R,RR} - P_{R'}\right] \cdot \text{TQ}_{R,R'} = 0$$

where $\text{TQ}_{R,R'}$ represents the quantity traded between regions (going from R to R'), such that

$\sum_{R'} \text{TQ}_{R',R} = \sum_{\text{other Regs}}$ (inflows) and $\sum_{R'} \text{TQ}_{R,R'} = \sum_{\text{other Regs}}$ (outflows) for any region R

Quota on exports

$$\left[\text{Quota}^R_{\text{Export}} - \sum_r X^R_r\right] \cdot \text{ImTax}_{\text{export}} = 0$$

Quota on imports

$$\left[\text{Quota}^R_{\text{Import}} - \sum_r M^R_r\right] \cdot \text{ImTax}_{\text{import}} = 0$$

where the implicit tax on exports or imports is nonzero if the quantity constraint on total exports or imports becomes binding. In such a case, the export and import price relationships have to account for nonzero values of both transportation costs and the implicit tax in the no-arbitrage condition.

This structure obviates the need for an explicit objective function (the maximization of joint consumer and producer surplus, for example)—but requires that the model have an equal number of equations and free variables. This requires us to exercise particular care in the preparation and checking of the data before use in the model.

Calibration of model in Supply, Demand, and Trade

Based on the specified structure of the model, that we've described, and the data with which it has been built, we now have a complete framework for doing simulation of supply, demand, and trade in the EAC region. To ensure that the results are congruent with what we would expect, we must validate the base model results against observed data. As a first step in the model validation process, we verify that the model is able to reproduce the observed supply, demand, and trade that is observed in the data, in order to gain confidence in its proper functioning. This will verify that the calibration of transportation costs and border prices (with the rest of the World) were done correctly.

In Table 13, we show that the production and demand of the three key commodities in the EAC countries match the data from FAO exactly when the market equilibrium is simulated by the model.

Similarly, in Table 14, we show that the imports and exports of the three key commodities in the EAC countries also match the data found in the FAO trade balances.

Based on this, we can now have confidence that the national-level supply, demand, and trade balances can be captured correctly by the model, which allows us to proceed in carrying out alternative policy simulations. In addition to this level of calibration, we must also obtain sub-national values of area and yield for the three key staple commodities that adds up to the national-level totals for production that are shown (Table 13). This process of calibration is described in the next subsection.

Annex 2: Calibrating Base Areas and Yields to GIS-Based Modeling Outputs

In order to ensure that the model is completely consistent with the key underlying data, we had to undertake an extensive exercise to disaggregate the national-level production for each EAC country into sub-national areas and yields that are consistent with both (1) the administrative region-level statistics on irrigated and rainfed areas and yields provided by the *HarvestChoice* project database and (2) the division of total cropland into zones of suitability provided by the GIS analytical component of WaLETS.

Table 13 Calibration of model to observed production and consumption at the national level of each EAC country ('000 tons)

		Production		Food demand	
		Model solution	FAO data	Model solution	FAO data
Rice	Kenya	39.3	39.3	275.7	275.7
Rice	Tanzania	763.3	763.3	792.0	792.0
Rice	Uganda	95.3	95.3	130.3	130.3
Rice	Rwanda	38.0	38.0	45.0	45.0
Rice	Burundi	44.7	44.7	42.3	42.3
Maize	Kenya	2920.0	2920.0	2942.0	2942.0
Maize	Tanzania	3735.3	3735.3	2348.3	2348.3
Maize	Uganda	1169.3	1169.3	750.7	750.7
Maize	Rwanda	92.3	92.3	110.0	110.0
Maize	Burundi	125.0	125.0	163.0	163.0
Beans	Kenya	397.3	397.3	344.3	344.3
Beans	Tanzania	626.0	626.0	498.3	498.3
Beans	Uganda	452.3	452.3	377.3	377.3
Beans	Rwanda	227.0	227.0	236.7	236.7
Beans	Burundi	226.3	226.3	201.7	201.7

These two sources of information provide the basis for calculating a plausible sub-national distribution of areas and yields for the three staple crops that meet these criteria:

(1) They add up (across administrative units) to the national totals, reflected in the national-level supply—that must balance with total demand, stock change, and trade.
(2) They add up (across suitability classes) to the same national totals.
(3) They constitute a share of total cropland that is consistent with the GIS analytical outputs.
(4) They reflect the relative yields across suitability classes that was suggested by the expert agronomists working on the WaLETS team.

In order to satisfy all of these criteria simultaneously, we had to use a model-based method to process the inputs from *HarvestChoice*, the WaLETS GIS team, and the agronomy experts, so that the outputs used for the economic model could reflect the best use of that information, and satisfy the necessary trade balance and adding-up conditions that are embedded in the trade model structure.

Since we are trying to obtain unique values for area and yield across sub-regions (r), crops (j), suitability classes (s), and countries (R), we end up solving a problem which is **ill-posed**. In other words, it contains more unknowns than data points and has *negative* degrees of freedom.

The maximum entropy formalism provides an efficient and robust way of solving ill-posed estimation problems and has been applied widely by scientists (ranging from the social to the physical sciences) to complex empirical problems.

Maximum Entropy/Minimum Cross-Entropy Approach

In order to meet the challenge of simultaneously satisfying various balancing and adding-up constraints, while trying to get the estimates for shares of irrigated areas and yields in total area and production that are most consistent with the various (although not always compatible) sources of information that we have, requires an optimization-based approach. While a least-squares type curve-fitting approach can be used, in order to minimize the distance between a target level of 'closeness' to the existing data, while satisfying certain constraints that have to be met exactly can be used for problems like this, there is often the issue of how to satisfy several different targets simultaneously, without imposing undue weight on any particular objective criterion over another. The problem becomes particularly vexing when one is confronted with a relative scarcity of reliable data—which might be far fewer in number than the many unknowns which have to be determined. This situation leads to the type of 'ill-posed' problems, where the degrees of freedom are non-positive, and solving a classical inverse-type problem, where the unknowns may be solved for by linear algebraic inversion of a data matrix with respect to a vector of variables, becomes impossible.

Table 14 Calibration of model to observed trade at national level

		Imports		Exports	
		Model solution	FAO data	Model solution	FAO data
Rice	Kenya	241.3	241.3		
Rice	Tanzania	112.7	112.7		
Rice	Uganda	47.7	47.7		
Rice	Rwanda	11.7	11.7		
Rice	Burundi	6.7	6.7		
Maize	Kenya	161.7	161.7		
Maize	Tanzania	133.7	133.7		
Maize	Uganda			1.7	1.7
Maize	Rwanda	31.3	31.3		
Maize	Burundi	65.7	65.7		
Beans	Kenya	9.0	9.0		
Beans	Tanzania			3.3	3.3
Beans	Uganda			9.0	9.0
Beans	Rwanda	3.0	3.0		
Beans	Burundi	9.3	9.3		

In order to resolve this issue, we turn to cross-entropy-based techniques which are used to derive unknown distributions from fairly limited data and includes one's own 'prior' beliefs on the underlying nature of the distribution, where possible. Cross-entropy methods have been used successfully in many types of statistical analyses in both the physical and social sciences and have even been used in IFPRI's own work, such as balancing the social accounting matrix (SAM) of a computable general equilibrium model (Robinson et al. 2000), or in calculating the distribution of irrigated and rainfed crops based on global data from a variety of (sometimes inconsistent) datasets (You and Wood 2004).

The cross-entropy method is built upon the same 'information-theoretic' principles that underlie the principle of 'maximum entropy' that was introduced by Shannon (1948a, b) to describe the degree to which a distribution differs from a uniform and un-informative profile—thereby capturing the 'surprise' that is embodied in a (random) outcome. In juxtaposition to Shannon's entropy measure $H = \sum_n -p_n \log(p_n)$ for n discrete, random events, we can also express the *cross-entropy* of a distribution by the measure $\mathrm{CE} = \sum_n -p_n \log\left(\frac{p_n}{\bar{p}_n}\right)$, which includes the *prior* distribution of weights (or probabilities) $\{\bar{p}_n\}$ that can be assigned for each random outcome. As shown by Kullback (1959), the maximization of the Shannon criterion with respect to the adding-up constraint $\sum_n p_n = 1$ is equivalent to the minimization of the cross-entropy criterion, similarly constrained, if the prior distribution is uniform (i.e., assigns an equal likelihood to each outcome). The divergence of a calculated distribution from prior beliefs as calculated by the cross-entropy criterion conveys information content in a similar way to that calculated by the Shannon measure of information (Kullback and Leibler 1951).

Implementation of Optimization Procedure

In applying this method to the problem of disaggregating areas and yields to the sub-national-level, based on the various sources of information that were available, we constructed an entropy-based estimation procedure which has the additional following features:

- It specifies the 'shares' of crop harvested area that belong to each suitability class, in each region—thus allowing them to act as 'weights' within the entropy framework.
- It specifies the shares of total cropland that belong to the suitability classes indicated in the outputs of the WaLETS GIS-analysis.
- It forces all shares to add up to one.
- It enforces all the adding-up relationships between total harvested crop and total cropland area that we would expect.

The key 'known' sources of data and 'unknown' variables that we have to solve for are summarized in Table 15.

Some of the 'known' parameters are outputs from the GIS modeling exercise and expert opinion (and, therefore, perhaps not 100% consistent with the underlying reality). But we treat these as inputs to the data processing program, which seeks to divide total cropland area into the area into:

Table 15 Key known and unknown variables in calibration model

Data and 'known' parameters	• Areas by crop and admin sub-region • Total cropland areas by admin sub-region • Total cropland areas by suitability class[a] • Shares of total cropland in admin sub-regions that are occupied by maize, beans, and rice • Total production by country and sub-region • Shares of maximum potential yield attainable over the suitability classes in each country[b]
'Unknown' parameters to be solved for	• Areas by crop, admin sub-region, disaggregated to suitability class • Total cropland area by suitability class disaggregated to admin sub-regions • The total cropland area (excluding maize, rice and beans) disaggregated to sub-region and suitability class • Yields by crop and sub-region, across all suitability classes

[a]At the time of writing this chapter, only the suitability classes for Tanzania and Kenya were available. Therefore, we assume that Rwanda and Burundi have the same area share of suitability classes in total cropland as Tanzania, and that Uganda has the same shares as those for Kenya. As further information is made available, these can be updated
[b]At the time of writing this chapter, there were no estimates of attainable yield shares for Rwanda—so the values for Burundi were assumed to apply for Rwanda

(1) the area occupied by the three key staples of interest and
(2) the area occupied by all other crops that are not of interest to the modeling exercise.

Therefore, we seek to enforce the following identities:

$$A_{r,s}^{\text{Totcropland}} = \underbrace{\sum_j A_{j,r,s}}_{\text{staples}} + A_{r,s}^{\text{Other}}$$

$$A_{r}^{\text{Totcropland}} = \underbrace{\sum_j A_{j,r}}_{\text{staples}} + A_{r}^{\text{Other}}$$

$$A_{s}^{\text{Totcropland}} = \sum_r A_{r,s}^{\text{Totcropland}}$$

where the quantities $A_s^{\text{Totcropland}}$, $A_r^{\text{Totcropland}}$, $A_{j,r}$ are known prior to the estimation, and the values of $A_{r,s}^{\text{Totcropland}}$, A_r^{Other}, $A_{r,s}^{\text{Other}}$, $A_{j,r,s}$ must be solved for.

In order to disaggregate the areas of staples and croplands across the appropriate suitability classes, we must find the shares alpha beta, such that the following relationships hold

$$
\begin{aligned}
&A_{j,r,s} = \alpha_{j,r,s} \cdot A_{j,r} && \alpha_{j,r,s} \in (0, 1] \\
&A_{r,s}^{\text{Totcropland}} = \beta_{r,s} \cdot A_{r}^{\text{Totcropland}} && \beta_{r,s} \in (0, 1] \\
&\sum_{s \in S_j} \alpha_{j,r,s} = 1
\end{aligned}
$$

where we carry out the summation of shares for staple crops ($\sum_{s \in S_j} \alpha_{j,r,s} = 1$) only over a defined sub-set of suitabilities that match with the crop type. In other words, we don't sum values for maize over bean suitabilties, and likewise for others. Since we know that the land areas defined by the maize, bean, and rice suitabilties overlap strongly with each other, we do not enforce the summation over the 'beta' values (i.e., $\sum_{s} \beta_{r,s} = 1$).

Since we must also enforce the adding up of total production to the national and sub-national totals provided by the multiplication of areas and yields, across sub-regions and suitability classes, we must consider these identities

$$
\begin{aligned}
&\sum_{r} Q_{j,r} = \sum_{s} \sum_{r} A_{j,r,s} \cdot Y_{j,r,s} \\
&Y_{j,r,s} = Y_{j}^{\text{MaxPotential}} \cdot \theta_{s}^{\text{attain}} \cdot \Delta_{j,r,s} \quad \theta_{s}^{\text{attain}} \in (0, 1], \Delta_{j,r,s} \in (0, +\infty)
\end{aligned}
$$

where we allow the sub-regional, suitability-specific yields to be informed by expert opinion, but ultimately allowed to deviate from it to the extent necessary to make the overall balance of area and production hold. The country-level agronomy experts from the WaLETS team described the maximum potential yield for each crop in each country $\left(Y_{j}^{\text{MaxPotential}}\right)$, and the share of that maximum potential that is attainable in each suitability class $\left(\theta_{s}^{\text{attain}}\right)$. We allow the yields we ultimately solve for to deviate from that amount by an allowable margin $\left(\Delta_{j,r,s}\right)$ that we try and keep as close to 1 as possible—so as to preserve as much of the expert information as possible. So, we place a 'penalty' upon any deviations in the value of $\Delta_{j,r,s}$ from one, using the cross-entropy principle in the objective function of the estimation program.

Now that we've described the key components of the program, we can now write out the entropy-based mathematical programming problem we are trying to solve as follows:

$$
\max_{\substack{\alpha_{j,r,s}, \beta_{r,s}, Y_{j,r,s}, \Delta_{j} \\ A_{j,r,s}, A_{r,s}^{\text{Totcropland}} \\ A_{r,s}^{\text{Other}}, A_{r}^{\text{Other}}}} \sum_{j} \sum_{r} \sum_{s} \left[-\alpha_{j,r,s} \cdot \log\left(\alpha_{j,r,s}\right)\right] + \sum_{j} \sum_{r} \sum_{s} \left[-\beta_{r,s} \cdot \log\left(\beta_{r,s}\right)\right]
$$

$$
+ \sum_{j} \left[\Delta_{j} \cdot \log\left({}^{\Delta_j}/_{1}\right)\right]
$$

s.t.

$$
\sum_{r} Q_{j,r} = \sum_{s} \sum_{r} A_{j,r,s} \cdot Y_{j,r,s}
$$

$$
\begin{aligned}
&Y_{j,r,s} = Y_j^{\text{MaxPotential}} \cdot \theta_s^{\text{attain}} \cdot \Delta_j \\
&A_{r,s}^{\text{Totcropland}} = \underbrace{\sum_j A_{j,r,s}}_{\text{staples}} + A_{r,s}^{\text{Other}} \\
&A_r^{\text{Totcropland}} = \underbrace{\sum_j A_{j,r}}_{\text{staples}} + A_r^{\text{Other}} \\
&A_s^{\text{Totcropland}} = \sum_r A_{r,s}^{\text{Totcropland}}
\end{aligned}
$$

$$
\begin{aligned}
&A_{j,r,s} = \alpha_{j,r,s} \cdot A_{j,r} \\
&A_{r,s}^{\text{Totcropland}} = \beta_{r,s} \cdot A_r^{\text{Totcropland}} \\
&\sum_{s \in S_j} \alpha_{j,r,s} = 1, \\
&\alpha_{j,r,s} \in (0, 1], \beta_{r,s} \in (0, 1], \theta_s^{\text{attain}} \in (0, 1], \Delta_j \in (0, +\infty)
\end{aligned}
$$

where the objective function is a hybrid of the maximum entropy problem (with respect to the $\alpha_{j,r,s}$, $\beta_{r,s}$ variables and a cross-entropy problem (with respect to the Δ_j variable).

This program is run for the existing data in order to obtain a distribution of areas and yields over each country's sub-regions and suitability classes that can be used as base data for the economic market model. Given that production was constrained, within the program, to match the national-level production of the three staples, the model will replicate the base-year supply, demand, and trade when it is simulated.

As better information becomes available—either from the GIS-based analysis or the yield potential assessments of the agronomy experts—this data can be put into the calibration model so that it can be rerun and generate a new base data set of disaggregated areas and yields.

Table 16 shows the results from the model—aggregated across administrative sub-regions for each country—to display in a more convenient fashion.

Table 16 Calibration of base area and yield levels across suitability classes for each EAC country

		Area ('000 ha)			Yield (mt/ha)		
		Rice	Maize	Beans	Rice	Maize	Beans
Kenya	Maize/high suitability		85.2			2.1	
Kenya	Maize/moderate suitability		1520.1			1.8	
Kenya	Maize/low suitability		45.9			1.3	
Kenya	Rice/high suitability	3.7			3.1		
Kenya	Rice/moderate suitability	4.6			2.7		
Kenya	Rice/low suitability	4.7			2.1		
Kenya	Rice/unsuitable	4.5			1.3		
Kenya	Beans/high suitability			279.7			0.5
Kenya	Beans/moderate suitability			377.3			0.4
Kenya	Beans/low suitability			248.5			0.3
Tanzania	Maize/moderate suitability		1477.1			1.5	
Tanzania	Maize/low suitability		1373.1			1.1	
Tanzania	Rice/high suitability	13.1			1.7		
Tanzania	Rice/moderate suitability	201.0			1.7		
Tanzania	Rice/low suitability	203.4			1.3		
Tanzania	Rice/unsuitable	178.2			0.8		
Tanzania	Beans/high suitability			38.2			1.1
Tanzania	Beans/moderate suitability			392.3			0.8
Tanzania	Beans/low suitability			395.1			0.7
Tanzania	Beans/unsuitable			1.1			0.4
Uganda	Maize/high suitability		177.5			2.3	
Uganda	Maize/moderate suitability		434.9			1.5	
Uganda	Maize/low suitability		163.6			0.7	
Uganda	Rice/moderate suitability	33.9			1.5		
Uganda	Rice/low suitability	34.1			0.9		
Uganda	Rice/unsuitable	33.7			0.5		
Uganda	Beans/high suitability			238.3			0.6
Uganda	Beans/moderate suitability			330.4			0.56
Uganda	Beans/low suitability			255.7			0.45

(continued)

Table 16 (continued)

		Area ('000 ha)			Yield (mt/ha)		
		Rice	Maize	Beans	Rice	Maize	Beans
Rwanda	Maize/moderate suitability		54.2			1.1	
Rwanda	Maize/low suitability		53.4			0.6	
Rwanda	Rice/moderate suitability	5.1			6.3		
Rwanda	Rice/low suitability	1.7			2.3		
Rwanda	Rice/unsuitable	1.4			1.4		
Rwanda	Beans/high suitability			32.8			1.2
Rwanda	Beans/moderate suitability			135.1			0.9
Rwanda	Beans/low suitability			142.7			0.5
Burundi	Maize/moderate suitability		56.8			1.4	
Burundi	Maize/low suitability		56.6			0.8	
Burundi	Rice/moderate suitability	8.7			4.1		
Burundi	Rice/low suitability	4.0			1.5		
Burundi	Rice/unsuitable	3.4			0.9		
Burundi	Beans/moderate suitability			119.8			1.2
Burundi	Beans/low suitability			125.2			0.5

References

Anyango, G. (1997). *Comparative transportation cost analysis in East Africa*. Technical Paper no 22, SD Publications Series, Office of Sustainable Development, Bureau for Africa, United States Agency for International Development (USAID), Washington D.C.

Barkaoui, A., & Boutault, J.-P. (2000). Cereal and oil seeds supply with EU under agenda 2000: A positive mathematical programming application. *Agricultural Economics Review, 1*(2), 7–17.

Doole, G., & Marsh, D. (2014). Use of positive mathematical programming invalidates the application of the NZFARM model. *Australian Journal of Agricultural and Resource Economics, 58*(2), 291–294.

East African Community (EAC). (2011). *EAC food security action plan (2011–2015)*. East African Community Secretariat, Arusha, Tanzania. Available at http://www.agriculture.eac.int/.

Foster, V., & Briceño-Garmendia, C. (2010). *Africa's infrastructure: A time for transformation*. Washington, D.C.: The World Bank.

Gohin, A., & Chantreuil, F. (1999). La Programmation Mathématique Positive dans les Modèles d'Exploitation Agricole. Principes et Importance du Calibrage. *Cahiers d'Économie et de Sociologie Rurales, 52,* 59–79.

Heckelei, T., & Britz, W. (2000). Positive mathematical programming with multiple data points: A cross-sectional estimation procedure. *Cahiers d'Economie et Sociologie Rurales, 57,* 28–50.

Heckelei, T., & Wolff, H. (2003). Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy. *European Review of Agricultural Economics, 30*(1), 27–50.

Heckelei, T., & Britz, W. (2005). Models based on positive mathematical programming: State of the art and further extensions. In F. Arfini (Ed.), *Modelling agricultural policies: State of the art and new challenges: Proceedings of the 89th European Seminar of the European association of agricultural economists* (pp. 48–73). Parma: Monte Università.

Heckelei, T., Britz, W., & Zhang, Y. (2012). Positive mathematical programming approaches—recent developments in literature and applied modelling. *Bio-based and Applied Economics, 1*(1), 109–124.

Henry de Frahan, B., Buysse, J., Polomé, P., Fernagut, B., Harmignie, O., Lauwers, L., et al. (2007). Positive mathematical programming for agricultural and environmental policy analysis: Review and practice. In A. Weintraub, C. Romero, T. Bjørndal, & R. Epstein (Eds.), *Handbook of operations research in natural resources* (pp. 129–154). New York: Springer.

Howitt, R. E. (1995a). Positive mathematical programming. *American Journal of Agricultural Economics, 77*(2), 329–342.

Howitt, R. E. (1995b). A calibration method for agricultural economic production models. *Journal of Agricultural Economics, 46*(2), 147–159.

Howitt, R. E., Medellín-Azuara, J., MacEwan, D., & Lund, J. R. (2012). Calibrating disaggregate economic models of agricultural production and water management. *Environmental Modelling and Software, 38,* 244–258.

Judez, L., Chaya, C., Martınez, S., & González, A. (2001). Effects of the measures envisaged in "Agenda 2000" on arable crop producers and beef and veal producers: An application of Positive Mathematical Programming to representative farms of a Spanish region. *Agricultural Systems, 67*(2), 121–138.

Kaaya, A. K., Msanya, B. M., & Mrema, J. P. (1994). Soils and land evaluation of part of the Sokoine University of Agriculture farm (Tanzania) for some crops under rain-fed conditions. *African Study Monographs, 15*(2), 97–117.

Kanellopoulos, A., Berentsen, P., Heckelei, T., Van Ittersum, M., & Oude Lansink, A. (2010). Assessing the forecasting performance of a generic bio-economic farm model calibrated with two different PMP variants. *Journal of Agricultural Economics, 61,* 274–294.

Kullback, S. (1959). *Information theory and statistics*. New York: John Wiley.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22,* 79–86.

Mérel, P., & Howitt, R. (2014). Theory and application of positive mathematical programming in agriculture and the environment. *Annual Review of Resource Economics, 6*(1), 451–470.

Mérel, P. R., & Bucaram, S. (2010). Exact calibration of programming models of agricultural supply against exogenous sets of supply elasticities. *European Review of Agricultural Economics, 37*(3), 395–418.

Mérel, P. R., Simon, L. K., & Yi, F. (2011). A fully calibrated generalized constant-elasticity-of-substitution programming model of agricultural supply. *American Journal of Agricultural Economics, 93,* 936–948.

Mérel, P. R., Yi, F., Lee, J., & Six, J. (2014). A regional bio-economic model of nitrogen use in cropping. *American Journal of Agricultural Economics, 96*(1), 67–91.

Minot, N. (2009). *Using GAMS for agricultural policy analysis. Technical Guide*. Washington D.C.: International Food Policy Research Institute. Aavailable at: http://www.ifpri.org/publication/using-gams-agricultural-policyanalysis.

Omamo, S. W., Diao, X., Wood, S., Chamberlin, J., You, L., Benin, S., et al. (2006). *Strategic priorities for agricultural development in Eastern and Central Africa. Research Report 150*. Washington D.C.: International Food Policy Research Institute. https://doi.org/10.2499/9780896291584rr150.

Paris, Q. (2010). *Economic foundations of symmetric programming*. Cambridge: University Press.

Robinson, S., Cattaneo, A., & El-Said, M. (2000). *Updating and estimating a social accounting matrix using cross entropy methods*. Trade and MacroEconomics Division Discussion Paper No. 58. Washington, D.C.: IFPRI.

Röhm, O., & Dabbert, S. (2003). Integrating agri-environmental programs into regional production models: An extension of positive mathematical programming. *American Journal of Agricultural Economics, 85*(1), 254–265. https://doi.org/10.1111/1467-8276.00117.

Rosegrant, M. W., Paisner, M. S., Meijer, S., & Witcover, J. (2001). *Global food projections to 2020: Emerging trends and alternative futures*. Washington, D.C.: International Food Policy Research Institute.

Rosegrant, M. W., Cai, X., & Cline, S. A. (2002). *Water and food to 2025: Dealing with scarcity*. Washington, D.C.: International Food Policy Research Institute.

Rosegrant, M. W., et al. (2012). *International model for policy analysis of agricultural commodities and trade (IMPACT): Model description*. Washington, D.C.: International Food Policy Research Institute.

Sadoulet, E., & de Janvry, A. (1994). *Quantitative development policy analysis*. Baltimore, MD: Johns Hopkins University Press.

Salami, A., Kamara, A. B., & Brixiova, Z. (2010). *Smallholder agriculture in East Africa: Trends, constraints and opportunities*. Working paper no. 105, African Development Bank Group, Tunisia. Available at https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/WORKING%20105%20%20PDF%20d.pdf.

Shannon, C. E. (1948a). A mathematical theory of communication. *The Bell System Technical Journal, 27*(July), 379–423.

Shannon, C. E. (1948b). A mathematical theory of communication. *The Bell System Technical Journal, 27*(October), 623–656.

Teravaninthorn, S., & Raballand, G. (2009). Transport prices and costs in Africa: A review of the main international corridors. Washington D.C.: The World Bank.

USAID. (2017). *East Africa Regional profile: Feed the future country profile*. United States Agency for International Development (USAID). Available at https://www.feedthefuture.gov/country/east-africa-regional-0.

Waithaka, M., Nelson, G., Thomas, T. S., & Kyotalimye, M. (2013). *East African agriculture and climate change: A comprehensive analysis*. IFPRI monograph, International Food Policy Research Institute. Available at http://www.ifpri.org/publication/east-african-agriculture-and-climate-change-comprehensive-analysis.

Were, K., Musana, B., Mudiope, J., Tanui, L., & Muhutu, J.-C. (2016). *A GIS-based analysis of agro-ecosystems suitability for production of staple crops in the East African Community region*. A technical report for the project on Water, Land, Ecosystems and Trade in Staples (WaLETS). Available at https://www.kilimotrust.org/documents/reports/2017/walets/WaLETS_Final_Reports/WaLETS_GIS_TechnicalReport.pdf.

You, L., & Wood, S. (2004). *Assessing the spatial distribution of crop production using a cross-entropy method*. Environment and Production Technology Division Discussion Paper No. 126. Washington, D.C.: IFPRI.