Arie Arnon
Maria Cristina Marcuzzo
Annalisa Rosselli *Editors*

Financial Markets in Perspective

Lessons from Economic History and
History of Economic Thought

Springer Studies in the History of Economic Thought

This series offers an outlet for research in the history of economic thought. It features scholarly studies on important theoretical developments and great economic thinkers that have contributed to the evolution of the economic discipline.

Springer Studies in the History of Economic Thought (SSHET) welcomes proposals for research monographs, edited volumes and handbooks from a variety of disciplines that seek to study the history of economic thinking and help to arrive at a better understanding of modern economics. Relevant topics include, but are not limited to, various schools of thought, important pioneers and thinkers, ancient and medieval economic thought, mercantilism, cameralism and physiocracy, classical and neoclassical economics, historical, institutional and evolutionary economics, socialism and Marxism, Keynesian, Sraffian and Austrian economics, econometrics and mathematical studies as well as economic methodology and the link between economic history and history of economic thought.

All titles in this series are peer-reviewed.

For further information on the series and to submit a proposal for consideration, please contact Johannes Glaeser (Senior Editor Economics) Johannes. glaeser@springer.com.

Arie Arnon • Maria Cristina Marcuzzo •
Annalisa Rosselli

Editors

Financial Markets
in Perspective

Lessons from Economic History
and History of Economic Thought

 Springer

Editors
Arie Arnon (ID)
Department of Economics
Ben-Gurion University of the Negev
Beersheba, Israel

Maria Cristina Marcuzzo (ID)
Department of Statistical Sciences
Sapienza University of Rome
Rome, Italy

Annalisa Rosselli (ID)
Department of Economics and Finance
Tor Vergata University of Rome
Rome, Italy

Foreword

The Thomas Guggenheim Program in the History of Economic Thought has been created in 2007 in the Department Economics at Ben-Gurion University of the Negev thanks to the generous support of Professor Thomas Guggenheim from Geneva, Switzerland.

The **Program** promotes scholarship in the **History of Economic Thought** and **Economic History**, and it organizes international conferences and funds seminars and fellowships through the **Thomas Guggenheim Fellowships in Economics**.

At each conference, the Program chooses a recipient of the **Thomas Guggenheim Prize for Outstanding and Original Research in the History of Economic Thought**. The prize is awarded to a distinguished scholar for his or her life's work. The prize is awarded in a public lecture delivered by the winner at the program's conference.

The first conference in 2009 in Beersheba, Israel, was on **"Perspectives on Keynesian Economics."**

The second conference in 2011 in Beersheba, Israel, was on "**David Hume and the Scottish Enlightenment: Economic and Philosophical Studies"—Celebrating 300 Years of Hume's Birth**.

The third conference in 2015, in Geneva, Switzerland, was organized together with the Graduate Institute (University of Geneva) on **"Economic Research and Policy Making at the Federal Reserve in Historical and International Perspective."**

The fourth conference in 2017 in Beersheba, Israel, was on **"Expectations: Theory and Applications in Historical Perspectives."**

The fifth conference in 2019, in Rome, Italy, was organized together with the Accademia Nazionale dei Lincei on **"Financial Instability, Market Disruptions and Macroeconomics: Lessons from Economic History and the History of Economic Thought."**

As may be seen from the above list of subjects, an attempt is made to choose topics that are related either to actual events or to the modern economic literature. This is clearly the case with the first conference on Keynesian Economics. It was not a coincidence that it took place shortly after the beginning of the 2008 global

financial crisis, when Keynesian economics turned to be more relevant and issues of fiscal and monetary policies became of crucial importance.

The topic of the 2011 conference on David Hume was selected to celebrate his 300th birthday (1711–1776). It turned out to be not only an intellectual event on Hume's philosophy, but a discussion of Hume's genuine ideas on international trade balance and money supply in times of continuous disruption.

The third conference on Central Banking in 2015 was designated to celebrate the centennial of the Federal Reserve (1913) and the sixty-year anniversary of the Bank of Israel (1954).

The fourth conference in 2017 on Expectations in Historical Perspective covered the literature on this extremely important element of economic theory from times of old history to the most recent literature in macroeconomics and its micro-foundations.

The fifth conference in 2019 on Financial Instability, Market Disruptions and Macroeconomics: Lessons from Economic History and the History of Economic Thought dealt with the topical issue of the financial fragility of present and past capitalist economies and the impact on the real sector.

This volume contains the public lecture delivered by the winner of the Guggenheim Prize at the program's fifth conference, Prof. Alessandro Roncaglia, and thirteen papers presented in the conference.

Recipients of the Thomas Guggenheim Prize were:

2009   Prof. Bertram Schefold (Goethe-Universität, Frankfurt, Germany)
2011   Prof. Sam Hollander (Toronto University, Canada, and Ben-Gurion University, Israel)
2015   Prof. David Laidler (University of Western Ontario, Canada)
2017   Prof. Duncan K. Foley (New School for Social Research, New York, USA)
2019   Prof. Alessandro Roncaglia (Sapienza University of Rome, Italy)

The **Thomas Guggenheim Program in the History of Economic Thought** operates under an International Advisory Committee, comprising:

Prof. Thomas Guggenheim (Emeritus, University of Geneva, Switzerland)
Prof. Arie Arnon (Emeritus, Ben-Gurion University of the Negev, Israel)
Prof. Maria Cristina Marcuzzo (Sapienza University of Rome, Italy)
Prof. Joel Mokyr (Northwestern University, Illinois, USA)
Prof. Jacques Silber (Bar-Ilan University, Israel)
Prof. Warren Young (Bar-Ilan University, Israel)
Prof. Jimmy Weinblatt (Emeritus, Ben-Gurion University of the Negev, Israel)
Prof. Amos Witztum (London School of Economics, UK)

This committee acts as a steering and prize committee and thus is also responsible for the selection of the winner of the Guggenheim Prize.

Ben-Gurion University of the Negev,   Jimmy Weinblatt
Beersheba, Israel

Introduction

Alessandro Roncaglia's chapter, which opens this volume, rightly takes pride of place since it records his lecture as the 2019 Guggenheim Prize winner. It was the highlight of the Rome conference, the latest in a series that the Guggenheim Foundation has been holding since 2009. However, it also makes an appropriate opening since it effectively sums up the issue that runs through all the essays in the volume, namely the connection between money and finance in relation to the real economy, as economic theory has interpreted it and as it has developed historically with the evolution of capitalism and its institutions.

This book explores both sides of the relationship between the financial and the real sector. It considers how money, or rather lack of it as shortage of liquidity, can generate collapse of the real economy, and how a fall in the prices of financial assets can communicate to the whole system. But it also points out how the difficulties of the real economy, especially changes in income distribution, have an impact on the structure of financial markets, for example by inducing over-indebtedness and creating a superstructure of highly unstable debt and credit relationships.

The volume covers 150 years of history of ideas and events, starting from the second half of the nineteenth century, when large and complex financial markets developed in the Western world and investment techniques became a specific object of study. The evolution of financial markets is in fact the background to many essays. Some deal with it directly up to its most recent developments, underlining the changes in the way of thinking of economic agents and economists that the constant flow of innovations has brought about. Others bring it in as background to the reflections of some of the authors considered: first of all Keynes, as the one who placed the characteristics of a monetary economy at the core of his analysis, but also Kalecki, Marx, and Minsky.

What all the essays in this volume have in common, as the title suggests, is of course the historical perspective, whose fruitfulness is once again demonstrated. It can bring revealing light to bear on the present debate by tracing its historical roots, and on the failure of past theories to keep pace with the times and, if justifiable in another historical context, proving inadequate in interpreting more recent economic

phenomena. From this point of view, economic theories should be counted among the victims of crises not because they were unable to foresee them but for not even being able to envisage their possibility. However, the 2008 crisis seems to have prompted adjustment, and not the emergence of a different paradigm.

In this Introduction, we present the themes running through the different chapters, which we have organized into Parts to highlight the perspective from which the topics are addressed.

Alessandro Roncaglia focuses on the rise and fall of the twin ideas—myths of money as a veil and the invisible hand of the market. He stresses that not all classical economists shared Smith's and Ricardo's "received view," including Hume's international trade mechanism, taking Antonio Serra and William Petty as examples. At the core of his account is the methodology of the classical economists, arguing that their notion of money as a veil is compatible with the idea of an influence exerted by money on the real economy. Roncaglia then explores the connection between the notion of money as a veil and the myth of the invisible hand in neoclassical/marginalist economics. The two pillars of a reconstructed modern monetary theory are considered, identified by Roncaglia as the Keynesian notions of uncertainty and liquidity preference and Minsky's money-manager economy, or in other words a situation in which, to use Keynes's metaphor, the tail of finance wags the dog of the real economy. The new paradigm may be summarized in three points: money and finance cannot be kept separate and crucially affect the real economy, the Keynesian notions of uncertainty and liquidity preference constitute the foundations on which theorizing over money and finance should build, and the role of finance increases over time, with a change of regime to a "money-manager capitalism."

The Guggenheim Prize Lecture is followed by the Part "Financial History," which contains four papers. The first, by Janette Rutterford, Dimitris Sotiropoulos, and Antonis Kyparissis, deploys an impressive dataset to show how investment trusts emerged and operated in their early days. The authors take a look at the way the now well-known institutional investment approaches to adding value evolved in the late nineteenth century and first half of the twentieth century. The chapter explores how ways to optimize returns while minimizing risk for long-term investors were developed, promoted through financial advice manuals and texts, and put into practice by investment trusts from the 1860s onward. It documents how the concepts of diversification and yield enhancement dominated investment discourse and how investment trusts offered individual investors a low-cost means to maximize return relative to risk by implementing a range of investment strategies—in particular, active asset allocation, portfolio diversification, stock selection, market timing, and leverage—which form the bedrock of today's investment strategies.

Paolo Paesani and Annalisa Rosselli take us on a journey through the ideas advanced in defense and praise of speculation in the history of economic thought. The early scientific literature which developed with the emergence of organized speculation on financial and commodity markets in the second half of the nineteenth century was built on the identification of several categories of traders, from professional and amateur speculators to rentiers and members of the Haute finance. Based on analysis of the interaction between these agents, this literature recognized the

advantages of professional speculation (price discovery, risk hedging) as well as the problems posed by the widespread presence of amateurs. Reflections on these themes were closely connected with debate on the advantages and disadvantages of instruments which made it easier for inexperienced traders to access the trading floor. The chapter reconstructs these debates, which took place in Europe and the USA, involving Proudhon, Weber, Ehrenberg, and Marshall, among others. It shows how opinions hovered between two poles. On the one hand, expert observers identified enhanced liquidity and diversity of view as the main advantages deriving from widespread market participation. And at the same time, the stock exchange appeared as a means to empower the middle classes, offering opportunities for economic progress to all. On the other hand, as the number of traders increased, so did the share of amateurs and the incentive for professional speculators to reap profits by "fleecing" them, with destabilizing effects on the market. Pursuit of the right balance between these two poles raised interesting regulatory problems, not devoid of ethical considerations. The chapter goes back to the origins of current debate on the possibility and merits of the democratization of finance and their social implications—a problem that economists and regulatory authorities have been grappling with to the present day.

There is no question that quite possibly the most renowned investor among the economists enjoying an academic reputation is Keynes, often credited with a stellar ability, second perhaps only to that of Ricardo. The chapter by Maria Cristina Marcuzzo and Eleonora Sanfilippo focuses on one part of his portfolio investment, namely commodities. While Keynes is well known for his theory of normal backwardation in futures markets, he is rather less known as a trader in these markets. The question arises whether his theory—based on normal backwardation—acted as a guide to his trading activity, or rather, it was his trading activity that shaped his theory. Backwardation was not viewed by Keynes as a permanent feature of futures market, but rather a situation which comes about only if prices do not increase by more than the risk premium and only when supply and demand are balanced. For the speculator, it will not suffice to pursue a constant strategy of being systematically long, as would be the case if backwardation were the norm. Keynes's activity confirms this, since he engaged actively in collecting market information and assessing supply and demand conditions, thus acting as an informed trader. But knowledge and information are by definition limited, which may explain why his performance in commodity futures was not as stellar as is commonly believed. The authors provide evidence of Keynes's behavior, taking the example of his dealings in wheat, cotton, and tin, the commodities most traded by him.

Part II closes with a chapter by Donald MacKenzie, who explores the latest developments in the way trade is conducted and orders are executed in the financial markets. Over the last three decades, most face-to-face dealings have shifted to electronic trading. Since around 2000, many markets have made the further transition to automated trading, including ultrafast, algorithmic high-frequency trading, or HFT. This chapter applies a perspective that the author calls "material political economy" to HFT, focusing on (a) the "signals" (patterns of data) that inform how

HFT algorithms trade; (b) algorithms that usually "make" or "provide liquidity" (i.e., add orders to the exchanges' electronic order books which other algorithms or human beings can execute against); (c) algorithms that "take liquidity" (execute against orders already in order books); and (d) how HFT algorithms interact, and the material interventions of exchanges in that interaction. The chapter examines how the material features of HFT, its "signals," and the technical systems that make HFT possible interweave in complex ways with power, politics, and finance's everyday money-making.

Part III draws on the lessons which can be learnt from such great economists as Marx, Hayek, Keynes, and Kalecki. The chapter by Arie Arnon contrasts the studies of system's fragility by Marx and Hayek who, unlike Marx, was a dedicated supporter of capitalism. It argues that Marx's analysis of the capitalist system in *Capital, Vol. I*, shows that the fragility is rooted in "real" causes. Marx also analyzed significant contributing causes entrenched in the banking and financial spheres in *Capital, Vol. III*, and other writings, which were published posthumously from drafts he left. The chapter utilizes drafts only recently made available to shed light on Marx's evolving ideas about the "real" versus "less real" causes of fragility in the system. On the contrary, Hayek argued that the fragility of the capitalist system was rooted in "monetary" causes, i.e., in the "less real" rather than "real" causes. The chapter reviews the two scholars' divergent explanations of fragility, emphasizing a focal point concerning possible remedies. Marx, as we know, argued that there was no such remedy. Hayek, blaming the banking sector (and finance), argued in his 1930s theory of business cycles that more passive banks and central banks could provide an answer to the fragility in the system. However, it was not until the mid-1970s that Hayek supported a radical reform which would leave the monetary system to be ruled by "free banking."

The next two chapters take us to consider the contributions of two economists belonging to a completely different tradition, who made pathbreaking analyses of the link between the "real" and the "monetary" sides of capitalist economies.

Joerg Bibow takes another look at Keynes's monetary thought with a focus on financial instability and crises. Keynes's early monetary work, *Indian Currency and Finance* (1913), shows his keen concern over financial instabilities and his appreciation of the central banks' lender-of-last-resort role. In his attack on the Versailles Treaty in *The Economic Consequences of the Peace* (1919), Keynes analyzes the challenges posed by debt overhangs in the aftermath of World War I and the deflationary strategy for their resolution which lies at the core of the infamous "peace" treaty. The unfolding of events in the 1920s and 1930s then shaped Keynes's three major monetary works. *A Tract on Monetary Reform* (1923) concerns the price level instability experienced in Britain and elsewhere in the aftermath of WWI while *A Treatise on Money* (1930) is set against the backdrop of Britain's stagnation in the second half of the 1920s following the country's return to gold at sterling's pre-war parity. A close look at Keynes's assessments and writings during the Great Depression—the worldwide deflationary environment that is the background to *The General Theory* (1936)—reveals Keynes's deep understanding of the

havoc wreaked by deflation in banking systems—an issue subsequently pushed aside by assumption in that work.

The tendency to run into private debts and government deficits is a cyclical feature of capitalist economies and the issue of debt management is a central one. Jan Toporowski shows how Kalecki explained debt resolution by means of a fable about monetary circulation between mutually indebted individuals. This chapter generalizes the processes of debt payment in the fable to show how interest and debt payments depend on the financial circulation of money, rather than profits or income derived from the non-financial economy. Financial circulation makes debt payments dependent upon the liquidity in the financial system and marks a break with the classical theory of interest in which interest is tied to the rate of profit. Bank lending, securitization, and central bank buying of financial assets all have their part to play in maintaining debt payments to avoid default and financial crisis.

A related question in any debt-prone economy is how the burden is shared among debtors and creditors, and this leads to the issue of how income and wealth are distributed. Part IV has three chapters dealing with income distribution and the social roots of economic crisis. We start with the contribution by Amos Witztum, who points out that the recent financial crisis was preceded by a prolonged and severe increase in inequality. This brings to the fore the possible relationships between the distribution of income and instability. As financial crises are typified by excessive risk-taking behavior, the author argues that for a crisis to ensue this excessive risk-taking should be engaged in by those who are more likely to default: the poor. The obvious remedy to this would simply lie in redistribution of income. However, for modern economics, there are two difficulties in both aspects of the argument. Firstly, excessive risk-taking, which is a departure from rationality, is not something the theory takes into account, or associates with income distribution. Secondly, the system of competitive decentralization has lost its freedom with regard to income distribution since the problems of incompleteness in a world of uncertainty limit ownership structures and hence income distributions consistent with efficiency. Therefore, correcting income distribution for the sake of stability may clash with the inequality required for efficiency.

This chapter is followed by the analysis by Orsola Costantini. She surveys US household-sector micro data since 1989 to explore the relation between income inequality and financial fragility from a macro perspective. In contrast with a portfolio (and wealth inequality) approach to household financial instability, which focuses on the investments of the upper-middle class, the author argues that the transformation in the financial conditions of the bottom 50% of the equivalent income distribution has been crucial for macroeconomic stability. In the past three decades, households in that group saw their debt-to-income ratio dramatically increase on average, together with financial strain and debt service payments. Being particularly exposed to the compression of real available incomes, their unprecedented access to credit contributed to producing a smooth flow of revenues to the corporate and rentier sectors, with systemic consequences. In this chapter, these consequences are analyzed using insights from Kalecki and Luxemburg to highlight how alternate and specular movements in household net borrowing and

public net spending subsidized the US economy, but also produced decades of low growth and stagnant productivity, punctuated by financial excesses. The chapter stresses the historical and policy-driven nature of the phenomenon, to conclude that a policy attempting to reduce fragility must tackle the structure of aggregate demand and the sources of income inequality.

The concluding chapter in this part, by Anwar Shaikh, takes us back once again to the classical approach of Keynes and Kalecki to understand the interlinks between profitability, stimulus policy, and finance. In both schools, the interest rate plays a familiar role: other things being equal, its increase will have a negative impact on the net rate of profit and hence on the rate of growth. However, in the classical tradition the expected rate of profit is linked to the actual rate of profit in the manner of Soros's theory of reflexivity, whereas in Keynes's *General Theory* the expected rate of profit is left "hanging in the air." In addition, in classical theory, the interest rate is tied to the profit rate, whereas in Keynesian theory it is tied to liquidity preference. Finally, we know that Keynes grounded his macroeconomic argument on the notion of "atomistic competition" and that even Kalecki's initial formulation of his theory of effective demand assumed "free competition." The classical analogue is the theory of "real competition," which—it is argued—is the appropriate foundation for Keynesian macroeconomics. In real competition, firms set prices, are demand conscious, and compete for market shares while subject to intra-industrial price competition and inter-industrial investment flows motivated by profit rate differences. A deficit-financed stimulus can have an immediate positive impact on output, employment, and growth, and real wages are likely to rise as the labor market tightens. If the rise in real wages exceeds the growth in productivity, i.e., if the wage share rises, the profit rate falls and the stimulus can give way to a slowdown in growth. Repeated stimuli can then lead to stagnation with inflation (stagflation). And with inflation will come rising interest rates, further reducing the net rate of profit and slowing down growth. In this sense, "wage-led" growth can *induce* a "profit-led" decline.

The final part of this volume addresses issues dealt with in current macroeconomic models, raising the question as to whether they are equipped to deal with financial matters. Muriel Dal Pont Legrand investigates how the macro agent-based literature, which has been showing lively development since 2000, analyzes the issue of financial instability. Within this new paradigm, attention is focused on two research communities engaged in investigating this question: the Keynes and Schumpeter (K&S) research program at the Sant'Anna School in Pisa and the Computational Adaptive System (CATs) research program being developed at the Università Cattolica del Sacro Cuore in Milan. The chapter examines their common analytical foundations and traces out the links to earlier research programs, like those prompted by Minsky, Leijonhufvud, and, more recently, Stiglitz. The chapter identifies differences in their respective modeling strategies and examines how they pursued different objectives.

Along similar lines, Hans-Michael Trautwein shows that the new Keynesian economics that relies on dynamic stochastic general equilibrium (DSGE) technology has been criticized as being hopelessly inadequate for dealing with financial crises of

the type that we have seen in recent years, or—more generally—with failures of the intertemporal price mechanism to coordinate investment and saving. However, the "financial frictions" literature has made considerable progress in using DSGE frameworks to model (shadow) banking, credit cycles, and financial crises. This chapter provides an assessment of what has been achieved with the reformulations of the DSGE approach and what inadequacies remain in the light of the approaches of Keynes, Fisher, and Minsky, among others.

The closing chapter opens up a somewhat more optimistic perspective with an approach that appears to be more fruitful. Roberto Scazzieri examines the link between the instability features of a profit-driven economy and the structural changes inherent to that type of economy. Moving from consideration of the profit motive as search and exploitation of price (or cost) differentials, the chapter considers the structural dynamic of this type of economy as a process driven by the generation of differentials that attract investment in real *or* financial activities. The chapter calls attention to the dual character of that dynamic depending on whether it takes place through the "horizontal" reshuffling of activities in the real economy or the "vertical" shifting of liquid funds between real and financial activities. Policymaking constraints and opportunities are generated by these dynamics. Taking a comprehensive view of both the horizontal and vertical changes in the economic structure, the chapter highlights the dual role of finance, which may alternatively trigger "patient" long-term investment in the real sphere, or generate instability under conditions of increased financialization and market volatility. This open-endedness of structural dynamics has important implications for economic policy, which can only overcome the trade-off between incentive to invest and economic expansion by triggering structural changes along a path compatible with investment of liquid funds in the real economy.

In conclusion, this volume presents a variety of approaches and a wide-ranging scope of investigation, but all the chapters share a common concern, giving centrality to the relationship between real and monetary phenomena. Moreover, they all account for financial instabilities within the context of the systemic features of capitalist economies: unequal income distribution, insufficient aggregate demand, market imperfections, and unregulated financial sectors. Case studies of actual financial markets, both past and present, afford further insight into the actual forces at work.

Past theories and facts are revisited to shed light on the present, calling for a better understanding of the real and monetary factors equally operating in the economic sphere. In this regard, the paradigm grounded on the classical political economy extending to Keynes–Kalecki–Minsky proves better equipped for the task.

<div align="right">

Arie Arnon
Maria Cristina Marcuzzo
Annalisa Rosselli

</div>

Contents

Part IV   Income Distribution and the Social Roots of Economic Crisis

Part V   Today's Macroeconomics Confronts Economic Crises

Contributors

Arie Arnon Department of Economics, Ben-Gurion University of the Negev, Beersheba, Israel

Joerg Bibow Skidmore College, New York, USA

Orsola Costantini United Nations Conference on Trade and Development, Geneva, Switzerland

Muriel Dal Pont Legrand Université Côte d'Azur, Nice, France

Antonis Kyparissis The Open University Business School, Milton Keynes, UK

Donald MacKenzie University of Edinburgh, Edinburgh, Scotland

Maria Cristina Marcuzzo Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

Paolo Paesani Tor Vergata University of Rome, Rome, Italy

Alessandro Roncaglia Sapienza University of Rome and Accademia Nazionale dei Lincei, Rome, Italy

Annalisa Rosselli Department of Economics and Finance, Tor Vergata University of Rome, Rome, Italy

Janette Rutterford The Open University Business School, Milton Keynes, UK

Eleonora Sanfilippo Department of Economics and Law, University of Cassino and Southern Lazio, Cassino, Italy

Roberto Scazzieri Accademia Nazionale dei Lincei and University of Bologna, Bologna, Italy

Anwar Shaikh New School for Social Research, New York, USA

Dimitris P. Sotiropoulos The Open University Business School, Milton Keynes, UK

Jan Toporowski SOAS University of London, London, UK

Hans-Michael Trautwein Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany

Amos Witztum Centre for the Philosophy of Natural and Social Sciences (CPNSS), London School of Economics, London, UK

Part I
Guggenheim Prize Lecture

The Myth of Money as a Veil

Alessandro Roncaglia

1 Introduction

This chapter focuses on the rise and fall (alas, not complete) of the twin ideas-myths of money as a veil and the invisible hand of the market.

Money might be considered a veil only if defined in a very restrictive way and only under very restrictive assumptions, largely similar to those required for the thesis of the invisible hand of the market. As a matter of fact, its links with finance are decisive, to the extent that the distinction between money and finance is practically impossible to draw, and finance is certainly not a veil.

Over the centuries, monetary and financial institutions have undergone fundamental changes in nature, and their complexity has increased. The theoretical debate, too, has seen important developments. However, the use of increasingly complex analytical tools is accompanied by a drastically simplified conceptual framework, underlying what is now considered the mainstream, with simplifying assumptions that deny certain basic characteristics of money and finance.[1]

In dealing with this subject matter—as with so many others—recourse to the history of economic thought (HET) is useful: first, by pointing out the different views of the world and thus the different conceptual categories that underlie theoretical differences and, second, by clarifying possibly subtle but certainly important specific characteristics of concepts designated by the same name but acquiring different meanings in different analytical contexts.

In HET, we can find theories maintaining that money is a veil but finance is not, that both money and finance are a veil, and that neither money nor finance is a veil—

---

[1] For a general illustration of this point, cf. Roncaglia (2019).

A. Roncaglia (✉)
Sapienza University of Rome and Accademia Nazionale dei Lincei, Rome, Italy
e-mail: alessandro.roncaglia@uniroma1.it

whereby with the term veil we mean that money and/or finance does not affect real variables (production levels and employment, relative prices, and income distribution) but only the general price level.

The history of money and finance theories unfolds in a continuous stream of debate, but we may read it as a sequence of different stages. First, we have a preparatory stage, focused on the role of money as a standard of measure and as a means of exchange, corresponding to relatively primitive monetary and financial institutions. Second, there is what we can designate as the classical period, characterized by a lively multiplicity of views (bullionists versus anti-bullionists, currency school versus banking school) but still addressing a relatively simple financial sector; within this period, the suspension of cash payments by the Bank of England (1797) marks a turning point. Third comes the marginalist approach, with a central notion of equilibrium leading to the tenet of the invisible hand of the market, extended from micro- to macroeconomic equilibrium and to finance as well. This period is characterized by increasing formal sophistication in theoretical analysis and simultaneous reliance on a simplistic view of the world, dominated by supply and demand market mechanisms and simplifying hypotheses such as the absence of uncertainty and a one-commodity world; this favors adoption of financial deregulation policies, in turn favoring development of increasingly sophisticated financial products and markets. Finally and partly overlapping in time with the third stage, we have Keynes's views (and subsequent developments, in particular by Minsky). Keynes brings into play a sophisticated notion of uncertainty to illustrate the central role of finance in the economy, engendering instability and crises; Minsky stresses an acceleration in the financialization of the economy leading to a stage of money manager capitalism.

2   Early Debates

Aristotle attributes two characteristics to money: that of being an abstract standard of measure and that of being a concrete commodity utilized as means of payment (and, as a commodity, also as store of value).

These two elements gave rise, in the Middle Ages, to two different theoretical streams (cf. Wood, 2002, p. 73). On the one side, we have the "sign" theory, also known as "feudal" or "ghost" theory. It focuses on money as a standard of measure, validated by the political authorities, thus constituting an important institutional advance, in that it "unifies" the market, allowing for easy comparisons between different (actual or potential) exchange acts.

On the other side, we have the "metallistic" theory. It focuses on money as a specific commodity, not useful as such but as a "fungible" for what is directly useful, since it is commonly accepted in exchange of commodities and services. It, too, is considered an important institutional progress, in that it enables transition from a still rather primitive barter economy to a complex economy based on the division of labor, thus requiring a network of exchanges.

In the Middle Ages, precise conceptual definitions and conceptual differences were the object of sophisticated analyses. This is also true for the distinction between the sign theory and the metallistic theory. After the Middle Ages, these two elements were once again brought together in discussion on money, its nature, and its origins. But in fact, they have different natures. The first—money as a sign—is an abstract characteristic that can be attributed to a physical commodity as well as to paper money and even to purely imaginary money.[2] The second—the metallistic theory—concerns the origins of monetary systems based on convertibility of paper money into gold (and/or silver). Both approaches are an important part of our past history and still influence present-day theorizing, though the sign notion appears more adequate to the monetary and financial institutions of the present.

The idea of money as a veil has its origins in metallistic theorizing. In it, we should distinguish two aspects: the irrelevance of (changes in) the quantity of money (i) for aggregate income and production levels and (ii) for relative prices and income distribution. It is logically connected to the idea that the invisible hand of the market is capable of guiding the economy to an optimal equilibrium characterized by full employment of resources, labor included.

In this specific meaning, the "veil" thesis cannot be attributed to the classical economists. They held, rather, that when focusing on the elements determining the size, distribution, and growth of the wealth of nations, money and finance could be left to separate treatment; in these separate analyses, logical links by which money and finance could affect "real" variables may be considered and may even prove important.[3]

Many classical political economists, from William Petty on, recognized that money has an impact on the "real" economy. Thus, according to Petty (1691, p. 113), "Money is but the Fat of the Body-politick, whereof too much doth as often hinder its Agility, as too little makes it sick." Here, Petty advances his view against the metaphor of money as blood, common in cameralist writings, seen as the very essence of the strength of an economy or its wealth: a role he attributes, instead, to commodities constituting the social product or the surplus product of the economy. Thus, Petty points to an influence of the quantity of money over the levels of production.

It is worth noting, in passing, that this implies an influence on relative prices as well, unless we introduce the assumption of all-pervading constant returns to scale. Of course, this is a point that Petty himself did not consider: like most if not all classical economists, Petty did not aim to construct an all-embracing model of the economy, as later general equilibrium theorists would aim to do. In fact, even Ricardo, the most systematic of the classical economists, builds "analytical bricks": analysis of the factors determining relative prices, the connection between profits and accumulation and hence between income distribution and the rate of growth, the

---

[2]For examples of imaginary money largely utilized in the late sixteenth and early seventeenth centuries, cf. Rosselli (1995, p. 42) and Goldsmith (1990, p. 246).

[3]Schumpeter ([1954] 1994, pp. 277–278) hints in this direction.

machinery issue, the role of banks and the determination of the rate of interest, and so on; he then assembles some of these with other analytical bricks, such as Malthus's theory of rent, to arrive at policy proposals such as the expediency of abolishing the duties on corn imports. He is aware of the limits of some of these analytical bricks and in particular of his labor theory of value, but with his grasp of the working of the economy (as testified by his successes on the stock exchange), he can retain confidence in his policy views.[4]

A position similar to Petty's can be attributed to a number of participants in the centuries-long debates on the causes and consequences of high or low interest rates discussed in Tucker's (1960) masterly work, a PhD thesis supervised by Sraffa. Tucker highlights a lively debate between those who derive the interest rate from the rate of profits, attributing a passive role to the former, and those who consider the rate of interest as affecting the profit rate (with the interest rate possibly determined, or co-determined, by the supply of money). This debate is also connected to another one, concerning the view of interest rates (and more generally of financial conditions) as affecting investment and activity rates, with low interest rates favorable to growth.[5]

A complex debate on monetary and financial issues, encompassing quite different views, thus precedes, underlies, and surrounds the two leaders of classical political economy, Adam Smith and David Ricardo. It is only by focusing on them—and with some not insignificant interpretative forcing—that we get the "standard view" often attributed to the whole of classical political economy, based on a few tenets. First, interest is considered a part of profits, and the rate of interest is determined by (is strongly correlated with) the rate of profits. Second, the quantity of money is considered a sign (a consequence) of the wealth of an economy, not its cause. Third, the quantity of money affects the general price level and, through this, the distribution of money between different countries (the Hume mechanism). Whether this involves rejecting the idea of a subsidiary influence of the quantity of money on income distribution and relative prices is immaterial in this context—though we may remark that at least Ricardo incidentally (but only incidentally) rejects this influence.[6]

---

[4]Ricardo's great intelligence is fully policy-oriented: we should keep in mind that the professionalization of economics is far off in the future, and the method of axiomatic theoretical models (Bourbaki) is still farther off. Sraffa, as great an admirer of Ricardo as we might desire, commented (in answer to some queries by Gramsci), "Ricardo was, and always remained, a stockbroker with a mediocre culture" (Sraffa, 1991, p. 74, my translation). In fact, it was James Mill who pushed his friend David to write the *Principles* and give them a relatively compact structure.

[5]We should also keep in mind the role of usury laws in determining interest rates or at least maximum ceilings for them. On usury laws, cf. Tawney (1926).

[6]"An alteration in the value of money has no effect on the relative value of commodities, for it raises or sinks their price in the same proportion" (Ricardo, 1951–1955, vol. 2, p. 396: quite explicit but the only explicit reference I was able to find). At the same time, as Hollander (1979, p. 480) remarks, Ricardo stresses that in the transitory periods required for full adjustment after a change in the quantity of money, such effects are possible, indeed likely.

According to the "Hume mechanism," disequilibrium in foreign trade and the ensuing gold flows between countries set in motion re-equilibrating movements in their general price level. This mechanism is part of the "received view." However, it is by no means undisputed. Earlier on, though unknown to most classical authors, a particularly relevant exception had been represented by Antonio Serra (1613). In the third part of his short and splendid book (even today the least read part of that book), Serra points out that at that time, international financial transactions (and speculation) already had a possibly dominating influence on what we now call the balance of payments and hence on rates of exchange between currencies, on relative international competitiveness, and on economic growth or decay.[7] In fact, another leading economist who does not subscribe to the "Hume mechanism" in its entirety is Ricardo: his thesis concerns equality of the purchasing power of money over gold at home and abroad, not equality of the purchasing power of money over commodities.[8]

Another point on which pre-marginalist writers display a variety of positions concerns the definition of money and its distinction from finance. Apart from petty commerce, since antiquity, in nearly all major exchange deals, the time of settlement and the time when the commodities are handed to the buyer have been different; these intervals of time bring in uncertainty, involving an element of loan and possibly of insurance.[9]

Money is originally identified with gold coins and then with gold coins plus bank notes convertible into gold coins (with the consequence that the quantity of money in circulation turns out to depend on prices, rather than the other way round, as stressed by the Banking School and as Kaldor will recall to Friedman in a famous BBC debate). But bank notes are also included in "finance," together with bank deposits, commercial bills, and other forms of credit, and onto public debt certificates and shares of companies traded on the stock exchange. Thus, the classical economists already appear to have seen a continuum of financial assets ranging from commodity money to the most complex forms of financial contracts.[10]

Classical economists often attributed a role to interest rates in affecting investments and activity levels. This does not necessarily clash with the idea of "money as a veil," provided that the rate of interest is seen as determined by "real" factors alone.

---

[7]On Serra's monetary theory, cf. Rosselli (1995).

[8]As a consequence, the main policy target is equality between the market and the official price of gold (cf. Marcuzzo & Rosselli, 1991).

[9]For instance, in traditional oil or gas deals, both payments and commodity consignments may extend over years, even for decades, giving rise to financial derivatives.

[10]Let us recall the controversies between bullionists and anti-bullionists and then between the currency and the banking schools. With the development of banking, a precise definition of money, so as to make it fully exogenous, becomes more difficult to attain. It requires either focusing on coins (commodity money) or relying on a univocal relation between convertible bank notes and the underlying metallic reserves. Tooke's banking school rejects this tenet (cf. Arnon (1991), who also remarks that other classical authors do not strictly adhere to rigorous definitions of money; cf., e.g., p. 23 on Adam Smith).

There are hints in this direction in the work of some of the classical authors, but not in all of them, and in most cases, the situation turns out to be more complex, with various factors operating on different levels (for instance, changes in uncertainty due to institutional and political developments and, on a different level, capital flights and speculation on foreign exchange, as in Serra).

The strict idea of a rate of interest determined by real factors alone (with the supply of loans corresponding to savings and the demand for loans corresponding to investments),[11] though often attributed to the classical authors in their entirety,[12] was in fact only fully developed and became the "received view," with the advent of the marginalist approach, in the last decades of the nineteenth century.[13] It was only after the rise of marginalism, with the tendency to develop all-inclusive models of the economy, that the idea of a full "real" equilibrium of the economy emerged.

The point that needs to be stressed is that classical authors did not hold with axiomatic theories or in any case did not aim at overall, complete, theoretical models explaining the functioning of the economy in all its respects. In this latter kind of models, either money is a veil or it is not. For the classical authors, proceeding with "analytical bricks,"[14] money might well happen to be left aside, as if it was a veil, in some respects (when attention is focused on some specific issues, such as the role of the division of labor in fostering the wealth of nations as in Smith, or the relationship between income distribution and the growth of the economy, as in Ricardo, while aspects which are secondary in that respect—and in that respect only—are left aside), and is not a veil when the issue of its influence on real variables is considered directly.

In other words, money is essentially a veil only in the standard model of the marginalist theory of value and distribution (as in Jevons or Walras or, in the following generation, in Wicksteed, "the purist of marginal theory"), where the problem of allocating scarce resources between alternative needs and desires leads to determining relative prices (including distributive variables, considered as prices of the factors of production) that ensure an optimal equilibrium under competitive conditions implying full utilization of available resources.

---

[11]Of course, this equality only holds when international relations and the public sector are assumed away.

[12]Cf., for instance, Blaug, 1997, pp. 157 ff.

[13]Things become more complex with the increasing complexity of the national and international monetary and financial markets. Hawtrey's "Treasury View" with his adhesion to the loanable funds theory depends on his tenet of a separation between monetary and (long-term) financial markets (already criticized by Keynes) and is strictly speaking only valid for a stationary economy (cf. Tonveronachi, 2019).

[14]On the method of "analytical bricks," applied to the interpretation of Sraffa's analysis, cf. Roncaglia (2009b, in particular pp. 25–28 and 49–51).

3   Money and the Myth of the Invisible Hand of the Market

The myth of money as a veil thus turns out to be connected to the myth of the invisible hand of the market.

With respect to this latter myth, let us recall two points. First, attributing this idea to Adam Smith is a dramatic mistake in the history of economic thought. None of the many commentators on the *Wealth of Nations* attributed the idea to Smith until it was proposed in a paper by George Stigler (1951), aiming more at ennobling the neoliberal stream with a strong historical pedigree (and, at the same time, denying all conceptual differences between the classical and the marginalist traditions). Notwithstanding its feeble foundations (cf. Rothschild, 1994; Roncaglia, 2005, Chap. 5), this idea became a tenet for mainstream economists and the public at large.

Second, as a theory concerning the working of the economy, the theory of the invisible hand of the market is utterly wrong. It only holds for a one-commodity world, an assumption that is necessary in order to avoid reswitching of techniques and capital reversals—or, in the terminology of general equilibrium theory, the instability of the possibly multiple equilibriums—that impede the automatic tendency of the competitive economy to full-employment equilibrium and in a world where Keynesian uncertainty (to be discussed below) is ruled out.

Within mainstream macroeconomics, money has no longer been considered a veil since Pigou pointed out the so-called wealth effect (Pigou, 1941): when money wages fall under the pressure of unemployment, insomuch as real wages do not fall, there will be a fall in the general price level; as a consequence, the real value (purchasing power) of the money holdings of households increases, favoring an increase in their consumption and savings levels. Analogously, changes in the quantity of money in circulation, affecting the general price level, have a "real" effect.

Pigou's wealth effect is conceived as an equilibrating mechanism, additional to the flexibility of the wage rate: monetary policy can help the economy to move toward its full employment level, whenever money wages are downwardly sticky. This effect is then embodied in Modigliani's (1944, 1963) neoclassical synthesis and, through Samuelson's (1948) textbook, in mainstream macroeconomics. Thus, money is seen as a veil only in a limited sense: it does not affect relative prices, but monetary policy may ease attainment of full-employment equilibrium whenever the economy meets some impediments (such as downward money wage rigidity).

Let me stress that this is not Keynes. In fact, Keynes considered Pigou's analysis (cf., for instance, Pigou, 1933) as the prototype of the "classical" theory he set out to criticize. It can be summarized as made up of three elements: the quantity theory of money, the equality between real wage and marginal productivity of labor, and an aggregate production function, leading as a solution to a full-employment stable equilibrium. This model, with the addition of Pigou's wealth effect and the assumption that the money wage rate may be inflexible downward because of the trade unions' bargaining power, leads to Modigliani's (1944, 1963) neoclassical synthesis,

where an underemployment equilibrium is possible and a role for expansionary monetary policy is recognized (cf. Roncaglia & Tonveronachi, 1985).

In mainstream macroeconomics, finance, too, is substantially considered irrelevant to the definition of the "real" equilibrium of the economy. Together with the loanable funds theory of the rate of interest mentioned above, there are two other aspects to be recalled in this respect.

First, a theorem by Modigliani and Miller (1958) states that, under "perfect" competitive conditions, it is irrelevant for the firm whether to finance investments with retained profits or with external finance (bank loans, bonds, or shares placements). The necessary assumptions for this finding are clearly stated; Modigliani and Miller might thus have seen their theorem as demonstrating that, these assumptions being commonly violated in actual practice, entrepreneurs do have preferences over the different forms of investment (internal finance from retained profits preceding, in order, bank loans and bonds and new shares).[15]

Second, there is the efficient financial markets theory (Fama, 1970): when uncertainty is ruled out (not necessarily by assuming certainty: a stochastic uncertainty may be admitted, provided it is not accompanied by cumulative effects, hence assuming stochastic independence), then financial variables such as share prices reflect the underlying "real" variables, such as expected profits. This theory has an important policy impact, being utilized by neoliberal economists and policymakers in support of financial liberalization policies.[16]

4 The Keynesian Notions of Uncertainty and Liquidity Preference

We might say that, compared to Modigliani's attitude, Keynes's is just the opposite. Since the world is far from "perfect," we have better start from the world as it is, taking into account its main characteristics, beginning with uncertainty.

On this count, Keynes is often misinterpreted, by assimilating his notion of uncertainty to the Knightian dichotomy between risk and what is called fundamental uncertainty. Keynes's 1921 book on probability points in a different direction: complete certainty (which includes probabilistic risk) and complete ignorance are the two limit states of the world, with a greater or lesser uncertainty present in most if not all human affairs. Thus, in Keynes's "vision," an evaluation of the probability of

---

[15] In fact, Modigliani and Miller (1958, p. 197), after stressing the assumption of perfect competition (and, earlier on, of certainty), call for a follow-up to their analysis where the simplifying assumptions could be abandoned. The preference for internal financing in more realistic analyses is stressed by post-Keynesian economists (Eichner, 1976 and others) who build on this foundation a theory of income distribution whereby investments determine financing requirements and hence pricing decisions, profits, and the profit rate.

[16] Some mainstream economists (for instance, Blanchard & Summers, 2019) recently appear to be reconsidering their policy standing but without putting in doubt their theoretical foundations.

any given state of affairs is to be accompanied by an indication of our confidence in our evaluation—a degree of confidence ranging, so to say, from zero (perfect ignorance) to one (perfect knowledge), both theoretical limit boundaries with scant empirical correlate.

This also explains why Keynes rejects the axiomatic construction of the subjective probability theory developed by his friend Ramsey (1931) and simultaneously and with greater flexibility by de Finetti (1931), with Savage twenty years later (Savage, 1954) adding up in a strictly axiomatic formulation of subjective probability and the von Neumann-Morgenstern expected utility approach.

In Keynes's view, our evaluations of probability (for any given state of affairs) are subjective, in that they are made by each individual agent endowed with a specific personal set of information, but are at the same time and much more importantly objective, in that each individual agent's information set rationally leads to holding some views and not others. This implies that the agent is not free to hold whatever set of probability estimates she/he likes, only provided that it is internally consistent.

This is an important point and has been the object of debates between opposite interpretations of Keynes's probability theory, as either objective or subjective. Let me try to clarify my compromise proposal, sketched out above, with an example. How long would you bet my talk will last? A frequentist solution would be to look at the average length of past talks in this conference, about 20–30 minutes, and bet on that interval. As Keynes stresses, this solution ignores changes in the environment, producing what statisticians call a "break" in the time series, and this is clearly the case here, as a glance at the conference program immediately shows.[17] A purely subjective solution would consider any bet to be rational, provided that it is internally consistent, so that no Dutch book can be constructed against it, thus, for instance, even attributing probability one to the interval between three and four hours. An objective bet à la Keynes would imply relying on the program, no discussant and time available up to an hour, but if you have to bet on a ten-minute interval, your confidence in your bet will not be very high. It is here that the personal (but objective) information set acquires importance. Some of you may have heard Cristina Marcuzzo telling me to speak around 45 minutes. Thus, if you consider me sufficiently reliable, you would bet with some confidence on the 40–50-minute range. However, this afternoon, I am excited about the Guggenheim prize, so that I may lose control and talk for an hour. Some uncertainty is unavoidable, but in any case, a rational agent should behave like a good paterfamilias and keep into account the objective elements provided by her information set.[18]

Thus, considering the substantial differences in the kind of uncertainty involved in different situations, Keynes adopts the method of the "short chains of reasoning,"

---

[17]This important methodological element is often forgotten in financial risk analysis, with serious consequences in the setting of regulatory capital requirements for financial institutions (cf. Roncaglia, 2012).

[18]Actually, my talk lasted 43′25″.

with what we might call a "piecemeal" approach to the construction of economic theory: fields for which our "degrees of confidence" substantially differ (as in the case of financial speculation, investment decisions, or decisions on production levels) give rise to separate theories that are subsequently connected in longer "chains of reasoning." This interpretation, together with the stress on the specific Keynesian notion of liquidity discussed below, is in contrast with Hicks's "received" reading in terms of an IS-LL model, which unifies in a "general" theory all the components of a (modified) Keynesian analysis.

In other words, in Keynes's theoretical construction, his theory of financial markets is to be kept separate (and may be seen as a first logical step) from the theory concerning the determination of the level of investments, and both are better kept separate from the ensuing theoretical stage concerning the influence of investments on the levels of income and employment. Each of these theories involves a different kind of uncertainty: different sets of information are relevant.[19] The agent is concerned with three separate aspects of the economy: evaluation of movements of interest rates (and of other financial variables) in the immediate future, affecting demand for liquidity; evaluation of returns from alternative investment strategies, over a quite long spell of time corresponding to the durability of fixed capital; and evaluation of the prospects for sale of the product over a relatively short span of time (decisions on production levels can be changed quite rapidly, even if not instantaneously as is the case with decisions concerning financial assets).

This explains why Main Street is strictly connected to Wall Street, as Minsky used to say, or in other words why finance may rule the roost in the logical chain of reasoning with which Keynes describes the working of a monetary production economy.

It also explains Keynes's specific notion of liquidity preference, which constitutes the basis for his theory of money and interest—another notion quite often misunderstood. The preference for liquidity is connected to uncertainty (in the specific sense Keynes attributes to the term): I prefer to hold part of my wealth in a more "manageable" form allowing me to change my strategies quite rapidly at nil or very low costs, even if I have to accept a lower return on more liquid assets compared to other less liquid assets. Whenever my confidence in my evaluation of the situation diminishes, my preference for liquidity increases. If this change in the state of confidence has some general motivation (for instance, a political crisis), many financial operators more or less simultaneously sell assets characterized by higher returns but less liquidity; the price of those assets falls, and interest rates increase (in fact, the whole structure of interest rates is modified).

Liquidity is an attribute of all assets, not just of money; in evaluating the liquidity of different assets, much depends on the way the different markets work and so not only on the nature of what is traded in each market but also on institutions, customs, and the constantly changing state of affairs. As the 2008 crisis showed, the banks'

---

[19]There is an analogy here with Keynes's notion of "groups" proposed in the *Treatise of Probability* (Keynes, 1973, p. 134). On this, cf. Roncaglia (2009a, p. 498).

certificates of deposits may all of a sudden become illiquid—or at any rate less liquid than they were considered previously or less liquid than gold or diamonds. Thus, a drive to liquidity may take different forms in different circumstances.

5   Minsky's Money Manager Economy

Monetary and financial institutions change over time, from the origins to our own day. The changes are such that we can actually speak of a "regime change." Indeed, Hyman Minsky proposes the idea of a shift from managerial capitalism (the stage described by Berle and Means, which succeeds competitive capitalism and, earlier on, commercial capitalism) to money manager capitalism. By this, Minsky means a situation in which, to use Keynes's metaphor, the tail of finance wags the dog of the real economy.

The way of describing the economy adequate to money manager capitalism is, according to Minsky, that of considering economic agents' decisions as relying on streams of expected receipts and outlays.

Commonly, households are confronted with streams of expected income more or less equal to or a bit higher than expected expenditure. Only exception, for instance, when buying a house with recourse to a mortgage, is the expenditure at the moment of the acquisition far higher than the receipts, with the mortgage covering the difference, but in subsequent periods, the expected income is once again higher than expected outlays inclusive of mortgage interest and repayments. Thus, in general, households hold what Minsky calls "covered" positions and "speculative" positions only when resorting to a mortgage (or consumer credit).

Firms, when investing, commonly take on loans that they expect to repay with the proceeds from the sale of their products; expected outlays are thus initially higher than expected receipts, but overall, the present discounted value of the series of outlays and receipts is confidently expected to be positive over a sufficiently wide range of interest rates. Firms thus commonly hold "speculative" positions.

We have, instead, a "Ponzi" (or ultra-speculative) position whenever, after the initial investment, the receipts are expected to be insufficient to cover interest and repayment instalments on the original loan for a long period, so that new loans are continually required; the present value of the series is expected to be positive because of the increase in value of the asset, the acquisition of which motivates the original loan, with the price of the asset expected to increase over time at a rate higher than the current rate of interest. Thus, for instance, we have a Ponzi situation when assets such as a house or gold are acquired with the awareness that receipts over a long period will be absent or in any case insufficient to repay not only the amortization of the loan but also the yearly interest, so that the amount of the debt grows over time.

Obviously, when the price of the asset ceases to increase or the rate of interest increases enough to yield a negative present value of the flow of outlays and receipts, Ponzi speculators go bankrupt. If Ponzi situations are widespread, the economy

undergoes a systemic crisis, originating in the financial sector of the economy and spreading to the real economy.

According to Minsky, the financialization of the economy—being not only a quantitative but also a significant qualitative phenomenon—brings with it the risk, indeed the likelihood, of a growing share of Ponzi situations, leading to increased fragility of the economy. At each crisis, the authorities intervene, and when, thanks to the interventions, the situation has been stabilized, the share of speculative and Ponzi finance increases once again—and this takes place time after time with a greater propensity to risk-taking, since the speculators become used to being rescued by state intervention enacted to avoid the worst pitfalls of the crisis. Thus, with each successive episode, Ponzi and speculative finance tend to cover an ever larger part of the economy, and the financial crises may increase in magnitude.

6   Some Provisional Conclusions

I have considered four distinct stages in the history of theorizing over money and finance. After an initial, "preparatory" stage, in the second period, that of classical political economy, the growth of banking saw transition from a simple "commodity" theory of money to a more complex view of interrelation between money and finance, in the context of debates ranging over a variety of views. In the third stage, following the "marginalist revolution," the associated ideas of the invisible hand of the market and of money as a veil were embodied in all-embracing theories of competitive equilibrium as an optimal solution to the problem of utility maximization under the constraint of given resources; analyses of money and finance were built on such "real" foundations and relied on the assumption of certainty (or, at most, stochastic risk) and on a—faulty—underlying theory of value ensuring a basic tendency to a full-employment equilibrium, hindered only by obstacles to perfect competition. More recently, while mainstream views still dominate economics teaching and policy advice, the Keynesian-Minskyan revolution with the focus on uncertainty while recognizing the driving role of finance in the contemporary economy has proposed a new paradigm shift.

The basic aspects of the new paradigm may be summarized in three points. First, money and finance cannot be kept separate; they are not a veil, but crucially affect the real economy. Second, the Keynesian notion of uncertainty and the associated notion of liquidity preference constitute the foundations on which theorizing over money and finance should rely. Third, the role of finance increases over time, with a change of regime to a "money manager capitalism."

Clearly, the policies designed on the basis of the ideas of money as a veil and of efficient financial markets lead to financial deregulation and so to financial fragility; they have been responsible for the most severe crises of the past nine decades, and they risk paving the way to even more dramatic crises. About a century ago, economic depression led to the rise of fascism and Nazism; today, we are confronted with a—possibly equally dangerous—rise of populism and nationalistic political

movements. Mainstream economics must revise their interpretation of the monetary production economy and their policy strategies before it is too late.

References

Arnon, A. (1991). *Thomas Tooke. Pioneer of monetary theory*. Edward Elgar.

Blanchard, O., & Summers, L. H. (2019). *Evolution or revolution? Rethinking macroeconomic policy after the Great Recession*. MIT Press.

Blaug, M. ([1962] 1997). *Economic theory in retrospect*. Richard Irwin. 5th edition, Cambridge University Press.

de Finetti, B. (1931). *Probabilismo. Saggio critico sulla probabilità e sul valore della scienza*. Perrella.

Eichner, A. S. (1976). *The megacorp and oligopoly*. Cambridge University Press.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*(2), 383–417.

Goldsmith, R. W. (1990). *Sistemi finanziari premoderni*. Laterza.

Hollander, S. (1979). *The economics of David Ricardo*. Toronto: University of Toronto Press.

Marcuzzo, M. C., & Rosselli, A. (1991). *Ricardo and the gold standard. The foundations of international monetary order*. Macmillan.

Keynes, J.M. ([1921] 1973). *A treatise on probability*. Macmillan. As reprinted in *The collected writings of John Maynard Keynes*, vol. 8. Ed. by D. Moggridge. Cambridge: Cambridge University Press for the Royal Economic Society.

Modigliani, F. (1944). Liquidity preference and the theory of interest and money. *Econometrica, 12*(1), 45–88.

Modigliani, F. (1963). The monetary mechanism and its interaction with real phenomena. *Review of Economics and Statistics, 45*(supplement), 79–107.

Modigliani, F., & Miller, M. (1958). The cost of capital, corporation finance and the theory of investment. *American Economic Review, 48*(3), 261–297.

Petty, W. ([1691] 1899). *Verbum sapienti*. In appendix to W. Petty, *The political anatomy of Ireland*. D. Brown and W. Rogers. As reprinted in *The economic writings of Sir William Petty*. Ed. by C. Hull, vol. 1 (pp. 99–120). Cambridge University Press.

Pigou, A. C. (1933). *The theory of unemployment*. Macmillan.

Pigou, A. C. (1941). *Employment and equilibrium*. Macmillan.

Ramsey, F. (1931). *The foundations of mathematics*. Routledge and Kegan Paul.

Ricardo, D. (1951–55). *Works and correspondence*. 10 vols. Ed. by P. Sraffa. Cambridge University Press.

Roncaglia, A. (2005). *The wealth of ideas*. Cambridge University Press.

Roncaglia, A. (2009a). Keynes and probability: An assessment. *European Journal of the History of Economic Thought, 16*(3), 485–510.

Roncaglia, A. (2009b). *Piero Sraffa*. Palgrave Macmillan.

Roncaglia, A. (2012). Keynesian uncertainty and the shaky foundations of statistical risk assessment models. *PSL Quarterly Review, 65*(263), 437–454.

Roncaglia, A. (2019). *The age of fragmentation. A history of contemporary economic thought*. Cambridge University Press.

Roncaglia, A., & Tonveronachi, M. (1985). The pre-Keynesian roots of the neoclassical synthesis. *Cahiers d'économie politique, 10*(11), 51–65.

Rosselli, A. (1995). Antonio Serra e la teoria dei cambi. In A. Roncaglia (Ed.), *Alle origini del pensiero economico in Italia* (*Moneta e sviluppo negli economisti napoletani dei secoli XVII–XVIII*) (Vol. 1, pp. 37–58). Il Mulino.

Rothschild, E. (1994). Adam Smith and the invisible hand. *American Economic Review*. *Papers and Proceedings, 84*(2), 319–322.

Samuelson, P. A. (1948). *Economics*. McGraw Hill.

Savage, L. J. (1954). *The foundations of statistics*. Wiley.

Schumpeter J. (1954). *History of economic analysis*. Ed. by E. Boody Schumpeter. Oxford University Press. Reprinted (1994). Routledge.

Serra, A. (1613). *Breve trattato delle cause che possono far abbondare li regni d'oro e d'argento dove non sono miniere con applicazione al Regno di Napoli*. L. Scorriggio. Reprinted (1994). Verlag Wirtschaft und Finanzen GmbH.

Sraffa, P. (1991). *Lettere a Tania per Gramsci*. Editori Riuniti.

Stigler, G. J. (1951). The division of labor is limited by the extent of the market. *Journal of Political Economy, 59*(3), 185–193.

Tawney, R.H. (1926). Introduction. In T. Wilson (1572). *A discourse upon usurye*. Ed. by R.H. Tawney. Bell. Reprinted (1963). Frank Cass.

Tonveronachi, M. (2019). Ages of financial instability. Paper presented at the Fifth Thomas Guggenheim conference in the history of economic thought *Financial instability, market disruptions and macroeconomics: Lessons from economic history and the history of economic thought*. Rome, Accademia Nazionale dei Lincei, 17–18 December.

Tucker, G. (1960). *Progress and profits in British economic thought 1650–1850*. Cambridge University Press.

Wood, D. (2002). *Medieval economic thought*. Cambridge University Press.

Part II
Financial History

British Investment Trusts 1868–1928: Portfolio Diversification and the Beginnings of Institutional Investment

Janette Rutterford, Dimitris P. Sotiropoulos, and Antonis Kyparissis

1 Introduction

UK investment toward the end of the nineteenth century was heavily influenced by a reduction in the supply of officially listed national debt. Yields on British government consols fell below 3% by the last decade with only a few alternatives available to risk-averse investors, such as trustees. For example, the risk of investing in domestic corporate securities was high: in the 1860s and 1870s, after the introduction of limited liability regulation in the 1850s and 1860s, two out of three newly incorporated companies failed within 3 years. By 1914, the average life of a joint stock company was still only 10 years (Michie, 1981). Many publications aimed at individual investors urged them not to invest in ordinary shares of companies at all (Rutterford, 2004). Investors did, though, have the option to invest overseas, in foreign government, municipal, provincial, or corporate securities. These securities typically had higher yields than did equivalent domestic securities (Sotiropoulos & Rutterford, 2018). But overseas investment was also riskier, as evidenced by the number of foreign bondholder associations which lobbied on behalf of British bondholders whose bonds were in default (Flandreau, 2013). Despite such risks, by 1913, British investors' overseas investments were more than double those of any other country and amounted to the sum of £3.1 billion.[1]

At that time, British investors concentrated primarily on income yield as a valuation tool (Rutterford, 2004). The level of risk was priced in the market yield, with the yield on consols that of the British risk-free benchmark. The riskier the

---

[1] The estimation is from Platt (1986, p. 60).

J. Rutterford (✉) · D. P. Sotiropoulos · A. Kyparissis
Department of Accounting and Finance, The Open University Business School, Milton Keynes, UK
e-mail: j.rutterford@open.ac.uk; dimitris.sotiropoulos@open.ac.uk; antonis.kyparissis@open.ac.uk

security, the higher the required yield, or as Beeton (1870, p. 26) put it, "The higher the rate of interest, the worse the security."[2] Once the required level of yield had been determined, the investor could minimize risk in a number of ways. The first was to avoid investing in categories of security that were considered too high up the risk scale, that is, the higher yield being deemed not worth the risk of interrupted income and/or capital loss. The second was to spend time investigating each security in depth, by studying the accounts and reading newspapers or by consulting advisers.[3] The third method of reducing risk was to spread risk across different securities, via explicit diversification. The ability to diversify was a game changer for the developed financial markets in the wake of limited liability.

This chapter looks at how the now well-known institutional investment approaches to adding value, such as diversification, evolved in the late nineteenth century and first half of the twentieth century. The chapter explores how ways to optimize returns while minimizing risk for long-term investors were developed, promoted through financial advice manuals and texts, and put into practice by investment trusts from the 1860s onward. We document how the concepts of diversification and yield enhancement dominated the investment discourse and how investment trusts offered individual investors a low-cost means of maximizing return relative to risk. We also show how investment trusts implemented a range of investment strategies—in particular, active asset allocation, portfolio diversification, stock selection, market timing, and leverage—which form the bedrock of investment strategies today.

2   Risk Reduction Through Diversification

Spreading risk across a number of securities was widely promoted as early as the 1870s. Advisers offered recommendations as to how to combine a number of investments in a portfolio and thereby improve the risk return trade-off. *Chadwicks' Investment Circular* argued in 1870 (28 December, p. 30):

> We hold that, by a careful selection from the various media of investment, very remunerative returns in the shape of interest may be obtained; while, by a proper division of risks, not only may the *security* for the principal be rendered perfectly satisfactory, but there may be a good prospect that the invested capital will steadily increase in value.

Similar advice was also offered in *Beeton's Guide Book to the Stock Exchange and Money Market*, published in the same year (Beeton, 1870, p. 26):

---

[2]For further discussion of how investors, in particular British investors, valued securities before World War I, see Rutterford (2004).

[3]Another way to improve information flow was to live close to the company's headquarters, area of operations, and/or location of annual general meetings. For more discussion on local investment, see Rutterford et al. (2017).

> If an investor wishes to secure a high rate of interest, he should divide his capital among a number of stocks that can be bought to pay a high rate of interest—the more the better. Supposing he has £500 to invest, let him invest £100 in each of the following—Turkish, Italian, Spanish, Egyptian, Guatemalan, or Argentine. By dividing his capital in this way, the investor reduces risk to a minimum, as it is unlikely that all these countries could stop paying their interest, although it is not unlikely that anyone might do so.

Although appearing to limit investor choice to foreign government bonds, *Beeton's Guide* suggested that investors choose from a wide range of countries and types of security but, as with Chadwicks, preferred a mix of low-yield, low-risk domestic securities and higher-risk, higher-yield foreign securities. These recommendations allowed for three to five different holdings of standard—but not equal—sizes, with a minimum portfolio size of £500. The amount to be invested in each security was a function of the desired average yield.

This approach to diversification continued until the early 1900s when a key development was a more *scientific* approach to portfolio diversification. Instead of adding as many risky securities as required to generate the required yield, some investors began to realize that a more top-down approach to portfolio construction was desirable, targeting a particular level of yield and minimizing capital risk through the choice of relatively uncorrelated securities which operated in different geographical regions. Such a diversification strategy was developed by Henry Lowenfeld, author of numerous investment texts, and actively promoted by the *Financial Review of Reviews*, a monthly magazine first published in 1905 (Rutterford & Sotiropoulos, 2016). Lowenfeld (1909, p. 11) recommended the following simple rules for portfolio diversification:

> The safety of Capital is obtained by dividing it (1) equally among a number of sound stocks (2) of identical quality, but (3) every stock held must be subject to an entirely different market and trade influence.

Lowenfeld's approach differed from those of Chadwicks and Beeton in two key ways. First, he required equal amounts[4] to be invested—an approach we now call "naïve diversification."[5] Second, he proposed a total of ten securities (less for the less wealthy investor) which would be subject to different risks since they would be spread across the entire globe or, in modern parlance, a "top-down" approach. To achieve this, Lowenfeld split the world into nine regions: Britain, British colonies, Asia, Africa, North, Central and South America, and North and Southern Europe

---

[4]The emphasis on nominal rather than market value reflected the relative disregard for capital gain or loss compared with yield as a source of return. Some publications were unsophisticated as to the number of securities to choose and the difference between nominal and market values as far as diversification was concerned. For example, the weekly *Investors' Review*, in 1905, recommended a model trust with four securities of nominal value £100 each, with market prices varying from £102 ½ for Buenos Ayres Railway Debentures paying 5% nominal to £280 for Nobel Dynamite shares paying 10% nominal yield (11 November, p. 594).

[5]However, some allowance was made for the amount of money to be invested: for example, Lowenfeld recommended holding 5–6 stocks for an investment of £500–1000 and 8–10 stocks for £5000–20,000 (Lowenfeld, 1907, p. 85).

with a tenth category being "international" securities such as shipping, telegraph, and marine insurance (see Lowenfeld, 1907). Thus, a maximum of two-tenths of the portfolio was to be invested in Britain and its colonies. As with Beeton and Chadwicks, though, the securities would be chosen to achieve a particular desired weighted average yield, for example, 4, 5, or 6%. In this way, the diversification question was translated into investing equal amounts in a range of securities which ensured a so-called geographical distribution of capital.

3   The Role of Investment Trusts in Portfolio Diversification

For investors who did not have enough savings to be able to build diversified portfolios themselves, investment trusts provided the opportunity to acquire rights to a share of a diversified portfolio held in a trust. The first investment trust, Foreign and Colonial Government Trust (F&C) was launched in the United Kingdom in 1868 (see Rutterford, 2009, pp. 161–162). Promoted by Philip Rose, Disraeli's personal financial adviser and a partner in a law firm (Chambers & Esteves, 2014), and hence familiar with the legal structure of trusts, the trust form was initially preferred to that of the limited liability company to avoid "the now unpopular name of the company" (McKendrick & Newlands, 1999, p. 26).[6] Only 2 years earlier, in 1866, the Overend Gurney Bank—and a further six banks—had failed (Turnbull, 2018).

The principle of diversification was a key investment objective. For example, the first F&C prospectus (cited in Powell, 1915, p. 469), seeking to raise £1 m, stated:

> The object of this trust is to give the investor of moderate means the same advantages as the large capitalist in diminishing the risk of investing in Foreign and Colonial Government stocks, by spreading the investment over a number of different stocks and reserving a portion of the extra interest as a sinking fund to pay off the original capital.

A minimum amount of diversification was guaranteed by requiring that the percentage holding in any one stock was a maximum of 10% and hence that the minimum number of holdings was ten. In the initial portfolio, outlined in the prospectus, there were to be 18 holdings in total, of which three were at the maximum holding of 10%, or £100,000 in this case: Spanish new three per cents, Peruvian five per cents, and Italian five per cents (1861). The holdings were not equal in size, ranging from 1.5% to 10% of the portfolio. The idea was to target a specific market yield, by putting together a judicious mix of domestic and foreign securities (Rutterford, 2009, p. 159). As *The Times* (1868, 20 March, p. 10) commented:

> The scheme in its principle supplies a want that has long been felt, since it not only gives to that large number of persons who are always disposed to encounter the risk of foreign

---

[6]The failure of the newly floated Overend Gurney Bank in 1866 had led to a loss of confidence in the public company. The chairman of the trust, Lord Westbury, had, as attorney general, carried through the Fraudulent Trustees Bill in 1857 and the Bankruptcy and Insolvency Bill in 1861.

investments the means of restricting that risk to the smallest amount, but will also to a great extent provide an insurance against it by limiting the yearly dividends to a sum which, with the gains from sinking funds, will admit of an accumulation to meet any untoward contingencies.

In the case of F&C, the diversification was spread across eighteen different foreign and colonial government bonds, whose coupons ranged from 3% to 8% and whose yields ranged from 5.1% for New South Wales stock to 13.7% for Turkish five per cents. These were certainly not all risk-free investments. Indeed, the proportion of the portfolio to be invested in colonial securities—there were no domestic securities—was a paltry 5% of the total. This was much less than Chadwicks' or Beeton's recommendations to individual investors, as we discussed above. The remaining 95% of the portfolio included some high-risk choices. For example, *The Economist* referred to Austria as a "dishevelled" state and Italy as "inchoate" (cited in McKendrick & Newlands, 1999, p. 37). In 1868, the Turkish five per cents were priced at £$36^{1}/_{8}$. They rose to £53 in 1873, a rise of 31.8%, only to fall back to £39½ a year later (Scratchley, 1875, p. 16). Trustees and investors expected defaults; as early as 1871, F&C was reporting nonpayment of interest on Turkish six per cents of 1865, although the chairman was confident of payment as "he had always found the Turks very honourable in their commercial dealings" (McKendrick & Newlands, 1999, p. 42).

However, there was provision for the setting up of a reserve to cover irrecoverable losses. The trustees promised investors a nominal yield of 6% on the certificates being issued, equivalent to a market yield of 7% on the issue price of £85 per cent. On the other hand, the portfolio described in the prospectus would provide a weighted average market yield of exactly 8%. The 1% difference between the yield to be received and the yield paid out (8%–7%) was to be retained as a reserve against unforeseen events and used as a sinking fund to pay off the certificates using annual drawings. The securities listed in the F&C portfolio were undated government bonds which could mature at any time through a drawing, as in a lottery, but otherwise were expected to be still outstanding after the closure of the trust; this was to be in 24 years' time. With 17 out of 18 of the bonds in the portfolio priced below par, early redemption would generate a capital gain which could be used in the event of a bond default or provide a surplus payable to certificate holders on closure of the trust. Thus, investors would have the benefits of diversification, the buildup of reserves against possible losses, and some possibility of capital gain. These benefits, together with relatively low-cost professional management, were attractive to all wealth levels.

The popularity of the F&C issues (there were four in total by 1872) led to a rash of imitations of what became known as "average investment trusts," that is, trusts aiming to benefit from diversification (Scratchley, 1875, title page). For example, The Share Investment Trust, floated in 1872, drew directly on the success of the F&C:

The principle of distribution of risk by embodying in a Trust a number of undertakings, yielding high rates of interest, introduced by the F&C Trust, has been fully recognised to be

> of great advantage to investors [. . .]. The present scheme proposes to embrace a number of well-selected industrial undertakings yielding high rates of interest.[7]

For the Share Investment Trust, the structure was the same as for F&C, but the portfolio preference was for domestic securities rather than foreign government bonds, including shares as well as bonds in "submarine cables, tramway companies, iron and engineering companies, telegraph and construction companies, and other industrial undertakings yielding high rates of dividend." Preference shares offered higher yields than bonds to reflect their higher risk, with ordinary shares the riskiest. The Scottish American Trust was launched in 1873 by Robert Fleming, who had experience of investing in North America. He preferred instead to invest 80% of Scottish American's portfolio in railroad securities. Over time, two centers for investment trusts developed, Scotland (in particular, Aberdeen, Glasgow, Edinburgh, and Dundee) and London (Sotiropoulos et al., 2019). Thus, although the "averaging" principle was the same for all such trusts and the investment objective of achieving higher expected yields for their investors was the same for all trusts, very different asset allocation decisions were taken to achieve these aims by the directors of individual trusts, with the choice depending largely on the timing of the issue and the individual preferences of the managers. A major change, however, was the switch from trust status to corporate status which took place in the 1870s and 1880s (Rutterford, 2009).

4 Investment Trusts as Companies

There were three important differences between investment trusts (hereafter, ITs) structured as trusts and those structured as companies. One key difference was the ability of incorporated trusts to have a capital structure—that is, fund the portfolio—with more than one type of security. ITs took advantage of this opportunity on conversion from trusts to companies, with very few having just ordinary shares. For example, on conversion of the four F&C trusts into a single company in 1879, certificate holders received both preferred stock and deferred stock in lieu.[8] Many trusts chose to have debenture stock as well as preferred and deferred (ordinary) as this reduced the overall cost of capital still further—or, rather, enhanced the potential dividend to ordinary shareholders. For example, the chairman of the Railway Debenture Trust commented at the 1875 Annual General Meeting that every increase of £500,000 in the borrowed money at 5% interest, with an additional ½% for a sinking fund, would add 1½% to the dividend to the share capital, so that with borrowed capital of £2,000,000, they would be able to pay a steady dividend of 10% and the shares would be worth a considerable premium (Scratchley, 1875, p. 38). But

---

[7] Prospectus, Guildhall Library, MS 14235.

[8] Essentially 5% cumulative preference shares and ordinary shares (McKendrick & Newlands, 1999).

the objective was the same. Given a weighted average portfolio yield, the problem was how could the capital structure be tweaked in order to maximize the dividend yield on the ordinary shares. A common target, as for the Railway Debenture Trust, was 10% yield for the ordinary shareholders, substantially more attractive than consols.

Table 1 shows the size and capital structure of both English and Scottish ITs throughout the period. The average size of these ITs was not very different for English and Scottish ITs and rose only slowly—from just under £1 m to £1.5 m—between 1880 and 1929. There were also regional preferences with respect to capital structure. Scottish ITs preferred short-term debentures of, say, 3–7 years, whereas English trusts opted for long-term debentures of up to 50 years, locking in low rates until after WWI. Scottish ITs also relied more on debentures than did their English counterparts; Scottish debentures represented 40% of total capital compared to 25% for English ITs in 1929.[9] This led to higher leverage for Scottish trusts, with ordinary shares representing less than one-quarter of nominal capital as compared to more than one-third for English trusts in the same year. However, all trusts increased the amount of leverage—and reduced their cost of capital—over time. As ITs became more established, directors felt able to increase the borrowing powers and hence leverage ratios for new ITs. For example, the articles of the Scottish Investment Trust allowed for a 50/50 preferred/deferred (ordinary) split and debenture issues up to 50% of capital; the second Scottish Investment Trust had a 60/40 preferred/deferred split and a 100% borrowing limit (Robinson, 1923, p. 19).

A second major difference between ITs as trusts and ITs as companies was that the finite life of a trust was replaced by the unlimited life of investment trust companies.[10] In strategic terms, unlimited life meant that trustees—turned directors—could reinvest the proceeds of bond redemptions either on maturity or if bought in early (e.g., US corporate bonds often had sinking funds and annual random drawings so that many bonds were redeemed early). Any capital gains on realization were used to create reserves. These reserves could be "outer" reserves, visible on the balance sheet, or off-balance-sheet "inner" reserves which were used to write down existing or new holdings to a lower book value (Sotiropoulos et al., 2019). This meant that a portion of the net revenue could be retained and reinvested (accumulated "at compound interest"), in effect setting up a reserve to cope with potential future losses (Kilborne, 1925, p. 170). These reserve cash inflows also gave IT directors the ability to change their portfolio strategy while remaining fully invested in the markets. These cash flows thus forced directors of ITs to actively manage their portfolios over time.

---

[9]English ITs were later criticized for not having included options to redeem early for their long-term debentures, preventing them from benefitting from the lower interest rates of the 1920s and 1930s (Investors' Chronicle, 1949, p. 158).

[10]Many ITs, including F&C, are still in existence today.

Table 1 Size and capital structure of British investment trusts

| | Investment trusts | | | Average paid-up capital (£ m) | | | Average capital structure (% of paid-up capital) | | | | | |
| | | | | | | | English | | | Scottish | | |
	Total	English	Scottish	Total	English	Scottish	Ordinary	Preferred	Debenture	Ordinary	Preferred	Debenture
1880	9	5	4	0.837	0.991	0.646	67.5	20.1	12.4	80.4	0.0	19.6
1885	12	8	4	1.118	1.344	0.667	57.9	29.9	12.2	80.4	0.0	19.6
1890	46	37	9	1.075	1.153	0.757	46.6	29.9	23.5	50.3	18.1	31.6
1895	50	40	10	1.122	1.173	0.920	40.8	32.7	26.5	45.6	19.9	34.4
1900	53	41	12	1.082	1.148	0.857	41.1	32.2	26.7	48.0	19.3	32.7
1905	58	44	14	1.083	1.132	0.932	40.9	33.3	25.8	43.3	25.8	30.9
1910	65	46	19	1.133	1.211	0.946	39.6	32.2	28.2	40.1	25.0	35.0
1915	85	53	32	1.198	1.348	0.950	39.0	32.8	28.2	33.0	32.8	34.2
1920	90	55	35	1.242	1.419	0.964	40.6	31.7	27.7	32.2	34.2	33.6
1925	114	68	46	1.311	1.411	1.162	41.5	32.1	26.5	28.2	34.3	37.5
1929	186	112	74	1.502	1.490	1.521	37.6	37.9	24.6	24.2	36.2	39.6

Notes: The British investment trust companies are identified by the three studies of Glasgow (1930, 1932, and 1935). Our calculations of capital structure are based on data collected from Glasgow (1935) and the *Stock Exchange Yearbook*

The calculations in the table are for all UK investment trusts, as they were defined in the text

5   Corporate Governance

The third difference between trust and company ITs was that of corporate governance. Trustees for life responsible for the original trusts were replaced, in some cases gradually, by company directors.[11] ITs became companies, required to produce accounts, report to the shareholders, and obtain shareholder approval for dividend payments and issues of new capital. They were also listed on the London Stock Exchange and/or Scottish stock exchanges and were thus required to adhere to stock exchange regulations.

From the start, ITs required management. Sotiropoulos et al. (2019) show how the main professions from which directors were drawn were merchants, lawyers, accountants, and directors with links to particular asset classes, e.g., Argentinian or US railways, and brokers familiar with stock markets. As Cassis (1987) pointed out, in 1890, Lord Eustace Cecil was chairman of a railway company and director or chairman of five ITs, while, in 1912, Lord St David's was a director or chairman of six ITs as well as of six overseas companies (Cottrell, 2012). Experts in trusts and portfolio management for individuals, they had the skills required to create diversified portfolios, devise capital structures, and buy and sell securities as needed over time.

Boards of directors were not large. For the ITs in our sample, described below, there were on average five directors for English and Scottish ITs on incorporation in 1914. Keynes, for example, was one of three directors of a London-based IT, the Independent Investment Trust, launched in 1924 to adopt a market-timing strategy related to bonds and common stock in the United States. As a result, ITs had low management costs: the norm was ½% of the portfolio value including director fees (Robinson, 1923, p. 17). However, directors were often on the board of more than one IT, thereby boosting their income. The trusts did not employ specialist stock market analysts, but the boards of directors—and their contacts—included stockbrokers, lawyers, accountants, directors of insurance companies, and directors of other ITs. It was common for directors to meet regularly to decide on sales or purchases, relying on their own knowledge or that of specialist brokers or occasionally asking for or being given suggestions by shareholders (as was the case for the Share and Debenture Trust; see Rutterford, 2009).

There were numerous administrative issues to be dealt with by IT directors and managers, exacerbated by the overseas nature of much of their portfolios. Foreign bonds were often bearer; might be denominated in, say, US dollars; were bound by trust deeds relating to the security underlying the debt, such as gold or mortgages; and included sinking funds which paid for regular drawings and hence redemption of the bonds. Taxes on income from US securities were also payable. There was often no fixed maturity date for foreign government bonds, whereas corporate bonds could be called before their stated redemption date. Preferred stock and equities had no specified maturity but were subject to the risk of reconstruction or liquidation. As

---

[11]For example, this role was only abolished for F&C in 1913 (Chambers & Esteves, 2014, p. 5).

Steiner (1929, p. 6) remarked, "distance, unfamiliarity, varying currencies, regulations, business laws, and languages" meant that the individual investor did not have the skills, knowledge, and connections needed to invest across the globe. ITs, on the other hand, did have directors "trained in political and economic events" with local knowledge and financial expertise (Williams 1928, p. 28).

In addition, there was a long history of defaults by foreign issuers on bonds listed on the London Stock Exchange, and this had led to the setting up of a number of foreign bondholder associations—such as Spanish, Mexican, Greek, Peruvian, Colombian, and Venezuelan—from as early as the 1820s and 1830s (Flandreau, 2013). The powerful British Corporation of Foreign Bondholders (CFB) was founded the same year as the F&C was launched in 1868, and one director of the F&C, William Trotter, was also a member of the CFB. Lord St David's London Correspondent of the Aberdeen Trust Company Limited, founded in 1875, was also on the Spanish Bondholders' Committee. These connections were put to good use:

> Since the boards of investment trusts are often behind the scenes in regard to what is happening to defaulted bonds and debentures, there are occasions when it is wise for them to purchase these silent securities for the sake of capital profit which is sometimes a practical certainty within a comparatively short time (Robinson, 1923, p. 9).

6   The Investment Trust Sample

We now turn to how these very early institutional investors managed their investment portfolios. Our aim is to explore the ITs' investment strategies, in particular their approach to asset allocation as a means of diversification. We also examine how active ITs were in terms of enhancing returns via stock selection and market-timing strategies.

To do so, we study a sample of those investment trusts which applied the so-called averaging (diversification) principle as determined by Glasgow (1930, 1932, 1935; see notes to Table 1),[12] and that also, within the period 1886–1928, provided details of their portfolios in their annual reports and accounts. The years sampled are at 4–6-year intervals. Figure 1 shows our sample, which rises from five ITs in 1886 to 33 ITs in 1928, and compares with the total population of English and Scottish ITs. All but one of Scottish ITs and half of English ITS did not provide portfolio details. Scottish trusts did, though, provide summaries of their asset allocation strategies, rather than full details, in the Chairman's Report (Williams, 1928, p. 10). However, comparing corporate variables, such as size, leverage, number of directors, and performance, provides no evidence that our sample of ITs which disclosed portfolio details is statistically different from those that did not in 1914 (Sotiropoulos et al., 2020). As we see in Fig. 1, the structure of our sample is an

---

[12]Glasgow made three exhaustive studies of average investment trusts: in England (1930), in Scotland (1932), and in both countries (1935). He excluded from his analysis investment trusts which had activities other than managing an investment portfolio.
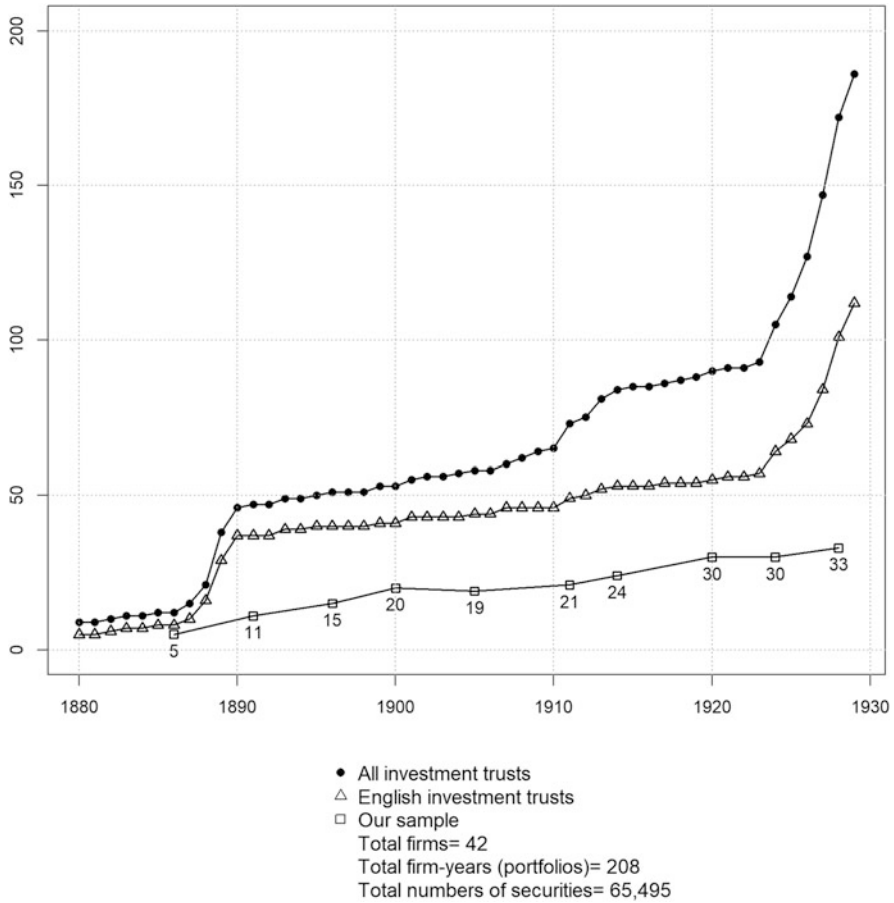
Fig. 1 Size of our sample of investment trusts with disclosed portfolio holdings (*source*: our dataset)

unbalanced panel with 42 different ITs. This gives us 208 portfolios over time to study, which include a total of 65,495 portfolio holdings.

The ITs in our sample all included a list of portfolio holdings attached to their annual report and accounts and filed in the Guildhall Library in London. Some ITs did not consistently report their portfolio holdings in every sampling year. For those portfolios that were provided, details given included the full description of each security held accompanied by the total investment in the security at nominal value. From the description of each security, we were able to identify the security type (ordinary, preference, or fixed interest), its geographical origin, and its sector. For example, from the description "Buenos Ayres and Pacific Railway 7% Debenture Stock," we can assume that this was an Argentinian fixed-interest railway security.

These data allow us to analyze the asset allocation strategies of ITs, cross-sectionally and over time. We can explore the relative emphasis on geography,

sector, and type of security for these ITs and whether these strategies changed over time. Many ITs had names that indicated their asset allocation preference, such as Share and Debenture Trust, Foreign and Colonial Investment Trust, and Brewery and Commercial Investment Trust. Others were vaguer, such as the Omnium Trust, or included words such as "general" or "international" in the name to allow maximum flexibility. Another way to show flexibility of investment policy was to choose a name which reflected where the money came from rather than where they were invested, such as the Scottish Investment Trust—or both, such as the Scottish American Trust. ITs were vague as to their investment strategy in their prospectuses and imposed few constraints in their articles of association (see Sotiropoulos et al., 2020). Thus, the name of the IT does not appear to have constrained its investment strategy in practice.

Nor did the way in which ITs listed the securities held in their portfolios provide much indication of which asset classes they prioritized. The majority of ITs in our sample that provided portfolio details (just over 50%) chose to list their holdings in alphabetical order. An alphabetical listing allowed the investor to search for a particular security with ease and grouped issues by the same issuer but gave no indication as to how or why particular securities had been chosen nor how they related to other securities in the portfolio. Almost half of the ITs in our 1920 sample, though, categorized holdings first by geography, that is, by country or region, choosing to emphasize the "geographical distribution of risk," and then by sector or security type. A minority of trusts listed securities by security type and hence risk level or, more often, by sector, such as breweries or railroads. Another common approach was to put the dominant holdings first, such as British government bonds acquired during and after WWI, together with the railroads sector. These choices imply that asset allocation strategies of these ITs were put geography first while choice of sector and security type were interlinked. Was this emphasis on geography along the lines of the Lowenfeld geographical distribution of capital—that is, equally distributed across the globe and also in similar risk securities? Or were the IT strategies more active in terms of asset allocation?

7 Investment by Geographical Region, Sector, and Security Type

Figure 2 presents the investment profile (total average and boxplots[13] to capture dispersion per available year) of the ITs in our sample across six regions of the world. The figure highlights a number of key points. First, there are clear preferences for regions which persist (with the exception of North America in the 1920s); second, there are major changes in asset allocation over the period 1886–1928;

---

[13] A boxplot displays the variation in a sample, depicting data via their quartiles. The spacings between the different parts of the box illustrate the degree of dispersion and skewness in the data.
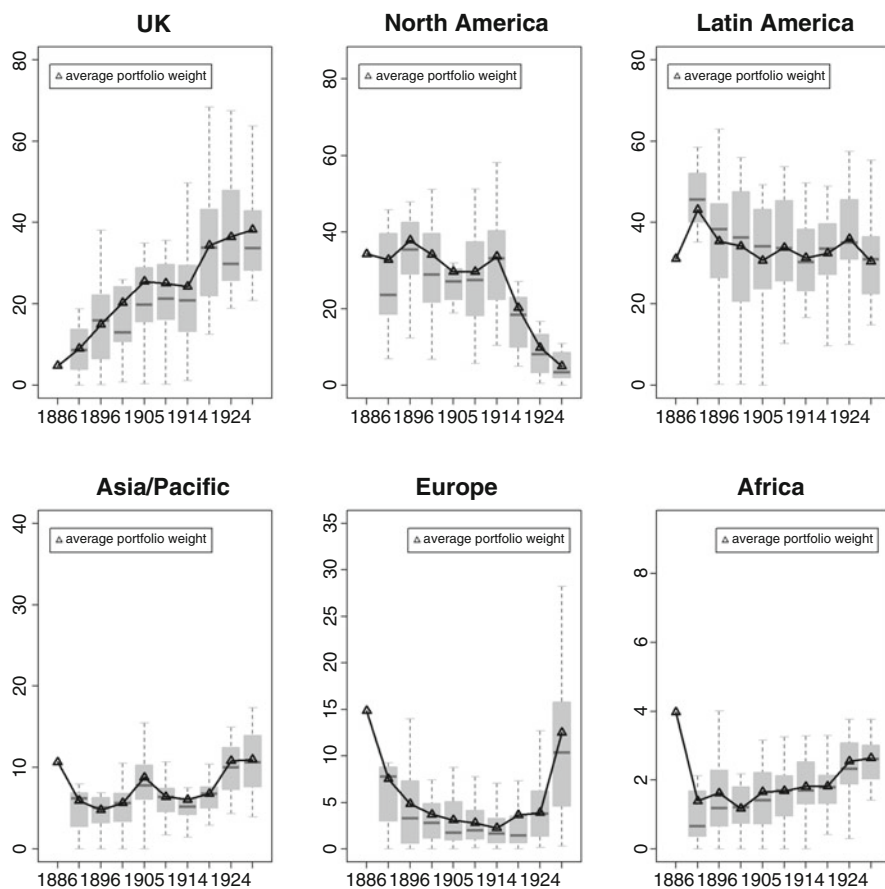
Fig. 2 Geographical distribution of investment trust portfolios in our sample (percent of portfolio nominal value) (*source*: our dataset). *Notes*: Our calculations are based on the reported nominal portfolio values. The boxplot for the year 1886 is omitted because there are only five investment trust portfolios in our dataset. Russia has been included in the Asia/Pacific region. North America includes the United States and Canada and Latin America the rest of the American continent

and third, there is significant variation between different ITs in their asset allocation strategies, as shown by the spread of the boxplots. Before WWI, North America and Latin America were the two preferred regions. Europe was of limited interest except at the beginning and end of the period. The Asia/Pacific and Africa regions are steady over time, at 10%–15% combined. The major changes over time are the rise in UK securities and the decline in North American holdings, both after WWI. Holdings of domestic securities rose on average from under 5% in 1886 to 24% by WWI, rising further to 38% by 1928. In contrast, holdings of North American securities fell from 34% at the outbreak of war to below 5% by 1928. Steiner (1929, p. 34) explains how ITs were "forced" during WWI to dispose of their dollar-denominated American holdings to help the British government obtain credit
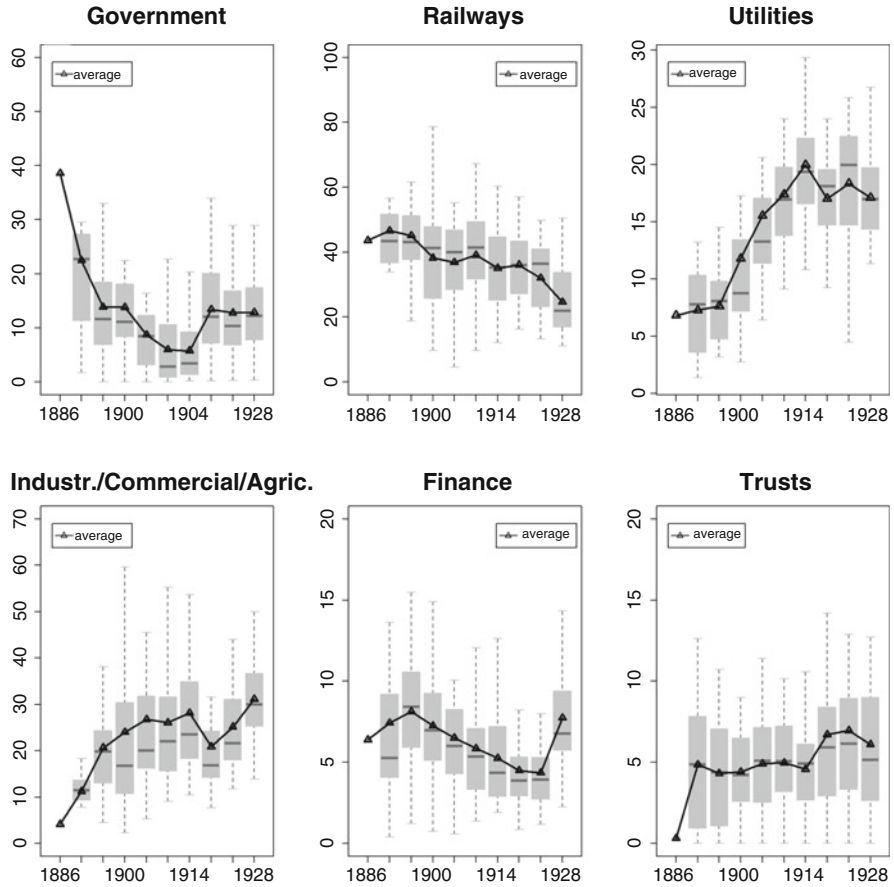
Fig. 3 Sector distribution of investment trust portfolios in our sample (percent of portfolio nominal value) (*source*: our dataset). *Notes*: Our calculations are based on the reported nominal values in the investment trust annual reports. The boxplot for the year 1886 is omitted because there are only five investment trusts in our dataset

in the United States. Within Latin America, the most popular destinations for investment were Argentina, averaging between 22.3% in 1891 and 11.2% in 1928, respectively, of IT portfolio share, second only to the larger and more developed United States.

Figure 3 shows the sector allocation of the IT sample for the period 1886–1928. Government bonds fell out of favor after their 1886 peak of 38%. By 1914, the IT portfolio exposure to government bonds was only 6%. After WWI, the percentage rose to between 10% and 15% for the rest of the 1920s. Railways were the most important sector for ITs up to WWI but showed a steady decline over time, from a high of 45%. By the late 1920s, railways still represented around 20% of total IT portfolios, despite poor performance from the sector in both the United Kingdom
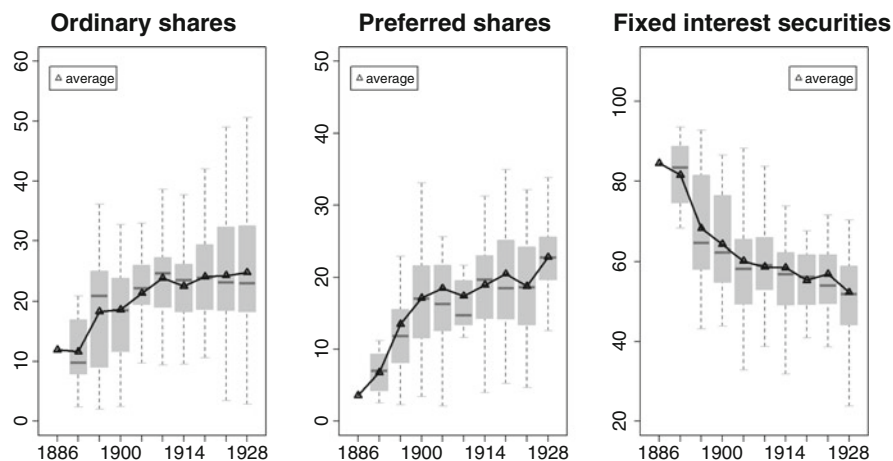
Fig. 4 Security-type distribution of investment trust portfolios in our sample (percent of portfolio nominal value) (*source*: our dataset). *Notes*: Our calculations are based on the reported nominal values in the investment trust annual reports. The boxplot for the year 1886 is omitted because there are only five investment trusts in our dataset

and the United States. There were significant changes in sector allocation over the period, with utilities and industrial/commercial/agriculture sectors growing rapidly in importance. Utilities rose from 7% to peak at 20% of portfolios by WWI, and investment in the industrial, commercial, and agriculture sector rose from under 5% to over 30% by 1928. Financial securities represented a steady 5% throughout, with crossholdings in other ITs also averaging 5%, a low figure considering the common directorships outlined by Sotiropoulos et al. (2019). The boxplots show significant variation between ITs as to holdings in other ITs, but we find no particular preference for ITs with which there were joint directorships (see Sotiropoulos et al., 2020).

Figure 4 shows portfolio asset allocation across different types of securities for our sample of ITs: ordinary shares, preferred shares, and fixed-income securities. The charts confirm the view of Robinson (1930, p. 287) that "from earliest days the British investment trusts have been primarily buyers of bonds, and this is true today, although a growing appreciation of equities is evident." However, from an average of more than 85% of portfolios in 1886, the proportion of fixed-income securities fell rapidly by the 1900s and hovered around 50%–60% thereafter. There was a corresponding increase in corporate preference and ordinary shares over the period, with both types representing around one-quarter each of the average portfolio by the late 1920s.

It is perhaps surprising that ITs had on average almost a quarter of their portfolios invested in ordinary shares, particularly as we know that throughout this period, insurance companies preferred bonds and mortgages. There was a significant increase in holdings of ordinary shares to nearly 25% before 1914, compared with 4% for life offices (Scott, 2002). This was *before* the period of high inflation post WWI and *before* equities were shown to have outperformed bonds in the long term

by Smith (1926) for US common stocks and by Raynes, chief actuary of Legal & General, for the United Kingdom in 1928 (Rutterford, 2009). Post WWI, however, there is significant variation in IT portfolios' holdings of ordinary shares, with some ITs holding more than 50% in ordinary shares and others almost nothing. Scott (2002) reports how some of the smaller insurance companies in the 1920s and 1930s preferred to invest in equities by buying shares in ITs which held them, rather than invest directly.

The choice of type of security was closely linked to the ITs' preferred sectors. For example, the rapid decline in holdings of overseas government bonds before WWI meant fewer fixed-income securities. Similarly, there were falls in holdings of railway fixed-interest securities reflecting sales of US dollar-denominated securities during WWI. Also, one explanation for the rise in preference shares and ordinary shares in portfolios is the switch to industrial, commercial, and agriculture securities which rose from an average of 5% to 30% of portfolio nominal value during the period. Before WWI, ITs were able to buy securities in developing markets such as Argentina for higher yields than were available on equivalent domestic securities. After WWI, they turned to UK government war loans and more "junior stocks and shares bearing no fixed interest and having no foreclosure rights" in order to enhance portfolio yields (Robinson, 1923, p. 20).

Figures 2, 3, and 4 have shown us how, despite names and aims and objectives which argued for a particular asset allocation strategy, individual ITs were not constrained by their prospectuses and articles of association, nor did they feel constrained to either follow such strategies from inception or maintain the same strategy over time (Kilborne, 1925, p. 162). Our analysis has also shown how ITs tended to concentrate on a limited number of regions, albeit with some exposure to less popular markets. Their preference for North and Latin America persisted for many years, but WWI forced some changes to asset allocation.

This is evidence of an active asset allocation strategy, in contrast to Lowenfeld's "naïve diversification" approach to global markets. The way in which portfolios were constructed also differed. ITs adopted a bottom-up approach, choosing preferred regions, preferred sectors, and then a wide variety of securities. Leibson, in 1930 (p. 15), argued that the possibilities for ITs were extensive: he asserted that there were 100 stock exchanges and 200,000 marketable securities to choose from. Figure 4 has shown that ITs held all types of financial security with varying risk levels. Robinson (1923, p. 20) refers to a "lack of uniformity" in holdings of ordinary, preferred, and debenture stock. Lowenfeld's approach, by contrast, was to find approved securities of *similar* risk in *all* regions of the world—with the required level of yield and the maximum capital safety for that yield. Securities which passed the test were then put on an investment list from which to choose (Rolleston, 1909, pp. 12–13).

Size was doubtless a factor in portfolio choice. In 1908, the Investment Registry, which managed portfolios using the Lowenfeld approach, was managing 454 portfolios, with an average value of £11,531 (ibid., p. 34). Ten or twenty individual holdings of £500 or £1000 were appropriate for these portfolios. As we saw in Table 1, ITs were typically around £1 m in size. How did ITs manage much bigger

portfolios? Did they have larger—or a greater number of—holdings? In other words, did ITs manage stock selection?

8   Stock Selection

Figure 5 shows the striking rise in number of holdings and the relatively small average holding value over time. Although there is substantial cross-sectional variation in the number of portfolio holdings of an IT, as shown in Fig. 5, this number was never lower than 68 (which was in the very first year in our sample). The very early trusts had larger and fewer individual holdings: for example, F&C had 18 holdings with an average nominal value of over £50,000 when it first launched in 1868. Its top ten holdings accounted for 80% of the proposed portfolio in nominal terms and 73% of market value. This did not last for long. By 1900, the average nominal value holding of F&C had fallen to under £15,000 (Chambers & Esteves, 2014). For our sample of ITs, as shown in Table 2, the average number of holdings by 1900 was 276 with an average size of £5273. By 1928, the equivalent figures were 383 and £4761.

Three-quarters of IT portfolios in the sample included more than two hundred securities, with some holding over 500. For instance, the Mercantile Investment and General Trust had a portfolio of 571 securities in 1900, and the Industrial and General Trust reached 717 holdings in 1914. The chairman of the International
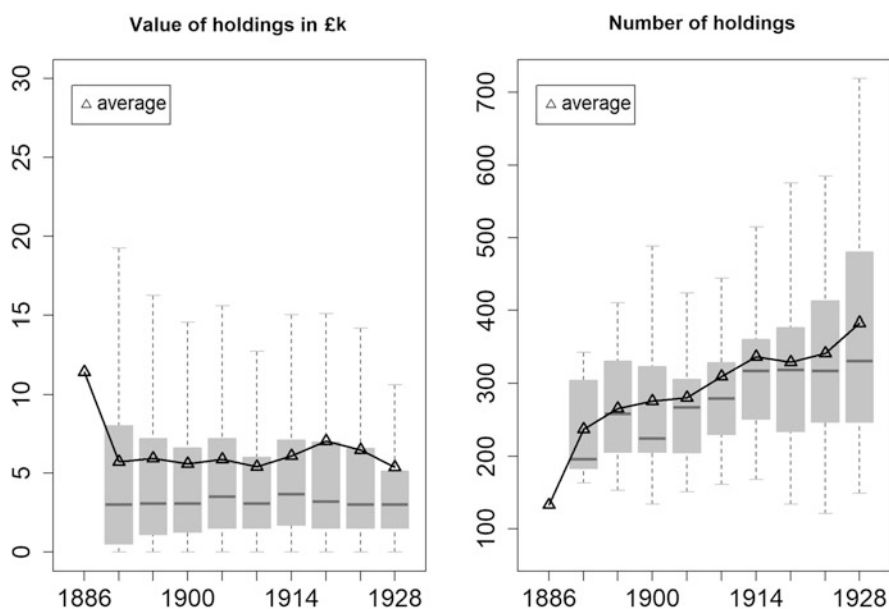
Fig. 5 Number and value of holdings for the investment trusts in our sample (*source*: our dataset)

Table 2 Average portfolio concentration and number and value of portfolio holdings in our sample (*source*: our dataset)

	1886	1891	1896	1900	1905	1911	1914	1920	1924	1928
Number of holdings per portfolio	133	237	266	276	281	310	337	330	342	383
Ordinary shares	33	84	69	76	75	90	90	96	92	117
Preferred shares	7	30	38	55	58	59	70	78	79	95
Fixed-interest securities	93	123	158	145	148	160	177	156	171	171
Value of the individual holdings in £	12,264	6266	5715	5273	5754	5354	5766	7422	6358	4761
Ordinary shares	8547	3360	5017	4387	5260	5292	5109	6557	5634	3982
Preferred shares	32,155	4558	5139	4590	4842	4878	5082	6004	4689	4317
Fixed-interest securities	13,054	8800	6558	6062	6551	5837	6271	8715	7403	5715
Portfolio share (%) of top ten holdings by value	33	22	21	21	20	18	18	23	21	14
Portfolio share (%) of top 10% holdings by value	36	38	36	36	36	35	35	40	37	31
Portfolio share (%) of top 25% holdings by value	61	67	60	60	60	58	58	61	59	53
Portfolio share (%) of top 50% holdings by value	83	90	83	83	83	81	80	83	81	78

Investment Trust, in 1899, commented on how they had worked to increase the number of holdings from 363 to 469 over the past 4 years and that, although this was extra work for the staff, it would remove "the great fault of the trust" whose "large amount [was] not sufficiently scattered" (*Financial Times*, 1899, 7 March, p. 2) On average, portfolios comprised 213 different issuers. Corporate activity such as frequent new issues to fund, say, railway construction, bonus shares, and capital reconstructions increased the number of holdings per issuer.

Part of the trend to more and smaller holdings can be explained by a switch into domestic shares and debentures, away from the more liquid government bonds. Keynes, in his capacity as chairman of the insurance company National Mutual, complained in 1928 of the "narrow market" for domestic equities and how information could only be had on 50 of the top 250 commercial and industrial enterprises. The equity market in the United Kingdom was also skewed, as the ordinary shares of

only two companies, Courtaulds and Imperial Chemical Industries, were worth, in 1928, more than the entire British railway industry (Keynes, 1983, p. 158). Keynes' solution was to buy not one share, but a number of shares in a favored sector, "dividing investment between larger firms in the business even though we do not know much about them individually" (ibid., pp. 158–159). There is some evidence of this scattergun approach in our sample of ITs, where the average holding for ordinary shares fell to under £4000 by 1928 compared with £5700 for fixed-interest holdings. Robinson (1923, pp. 19–20) observed that, even for large trusts, it was best practice to invest no more than £7000 or £8000 in any one issue, with the average investment "a few thousand pounds." This led to very diffuse portfolios, as shown in Table 2. Instead of the original F&C portfolio with 80% of the portfolio in only ten securities, here, investment trusts' top ten securities accounted for 33% in 1886 falling to 14% in 1928. The top 50% (not number) of holdings accounted for around 80% of portfolio value. These highly diversified portfolios were the norm in the IT sector. And yet, very few securities were held in common, despite personal links via directorships and common addresses (Sotiropoulos et al., 2019). In 1914, for our sample of 24 ITs, 75% of securities were held by only one IT. Only 15% were held by one or two ITs, and only 2.5% of holdings were held by more than ten ITs.

Table 3 shows those securities which were held by two-thirds or more of ITs, that is, 16 or more. In 1914, the year sampled, only 12 securities, out of hundreds held, were in more than two-thirds of our sample portfolios.

9   Market Timing

Even though ITs were closed-end funds, IT directors were keen to maintain the liquidity of their portfolios and the marketability of individual securities so that they won't glut the market. It was better, in a depressed market, to try to sell 10 or 20 holdings of $25,000 rather than one at $500,000 (Leibson, 1930, p. 17). As closed-end funds, ITs could remain fully invested, with no need to keep a percentage in cash to meet requests for repayment.[14] After the initial allocation, therefore, new securities could only be bought for the portfolio from the proceeds of sales of existing holdings or from bond redemptions with bank loans used to manage time lags between sales and purchases (Sotiropoulos et al., 2020).

It is clear that ITs were not passive with respect to buying what they considered "cheap" securities and selling those they believed were overpriced. Robert Fleming, at the 1924 AGM of Metropolitan Trust, commented, "we have seized such opportunities as offered to make what we deemed to be judicious exchanges" (*Investors' Monthly Manual*, 1924, March, p. 120). Robinson (1923, p. 21) concurred, "the

---

[14]Shareholders could sell their investment trust securities if they wished to get their money back, although this might be at a premium or discount to the net asset value (Sotiropoulos et al., 2019).

Table 3 Securities held in 1914 by 16 or more of investment trusts in our sample (*source*: our dataset)

Trust portfolios in 1914 (24 trusts in our sample)				
Frequency	Security name	Security type	Country	Sector
19	Central Argentine Railway Company, new shares	Ordinary share	Argentina	Railway
18	Cordoba Central Railway Company, 4.5%. Second debenture stock	Fixed interest	Argentina	Railway
17	Cordoba Light, Power, and Traction Company, 6%. Five-year notes	Fixed interest	Argentina	Utility
17	Cordoba Light, Power, and Traction Company, £1 shares, fully paid	Ordinary share	Argentina	Utility
17	Otis Steel Company, 5% prior lien debenture stock	Fixed interest	United States	Industrial
17	New York Breweries Company, 6% perpetual debentures	Fixed interest	United States	Brewery
16	Buenos Ayres Great Southern Railway Company stock	Ordinary share	Argentina	Railway
16	Seaboard Air Line Railway, 5% adjustment mortgage bonds (1949)	Fixed interest	United States	Railway
16	Hudson and Manhattan Railroad, 5% first mortgage bonds	Fixed interest	United States	Railway
16	Mexico Tramways Company, 5% general consolidated first mortgage gold bonds (1956)	Fixed interest	Mexico	Utility
16	Missouri Pacific Railway Company, extended second mortgage, 5% gold bonds (1938)	Fixed interest	United States	Railway
16	Buenos Ayres Lacroze Tramways, 5% debenture stock	Fixed interest	Argentina	Utility

investment trust is alive to opportunities for profitable purchase and sale." He argued that ITs had special privileges when buying securities:

> many of the best things are offered privately to the investment trusts before they are put upon the market; and [. . .] as regards a great number of securities, the investment trusts are able to get them by underwriting [. . .] at lower prices than the outside investor is obliged to pay (*Investors' Monthly Manual*, 1925, April, p. 171).

Nor were ITs averse to market timing, particularly after WWI when, with Britain off the gold standard, overseas investment became more volatile in what Steiner (1929, p. 4) described as "deranged foreign exchanges." The chairman of the Railway Share Trust and Agency (Company) Limited reported, at the 1920 AGM, that:

> [w]hen the exchange in favour of dollar securities went, as it did last year, to such unprecedented figures, the board had felt it would be unwise to let slip the opportunity of realizing many of the American securities at figures which showed large profits on their book values. [. . .] Such profits could only be regarded as abnormal (*The Times*, 1920, 28 February, p. 23).

The profits on sale of securities during that year, at £61,091, were compared to a gross profit from investment income of £75,811. The capital gains were placed to reserves. The directors' aim was to have sufficient reserves to be able to pay the desired dividends regardless of market movements. This meant constant scrutiny of the portfolio. Those securities which had accrued capital gains were sold to bolster reserves and were replaced by securities deemed to be underpriced on issue or undervalued or in a temporarily depressed market. The preference was for securities which were "marketable, seasoned and of definite merit" and "with values greater than prices" (Williams, 1928, pp. 4, 28; Robinson, 1923, p. 21). The need to generate capital gains to boost reserves meant that ITs were looking—in modern terminology—to buy at the low and sell at the high. They were not passive investment managers.

10 Conclusions

This chapter has discussed portfolio management strategies as practiced by investors in the late nineteenth and early twentieth centuries, with a special emphasis on the role of investment trust companies (ITs). We find ample evidence of a sophisticated approach to asset allocation by ITs, which built a wide variety of portfolios around different types of securities from different regions and sectors, with the aim of enhancing yields as well as bolstering reserves. These trusts were not passive "buy and hold" investors and altered their asset allocations as world events impinged on the financial markets, although they retained a preference for overseas investment throughout the period studied, 1886–1928. Individual ITs adopted independent portfolio strategies, with very little overlap of portfolios. ITs also sought to add value by stock selection—involving hundreds of securities per portfolio—and judicious purchases and sales. They were also not afraid to buy equities, when available, well ahead of other institutional investors, such as insurance companies.

The management of ITs gained—and retained—a reputation for being skilled and professional throughout the period, despite a world war and major market movements. The sheer complexity of the securities included in portfolios—some of which were not listed in the United Kingdom—indicates significant asset management skills and knowledge of the market. Indeed, many studies of the investment trust sector before the 1930s made the explicit point that the mere "machinery" of diversification was by no means enough to guarantee successful investment performance. Management skills were equally, if not more, important (see Scratchley, 1875; Parkinson, 1932; Campbell, 1924; Glasgow, 1935, p. 19). Or as Sturgis remarked in 1924, the success of an English investment trust "is entirely and absolutely dependent upon the character of its management" (Sturgis, 1924, p. 171).

This early foray into fund management by UK ITs was deemed a success, but they remained a tiny part of total London Stock Exchange capitalization. It is an as yet unanswered question as to why it took so long for the asset management industry as a whole to emulate the sophisticated strategies first adopted by ITs well before

WWI. A focus on different episodes in the history of investment trusts can help shed more light on the—under-researched—evolution of the asset management industry. This will allow not only economic historians but also professionals in finance and policy makers to draw lessons from how history affects the evolutionary path of modern financial practices.

References

Beeton, S. (1870). *Beeton's guide book to the stock exchange and money market: With hints to investors and the chances to speculation*. Ward, Lock & Tyler.

Campbell, E. M. (1924). Some management problems of investment trusts. *Harvard Business Review, 2*(3), 296–302.

Cassis, Y. (1987). *La city de Londres*. Belin.

Chadwicks' Investment Circular (1870, 28 December). Which investments pay best in the long run?. *Chadwicks' Investment Circular*, pp. 30–31.

Chambers, D., & Esteves, R. (2014). The first global emerging markets investor: Foreign & Colonial Investment Trust 1880–1913. *Explorations in Economic History, 52*(April), 1–21.

Cottrell, P. L. (2012). *Investment banking in England 1856–1881* (Vol. 2). Routledge.

Financial Times (1899, 7 March). International investment trust. *Financial Times*, pp. 2–3.

Flandreau, M. (2013). Sovereign states, bondholders committees and the London Stock Exchange in the Nineteenth Century. *Oxford Review of Economic Policy, 29*(4), 668–696.

Glasgow, G. (1930). *The English investment trust companies*. Eyre & Spottiswoode.

Glasgow, G. (1932). *The Scottish investment trust companies*. Eyre & Spottiswoode.

Glasgow, G. (1935). *Glasgow's guide to investment trust companies*. Eyre & Spottiswoode.

Guildhall Library, MS14235, *Prospectus of Share Investment Trust.*

Investors' Chronicle. (1949). *Practical investment*. Eyre & Spottiswoode.

Investors' Monthly Manual. (1924). Investment trust companies. March, pp. 119–121. Available from: https://som.yale.edu/imm-issues

Investors' Monthly Manual. (1925). Investment trust companies. April, pp. 171–172. Available from: https://som.yale.edu/imm-issues

Keynes, J.M. (1983). *The collected writings of John Maynard Keynes*, vol. 12: *Economic articles and correspondence: Investment and editorial*, ed. by E. Johnson and D. Moggridge. Cambridge University Press for the Royal Economic Society.

Kilborne, R. D. (1925). American investment trusts. *Harvard Business Review, 3*(2), 160–170.

Leibson, I. B. (1930). *Investment trusts—How and why*. Financial Publishing Company.

Lowenfeld, H. (1907). *Investment an exact science*. The Financial Review of Reviews.

Lowenfeld, H. (1909). *All about investment*. The Financial Review of Reviews.

McKendrick, N., & Newlands, J. (1999). *F&C: A history of Foreign & Colonial Investment Trust*. Chappin Kavanagh.

Michie, R. C. (1981). Options, concessions, syndicates and the provision of venture capital, 1880–1913. *Business History, 23*(2), 147–164.

Parkinson, H. (1932). *Scientific investment*. Pitman and Sons.

Platt, D. C. M. (1986). *Britain's investment overseas on the eve of the first world war*. Macmillan.

Powell, E. T. (1915). *The evolution of the money market, 1385–1915*. The Financial News.

Robinson, L. R. (1923). British investment trusts. *Trade Information Bulletin, 88*, 1–32.

Robinson, L. R. (1930). Investment trusts. *The Journal of Business of the University of Chicago, 3*(3), 279–316.

Rolleston, J. F. L. (1909). *Scientific investment in daily practice*. Financial Review of Reviews.

Rutterford, J. (2004). From dividend yield to discounted cash flow: A history of UK and US equity valuation techniques. *Accounting History Review, 14*(2), 115–149.

Rutterford, J. (2009). Learning from one another's mistakes: Investment trusts in the UK and the US, 1868–1940. *Financial History Review, 16*(2), 157–181.

Rutterford, J., & Sotiropoulos, D. P. (2016). Financial diversification before modern portfolio theory: UK financial advice documents in the late nineteenth and the beginning of the twentieth century. *The European Journal of the History of Economic Thought, 23*(6), 919–945.

Rutterford, J., Sotiropoulos, D. P., & van Lieshout, C. (2017). Individual investors and local bias in the UK, 1870–1935. *Economic History Review, 70*(4), 1291–1320.

Scott, P. (2002). Towards the "cult of the equity"? Insurance companies and the interwar capital market. *Economic History Review, 55*(1), 78–104.

Scratchley, A. (1875). *Investment trusts*. Shaw and Sons.

Smith, E. L. (1926). *Investment trust fund. A sequel to common stocks as long-term investments*. Investment Managers.

Sotiropoulos, D. P., & Rutterford, J. (2018). Individual investors and portfolio diversification in late Victorian Britain: How diversified were Victorian financial portfolios? *The Journal of Economic History, 78*(2), 435–471.

Sotiropoulos, D. P., Rutterford, J., & Keber, C. (2020). UK investment trust portfolio strategies before the first world war. *Economic History Review, 73*(3), 785–814.

Sotiropoulos, D.P., Rutterford, J., & van Lieshout, C. (2019). The rise of professional asset management: The UK investment trust network before World War I. *Business History*. https://doi.org/10.1080/00076791.2019.1656197.

Steiner, W. (1929). *Investment trusts: The American experience*. Adelphi.

Sturgis, H. (1924). *Investment: A new profession*. Macmillan.

The Times. (1868, 20 March). Money market and city intelligence. *The Times*, p. 10.

The Times. (1920, 28 February). Railway share trust and agency company. Record income from investments. *The Times*, p. 23.

Turnbull, C. (2018). *200 years of asset allocation*. Aberdeen Standard Investments.

Williams, M. H. (1928). *Investment trusts in America*. Macmillan.

Early Reflections on the Democratization of Organized Markets and Their Regulations

Paolo Paesani and Annalisa Rosselli

1 Introduction

Around the 1850s, organized markets for commodities and securities underwent deep-reaching changes. Innovation in transport (railways, transoceanic shipping) and communication (telegraph, transatlantic cable) allowed goods, people, and information to move ever more rapidly. The markets for wheat, cotton, and other staple commodities became part of multinational networks. This also applied to many classes of securities, issued in ever larger numbers to meet the needs of new industries and foreign states. As trade in commodities and securities grew, so did the importance of stock exchanges at the global level.

In Western Europe and the United States, where formally organized stock exchanges had been in place before 1850, the volume and value of traded securities increased. As Michie (2006, p. 84) recalls:

> In France the amount of securities owned by French investors rose from 9.1 billion francs in 1850 to 86.9 billion in 1899, an almost tenfold increase. [. . .] Similarly, the value of the securities quoted on the London Stock Exchange, the world's largest and most international throughout the period, rose almost six-fold, from £1.2 billion in 1853 to £7 billion in 1903.

The composition of traded securities also changed significantly during this period. In 1853, for example, 70% of securities traded on the London Stock Exchange consisted of government bonds, 18% of railways stocks, and 6% of securities issued by foreign governments. Forty years later, the weight of domestic public debt was 18%, while the shares of railways and foreign securities had risen to 49% and 21%, respectively. A similar change occurred on all the major stock exchanges in Europe and the United States (Michie, 1999, p. 89).

P. Paesani · A. Rosselli (✉)
Department of Economics and Finance, Tor Vergata University of Rome, Rome, Italy
e-mail: paolo.paesani@uniroma2.it; annalisa.rosselli@uniroma2.it

Between the 1850s and the 1890s, then, a complex and integrated network of markets emerged at the national and international level, with London, Paris, and, to a lesser extent, Berlin as the principal nodes in Europe, together with New York for securities and Chicago and Liverpool for staple commodities.[1] As this network grew in size and complexity, so did the importance of registered stock exchange members, acting as intermediaries between outside investors and companies in their capacity as dealers and brokers.

Repeated speculative outbursts punctuated the growth of organized markets, from the Canal and the Railway manias to speculation on securities issued by mining companies and foreign states and the *Krach* of the Paris Bourse in 1882. These outbursts affected members of all social classes and influenced legislation seeking to prevent their occurrence by regulating the activities of the stock exchanges.

In this context, opinions about "market democratization," defined as the possibility for amateur speculators and savers to access the stock exchange freely, differed significantly.[2] The traditional condemnation of greed and folly that had accompanied early speculative episodes in the seventeenth and the eighteenth centuries was tempered by new considerations of both a moral and material nature. Free access to markets was justified by many as part of natural liberty and as a way to promote individual enrichment in the spirit of egalitarianism. Small investors were drawn into the market by the mirage of quick riches but also, by "a Utopian yearning for freedom and equality" as Chancellor puts it (1999, p. 29). "The Stock Exchange is the monument *par excellence* of modern society," Proudhon said (1857, p. 65; our translation), contributing, through the rapid success and enrichment of some members, to demolish the old order of the *Ancien Régime* based on land ownership and inherited wealth. Against this view, critics claimed that free market access exposed small investors to fraudulent manipulation by unscrupulous professionals, aiming to enrich themselves by "fleecing the sheep" rather than by applying their mind to identifying sound investment opportunities.

At the same time, the need to raise hefty amounts of capital to finance railways, large-scale manufacturing activities, and transoceanic enterprises provided arguments of an economic nature in favor of free access to trading on the stock exchange. In fact, large-scale investor participation would enhance liquidity, reduce the cost of raising capital and thus foster innovation and prosperity. Those who held a contrary view recommended alternative arrangements (e.g., investment banks or cooperative credit) to finance capital accumulation on a large scale in a more orderly and cost-efficient way:

> It is this branch of credit that deals with the transformation of savings into new capital, i.e. capitalisation. Capitalisation is the goal, speculation is the means. It remains to be shown that it is an indispensable means (Walras, 1898, p. 404; our translation).

---

[1]The literature on the organization and evolution of these centers during the nineteenth century is very extensive. See, for a good recent overview, Cassis (2010, Chapters 2 and 3).

[2]For a recent definition of market democratization, see Erturk et al. (2007), among others.

Focusing on Britain, France, and Germany, between the 1870s and the 1890s, our research aims to contribute to the literature that explores the role of small investors in financial markets (e.g., Hansen, 2017; Rutterford & Sotiropoulos, 2017), reconstructing and contextualizing early opinions on market democratization. This issue has never ceased to attract attention since the early developments of stock exchanges, especially in the wake of periods of financial distress. Reflecting on the different organization of British and American financial markets in the 1930s, Keynes (1936, pp. 159–161) famously argued about the problems emerging in connection with easy market accessibility that favored traders interested in investing not "for income" but with the intention of obtaining immediate speculative gains. This problem has attracted renewed interest in recent years in connection with the development of new investment facilities (Barber & Odean, 2001) and the repeated occurrence of speculative bubbles driven by inexperienced and overconfident retail investors (Greenwood & Nagel, 2009; Barber & Odean, 2013). Our chapter is a contribution to reconstruction of the developments of the long-standing problem of finding the correct balance between the advantages of free market access and investor protection. We show how central it had been to reflections on speculation as from a very early stage, and we offer—to the best of our knowledge—the first systematic taxonomy of early opinions on market democratization and its normative implications.

In Sect. 2, we identify three main opinions regarding the merits of market democratization, extracting them from the nascent literature taking a scientific approach to speculation and organized markets. There were the optimists, who favored large-scale participation in the belief that the benefits outweighed the costs for both society and individuals, and the pessimists, who thought that the stock exchange provided an expensive way of raising capital and was a form of legalized fraudulent gambling. Between these extremes, others pursued a third way, recognizing the benefits of stock exchanges but recommending reforms to limit access to them to professional traders only.

At a deeper level, we find that these differences in opinion reflect different views about the price formation mechanism on the stock exchange. The pessimistic view, opposed to the spread of speculation, can be considered a legacy in the nineteenth century of an earlier tradition to the disappearance of which the prevalence of classical political economy, with its reliance on natural prices and the spontaneous readjustment of markets, made no small contribution (Sect. 3).

Opinions about the market democratization and its implications for the pricing mechanisms formed in diverse institutional environments. The various stock exchanges coped with the problem of regulating the access of outside investors to the trading floor in their own various ways (Sect. 4). Where barriers proved more effective in limiting access, in particular for financially unsound speculators, the stock exchanges prospered. Where those barriers were less effective, attempts to contain mass speculation led the regulatory authorities to clamp down on the activities of the stock exchanges, stunting their growth. The 1870s and the 1890s, in particular, saw widespread hard thinking about the best way to regulate stock markets in terms of access. Section 5 addresses some relevant contributions on this

issue, exploring the normative implications and different recommendations in the light of our taxonomy of theoretical positions.

2 Early Reflections on Free Access to Organized Markets and Market Democratization

Moral and economic considerations on the merits of free access to organized markets by outside investors developed in connection with early theoretical views on speculation and its impact on society and economic activity. Organized speculation was still a relatively new phenomenon in the mid-nineteenth century. It took time and a considerable intellectual effort by economists to dispel popular doubts about the constructive role of speculation, as we reconstruct elsewhere.[3] As part of this effort, considerable attention was devoted to the distinction between professional speculators and other investors and to analysis of their respective contributions to the efficiency of resource allocation. Three opinions emerged in this context.

The first opinion emerges from the works of different authors, many of them belonging to or drawing inspiration from the Anglo-Saxon context, where the activities of the stock exchanges, including decisions on the number of members, commissions, and penalties for wrongdoing, were a matter of self-regulation (Ellis, 1876; Giffen, 1877; Leroy-Beaulieu, 1888; Gibson, 1889, Emery, 1896; but also Lexis, 1896, quoted by Van Antwerp, 1913, p. 21; Courtois, 1902). The London Stock Exchange, in particular, admitted people who applied for membership and companies and governments willing to quote their securities, based on relatively light control mechanisms, which allowed the exchanges to expand over time.[4] This encouraged an optimistic view of maximum freedom of access to the stock exchange and confidence in the rules that regulated the relationship between outside investors and members of the stock exchange. It was hoped that these rules, together with the reputational mechanisms underlying them, would provide sufficient protection to market participants, without any need for government intervention beyond ordinary judicial guarantees:

> High character and inflexible rectitude among members is necessary to inspire and merit public confidence, and they are also indispensable in the inter-relation of members. It must be remembered that enormous transactions in all Exchanges are entered into in the most hurried manner, frequently closed by a nod or gesticulation, without written contract, witnesses, notarial acknowledgement (Gibson, 1889, p. 32).

The presence of a body of reliable professionals, willing to ensure price continuity and regular transactions, encouraged outside investors, including amateur speculators and savers, to access stock exchanges in ever larger numbers. In turn, as trading

---

[3] See Paesani and Rosselli (2020) for a focus on the Anglo-Saxon literature.

[4] The situation in the United States was similar to that of the United Kingdom, with some significant differences, which Michie (1986) analyses comparing the London and New York stock exchanges.

volumes and brokers' income rose with the number of investors, increasing numbers of professionals were attracted into the market. To the eyes of supporters of market democratization, two major advantages are derived from free market access and rising numbers of investors. Firstly, diversity of opinion increased with the number of market participants, which was regarded as having a stabilizing effect on market prices:

> The participation of the public, however, does increase numbers, and in normal times numbers themselves are a steadying influence in the market. The more buyers and sellers the less likelihood, in the long run, of wide fluctuations. Every movement of price has a more powerful body of opinion to resist. Furthermore, manipulation in a wide and active market is probably more difficult than in a narrow market (Emery, 1896, p. 190).

The second advantage deriving from free market access was the possibility to finance large-scale investment projects by collecting resources from a multitude of small investors. The financing of railways and utilities was a case in point, given the size of the financial flows involved and the complexity of the problems it raised:[5]

> The subdivision of the title of ownership of such great engines of progress as manufacturing, shipping, mining, banking, trading, telephone, telegraph, cable, water, and gas companies, is indispensable to their organization and operation. The fortune of no individual would be ample to set these forces into motion on a gigantic scale (Gibson, 1889, p. 7).

That the large number of outside investors attracted into the market would include amateur speculators,[6] driven by a gambling instinct, was regarded as natural and controllable. The reputational mechanisms, underlying the functioning of organized markets, compelled expert brokers to discern among their clients:

> Evil practices [. . .] are rife among persons of very limited means, who are not in such circumstances as to justify a broker in speculating for them. It is to some extent a matter of surprise that brokers should be found willing to undertake such business. [. . .] But we have been told that much of this gambling business, so far as it is transacted by members of the house, is done by the younger and more necessitous members, who, having little substantial business, are driven to accept business of this kind with its attendant risks (*Report from the Commissioners on the London Stock Exchange*, British Parliamentary Papers, 1878, p. 21).

Moreover, even gambling attitudes did not have lasting consequence on prices:

> It is unfortunately to be feared that certain portions of society and the general business world participate in this gambling, and there is a good deal of other business conducted on the gambling principle, which it is impossible to eradicate. The gambling in no case is possible without credit, and where there is credit, while human nature remains as it is, there will always be undue credit. But the Stock Exchange gambling will not, any more than the

---

[5] For an introduction to these problems, with a specific focus on North American railways in the nineteenth century, see Eichengreen (1995), among many others.

[6] Apart from mass advertising, often of a very dubious nature (see Cogdon, 1883), amateur speculation was encouraged in different ways. In London, these included the practice of allowing small down payments upon subscription (from 20% to 10% of face value), the possibility of trading subscription rights ("scrip") prior to the issuance of a new security, and the widespread use of options and margin trading. For an early description of these methods, see Crump (1875).

legitimate Stock Exchange speculation, materially affect general prices (Giffen, 1877, pp. 42–43).

Fears that market democratization might occasionally lead to instability, as a result either of manipulation through spreading of false information or of herd behavior, were easily allayed. As markets grew in size and accurate information circulated more rapidly through the new means of communication, the risk of manipulation diminished, especially in the case of commodities and the most widely traded securities.

At the same time, a large and expanding vernacular literature aimed to educate investors (Preda, 2009). It made them less susceptible to falling victim to sly speculative maneuvers and appeased public moral concerns by making naive investors responsible for their own ruin.

To this argument, the French literature added considerations in favor of democratization based on the observation that speculation did not respect social hierarchies and open and easily accessible organized markets offered chances of material progress to many people who would have no other opportunities to better their position:

> but this market is now democratized; while in the past it had looked like a monopoly profiting the financial aristocracy, it now embraces all the spheres of human activity and goes down to the people, this source of modern life, that transforms an investor's saving into a state bond or a railway stock, certain that he will find not only an advantageous income, but to see his capital prosper (*la Bourse* 1854, p. 3, quoted in Preda, 2009, p. 152).

Even Proudhon, fiercely critical of French finance as he was, acknowledged the egalitarian potential of the stock exchange, and like him, while attacking the Bourse, socialists from Lassalle down recognized it as a symbol of modern commerce, as the center of the vast industrial system (Emery, 1896, pp. 158–159):

> But soon the third state, in itself uniting all the productive faculties, [. . .] became master of society and the state, and expelled rival castes from their property. Since 1789 the merging of the economic forces has become law, and to a certain extent practice: every citizen has the right to be simultaneously a worker, a capitalist, an entrepreneur, a trader or a commissioner and a speculator, and a certain number are indeed so. However, in this respect the revolution of 1789 is still far from having produced all its consequences; the fusion has hardly begun; and the disturbances which our political state has been experiencing for sixty-five years are the symptoms of this laborious birth (Proudhon, 1854, p. 25; our translation).

Against the optimistic view, several critics spoke out, many of them of a socialist orientation, condemning the stock exchange as a fraudulent apparatus set up by the dominant class of capitalists to despoil amateur speculators and small investors, tricking them to put their saving into "fictitious capital" separate from genuine productive capital:[7]

---

[7] "Fictitious capital" is the expression used by Marx to contrast financial capital with real capital invested in physical means of production (on this, see De Brunhoff, 1990; Tomlinson, 1990; Bolbol & Lovewell, 2001; Haiven, 2014, pp. 27–35, among others).

> Speculation thus understood is nothing more than the art, always fortuitous, however, of getting rich without labour, capital, trade and genius; the secret of appropriating public or private wealth without giving any equivalent in exchange: it is the canker of production, the plague of corporations and states (Proudhon, 1857, p. 47; our translation).

Critics expressed decidedly pessimistic views about the possibility for small investors to profit from participation in the activities of the stock exchange. To their eyes, major information asymmetries allowed insiders (professional dealers, managers of newly floated companies) to exploit outsiders (amateur speculators acting on secondhand information, small investors). According to this view, the stock exchange appears as a monopoly where professionals can manipulate information to their advantage and the detriment of amateurs resorting to various stratagems including false prospects and misleading advertising published by an unprincipled press.

Privileged access to government officials, including members of parliament and of the aristocracy, increased the power of insiders over small investors. In this context, attempts at self-regulation were at best half-hearted and at worse useless. Nothing short of closure of the stock exchanges could hope to redress the balance between unscrupulous professionals and gullible amateurs.

Between optimists and pessimists, a third view of market democratization emerged from the scientific literature on stock exchanges toward the end of the nineteenth century. Advocates of this middle position recognized the positive contribution that open and freely accessible markets offered to resource allocation. In principle, economic theory would suggest that the greater the number of speculators and the wider the diversity of opinion, the greater should the probability be that the mean price is correct, as it is easier to determine the true price through a large number of people than a small number. However, this conclusion only applies to the activities of competent speculators, whose number and influence may be limited, as noted by Richard Ehrenberg (1883, p. 206; our translation), an influential observer of German financial markets:[8]

> Unfortunately, however, this elite class always represents such a minimal fraction of all fund speculation that it can only assert its correct views under certain circumstances. [...] The actual [competent] stock exchange speculation, i.e. the 500 or 1000 or 2000 people who are professionally engaged in speculation on the stock exchange itself, represent a qualitatively excellent but quantitatively only very small part of the total speculation. The latter is recruited from all classes and nationalities and, as already mentioned, exercises only to a very slight extent the kind of sound judgement, which could enable it to influence the price formation of the speculative values in a good, correct sense.

Amateur speculators, Ehrenberg noted, acting on rumors, secondhand news, and unreliable information, are a source of noise, distorting price signals and resource allocation, and there is no hope that they can learn from their mistakes or win against the big players. Professional speculation could also be exposed to bogus ideas, but to a much lesser degree than its amateur counterpart. When a new interest appears, the

---

[8]Ehrenberg (1883) is quoted extensively, together with other German writers, in Emery (1896), which has been credited since its appearance as being "without doubt the most thorough work on speculation written in English" (Ryan, 1902, p. 337).

number of speculators inevitably increases, both inside and outside the stock exchange. This leads to a general deterioration of market conditions, in monetary, moral, and intellectual terms. As the wave of amateur speculation mounts, two phenomena occur, both detrimental to the economy. On the one hand, a certain number of professional speculators, realizing they are powerless to oppose the wave, withdraw from the market, causing a further worsening of the quality of the pricing process. On the other hand, other competent speculators may decide to remain in the market and draw profit by riding the speculative wave rather than opposing it.

Ehrenberg maintained that in both cases, competent speculators abdicate from their vocation, leaving the market adrift. When this occurs, prices move away from their equilibrium levels. In this way, market exuberance leads to speculative bubbles, which feed on ever larger numbers of amateur speculators and outside investors. While the former trade on borrowed funds, using options and futures, the latter use their own capital. When the speculative bubble bursts, both amateur speculators and outside investors go under: the former reap the fruit of their hazard; the latter lose the savings of a lifetime.

Like Ehrenberg, several observers concluded that outside investors ought to be legally excluded from the stock exchange:

> Finally, we will consider the people who lose money on the stock market in margin-trading much more often than in bets, fathers [. . .] who squander their wives' dowries and children's inheritance on stock market games. To tell the truth, we would have liked to see these individuals give up their operations on their own initiative, if not for moral considerations or contempt of gains obtained without work, at least for intelligence and for having understood that playing the stock exchange against *crédit mobiliers* and the bankers who set the market rising and falling is playing a game unknown to you against opponents who know your cards. [. . .] But since, it seems, the moment has yet to come when morals will be able to do without the help of the law, we willingly consent to the closing of the stock exchange doors to those who have nothing to do with it (Walras, 1898, p. 436; our translation).

If this exclusion could not be achieved by law, because "speculation is the most spontaneous, the most incoercible, the most resistant to appropriation and privilege, the most indomitable to power" (Proudhon, 1854, pp. 31–32; our translation), the instruments reserved to pure speculation should be forbidden or significantly restricted. Walras recommended that small investors, intent on speculation, should be forbidden to deal in futures and limit themselves to cash transactions:

> Rather than buying safe securities to keep them indefinitely, a man who has or believes he has enough business knowledge can buy cheaply securities that are not yet known, appreciated and classified, to resell them at a higher price and buy back others, except to lose sometimes where he thought he was winning. But, under these conditions, one must buy and sell for cash, at most buy and sell at the end of the day, but not be deferred. An individual who operates by way of margin trading usurps a role and functions that belong exclusively to the securities dealer. There would therefore be no difficulty in prohibiting such transactions for any person whose status as a banker is not established by a declaration or patent (Walras, 1898, p. 436; our translation).

Options attracted similar mistrust, and indeed, although widely used, they did not enjoy the protection of the law anywhere until well into the twentieth century.

3   Speculation, Market Democratization, and Price Discovery

The pessimistic view against the spread of speculation can be considered a legacy, in the nineteenth century, of an earlier tradition, to the disappearance of which the prevalence of classical political economy contributed to a significant degree. At the dawn of the development of organized financial markets, two main traditions confronted each other: a preclassical tradition which focused on market fairness and a classical one which focused on market efficiency (Grenier, 2011; De Marchi & Morgan, 1994). The tradition which we can call preclassical considered prices as the outcome of a process of bargaining on the market where buyers and sellers confronted each other, bringing into play all their skills and powers of persuasion. The actual prices depended on the path that the bargaining process had followed. As Cantillon ([1755] 2015, p. 56; italics added) argued:

> Let us assume that there are butchers on one side and buyers on the other. The price of meat will be determined after some bargaining. The butcher sets his price by the number of buyers that he sees; the buyers, on their side, offer less if they feel that the butcher will have smaller sales: the price fixed by some is usually followed by others. Some butchers are better at marketing their merchandise; other buyers are more adept at running it down. Although this method of fixing market prices has *no exact or geometrical basis*, because it often depends on the eagerness or readiness of a small number of buyers and sellers, it does not appear that it could be arrived at in any more convenient way.

When the power of the sellers on the market was equal to that of the buyers, their bargaining gave rise to a price that was "just" both in the sense that it guaranteed the reproducibility of the system—the remuneration of all factors of production was perceived as adequate—and insofar as it did not favor one group of economic agents over another. In this context, the legislator's overriding concern was to set all market participants on an equal footing, in the belief that anyone would try to get the better of the other if they had the chance. The preclassical political economy made a very clear distinction between harassing speculation and the normal "good" speculation that arises from the legitimate desire of the merchant to make a profit. However, it was agreed that the profit seeking of "good" speculation could not be dissociated from harassing speculation, which is a natural phenomenon, not an exception. Therefore, the law must try to create real competition between economic agents and prevent "monopolies," i.e., positions of market control, and improve the moral and practical quality of the price formation process. A clear example was the market for wheat and subsistence goods in the *Ancien Régime*. It was regulated not in the sense of establishing a price but in the sense of guaranteeing equal opportunities to buyers and sellers. A plethora of rules aimed at creating competition and information symmetry: dissemination of information, prohibitions to resell within a certain time, prohibitions for certain categories to make purchases, prohibitions to operate outside the market, access reserved to consumers/artisans before the merchants, etc. The aim was clearly normative: only if the wheat price was the outcome of a process of evenhanded bargaining would it be fair and perceived as such by the public.

Similarly, the Dutch stock exchanges accepted forward contracts in commodities that insured producers or manufacturers against unforeseen variations in prices, but it was suspicious of forward contracts in securities because in these transactions information asymmetries prevailed and it was harder to rely on a benchmark (De Marchi & Harrison, 1994).

However, at the end of the eighteenth century, another tradition emerged, that of classical political economy, which shaped nineteenth-century economic thought in this regard. It fostered belief in "natural prices," expressions of a natural order, toward which market prices moved spontaneously in a free-trading environment. Natural prices reflected intrinsic fundamental values. Competition on the market did not *create* but *discovered* them, and the discovery was enhanced by ample market participation. The natural prices of classical political economy did not depend on the process which led to their formation and that economic theory could safely ignore. Free trading itself, allowing factors of production to move from one sector to a more profitable one, was part of a natural order. The concern of the legislator was to remove the obstacles to its action and not, driven by considerations of fairness, to set limits to its working. Once the idea of a natural price was established, it was possible to extend the concept of fundamental value from produced commodities to financial assets.

When the "natural" prices of classical political economy were superseded by the "equilibrium prices" of the marginalist approach, the same conclusions emerged even more forcefully. Prices were defined by the forces of supply and demand and by the rationality of economic agents that lay behind them. Economic theory assumed that these forces would always prevail or the concept of equilibrium would become meaningless. Any noise disturbing this process, including the noise produced by amateur speculators and outside investors, did not matter, if not transitorily, and could in no way alter the pricing mechanism. Middlemen, including professional traders, were part of this mechanism.

Optimists about market democratization drew support from this school of thought and in general from the idea that when prices deviate from their natural level, possibly as a result of market manipulation, compensating forces are set in motion that lead market prices back toward their equilibrium levels. For example, at the end of his analysis of syndicates, rigs, and corners on market prices, Giffen (1877, pp. 59–60; italics added) comments:

> In general, we conclude that the importance often attached to these syndicates is greatly exaggerated. At certain times, when securities all tend to rise, the syndicates and speculators have some power to concentrate the force of the upward current on one or two groups of old or newly-created securities. At other times, when securities all tend to fall, they have a certain power of inducing sales of special securities and so precipitating their collapse. But their power is exercised at great risks to themselves, does not upset *any general laws*, and does not interfere with the general levels of price, which these laws tend to establish at different times.

The legal framework therefore favored experienced speculators, who knew the natural or normal price and hastened the convergence toward it, punishing gamblers who harmed the market. For example, in France, the nineteenth-century penal code

well reflected the liberal economic ideas about the benefits of competition. It introduced article 419 which banned behaviors that aimed at distancing the price from what "would exist in a situation of competition," no matter how difficult it was to determine that price. Therefore, instead of artificially organizing the market to make it more egalitarian as it had done in the *Ancien Régime*, it left the task of providing wealth and justice to the market (Grenier, 2011, p. 61).

4 Market Democratization in Different Institutional Contexts

During the second part of the nineteenth century, stock exchanges, coping with booming trading volumes, found different solutions to the problem of regulating relations between their members and an increasing number of outside investors. Max Weber ([1894] 2000) provides an early comparison of these different arrangements, focusing on the London, New York, Paris, and German stock exchanges.

The London Stock Exchange (LSE) was, according to Weber, a closed club of professional exchange traders, a "monopoly of the rich," where the public could transact only through registered members.[9] Two main filters regulated the access of nonprofessional traders to the floor of the LSE: commissions and reputational mechanisms. Brokers competed over commissions, offering favorable rates, especially to clients with a large volume of business to transact, while providing an effective barrier against small investors. Reputational mechanisms also restrained members from dealing with small savers when their speculative intent was clearly revealed by obvious disproportion between their orders and income.

Apart from class considerations, dealing with financially fragile clients exposed members of the LSE to the risk and costs of default. As Dickens (1888, p. 253) recorded, "Members unable to fulfil their engagements are publicly declared defaulters [. . .]. Defaulters are only eligible for re-admission when they have paid at least one-third of the balance of the loss caused by their failure [. . .]. The names of defaulters are now officially communicated to the daily papers." These mechanisms, based on self-regulation, contributed to maintaining the reputation of the LSE as a "closed club model," capable of excluding outside investors financially and/or morally unfit to participate in its activities.

Against this model, the *Bourse* in Paris was "democratically organized" according to Weber. Everyone could access the trading floor and do business,

---

[9]Their number rose from 1433 in 1870 to 2850 in 1888. They were divided into three categories of approximately the same numerical importance: brokers, dealers (*jobbers*), and clerks to other members. No member could act in the double capacity of broker and dealer, and the list of members was reelected annually by the Committee for General Purposes, which thus retained the power to exclude unworthy members. Membership was acquired through an application procedure that required the endorsement of three current members, who pledged 750 pounds each to discharge any debt that the new member might fail to honor within 2 years from admission.

including workers in overalls. Inside each stock exchange, however, a *Parquet*, formed by *Agents de Change*, authorized by the government, brokered transactions. Those who wanted to do business quickly had to go through them. Unlike London, not the whole stock exchange, but its innermost part was under their control. The number of brokers, with their monopoly of the trade of government bonds and other tradable securities, was fixed at 60 in 1816 and was raised to 70 in 1898 (Riva & White, 2010).

The brokers formed a corporation. They were pure agents, forbidden to trade on their own account. Minimum commissions were fixed by their governing body and maximum commissions by the government (Riva & White, 2010, pp. 4–5). The monopoly of the *Agents de Change* gave rise to the *Coulisse*, a free innovative market, which existed alongside the official stock exchange where commissions were usually half the rate charged on the *Parquet*. Therefore, while the *Parquet* provided security, transparency, and effective settlement-delivery to unsophisticated investors trading on the spot market, the *Coulisse* provided liquidity, immediacy, and opacity to professional investors trading mostly forward. This juxtaposition had important virtues for market participants, since it allowed the exchanges to specialize in different investors and services and made the exchanges complementary to each other (Hautcoeur & Riva, 2012).

In Germany, security markets that had functioned under self-regulation during the first part of the nineteenth century were increasingly subjected to government control and from 1881 to a high turnover tax. Between 1815 and 1848, Frankfurt and Hamburg enjoyed greater importance than Berlin as trading venues due to the dwindling level of the Prussian government debt and the restrictions placed by the Prussian government on the chartering of new corporations. After 1848, "[. . .] Berlin rapidly became the leading German exchange" (Davis et al., 2003, p. 122).[10] Hamburg, however, retained a prominent role as well as a peculiar kind of organization based on the absence of privileged brokers and the possibility for everyone to trade as long as they kept their records in accordance with the city's rules.

The organization of the Berlin Stock Exchange was midway between that of London and Hamburg. Anyone could access the trading floor and do business, as long as they were native Prussians, paid a ticket, and were recommended by reputed members, who took no responsibility. Discipline was rather lax, and traders who had gone bankrupt were easily readmitted to the stock exchange. While anyone could broker transactions, in Berlin, there was a special group of sworn-in brokers (*vereidigte Maklern*), authorized by the state government. Registered brokers enjoyed the privilege of setting the official prices. They were not a homogenous class as in London, for they differed significantly in status and wealth, ranging from agents for big banks to small traders. In 1884, the appointment of brokers came

---

[10]On the development of German stock exchanges during the nineteenth century and their close connection with the banks, see Bersch and Kaminsky (2009) and the literature cited therein. On speculation as a trigger for regulation of capital markets in Germany in the late nineteenth century, see Baltzer (2013).

under regulations prohibiting them from trading on their own account or with other brokers while granting them the possibility of lifetime mandates to enhance their independence (Gehrig & Fohlin, 2006, p. 593). Despite these and other measures, external forces—from bankers and other interested parties, as well as outside investors—continued to exert a significant influence on price setting, often of a destabilizing nature. Large-scale market participation was favored by the low level of trading costs, especially in the Berlin Stock Exchange. This contributed to stoking speculation, triggering in turn a demand for regulation, which would eventually lead to significant restrictions on the activities of German stock exchanges.

5   Dealing with Market Democratization: Praise and Reforms

The financial crises of the 1870s and 1880s with the heavy losses they inflicted on the investing public prompted some hard thinking about the working of the stock exchange, involving policymakers and economists. In this context, optimists about market democratization reiterated their belief that government intervention on the stock exchange was not only harmful for economic growth and inconsistent with principles of natural economic liberty but also unnecessary, as markets were capable of effective self-regulation. Two official enquiries conducted in the United Kingdom in 1875 and 1877, respectively, reinforced this conclusion:

> The existing body of rules and regulations have been formed with much care, and are the result of long experience and the vigilant attention of a body of persons intimately acquainted with the needs and exigencies of the community for whom they have legislated. Any attempt to reduce these rules to the limits of the ordinary law of the land, or to abolish all checks and safeguards not to be found in that law, would in our opinion be detrimental to the honest and the efficient conduct of business (*Report from the Commissioners on the London Stock Exchange*, British Parliamentary Papers, 1878, p. 26).

The commissioners were well aware that brokers faced contrasting incentives in their relations with inexperienced investors. On the one hand, it was in their interest to exclude them from trading if they could not offer adequate guarantees, as the brokers were liable for any loss they might incur. On the other hand, the presence of a substantial mass of eager investors was essential to ensure them an adequate income. One solution to this problem was to charge small investors higher commissions and/or require higher margins. Another way to achieve the same result was to impose a tax on each transaction. However, rather than on these mechanisms, the LSE continued to rely on self-regulation and the disciplinary mechanisms enforced by its members. Robert Giffen, the only economist and only witness not belonging to the stock exchange who gave evidence to the 1877 commission, expressed serious concern about the actual commitment to stop suspicious transactions and dubious business by members of the stock exchange. Nevertheless, he did not propose any substantial change to the existing rules, but rather more consistent and resolute enforcement of them (Giffen, 1877, p. 303).

Recognition of the advantages of self-regulation went together with belief in the importance of financial education enabling the public to assess the quality of new securities, since the stock exchange generally did not possess the administrative resources to conduct investigations of its own:

> Stock Exchanges guard the public, in so far as they are able, in declining to admit to quotations the questionable enterprises of 'shady' promoters, but they do not in any manner thereby indicate any opinion, personal or official, as to the value of such issues, or their real genuineness or soundness. That is entirely beyond their province, and persons buying issues that have been 'listed' should scrutinise the property and investigate the value for themselves (Gibson, 1889, pp. 37-38).

By contrast, in continental Europe, governments often intervened to regulate organized markets with the aim of curbing speculative excesses. Cohn[11] (1868) provides early evidence of the need to create a corporation of professional traders under the auspices of the local and national government. Ehrenberg (1883) echoes his views as he reconstructs the attempts between the seventeenth century and the 1870s to curb speculation through government intervention. Often, anti-speculative measures consisted in the prohibition of practices (e.g., short selling, trading for differences) and instruments (futures, options), identified by the regulatory authorities (and public opinion) as responsible for disruptive activities. Ehrenberg, however, noted that historical experience shows that edicts against different types of speculative practices, to which markets had become accustomed, including short selling, had always been ineffective in spite of severe sanctions. A different, and more effective, approach to curbing speculation consists in combining preemptive action against speculative concerns with measures to exclude from the stock exchange those traders that do not meet the financial, moral, and intellectual requirements, evidencing a positive opinion of the London Stock Exchange.

However, Ehrenberg was wary of exclusive reliance on self-regulation and advocated external intervention by the governmental authorities. As an example, he discussed the measures introduced by the Prussian authorities in the 1850s to regulate, and in certain cases prohibit, the foundation of new investment banks that, following the example of the *Crédit Mobilier* in France, were at the forefront of organized speculation. Regarding examination of the new securities to be admitted to official listing, Ehrenberg recommended a public authority specifically appointed for this purpose.

These proposals reflect a reformist attitude toward market democratization. The most notable result in this direction was the German Exchange Act of 1896, preceded by the appointment, in 1891, of an Imperial Commission, instructed to investigate organized exchanges. The proposals of the German Commission formed the basis of the new legislation, which however not only advocated greater public supervision of traded securities and market rules but also restrained some widely used practices.

---

[11] Gustav Cohn, a reputed economist expert in financial markets, sat on the Imperial Commission on the Stock Exchange together with Gustav Schmoller.

In this way, the German Exchange Act tried to strike a compromise between self-regulation and government control over stock exchange members and new securities (Emery, 1898). This extension of governmental power combined with the prohibition of futures dealings in the securities of all mining and industrial companies and in grain. As a consequence, until 1908 when the law was revised (Emery, 1908), much trading activity in German securities either migrated to London or was internalized within large banks, where such activities were shielded from taxation and external scrutiny (Michie, 2006, p. 97). Thus, the 1896 Act had long-term consequences at both the theoretical and practical level. To the eyes of supporters of stock exchange and market democratization, Germany's experience proved that interfering with stock exchanges was dangerous and definitely not advisable, leaving the Anglo-Saxon model based on open markets and self-regulation as the only viable alternative. In practice, it stunted the growth of Germany's stock exchanges, enhancing the role of banks as the main source of funds for development of German economy. It is superfluous to recount what this meant for the evolution of the financial markets in Europe.

References

Baltzer, M. (2013). Spekulation als Anstoß für Kapitalmarktregulierung in Deutschland im ausgehenden 19. Jahrhundert. *Jahrbuch für Wirtschaftsgeschichte/Economic History Yearbook, 54*(2), 95–110.

Barber, B. M., & Odean, T. (2001). The internet and the investor. *Journal of Economic Perspectives, 15*(1), 41–54.

Barber, B. M., & Odean, T. (2013). The behavior of individual investors. In G. M. Constantinides, M. Harris, & R. M. Stulz (Eds.), *Handbook of the economics of finance* (Vol. 2B, pp. 1533–1569). North Holland.

Bersch, J., & Kaminsky, J. L. (2009). Financial globalization in the 19th century: Germany as a financial center. *George Washington University Working Paper* no. 15634. George Washington University.

Bolbol, A. A., & Lovewell, M. A. (2001). Three views on stock markets and corporate behavior: Tobin, Veblen, and Marx. *Journal of Post Keynesian Economics, 23*(3), 527–543.

British Parliamentary Papers. (1878). *Report from the commissioners on the London stock exchange with minutes of evidence appendix index and analysis*. Reprinted in Irish University Press Series of British Parliamentary papers.

Cantillon, R. ([1755] 2015). *Essay on the nature of trade in general.* Translated, edited, and with an introduction by Antoin E. Murphy. Liberty Fund.

Cassis, Y. (2010). *Capitals of capital*. Cambridge University Press.

Chancellor, E. (1999). *Devil take the hindmost. A history of financial speculation*. Farrar, Straus, Giroux.

Cogdon, C. T. (1883). The adulteration of intelligence. *North American Review, 136*(314), 88–99.

Cohn, G. (1868). *Die Börse und die Spekulation*. Lüderitz.

Courtois, A. (1902). *Traité des operations de Bourse et de change*. 12th edition, revised and updated. Librairie de Garnier Frères.

Crump, A. (1875). *The theory of stock exchange speculation* (4th ed.). Longmans, Green Reader and Dyer.

Davis, L., Neal, L., & White, E. N. (2003). How it all began: The rise of listing requirements on the London, Berlin, Paris, and New York stock exchanges. *The International Journal of Accounting, 38*(2), 117–143.

De Brunhoff, S. (1990). Fictitious capital. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Marxian economics* (pp. 186–187). Palgrave Macmillan.

De Marchi, N., & Morgan, M. S. (1994). Introduction. In N. De Marchi & M. S. Morgan (Eds.), *Higgling. Transactors and their market in the history of economics. Annual supplement to volume 26 of History of Political Economy* (pp. 1–24). Duke University Press.

De Marchi, N., & Harrison, M. S. (1994). Trading "in the wind" and with guile: The troublesome matter of the short selling of shares in seventeenth century Holland. In N. De Marchi & M. Morgan (Eds.), *Higgling. Transactors and their market in the history of economics. Annual supplement to volume 26 of History of Political Economy* (pp. 47–66). Duke University Press.

Dickens, C. (1888). *Dickens's dictionary of London*. Macmillan.

Ehrenberg, R. (1883). *Die Fondsspekulation und Gesetzgebung*. Springer.

Eichengreen, B. (1995). Financing infrastructures in developing countries: Lessons from the railway age. *World Bank Research Observer, 10*(1), 75–91.

Ellis, A. (1876). *The rationale of market fluctuations*. E. Wilson.

Emery, H. C. (1896). *Speculation on the stock and produce exchanges of the United States*. Columbia University.

Emery, H. C. (1898). The results of the German exchange act of 1896. *Political Science Quarterly, 13*(2), 286–320.

Emery, H. C. (1908). Ten years regulation of the stock exchange in Germany. *Yale Review, 17*, 5–23.

Erturk, I., Froud, J., Johal, S., Leaver, A., & Williams, K. (2007). The democratization of finance? Promises, outcomes and conditions. *Review of International Political Economy, 14*(4), 553–575.

Gehrig, T., & Fohlin, C. (2006). Trading costs in early securities markets: The case of the Berlin stock exchange 1880–1910. *Review of Finance, 10*(4), 587–612.

Gibson, G. R. (1889). *The stock exchanges of London, Paris, and New York: A comparison*. G.P. Putnam's and Sons, The Knickerbocker Press.

Giffen, R. (1877). *Stock exchange securities*. G. Bell and Sons.

Greenwood, R., & Nagel, S. (2009). Inexperienced investors and bubbles. *Journal of Financial Economics, 93*(2), 239–258.

Grenier, J.-Y. (2011). Valeur fondamentale et spéculation dans l'économie politique et le droit (fin XVIIe–XIXe siècle). In N. Levratto & A. Stanziani (Eds.), *Le capitalisme au futur antérieur. Crédit et spéculation en France, fin XVIIIe–début XXe siècles* (pp. 19–67). Bruylant.

Haiven, M. (2014). *Cultures of financialization. Fictitious capital in popular culture and everyday life*. Palgrave Macmillan.

Hansen, K. B. (2017). Crowds and speculation. A study of crowd phenomena in the U.S. financial markets 1890 to 1940. *CBS PhD series* no. 26.2017. Copenhagen Business School.

Hautcoeur, P.-C., & Riva, A. (2012). The Paris financial market in the nineteenth century: Complementarities and competition in microstructures. *Economic History Review, 65*(4), 1326–1353.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

Leroy-Beaulieu, P. (1888). *Précis d'Economie politique*. Ch. Delagrave.

Lexis, W. (1896). Speculation. In G. F. von Schonberg (Ed.), *Handbuch der politicher oekonomie*. H. Laupp.

Michie, R. C. (1986). The London and New York stock exchanges, 1850–1914. *Journal of Economic History, 46*(1), 171–187.

Michie, R. C. (1999). *The London stock exchange. A history*. Oxford University Press.

Michie, R. C. (2006). *The global securities markets. A history*. Oxford University Press.

Paesani, P., & Rosselli, A. (2020). How speculation became respectable: Early theories on financial and commodity markets. *The European Journal of the History of Economic Thought*, *28*(2), 273–291.

Preda, A. (2009). Rational investors, informative prices: The emergence of the "science of financial investments" and the random-walk hypothesis. In G. Poitras (Ed.), *Pioneers of financial economics* (Vol. 1, pp. 149–168). Edward Elgar.

Proudhon, P. J. (1854). *Manuel du spéculateur à la Bourse*. Librairie de Garnier Frères.

Proudhon, P. J. (1857). *Manuel du spéculateur à la Bourse* (5th ed.). Librairie de Garnier Frères.

Riva, A., & White, E. W. (2010). Danger on the exchange: How counterparty risk was managed on the Paris bourse in the nineteenth century. *NBER working paper* no. 15634. National Bureau of Economic Research.

Rutterford, J., & Sotiropoulos, D. P. (2017). The rise of the small investor in the United States and United Kingdom, 1895 to 1970. *Enterprise and Society, 18*(3), 485–535.

Ryan, J. A. (1902). The ethics of speculation. *International Journal of Ethics, 12*(3), 335–347.

Tomlinson, J. (1990). Finance capital. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Marxian economics* (pp. 188–192). Palgrave Macmillan.

Van Antwerp, W. C. (1913). *The stock exchange from within*. Doubleday, Page & Co..

Walras, L. ([1880] 1898). La bourse, la spéculation et l'agiotage. *Bibliothèque Universelle et Revue suisse*, *85*(5), March, 452–476 and *85*(6), April, 66–94; as reprinted in *Etudes d'économie politique appliquée. Théorie de la production de la richesse sociale* (pp. 401–445). F. Rouge and F. Fichon.

Weber, M. ([1894] 2000). Stock and commodity exchanges [die Börse]. Translated by Steven Lestition. *Theory and Society*, *29*(3), 305–338.

Keynes as a Trader in Commodity Futures

Maria Cristina Marcuzzo and Eleonora Sanfilippo

> *As regards the principle of investment in Commodities [...] a pretty long experience convinces me that so far as risk is concerned they are much safer than anything else, since of course intrinsic value remains and one is always protected from catastrophic losses. Held with obstinacy they are, in my experience, far the safest form of investment (KP PC/1/5/89, 1938–39).*[1]

1 Introduction

Keynes's theory of normal backwardation in futures markets, presented first in the 1923 issue of *The Manchester Guardian* (in the *Collected Writings of John Maynard Keynes*, henceforth CWK XII) in relation to commodities and later in the *Treatise on Money* (CWK VI), is well known; it illustrates Keynes's views *at that time* on the role and nature of speculative activity in those markets. He explicitly ruled out the possibility that speculators would be able to earn profits by anticipating price movements more accurately than other actors because, in general, speculators cannot forecast the future better than producers, traders, and consumers.

The systematic remuneration of speculators in commodity futures arises from the fact that "for the sake of certainty, the producer, not unnaturally, is prepared to accept a somewhat lower price in advance than what, on the balance of probability, he thinks the price is likely to be when the time comes" (CWK XII, p. 261; CWK VI, p. 128). From the interpretation of futures markets as a form of insurance and hence

---

[1] Keynes's Papers, kept at King's College, Cambridge, are here quoted as KP, followed by the catalogue reference numbers.

M. C. Marcuzzo
Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy
e-mail: cristina.marcuzzo@uniroma1.it

E. Sanfilippo (✉)
Department of Economics and Law, University of Cassino and Southern Lazio, Cassino, Italy
e-mail: e.sanfilippo@unicas.it

from the existence of a positive and systematic risk premium paid in *aggregate* by forward sellers to forward buyers, Keynes infers "that there is a 'backwardation' in the price of a commodity, or in other words that the forward price is below the spot price" (ibid.). However, backwardation is not viewed as a permanent feature of futures markets, but rather a situation which comes about only if prices do not increase by more than the risk premium. The issue is developed in the *Treatise on Money* (CWK VI, p. 128) where we find the "normal conditions" that give rise to backwardation explicitly stated, namely, only when supply and demand are balanced (see Fantacci et al., 2010).

Keynes started his investment activity very early, at least around 1905, and a bit more substantially by 1914, trading on his own behalf and on behalf of other people and institutions, speculating in foreign exchange[2] and sporadically in securities, but he became increasingly involved in commodity markets only after 1920. From 1921 onward, he dealt heavily in cotton and metals (tin, lead, copper, and spelter) and then in cereals (wheat, corn, maize), rubber, sugar, jute, lard, cotton oil, and linseed oil, mainly through futures contracts and also through options, especially in metals. His trading was particularly intense during the whole decade and extended late through the 1930s, although on a smaller scale, shifting the composition of his portfolio toward investment in shares in both the London Stock Exchange and Wall Street (see Cristiano et al., 2018; Marcuzzo & Sanfilippo, 2020; Sanfilippo, 2021).

It is therefore a plausible inference that Keynes's "extremely wide practical acquaintance with commodity markets and their habits" (Keynes to Hawtrey, 6 January 1936, in CWK XIII, pp. 627–628) shaped his ideas about the working of futures markets in particular and financial markets in general. Since there is no comprehensive study of Keynes's trading on commodity markets, this paper aims to improve on the extant literature, offering, on the basis of the available data, a new assessment of Keynes's behavior as speculator in futures commodity markets.

2   Keynes's Early vs Later Views on Speculation

Let us recall here the famous passage on the nature of the speculators in futures markets:

> [...] the speculator in the great organized "futures" markets is [...] not so much a prophet (though it may be a belief in his own gifts of prophecy that tempts him into the business), as a *risk-bearer*. [...] without paying the slightest attention to the prospects of the commodity he deals in or giving a thought to it, he may, one decade with another, earn substantial remuneration *merely* by running risks and allowing the results of one season to average with those of others; just as an insurance company makes profits [...] (CWK XII, pp. 260–261).

---

[2]Parallel to Keynes's speculation in foreign exchanges is his analysis of speculation in futures presented in his *The Forward Market in Foreign Exchanges* (1922), incorporated in the *Tract of Monetary Reform* (CWK IV), where some reference is also made to commodity futures.

In this description, Keynes was not entirely original; as Dardi and Gallegati (1992) argued, the theory of speculation as a counterpart of "hedging" had already been developed in Emery (1896) (Chap. 4, Sec. 4), a book which Marshall appears to have studied carefully. Marshall (1923) developed this theme in Chap. 5 of Book 2 of *Industry and Trade*, apparently taking it up from Emery.[3] Also, Knight (1921) argued that speculators are specialists in *risk bearing*, demanding a risk premium for providing that service to hedgers; Knight's interesting point, which was not taken up by Keynes in his early approach to the matter, is that the individual loss is reduced by virtue of the transfer while the probability of the occurrence of the event remains the same.

Knight explains:

> There is in this respect a fundamental difference between the speculator or promoter and the insurer, which must be kept clearly in view. The insurer knows more about the risk in a particular case—say of a building burning—but the real risk is no less because he assumes it in that particular case. His risk is less only because he assumes a large number. But the transfer of the "risk" of an error in judgment is a very different matter. The "insurer" (entrepreneur, speculator, or promoter) now substitutes his own judgment for the judgment of the man who is getting rid of the uncertainty through transferring it to the specialist. In so far as his knowledge and judgment are better, which they almost certainly will be from the mere fact that he is a specialist, the individual risk is less likely to become a loss, in addition to the gain from grouping (Knight, 1921, pp. 258–259).

However, Keynes's later views, possibly as a result of his experience in commodity markets and, especially in the 1930s, in shares both in Wall Street and in the London Stock Exchange, became more articulated including his well-known criticism of any speculative trading based on the "activity of forecasting the psychology of the market" (CWK VII, p. 158), distinguished from enterprise which is the "activity of forecasting the prospective yield of assets over their whole life." Thus, with his acquaintance with share markets, he developed the belief that speculation was a bet on a "favourable change in the conventional basis of valuation" (CWK VII, p. 159), the conventional basis being the average market opinion as described in the example of the "beauty contest."

This explains why speculation does not promote price stability in those markets. Unlike the efficient market theory, according to which by buying low and selling high speculators push up the low prices and push down the high prices, Keynes came to believe in the possible destabilizing nature of speculation. The destabilizing effects of speculation can be described as a sudden and large increase in open interest positions, unrelated to new information about fundamentals coming to the market: futures prices go up if the increase is in demand (an increase in long positions) and down if the increase is in supply (an increase in short positions). So accumulated net long positions in futures, constituting as they do a bet that prices will rise, actually make spot prices rise. Conversely, accumulated net short positions would make spot prices fall. Producers may base their supply decisions on the

---

[3]Lavington (1913, pp. 40–41) and also Young ([1924]1999, pp. 330–331) could be added to the list of Keynes's contemporaries who held similar views.

futures contract prices, while physical traders might use futures contracts as a reference to price their commodities. Thus, futures markets may dominate spot commodity markets.

In 1938, Keynes pointed out that for four commodities (rubber, cotton, wheat, and lead) "which are representative of raw materials marketed in competitive conditions, the average *annual* price range over the decade before 1938 was 67 per cent. An orderly programme of output, either of raw materials themselves or of their manufactured products is not possible in such conditions" (CWK XXI, p. 459). On the outbreak of war, the need to regulate the commodity markets becomes imperative, to Keynes's way of thinking, but this may also have reflected his changed view on the nature of speculation, which grew out of his experience as a speculator (see Fantacci et al., 2012).

This is another reason why close study of his activity in commodity futures may help to understand how Keynes became critical of the working of futures markets as a mechanism enabling price stability.

3   The Theory of Normal Backwardation

The theory of normal backwardation explains that the risk premium, which speculators demand, is the difference between the expected spot price and the forward price representing compensation for assuming the price risk that is shifted through the derivative contracts from the hedger to the speculator. In general, demands for risk transfer come from individuals both with long exposure (producers) and short exposure (consumers) in the underlying commodity. Wheat producers will be long while millers will be short. If market conditions are such that a short derivatives position prevails, there will be a downward pressure on forward/futures prices. Under the assumption that farmers are generally long the underlying commodity and short the derivatives—which is usually the case in agricultural markets—short hedgers would expect to lose in futures, while long forward purchasers expect to *gain* on average. The important implication of normal backwardation is that "the quoted forward price [...] must fall below the anticipated future spot price [...]" (CWK VI, p. 129).

Keynes explains:

> It is not necessary that there should be an abnormal shortage of supply in order that a backwardation should be established. If supply and demand are balanced, the spot price must exceed the forward price by the amount which the producer is ready to sacrifice in order to "hedge" himself, i.e., to avoid the risk of price fluctuations during his production period. Thus in normal conditions the spot price exceeds the forward price, i.e., there is a backwardation. In other words, the normal supply price on the spot includes remuneration for the risk of price fluctuations during the period of production, whilst the forward price excludes this (CWK VI, p. 128).

However, normal backwardation is only one possible market configuration, and while Keynes assumed that it was generally true in commodity markets, he made it clear that it was not a universal law. As he points out:

> [...] the existence of surplus stocks must cause the forward price to rise *above* the spot price, i.e. to establish, in the language of the market, a "contango"; and this contango must be equal to the cost of the warehouse, depreciation and interest charges of carrying the stocks. But the existence of a contango does not mean that a producer can hedge himself without paying the usual insurance against price changes. On the contrary, the additional element of uncertainty introduced by the existence of stocks and the additional supply of risk-bearing which they require mean that he must pay more than usual. In other words, the quoted forward price, though above the present spot price, must fall below the anticipated future spot price by at least the amount of the normal backwardation; and the present spot price, since it is lower than the quoted forward price, must be much lower than the anticipated future spot price. (CWK VI, p. 129).

In a market contango, we have a positive cost of carry (due to the cost of the warehouse, depreciation, and interest charges), indicating an expected increase in the spot price over the excess of physical and capital storage costs over the convenience and liquidity yields. But on top of that, the spot price is expected to rise still further to provide an additional compensation to the long speculators for accepting risk transfer.

In fact, in the *General Theory* it is clearly stated that "the relation between the 'spot' and 'future' contracts, as quoted in the market is notoriously different for different commodities" (CWK VII, p. 223).

Not unsurprisingly, therefore, there is little empirical evidence of normal backwardation, i.e., a price bias in futures; while Houthakker (1957a, b) and Cootner (1960a, b) supported the theory, their evidence was challenged by Telser (1958, 1960), Gray (1961), Gray and Rutledge (1971), Dusak (1973), and Kolb (1992), among others. Behind the normal backwardation, there are assumptions which may not hold: hedgers are not always and as a rule long the physical asset on average; they may be purchasers of a commodity on the spot market and hedge its price risk with a long derivatives position, and it is also possible that producers with long price exposure may not fully hedge, so that the impact on the future prices depends on the market participant composition.

Moreover, the existence of a risk premium presupposes that hedgers "as a group" are opposite to speculators "as a group." One can easily imagine a market in which long hedgers deal with short hedgers and long speculators deal with short speculators; classifying hedgers and speculators as opposing groups also ignores the fact that, even in the context of a single transaction, there can be both "speculative" and "hedge" components.

Finally, as argued by Culp (2004), the theory of normal backwardation can be viewed as a special case of the hedging pressure theory, to the effect that speculators are counterparties to hedgers and demand a risk premium. It is worth quoting the relevant passage in full:

> [...] this theory does not assume which side of the market hedgers will be on. Hedgers are not presumed a priori to have a long natural price exposure nor to hedge that exposure by

going short all the time. Instead, the *net* position of hedgers as a group may be long *or* short, and the speculative risk premium arises in response to that net hedging demand [...]. Whether forward prices are upward or downward biased depends on the asset market and the underlying composition of hedgers at any given time. [. . .] [So the theory of backwardation is] a special case in which hedgers are net short forwards/futures. But in the event that hedgers are net long, upward pressure on the forward price leads to upward biased forward prices and a positive expected payoff to *short* speculators [. . .] (Culp, 2004, pp. 285–286).

The main point is that since speculators are a mixed bunch, some of them may have superior knowledge and information, but having superior knowledge and information does not necessarily mean superior forecast skills. So, there is no inconsistency between the view of speculators as risk-bearing individuals and speculators with varying degrees of knowledge and information. In fact, the most important aspect which distinguishes the vulgate version of the normal backwardation theory from Keynes's approach (in its fully developed form) is the analysis of speculation allowing for individual differences in expectations, beliefs, and information evaluation. The set of conjectures upon which decisions are taken are subjective; they are related to individual varying degrees of confidence in those beliefs which in turn reflect how individuals experience reality. This is the approach which derives from Keynes's theory of probability. The variability in the confidence in expectations, both among individuals and within each individual, has an impact on how they are transmitted to markets, which in turn may convey signals to individuals to revise these expectations.[4] A "beauty contest" situation may then arise.

Coming back to the empirical verification of the theory, it is to be noted that the risk premium is paid as a reward and incentive for taking risk above the risk-free rate, not a reward for predicting market movements. It is a conjecture incorporated in the expected spot price, which calls forth a reward, making the future price—which is observable—diverge from the expected spot price, which is unobservable.

What we observe in any given moment is not the price expected to prevail at a future date, but the price which is sufficient to induce speculators to undertake the risk, given their expectations of what the spot price will be on maturity of the contract. The problem with verification of this theory, however, is that extracting this information is difficult in practice. Nevertheless, according to Keynes, the risk premium is "the best estimate we can make of probabilities" (CWK VII, p. 240). The concept is reiterated, even more explicitly, in a letter to Hugh Townshend, dated 7 December 1938:

> I am rather inclined to associate risk premium with probability strictly speaking, and liquidity premium with what in my Treatise on Probability I called "weight". An essential distinction is that a risk premium is expected to be rewarded on the average by an increased return at the end of the period. A liquidity premium, on the other hand, is not even expected to be so rewarded. It is a payment, not for the expectation of increased tangible income at the end of the period, but for an increased sense of comfort and confidence during the period (CWK XXIX, pp. 293–294).

---

[4]For discussion of the notion and nature of expectations in Keynes, see Marcuzzo (2020).

What we can observe, if the data are available, is the buying and selling of an individual speculator on the basis of which we infer his evaluation of the evidence and therefore the expectations guiding his decision-making in the futures market.

This is what we will attempt to do in this paper. We will not address the question whether Keynes as trader in commodity futures followed the standard backwardation theory implication—being systematically long—because in relation to an *individual* speculator it would be clearly nonsensical. We will, rather, try to reconstruct his behavior, to provide not a measure of success (performance approach) but a description of his decision-making process, as a basis to follow the development of his ideas in conjunction with his practice as investor in commodity futures.

4   Keynes as Investor in Commodity Futures Markets: An Overview

The relevant information useful to reconstruct Keynes's trading in the commodity futures markets is mainly contained in two different types of sources in his personal archives:[5] (i) his manuscript ledgers (KP SE/11/2/4–60), registering his dealings for the period 1921–1935, and (ii) the weekly statements of the Tilton Company (KP TC/4 and TC/5)—the company created by Keynes in 1926 for fiscal purposes—registering his open positions[6] in the various commodities he traded for the period 1926–1939. In the 1940s, the commodity markets experienced some restrictions due to the outbreak of the war, and Keynes abandoned his trading while continuing to be actively involved in the stock markets, both in the United Kingdom and the United States.

The ledgers register the commodity bought or sold; the date of the buying and selling operations; the type of contract, whether spot (very rare), future, or option (in the latter case also the type of option); the price (spot, future, or strike, accordingly); the date of delivery; and the total value of the contract. Other subsidiary, but important, sources of information lie in the correspondence with his main broker (Buckmaster & Moore) and various exchanges with other businessmen or experts whom Keynes approached for advice and general information on specific commodity markets or circumstances.

---

[5] Data on Keynes's personal investments are scattered throughout various different files, very often contained in manuscript documents of a different kind. Moreover, they are not always recorded in a regular manner.

[6] In particular, the cost and the market value of the different contracts; the book profits and losses by week; and, occasionally, a summary of the realized profits and losses. These weekly statements of accounts of the Tilton Company, since the exact date and price of the transactions are not indicated, cannot be readily used as records of individual transactions. This is why the data we have reconstructed for the 1930s are to be considered less reliable. On the other hand, the Tilton Company statements represent the unique source we have found for Keynes's activity on commodity futures from 1936 to 1939.

Table 1 Overview of Keynes's trading in commodities, 1921–1939 (*source*: our elaboration from Keynes's Papers SE/11/2/4–60, TC/4/2–3, and TC/5/2–3)

Commodity	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931	1932	1933	1934	1935	1936	1937	1938	1939
Cotton	■	■	■	■	■	■	■	■	■			■	■	■				■	
Wheat				■	■	■			■		■	■	■		■	■	■	■	■
Tin	■	■	■	■	■	■	■	■	■			■	■	■			■	■	■
Spelter		■	■	■	■	■	■	■					■						
Lead		■	■	■			■	■											
Copper	■	■	■	■					■								■		
Cotton oil															■	■	■	■	■
Linseed oil				■	■											■	■		
Lard																■	■	■	
Corn				■	■			■							■	■			
Maize								■	■				■	■	■	■			
Rubber				■	■	■	■	■	■	■									
Sugar		■		■	■	■		■											
Jute				■	■														

Keynes drew up seven *Special Memoranda on Stocks of Staple Commodities* between 1923 and 1930 (CWK XII, pp. 267–647), which show how well acquainted he was with the relevant statistics of these markets and are also a useful source of information.

As shown in the Table 1, the years in which his investments were more diversified by commodity were 1924 and 1925, when he traded ten different commodities. Of the 14 commodities traded by Keynes, only nine overlap with those that he analyzed in the *Memoranda* (cotton, wheat, tin, copper, spelter, rubber, jute, lead, and sugar), while the other five commodities (cotton oil, linseed oil, lard, corn, and maize) in which he invested are not discussed in the *Memoranda*. There, we find data on other five commodities (wool, coffee, tea, crude oil, and nitrate) in which he did not invest.

Although Keynes, as we have seen, traded in different commodities over the period 1921–1939, cotton, tin, and wheat appear, by far, the most important in terms of proportion of value out of the total purchases, as shown in Figs. 1 and 2.

More specifically, cotton loomed particularly large during the 1920s, wheat during the 1930s, and tin in both periods. Comparing the two periods, we can also observe that (i) the value of Keynes's purchases in the 1920s was about three times greater than in the 1930s; (ii) during the 1930s, there was a substantial increase in his involvement in wheat futures, with a parallel decrease in cotton futures and a shift from metals to cotton oil, lard, maize, and linseed oil; and, (iii) finally, during the 1920s, his portfolio in commodities was made up of only two commodities (tin and

Total £ 2,362,420
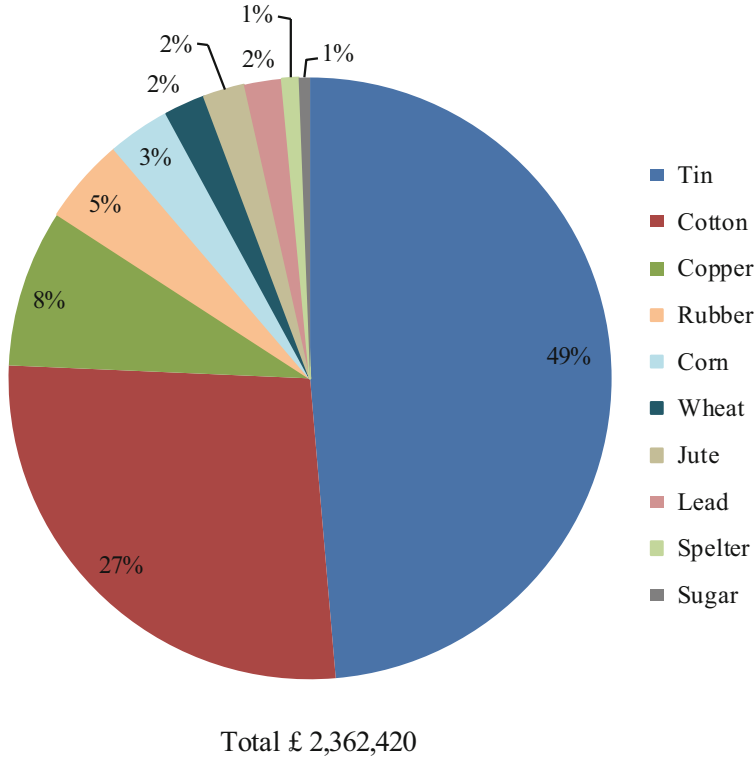
Fig. 1 Distribution by commodity as percentage of the total purchases in value (£) in the 1920s (*source*: our elaboration from Keynes's Papers SE/11/2/4–55)

cotton) which constituted the 75% of his investment, while in the 1930s, except for wheat futures which accounted for about 50% of the total value, his purchases were more evenly distributed among all the other commodities.

As shown again in Table 2, cotton, tin, and wheat took a prominent position in Keynes's investment activity in commodity futures also in terms of length of period of investment, number of contracts, and variety and sophistication of financial instruments used, which is why, in the following subsections, we focus on Keynes's dealings and investment behavior in these three commodities.

Total £ 846,709
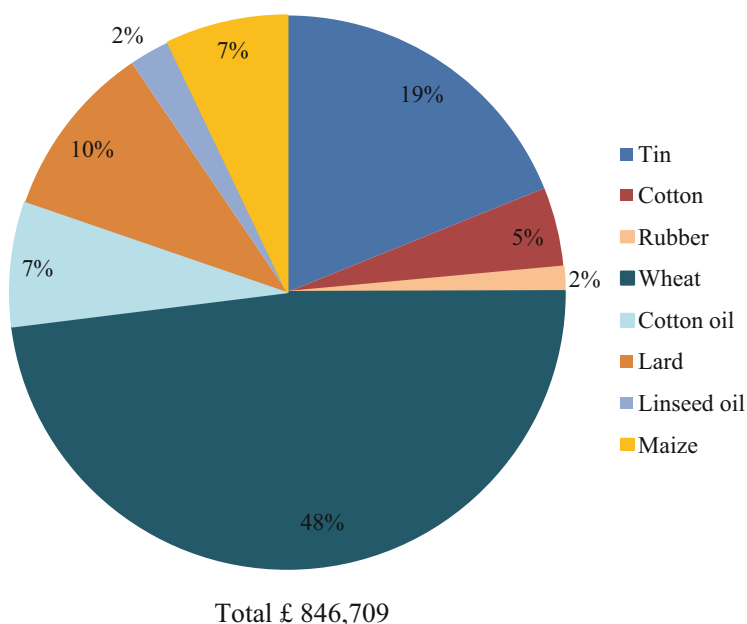
Fig. 2 Distribution by commodity as percentage of the total purchases in value (£) in the 1930s (*source*: our elaboration from Keynes's Papers SE/11/2/56–60, TC/4/2–3, and TC/5/2–3)

Table 2 Overview of Keynes's investments in cotton, tin, and wheat (*source*: our elaboration from Keynes's Papers SE/11/2/4–60, TC/4/2–3, and TC/5/2–3)

Commodity	Periods of investment	Contracts (no.)	Type of contract	Total purchases (value, £)
Cotton	1921–1929 1932–1934 1938	179	178 futures 1 put	676,328
Tin	1921–1933 1936 1938–1939	294	168 futures 70 call 15 BOD 20 put 14 double 7 SOD	1,318,459
Wheat	1924–1926 1929 1931–1933 1935–1939	89	89 futures	458,085

The purchases of the wheat futures contracts denominated in dollars traded in the North American markets and of the only contract denominated in pesos traded in Buenos Aires in 1929 have been converted into sterling using the relevant annual average exchange rate (Board of Governors of the Federal Reserve System, 1943).

Cotton[7]

Keynes started his investments in cotton (and also in tin, as we shall see) a couple of years before the publication of his 1923 article. This circumstance makes analysis of his behavior in this market of a particular interest, since his practical acquaintance with this specific futures market (together with his dealings in tin futures) could have been a real source of inspiration for his theory of futures markets.

At the time of Keynes's investment in cotton futures (mainly from 1921 to 1934, only with a break in 1930–1931, and some trading in 1938) the United States was the most important producer of cotton and Britain one of the leading markets for this commodity. Thanks to the characteristics of this trade and the establishment of a common grading system, as from the end of the nineteenth century, well-organized futures markets developed around the world. The most important exchanges for cotton futures were New York (established in 1870), New Orleans (established in 1880), Liverpool, and Le Havre (both established in 1882).While the crucial months for the cotton trade were July (when some more detailed information on the old crop started to be available) and October (when the same happened for the new crop), nevertheless, as far as the cotton futures contracts are concerned, they could be bought (or sold) at any date, with maturities (and potential delivery) in any of the following 11 months (Cristiano & Naldi, 2014).

Keynes mainly traded in American cotton during the 1920s on the Liverpool market but also in Egyptian Uppers, a cotton of a particularly high quality; occasionally, he also traded American cotton on the New York market during the 1930s. In this commodity, Keynes essentially traded through futures contracts, although the Liverpool market officially allowed options. Only in one case did he deal put options in cotton. Interestingly enough, while his broker Buckmaster & Moore refused to assist him in this type of highly speculative operations in the case of cotton (KP SE/2/1/143), in metals in the same years, Keynes heavily invested through options. Options on cotton, on the other hand, were expressly not admitted at the New York Cotton Exchange.

When Keynes entered the Liverpool cotton market on February 1921 being long, there was a situation of contango, as a consequence of the exceptionally high crop of 1920, which favored the accumulation of stocks. He traded cotton continuatively from the beginning of 1921 to fall 1929, even during the Big Crash, before abandoning the Liverpool cotton market in 1930. His investment behavior was characterized by a net prevalence of long positions (in situations both of contango and backwardation) but with some marked differences between the strategies adopted in the first 2 years of investment activity and the pattern he followed afterward (Cristiano & Naldi, 2014). Up to 1923, he traded only American cotton on the Liverpool market trading contracts for only small quantities at a time; except on two occasions when he assumed a short position, he went long, buying contracts

---

[7]For an earlier and detailed analysis of Keynes's investment activity in the cotton market but limited to the Liverpool market from 1921 to 1929, see Cristiano and Naldi (2014).

of long duration (between 6- and 11-month maturities), and in most cases, he closed his positions when maturity was approaching or having sometimes sold and repurchased a contract for the same quantity but for a longer maturity on the same day. Another characteristic that can be observed is that his positions were larger in the July than the October of each year, because in October there were more elements of uncertainty about the future crops (ibid.). In a few cases, nevertheless, he also carried out some very short-term operations of a few weeks' or 1 month's duration, which he closed as soon as prices went in the direction he had expected, allowing him to make some profits. During his first approach to this market, Keynes also ran into some losses when his bullish expectations proved wrong.

After 1923 and, in particular, starting from spring 1924, his strategy became more complex, and no routine behavior can be observed. First of all, the quantity traded increased; in the same year, Keynes started to extend his trading also to Egyptian Uppers cotton, even though there is still a prevalence of long positions, sometimes alternating, as in the period June–December 1924, with short positions of short duration, and in one case with a put option between April and October 1924. In 1925 and 1926, we observe an important reduction of his exposure in long positions, and this appears perfectly consistent with the analysis of the fundamentals made by Keynes himself and underlined in his *Memoranda* for those years since the situation turned from one of scarcity to one of relative plenty (Cristiano & Naldi, 2014). In 1928, he frequently shifted from long to short positions with mixed success. In this period, in conclusion, he followed all kinds of behavior, but when he was able to anticipate inversion in the market trend, he made substantial profits. He made huge profits in particular with the Egyptian Uppers being long, because in that period there was a major increase in the price of this quality of cotton. In 1928, he returned to trading American cotton, and this time, he gained, being short because of a decrease in the price of this quality. Subsequently, he stayed long on both qualities of cotton, and he lost some money up to the end of 1929, when he abandoned the cotton market for a couple of years (see Table 3).

During the 1930s, Keynes's investment in the cotton market was very occasional and not comparable with his dealings in the 1920s. He returned to the cotton market in July 1932, being mainly long, buying 6-month contracts that he closed soon after he had a chance of gaining some profits. Afterward, prices fell and he started to lose some money. In November 1933, he decided to try another market for cotton, the New York market (KP SE/11/2/58), being long. He stayed on the New York Cotton Exchange until May 1934, making some gains which he registered in dollars, since the price of cotton increased. In October of the same year, he abandoned the cotton market and started to trade in cotton oil futures and other commodities such as wheat, spelter, corn, and lard. He returned to the New York Cotton Exchange once again in 1938, trading only small quantities and being mainly long.

Table 3 Keynes's trading in cotton futures by year (*source*: our elaboration from Keynes's Papers SE/11/2/4–60 and TC/4/2)

Year	Contracts traded (no.)	Purchases (value, £)	Long positions (no.)	Short positions (no.)
1921	13	53,680	13	0
1922	30	109,044	30	0
1923	9	25,912	9	0
1924	25	180,654	21	4
1925	10	23,974	10	0
1926	23	58,202	23	0
1927	19	92,734	19	0
1928	16	63,839	12	4
1929	14	27,713	14	0
1930	0	0	0	0
1931	0	0	0	0
1932	4	4566	4	0
1933	3	6190	3	0
1934	3	16,694	3	0
1938	10	13,126	10	0

Tin[8]

In Keynes's times, the leading market for futures and options trading in metals was the London Metal Exchange (LME), which, established in 1877, had by the 1880s grown and developed to be the most important organized world market, in particular for exchanges on nonferrous metals, namely, tin, copper, lead, and spelter. Of metals, and all commodities in general, tin was the one most traded by Keynes. He entered the tin market in September 1921 and carried on trading in it almost uninterruptedly until 1939 (see Table 4). Between January 1922 and July 1929, he was very active in options, alongside large exposures in futures; he then suspended all trading in this market for some months and resumed it in April 1930, continuing until August 1931; he traded options occasionally again between April and June 1932 and between January and April 1933, when he practically left the tin option market. He continued to trade in futures between June 1936 and January 1937 and between April 1938 and April 1939, taking up mainly long positions (KP TC/4/3/124, 138, 216, and 257). When Keynes began trading tin futures and options, futures prices showed a high volatility. In fact, in the 1920s, several "tin pools" were formed to control production and prices, and in May 1925, Keynes himself took a share (SE/2/5/95) in a private pool.

Keynes invested in tin, trying out all the derivatives available to him at the time, and held the highest number of contracts than in any other commodity (see Table 2).

---

[8]This section draws on Marcuzzo and Sanfilippo (2016) and Marcuzzo and Rosselli (2018).

Table 4 Keynes's trading in tin futures and options by year (*source*: our elaboration from Keynes's Papers SE/11/2/4–60, TC/4/2–3, and TC/5/2–3)

Year	Contracts traded (no.)	Purchases (value, £)	Long positions (no.)	Short positions (no.)
1921	3	22,273	2	0
1922	5	26,745	5	0
1923	9	43,352	9	0
1924	34	221,326	33	1
1925	45	355,389	39	5
1926	44	220,809	30	8
1927	73	238,655	67	2
1928	9	24,437	4	4
1929	3	4599	0	3
1930	15	19,808	15	0
1931	12	7025	12	0
1932	5	2766	5	0
1933	4	3070	4	0
1934	0	0	0	0
1935	0	0	0	0
1936	10	55,800	9	0
1937	0	0	0	0
1938	19	61,740	18	0
1939	4	10,665	4	0

In this table, the number of contracts in each year (first column) does not always coincide with the sum of short and long positions for each year (third and fourth columns) for several reasons. First, in 1921 Keynes in one case took delivery (spot), while in 1926, for six contracts, Keynes sold the tin spot. Secondly, in 1925, 1936, and 1938, Keynes was also a writer of double options, while in 1927 and in 1928, he was a buyer of double options. Double options (bought or sold) should be classified as short/long positions at the same time, being in fact used when investors have no definite expectation about future increase or decrease in prices of the underlying commodity; therefore, we did not include them in the computation either of short or long positions. Keynes's dealings in them in some specific years can simply be taken as a signal of a high volatility in prices and/or of a scarce reliability of the information he possessed and therefore of a low "degree of confidence" he attributed to his own expectations about the direction of future changes in prices.

In 1926, he also took delivery of some of his futures and stocked tin in the LME warehouses, thus moving part of his operations onto the spot market. He behaved not only as a buyer of futures, call and double options, but also as buyer of buyer's option to double (BOD[9]) and seller of seller's option to double (SOD), and it is not easy to understand the reasons behind Keynes's choices between these different alternatives.[10] He was also a writer of double options and a writer of put options,

---

[9]The BOD (and SOD) were contracts that implied the purchase (sale) of a given quantity of the metal for future delivery together with the possibility of doubling the quantity to be bought (sold) at the same price.

[10]The price of options did not follow a precise rule, and lacking information, we are unable to compare the alternatives in terms of cost.

respectively, in February–May 1924 and from September 1925 to December 1926. The pattern of high volatility—which was typical of the price of tin—while providing scope for speculative activity exposed him to the risk of heavy losses when the timing of buying and selling did not match the price swings. Unfortunately, this is exactly what happened. So Keynes's activity in tin options was unsuccessful: he bore substantial losses throughout the whole period except for a few occasions in 1922 and 1926.

From April 1930 to August 1931, Keynes traded in both futures and call options, reaching a high exposure in June 1931 (TC/4/2/84–86). He progressively reduced his exposure in tin derivatives in the summer of 1931 and had closed all his positions by September 1931. In the same period, which coincided with the turmoil preceding UK exit from the gold standard, he started to be increasingly involved in investing in the Liverpool wheat market, which he saw as a form of hedging "against anything that might happen to sterling," as we will see. He returned to tin futures and call options in April 1932, being long. He increased his exposure up to July of the same year, when at the same time he started to further diversify his commodity futures trading by also buying cotton and spelter. More or less in the same period, he increased his investments in sterling securities at the LSE and started to regularly invest in dollar securities at Wall Street (see Cristiano et al., 2018; Marcuzzo & Sanfilippo, 2020). In September, he abandoned tin options and further reduced tin futures while at the same time increasing the number of cotton contracts. Afterward, he closed all his tin contracts. He returned to this market in January 1933, again buying futures, call options, and double options, again showing bullish expectations. He then stopped his dealings in tin for a couple of years, during which he invested mainly in corn, maize, spelter, cotton oil, and lard. He returned to tin in June 1936 and again in 1938 and 1939, mainly buying futures contracts.

Keynes's investment activity in tin prompts the question as to why Keynes was so fascinated by this metal, which took up such a large share in his portfolio. It is noteworthy that not a single year went by after 1921 without Keynes investing in some tin-related assets; the answer may be that information on tin was plentiful but of variable quality, unlike cotton and wheat—the other two commodities in which Keynes invested heavily. For the latter, reliable information was plentiful, but subject to considerable uncertainty due to the unpredictability of extra-economic factors (weather, parasites). Keynes described the characteristics of tin as follows: "Tin is a particular commodity in that both production and consumption are exceptionally insensitive to moderate changes of price, with the result that violent price fluctuations ensue whenever the difference between the two has to be absorbed into stock" (CWK XII, p. 377).

Although speculative activity is potentially profitable with price fluctuations, the price swings have to be anticipated correctly by monitoring the level of stocks, and this is what Keynes must have been doing constantly.

Wheat[11]

In the interwar years, wheat had become a world commodity of key importance for most countries, trade in which—on both the spot and futures markets—was subject to a number of legislative interventions and increasing regulation, especially in the 1920s and 1930s. The specific characteristics of this market (world commodity, standardization of quantities and qualities traded, the need for producers to hedge against fluctuations in prices due also to unpredictable atmospheric conditions and changes in the situation of stocks) had since the beginning of the twentieth century favored the establishment and flourishing of well-organized wheat futures markets around the world. The most important trading places for wheat futures contracts at the time of Keynes's investments were Chicago and Winnipeg in North America, Buenos Aires in South America, and Liverpool and also London in the United Kingdom, according to the main production and consumption areas of the underlying commodity. Keynes traded in wheat futures on all these different places, showing a thorough knowledge of the different economic and institutional environments.

He started to trade in the North American markets (Winnipeg and Chicago) in the mid-1920s, about a year and a half after his article on the theory of commodity futures markets had been published. In the first cycle of his investments (from December 1924 to May 1925), the duration of the cycle was quite short and the quantity traded relatively small. He took prevalently short positions, revealing his bearish expectations on both places. Given that from December 1924 to the end of January 1925 futures prices were increasing on both markets, it seems quite evident that his strategy aimed at anticipating reversal in the price trend. From August 1925 to August 1926, Keynes operated exclusively on the Chicago market, taking prominently short positions except in one case (see Table 5). In this second cycle of his investments, the quantity traded remained small, the short-term strategy still characterized his behavior, and the only change to be seen was extension of the length of the investment cycle. The turning points in his positions on both investment cycles occurred in the same period of the year: March and August 1925 on the Winnipeg market and March and July–August 1926 on the Chicago market. It must be remembered that in March, some reliable information on the future crops of the same year began to be available, and from July to August, wheat was tendered in the United States and Canada. Another important element to recall is that at that time, the gold standard system guaranteed fixed parity between dollar and sterling, so the exchange risk for an investor such as Keynes, who in 1924–1926—as we have seen—traded only in contracts denominated in dollars and on a small scale, was not so great.[12]

---

[11] This section partially draws on Foresti and Sanfilippo (2017).

[12] The risk was not completely eliminated because some slight oscillations around parity (the exchange rate that Keynes recorded for the wheat futures contracts he traded in May–August 1926 in the Chicago market was £1 = $4,86, KP TC/4/1/10–41) were usual and could easily affect an investment activity like future trading based on the gain (or losses) deriving from the differentials between the opening and closing prices of a given contract.

Table 5 Keynes's trading in wheat futures by year (*source*: our elaboration from Keynes's Papers SE/11/2/4–60, TC/4/3, and TC/5/2–3)

Year	Contracts traded (no.)	Purchases (value, £)	Long positions (no.)	Short positions (no.)
1924	1	0	0	1
1925	15	29,021	1	14
1926	7	22,229	0	7
1927	0	0	0	0
1928	0	0	0	0
1929	1	415	1	0
1930	0	0	0	0
1931	6	11,341	6	0
1932	3	5939	3	0
1933	2	3071	1	1
1934	0	0	0	0
1935	12	90,038	12	0
1936	18	247,296	18	0
1937	10	37,126	6	4
1938	8	7230	3	5
1939	6	4379	6	0

After this speculative activity in the North American markets, which brought him some profits, Keynes abandoned the wheat futures trading for a few years (see also KP TC/4/2/2 and TC/4/2/57) while he was still very active in other commodity markets such as metals, mainly tin and lead, as well as cotton, rubber, corn, and maize (see KP SE/11/2/14–52 and Table 1). On 17 October 1929, a few days before the Big Crash on the New York Stock Market, Keynes recorded in his ledgers a forward purchase of wheat for February delivery on the Buenos Aires futures market, and he closed the position a few days later, on 30 October (KP SE/11/2/55), losing some money in the process.

After this operation, there is another break in his investment activity in wheat futures from the end of October 1929 to July 1931. This break seems explicable with the turmoil on the financial markets brought about by the Great Depression, which extended throughout 1930.

A few months before Great Britain's abandonment of the gold standard in the middle of a financial crises, Keynes returned to wheat futures, this time trying the Liverpool market on July 1931, opening a series of long positions (see Table 5), up to June 1932, in situations of both backwardation and contango. The Liverpool market was considered the most important for wheat at the time and representative of the world prices.

Keynes clarified the rationale behind his choice in a letter to Case on the 29 July 1929, where he explained that "a purchase in terms of sterling of commodities having a world price may be in conceivable circumstances a hedge against anything that might happen to sterling. This also applies to the question of buying wheat in Liverpool" (KP BM/2/121).

On December 1931, he returned to the Winnipeg market, opening a long position in a situation of contango. From the end of February to June 1932, he closed his Liverpool positions when, approaching the month of delivery, his expectation of an increase in price proved correct and simultaneously opened new positions for later maturities. These operations made it clear that on the Liverpool market, Keynes followed a "roll-over" strategy of long positions and behaved strictly as a "risk-bearer" against the producers-hedgers. On the other hand, on the Winnipeg market, Keynes acted as an investor trying to anticipate inversion in the price trend, exactly as he did in the same market in the 1920s.

After this period of intense activity in the Liverpool and Winnipeg markets, Keynes abandoned the wheat market for a while, until July 1935. In this interval of time, he traded in cotton, spelter, corn, and maize and, in the first part of 1935, also in cotton oil (see KP TC/4/3/99 and Table 1).

On July 1935, Keynes returned to wheat on the Chicago market, making a huge profit, and 1 month later, he also returned to the Liverpool market. In the period January–March 1936, Keynes went on buying forward huge quantities of wheat, testifying his increased financial capacity due also to increased use of loans (see Moggridge, 1983, CWK XII, p. 11, Table 3). It must in fact be remembered that a cover between 25% and 30% was to be deposited to the broker for each given contract. The sequence of deliveries of the contracts he bought for this period once again clearly shows Keynes adopting the roll-over strategy of long positions.

Keynes had closed all his opened positions on the Liverpool market by mid-October 1937, and for a while, he left this marketplace and wheat in general, redirecting his investments in favor of other commodities. He returned to the Liverpool market only at the beginning of June 1939, for few operations of buying forward for small quantities, up to August 1939, when commodity market trading was restricted.

While Keynes was much involved in wheat futures trading on the Liverpool market (from March 1936 to October 1937), he also once again approached the North American markets, where he went short or long, according to his bullish and bearish views of the market prospects, instead of renewing long positions over time. At the end of March 1936, Keynes also tried another wheat futures market—London—and a specific quality standardized wheat, grown in the Canadian region of Manitoba. He took consistently long positions, applying a clear roll-over strategy. At the end of May, he started progressively reducing his positions, and at the beginning of October 1937, he abandoned this market.

5   Conclusions

Keynes relied heavily on information relative to each individual market and commodity, weighing up the quality and reliability of that information through calculation of the relevant data, the advice of experts, and his own assessment of market conditions and of other participants' opinions.

He showed a great versatility and flexibility in his investment behavior in commodity futures. He followed no routine or mechanical rule, except perhaps at the beginning of the 1920s, when he approached these markets, as the example of cotton shows. Yet, from 1923 onward, as far as cotton was concerned, he diversified his dealings over different qualities, tried greater quantities, and did not adopt simple renewal of long positions (roll-over strategy) but frequently alternated long and short positions (especially between 1927 and 1928) according to his expectations, firmly grounded on careful analysis of the fundamentals; as for tin, he extended his activity toward all kind of options, showing a high propensity for risk (very rarely he hedged his positions), a true fascination for this type of highly speculative instruments, and also a thorough knowledge of how they worked; finally, as far as wheat was concerned, from the mid-1920s, he invested in different North American markets (Chicago and Winnipeg)—mainly assuming short positions and following a strategy aiming at anticipating market trends—and at the end of 1929 also in Buenos Aires, as we have seen, dealing through contracts denominated in currencies other than sterling and therefore also facing an exchange rate risk.

His investment strategy changed in the 1930s, when, without losing his faith in commodities, he reduced the scale by two-thirds and almost abandoned the riskier component of his commodity investments (metals options) while at the same time increasing his investments in stock markets. If we take as an example the year 1936—the year of the publication of the *General Theory*—his dollar securities end-of-year portfolio amounted to a market value of about $1,400,000 (almost £300,000), his sterling securities end-of-year portfolio to about £420,000, and the value of his purchases only of two commodity futures (viz., wheat and tin) to about £300,000 (which implied a cover of about £90,000 deposited to the broker).

In those years, Keynes consistently stepped up his investment activity, operating in different markets and marketplaces: he traded in derivatives in several commodities at the same time in the UK markets and in North American markets and occasionally in exchange markets; he also traded on the LSE and Wall Street, dealing in shares of both American and British companies, for which he followed mainly, but not exclusively, a "buy and hold" strategy, although never adopting mechanical investment schemes.

Returning to commodities, his trading is emblematic of his capacity as an investor to diversify his choices among different markets and adapt to different circumstances.

As far as the evolution of Keynes's investment behavior over time is concerned, there is further confirmation of the view—shared in the literature (Chambers & Dimson, 2013; Marcuzzo & Sanfilippo, 2016)—that a change occurred in Keynes's speculative style (both in shares and in commodities) around the beginning of the 1930s, when he abandoned a short-term-type investment behavior in favor of a long-term investor perspective, but this is not confirmed in toto by all his dealings in commodities.

As far as the relation between Keynes's theory and practice in the commodity markets is concerned, analysis of his trading also seems to confirm the fundamental tenet that backwardation should not be considered a permanent feature of the commodity futures markets and, accordingly, the theory of "normal backwardation"

represented only a special case. Indeed, Keynes seems to have been well aware of the broader set of specific circumstances which render each commodity and even—as we have seen in this analysis—each marketplace "a special case."

Moreover, as argued by Cristiano and Marcuzzo (2018, p. 281), Keynes never lost sight of the complexity of factors behind the surface of price changes but progressively and increasingly lost confidence in the ability to predict their course in the short run, and he turned increasingly to the fundamentals of the economy and behind individual assets "to provide a reasonable basis for rational, and in the long run at least, successful choice."

References

Board of Governors of the Federal Reserve System. (1943). *Banking and monetary statistics, 1914–1941*. Federal Reserve Bank of St. Louis. Accessed July 4, 2021, from https://fraser.stlouisfed.org/files/docs/publications/bms/1914-1941/BMS14-41_complete.pdf

Chambers, D., & Dimson, E. (2013). John Maynard Keynes, investment innovator. *Journal of Economic Perspectives, 27*(3), 213–228.

Cootner, P. H. (1960a). Returns to speculators: Telser versus Keynes. *Journal of Political Economy, 68*(1), 396–404.

Cootner, P. H. (1960b). Returns to speculators: Rejoinder. *Journal of Political Economy, 68*(4), 415–418.

Cristiano, C., Marcuzzo, M. C., & Sanfilippo, E. (2018). Taming the great depression. Keynes's personal investment in the US stock market, 1931–1939. *Economia Politica, 35*(1), 13–40.

Cristiano, C., & Marcuzzo, M. C. (2018). John Maynard Keynes: The economist as investor. *Review of Keynesian Economics, 6*(2), 266–281.

Cristiano, C., & Naldi, N. (2014). Keynes's activity on the cotton market and the theory of the "Normal backwardation": 1921–1929. *European Journal for the History of Economic Thought, 21*(6), 1039–1059.

Culp, C. L. (2004). *Risk transfer. Derivatives in theory and practice*. Wiley.

Dardi, M., & Gallegati, M. (1992). Alfred Marshall on speculation. *History of Political Economy, 24*(3), 571–594.

Dusak, K. (1973). Futures trading and investor returns: An investigation of commodity market risk premiums. *Journal of Political Economy, 81*(6), 1387–1406.

Emery, H. C. (1896). *Speculation on the stock and produce exchange of the United States*. Columbia University.

Fantacci, L., Marcuzzo, M. C., & Sanfilippo, E. (2010). Speculation in commodities: Keynes's practical acquaintance with future markets. *Journal for the History of Economic Thought, 32*(3), 397–418.

Fantacci, L., Marcuzzo, M. C., Rosselli, A., & Sanfilippo, E. (2012). Speculation and buffer stocks: The legacy of Keynes and Kahn. *European Journal of the History of Economic Thought, 19*(3), 453–473.

Foresti, T., & Sanfilippo, E. (2017). Keynes's personal investments in the wheat futures markets, 1925–1935. *History of Economic Ideas, 25*(2), 63–90.

Gray, R. W. (1961). The search for a risk premium. *Journal of Political Economy, 69*(3), 250–260.

Gray, R. W., & Rutledge, D. J. S. (1971). The economics of commodity futures markets. *Review of Marketing and Agricultural Economics, 39*(4), 57–108.

Houthakker, H. S. (1957a). Can speculators forecasts prices. *Review of Economics and Statistics, 39*(2), 143–151.

Houthakker, H. S. (1957b). Restatement of the theory of normal backwardation. *Cowles foundation discussion papers* no. 44. Cowles Foundation for Research Economic, Yale University.

Keynes, J. M. (1923). Some aspects of commodity markets. *The Manchester Guardian Commercial, European reconstruction series*, Section 13, 29 March. As reprinted in CWK XII (pp. 255–266).

Keynes, J. M. (1971–1989). *The collected writings of John Maynard Keynes* (CWK). Managing Editors E.A.G. Robinson and D. Moggridge. 30 vols. Macmillan.

    CWK IV   *A Tract on Monetary Reform.*
    CWK VI   *A Treatise on Money. Part 2: The Applied Theory of Money.*
    CWK VII   *The General Theory of Employment, Interest and Money.*
    CWK XII   *Economic Articles and Correspondence, Investment and Editorial.*
    CWK XIII   *The General Theory and After. Part I: Preparation.*
    CWK XXI   *Activities 1931–1939. World Crises and Policies in Britain and America.*
    CWK XXIX   *The General Theory and After. A Supplement.*

Knight, F. H. (1921). *Risk, uncertainty, and profit*. Houghton Mifflin.

Kolb, R. W. (1992). Is normal backwardation normal? *Journal of Futures Markets, 1*(12), 75–91.

Lavington, F. (1913). The social interest of speculation on the stock exchange. *Economic Journal, 23*(1), 36–52.

Marcuzzo, M. C. (2020). Expectations, conjectures and beliefs. The legacy of Marshall, Kahn and Keynes. In A. Arnon, W. Young, & K. Van Der Beek (Eds.), *Expectations: Theory and applications from historical perspectives* (pp. 53–67). Springer.

Marcuzzo, M. C., & Rosselli, A. (2018). Trading in the "Devil's metal": Keynes's speculation and investment in tin (1921–1946). In J. Kregel, M. Corsi, & C. D'Ippoliti (Eds.), *Classical economics today* (pp. 167–188). Anthem Press.

Marcuzzo, M. C., & Sanfilippo, E. (2016). Keynes and the interwar commodity option markets. *Cambridge Journal of Economics, 40*(1), 327–348.

Marcuzzo, M. C., & Sanfilippo, E. (2020). Keynes's personal investments in the London Stock Exchange and his views on the transformation of the British economy. *Structural Change and Economic Dynamics*. Accessed Jun 30, 2021, from https://doi.org/10.1016/j.strueco.2020.02.002.

Marshall, A. (1923). *Industry and trade* (4th ed.). Macmillan.

Moggridge, D. E. (1983). Editorial introduction. In E. A. G. Robinson & D. Moggridge (Eds.), *Economic articles and correspondence, investment and editorial* (pp. 1–113), vol. XII. *The Collected Writings of John Maynard Keynes*. Macmillan.

Sanfilippo, E. (2021). Keynes's trading on Wall Street: Did he follow the same behaviour when investing for himself and for King's? *Financial History Review, 28*(1), 1–25.

Telser, L. G. (1958). Futures trading and the storage of cotton and wheat. *Journal of Political Economy, 66*(3), 233–255.

Telser, L. G. (1960). Returns to speculators. *Journal of Political Economy, 68*(4), 404–415.

Young, A. A. ([1924] 1999). Insurance and speculation. In *The book of popular science*, The Grolier Society; reprinted in P. G. Mehrling & R. J. Sandilands (Eds.), *Money and Growth. Selected papers of Allyn Abbott Young* (pp. 322–336). Routledge.

High-Frequency Trading and the Material Political Economy of Finance

Donald MacKenzie

High-frequency trading or HFT is "proprietary" automated trading that takes place at speeds far faster than an unaided human can trade and in which trading's profitability is inherently dependent on its speed. (The goal of "proprietary" trading is direct trading profit, rather than, e.g., earning fees by executing trades on behalf of others.)[1] HFT is trading *by* machines (trading firms' computer servers and other equipment such as the specialized silicon chips known as field-programmable gate arrays or FPGAs) *on* machines: all modern exchanges are, at their heart, computer systems.[2] The materiality of those machines and how it has changed through time are crucial to HFT. Light and other forms of electromagnetic radiation—which I regard as just as material as silicon chips ("materiality" does not refer only to solid objects)—are also central to HFT's material practices. HFT requires the fastest possible transmission of data and orders to buy or to sell, which nowadays is often wireless transmission, not the use of cables, even fiber-optic cables.

---

[1] This definition of HFT reflects useful discussion of the point with interviewee BU. This chapter draws upon MacKenzie (2018a and 2021). It corrects the former in respect to what my subsequent research suggests is an overly strong view of "making"/"taking" specialization; see MacKenzie (2021) for more detail than possible here, including further references to the literature on HFT.

[2] Other researchers within the "social studies of finance" (see below) who have focused on HFT include Robert Seyfert (2016) and the group around Christian Borch at the Copenhagen Business School, especially Ann-Christina Lange: see, e.g., Borch and Lange (2017) and Lange (2016). Most materialist in approach is Alexandre Laumonier, with his remarkable investigations of the precise paths, especially in Europe, of HFT's microwave links: see Laumonier (2019) and his blog https://sniperinmahwah.wordpress.com/.

D. MacKenzie (✉)
University of Edinburgh, Edinburgh, Scotland, UK
e-mail: DonaldMacKenziePA@ed.ac.uk

1 Einsteinian Materiality

I think of the materiality of HFT as "Einsteinian." A key postulate of Einstein's theory of relativity (the postulate that no signal can travel faster than the speed of light in a vacuum, which is a fixed constant) has, in HFT, become an ever-present constraint. An algorithm trading in one computer datacenter often needs to know what is going on in other datacenters, and this makes speed of data transmission between datacenters crucial. In the early years of HFT, data transmission was generally via laser-generated pulses of light in fiber-optic cables, but light in a fiber-optic cable travels at only two-thirds of its speed in a vacuum: it is slowed by the material of which the strands of the cable are made, which is usually a specialized form of glass.[3] In contrast, a wireless signal sent through the Earth's atmosphere travels at very nearly the speed of light in a vacuum.[4] Because wireless transmission for HFT involves specialized geodesic-hugging links, tailor-made radios, and the use of frequencies that are in high demand, it is much more expensive than the use of fiber-optic cable usually is. (The "geodesic" or "great circle" is the shortest path on the surface of the Earth between two given points.) One interviewee talked of trying to avoid what he called "radio-frequency markets": those in which an HFT firm has no alternative but to use signals transmitted through the atmosphere.

One way of gauging the speed of HFT is the response time of an HFT firm's system: the time delay between the arrival of a "signal" (a pattern of data that informs an algorithm's trading) and an action—the dispatch of an order or a cancellation of an order—in response to that signal. By March 2019, one of my interviewees was telling me that, although his own systems were slower than this, he had learned of the achievement of response times as low as 42 nanoseconds (a nanosecond is a billionth of a second).[5] In a nanosecond, light in a vacuum, or a wireless signal in the atmosphere, can travel no more than around 30 cm or roughly a foot. The need for nanosecond speeds in an Einsteinian world makes HFT exquisitely sensitive to the precise location of technical equipment such as wireless antennas and to how closely the path of a fiber-optic cable or wireless link follows the geodesic between datacenters. The materiality of HFT is, therefore, perhaps above all a *spatial* materiality.

---

[3] The refractive indices of the relevant materials are around 1.5, implying a maximum speed of 1/1.5 or two-thirds of the speed of light in a vacuum.

[4] The refractive index of the atmosphere varies with temperature, pressure, and density of water vapor but is very close to 1.0.

[5] A technical presentation by the futures exchange Eurex in September 2018 reported—on slides kindly passed to me by another of my interviewees—Eurex as having measured responses as fast as 84 nanoseconds. At the time of writing, that is the fastest that has been recorded quasi-publicly.

2 Material *Political* Economy

The academic field in which I work has become known as "social studies of finance," the application to financial markets not of economics (or individualistic "behavioral" finance) but of wider social science disciplines such as anthropology, sociology, and science and technology studies. Much of this research, however, has been criticized—justifiably—as paying insufficient attention to power (corporate and political), financial interests (in the ideological but also the grubby monetary sense), deception, the state, law, and so on: see, e.g., Mirowski and Nik-Khah (2007) and Fligstein and Dauter (2007).

How can we remedy this deficiency while maintaining the ethnographic strength and in-depth insights of the best work in the social studies of finance? It is possible to envisage multiple start points, but one is from "actor-network theory" (ANT). This theoretical perspective (developed by authors such as Bruno Latour, 2005, Michel Callon, 1998, and Annemarie Mol, 2002) highlights the materiality of social life—the role in it of objects, of technologies, of other "nonhuman actors," and of material human bodies—and ANT has underpinned much recent research in "social studies of finance." The contribution to ANT most relevant here is Law and Mol (2008), which discusses "material politics." The idea is in essence simple but is elegantly laid out by Mol and Law. As they argue, it is possible to arrange the material world in different ways, and at least sometimes, the issue of which of these ways becomes real has a political dimension. There are, for example, "roads not taken" in the development of technology, as the historian Ruth Schwartz Cowan (1983) points out, drawing the phrase from the poet Robert Frost: technologies that could have been developed but were not and not necessarily because they were simply less efficient than the successful alternative, but sometimes for reasons that have more to do with class, gender and ethnic divides, state power, and so on.

"Material politics" is a pivotal aspect of HFT. For example, the activity's history is characterized by incumbent-challenger conflicts, in which HFT firms have traditionally been in the role of challenger; some of those conflicts continue today. The corporate interests, political lobbying, and sometimes actions by state agencies that are involved in those struggles are instances of exactly the type of phenomena that the critics of research in the social studies of finance suggest are not given enough salience in that research. Phenomena of this kind are, however, emphasized by the sociological perspective known as "field theory," developed above all by Pierre Bourdieu (in, e.g., Bourdieu, 1997) and elaborated by, for instance, Fligstein and McAdam (2012). A "field" is a specific domain of social and economic life in which differently positioned actors compete to achieve rewards that are often specific to the field. Fields involve informal norms of behavior and sometimes also explicit rules. Because incumbents usually benefit from a field's existing norms and rules, challenges to those incumbents often involve attempts to alter them. I have argued elsewhere that the materiality of actor-network theory needs to be complemented by the field theory emphasis on such issues, despite the fact that the two perspectives

have sometimes clashed bitterly.[6] Challenger-incumbent conflicts in finance have been and are being played out in multiple dimensions, but—as MacKenzie (2021) shows—the material arrangements of trading are certainly one of those dimensions.

"Politics" is, of course, an elastic word. Its broad sense encompasses the full gamut of phenomena that shape and are shaped by actors' privileges, position, and epistemic authority; the status and respect actors enjoy or fail to receive; their economic resources; and so on, and when I write about "material political economy," the second of those three words is used mostly in that broad sense. However, "politics" in the narrower sense of governments, political parties, and members of Congress or, e.g., of the European Parliament and congressional committees has also played a part in the development of HFT, especially via the interaction between the political system and the regulation of finance. That interaction takes a variety of forms, including differences among political parties in their typical attitudes to regulation. In the United States, for instance, Democrats have generally (although by no means always) preferred stricter forms of regulation, and Republicans often have "de-regulatory" impulses. Politicians, in addition, are often lobbied by financial sector interests, and money from the financial sector can form an important part of campaign contributions to politicians (there is an example of this and of its consequences for HFT in MacKenzie, 2018b). Furthermore, structural features of the political system—in the United States, as we will see below, the separate committees of the Senate—can also shape important aspects of the financial system.

3 Material Political *Economy*

Why, though, do we need the third word, "economy"? Surely, everything to do with finance is self-evidently economic? The reason for emphasizing the economic (indeed, the monetary) aspects of finance is that—along with power and the state—they too have often been given insufficient attention in the "social studies of finance." What in particular has been neglected is what I think of as the "mundane" political economy of finance: the undramatic, everyday, and indeed sometimes grubby ways in which money is made—often individually fairly small amounts of money but time and time again.[7] That moneymaking is often made possible by the occupancy of favorable positions in what market practitioners often call "market structure," by which they mean the way in which a market is organized,

---

[6]See MacKenzie (2019a). There is an excellent introduction to field theory in Kluttz and Fligstein (2016).

[7]What I mean by the "mundane" political economy of finance is different from the "everyday international political economy" advocated by Hobson and Seabrooke (2007), in their (entirely justified) attention to the actions of the "bottom ninety per cent" (2007, p. 12). The traders, dealers, and so on I focused on are elite actors; my argument, in contrast to Hobson's and Seabrooke's, is that the mundane actions of these elite actors and the economic consequences of those actions have too often been neglected.

especially the formal and informal "rules of the game" that dictate matters such as who or what can trade with whom and on what terms, how information flows, where its flows are blocked, and so on. Precisely because the rise of HFT has often involved challenges to aspects of market structure such as these, it thereby renders them visible, and the activity does of course also have a mundane political economy of its own.

That much of finance's moneymaking is, in the above sense, "mundane" does not detract from its importance. Since around 1980, profits and remuneration in the financial sector in many countries have soared and in doing so have become an important contributor to growing inequality (see, e.g., Philippon & Reshef, 2012, Bell & Van Reenen, 2013, and Godechot, 2012). The process has many aspects, but among them is the remarkable phenomenon identified by Philippon (2015: see MacKenzie, 2016): the apparent failure of the US financial system to become more "efficient," in terms of the unit cost of financial intermediation, from the 1880s to 2012. One possible explanation is that much of the efficiency benefits of the development of finance's underlying technology, information processing, has either been captured within the financial system in the form of fees and other forms of mundane moneymaking or wasted in what are in effect zero-sum games (Philippon, 2019), one of which is discussed below.

After a brief account of data sources and a section sketching major features of HFT and the firms that engage in it, this chapter explores the material political economy of HFT in four main sections. The first, on HFT's signals, briefly examines the "political economy" roots of the historically most important of these signals: the predictive power of data from the share-index futures market for algorithms trading shares. The second and third sections explore the two main types of HFT algorithm—which I will refer to as "making" and "taking"—and the two chief roots of HFT's zero-sum game, its speed race. The two components are the race among "making" algorithms to get to the head of the queue for electronic execution of their orders and the race between "making" and "taking" algorithms to be the first to respond to a signal that has predictive value. The fourth section of the paper focuses on material interventions by exchanges—which are now no longer member-owned bodies that often claimed a quasi-public role but nearly always for-profit corporations or subsidiaries of such corporations[8]—in how HFT algorithms interact with each other.

4 Data Sources

It is not straightforward to find high-frequency traders prepared to be interviewed in any depth about HFT's material practices. One of my interviewees, for example, tried unsuccessfully to persuade an acquaintance to speak to me: "even though he

---

[8]The significance of this change was first pointed out to me by Phil Mirowski some 15 years ago.

left over a decade ago, he signed paperwork going in and out of the company to keep quiet and also he (nor anyone) wants to get back on [X]'s radar because he has the money and ability to create legal issues for former employees that speak" (email from interviewee MI, 26 June 2017; X is a founder of the HFT firm for which my interviewee's acquaintance had worked). Gradually, however, I identified, primarily via snowballing, 86 practitioners of HFT—mostly in the world's main centers of the activity: Chicago, New York, London, and Amsterdam—who were prepared to be interviewed in varying degrees of technical depth (22 on more than one occasion, three of them six or more times), along with 32 suppliers of technology or communication links to HFT. Other interviewees included exchange staff and regulators; in total, 337 people were interviewed. In publications from the research, interviewees are anonymized via two-letter labels.

I have also visited two datacenters and taken part in six traders' conferences and an algorithmic trading training course. Other sources include the specialist trade press and a modest amount of archival research (including, in particular, in documents from the early years of one of the first HFT firms, Automated Trading Desk). There is, in addition, now a reasonably large body of research by financial economists on HFT (reviewed, e.g., by Menkveld, 2016). As far as I can—in particular, concerning the nature of "signals" used by HFT algorithms—I have also cross-checked the results of the interview with the findings of that literature.

5   High-Frequency Trading

HFT firms are typically recently established and small. Only a handful date from before 2000 and even an HFT firm with no more than a few dozen staff can be a significant player. In particular niches, even firms with only a handful of staff can be important: in 2019, an interviewee calmly told me that his tiny European HFT firm is responsible for five percent of all the share trading in India. Some big banks used to be active in HFT, but their efforts were often less than fully successful: the rapid development of the fast, highly specialized software systems that are needed can be difficult in a large, bureaucratic organization. Banks are still engaged in market-making in some classes of financial instrument—such as foreign exchange and governments' sovereign bonds—albeit often using systems that are slow by HFT standards, but the use of other HFT strategies by banks has effectively been ended by the postcrisis curbs on banks' proprietary trading.

The HFT firms I have visited vary widely. Some had offices in unremarkable or even scruffy buildings; others had spectacular views over Lake Michigan, Manhattan, or Greater London. HFT firms' premises could often pass for those of a generic dot-com firm. The staff of HFT firms are mostly young and—at least in the roles closest to trading—mostly male. Almost none of the men wears a business suit—it is common for me, as the visitor, to be the only person wearing a tie—and the shouting and swearing that used to happen on banks' trading floors seem much less common in HFT firms. I have visited firms only in the United States and Europe, but there, at

least, white faces dominate, though often intermingled with people of South Asian or Chinese extraction, while African Americans, for example, seem rarer.

The internal organization of the HFT firms from which my interviewees come varies. Some operate as unified entities. Just as Lange (2016) discovered, though, other HFT firms are divided into strictly separate trading teams, with deliberate barriers to communication. One firm, for example, physically separates teams by placing a row of administrative staff between them and in its main offices even plays white noise between the rows to reduce the chance of members of one team overhearing what is said by members of another. At one compartmentalized firm, said a young trader (interviewee AC) who worked for it, "you [. . .] could get in trouble for being in the next room talking to someone you're not supposed to talk to."

High-frequency trading, though, does not actually happen in these rooms. Instead, it takes place in exchanges and other trading venues' computer datacenters.[9] These datacenters contain both the exchange's computer system and the systems of HFT and other algorithmic trading firms, of banks, of communications suppliers, and so on. Exchanges' datacenters are not generally to be found in city centers, but in suburban areas in which real estate is cheaper: the datacenters important to HFT are mostly large buildings and indeed usually look like suburban warehouses, with, for example, few windows. They are packed with tens of thousands of computer servers, typically on racks in wire mesh cages (although sometimes the cages have opaque walls, so that a trading firm's competitors cannot see the equipment it is using). The servers are interconnected by mile upon mile of cabling, typically running above the racks in what looks to an outsider like an incomprehensibly complex spaghetti of different types of cable. In aggregate, those servers consume very large quantities of electricity and generate large amounts of heat, making a powerful air-conditioning system also a requisite. Normally, few human beings are to be found in these datacenters: a small number of security and maintenance staff, along with (at least some of the time) engineers from the exchange, trading firms, or communications suppliers visiting to fix problems or install new equipment.

No more than around twenty datacenters globally host the bulk of the world's trading and the vast majority of its HFT. Most US share trading, for example, takes place in the four datacenters in Northern New Jersey shown in Fig. 1. One is owned by the New York Stock Exchange's parent company, the Intercontinental Exchange. Another is leased by Nasdaq, traditionally the main rival to the NYSE as a trading venue for US shares. Two further datacenters (NY4 and NY5) host the systems of multiple trading venues, including the third main group of US stock exchanges, now owned by the Chicago Board Options Exchange. NY4 and NY5 are close together and in practice are run as a single datacenter.

---

[9]By no means all electronic trading venues are registered exchanges. For brevity, however, I will simply refer to all exchange-like trading venues as "exchanges."

Fig. 1 The main US share-trading datacenters. In this and Fig. 2, locations are shown only approximately

6 HFT's Signals

All of the most important US stocks are traded on all of the exchanges in these datacenters. That makes the automated trading going on in one share-trading datacenter a vitally important source of data for algorithms trading shares in the other datacenters: a vital class of "signal" as market practitioners would call it. (To repeat, a "signal" is a pattern of data that informs an algorithm's trading, e.g., prompting it to bid to buy shares or offer to sell them or perhaps to cancel an existing bid or offer.) Also of great importance to algorithms trading US shares, and constituting another crucial class of "signal," is what is going on in the share-index futures market, which is to be found not in New Jersey but in a datacenter in the suburbs of Chicago: see Fig. 2. (A "future" is a standardized, exchange-traded contract economically close to equivalent to one party undertaking to buy, and the
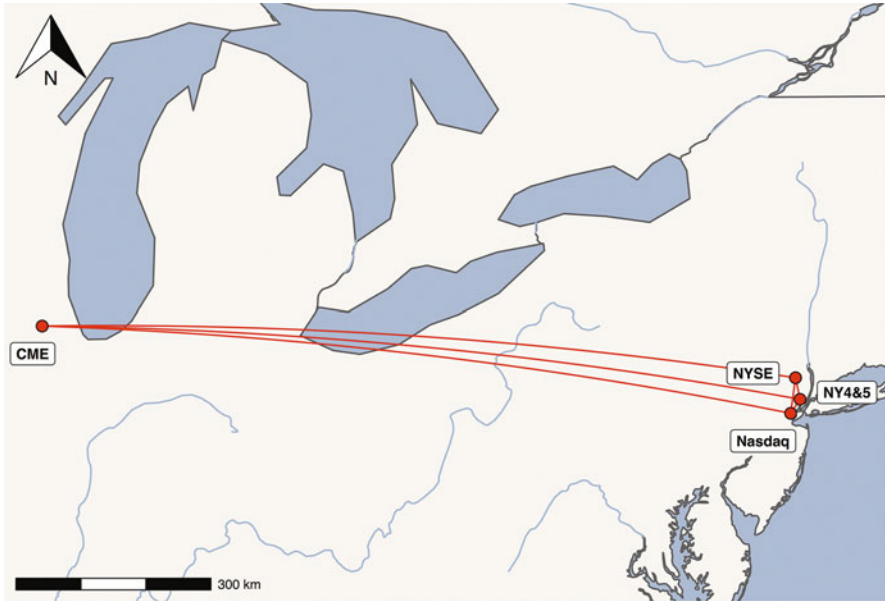
Fig. 2 Geodesics from Chicago to the New Jersey share-trading datacenters. "CME" is the main Chicago Mercantile Exchange datacenter

other to sell, a set quantity of some underlying asset, on a given future date but at a price agreed at the inception of the contract.)

There is a "material political economy" to HFT's signals of the kind just described. It is explored in much more detail in MacKenzie (2018b), but let me briefly summarize, using the example of what, for brevity, I call "futures lead": the tendency for the share-index futures market to move a tiny fraction of a second before the underlying shares (and therefore to form a crucial source of signals for HFT algorithms trading those shares). "Future lead" has its origins in efforts in the 1970s by leaders of Chicago's futures exchanges to circumvent the traditional legal distinction between legitimate futures trading and gambling, which had become a constraint on their development of financial futures because (as Levy, 2006 describes) it required physical delivery of the underlying asset to be possible. Leaders of the Chicago exchanges worked in the mid-1970s with the House and Senate Agriculture Committees, which traditionally had jurisdiction over futures trading, to set up a new futures regulator, the Commodity Futures Trading Commission or CFTC, in part in the hope that it would remove the requirement for physical delivery.

After the election of President Reagan and his appointment of a new chair of the CFTC, the latter permitted the launch of share-index futures with no physical settlement mechanism (such a mechanism would have been very clumsy, given the mathematical nature of an index) and allowed them to be traded with the high levels of "leverage" traditional in futures trading. ("Leverage" is the size of a trading

position relative to the capital needed to create and sustain it.) Leverage in share trading, in contrast, was tightly constrained by the stock market regulator, the Securities and Exchange Commission or SEC. The greater leverage possible in futures attracted traders, with the index futures traded by the Chicago Mercantile Exchange becoming typically the market that first reflected information relevant to US shares overall. That situation largely continues today. There have been multiple efforts in the decades since the 1970s to merge the two main US financial market regulators (the SEC and CFTC), but those efforts have always failed: regulators interviewed for this research suggest that the ultimate cause of the failure is the fact that the CFTC reports to the Senate Agriculture Committee, whose members are not prepared to give up jurisdiction to the Senate Banking Committee, to which the SEC reports.

7 Making

An HFT algorithm acts in essentially just three ways: by bidding to buy the shares or other financial instruments being trade, by offering to sell them, or by canceling or modifying one of its existing bids or offers. Market participants classify a bid or an offer into two broad categories, which I will call "making" and "taking." The divide sounds at first like a minor technicality, but (although never absolute) it is one of the two key drivers of the field's speed race, sometimes providing a defining characteristic of participants' skill sets and occasionally even of their self-identities.

What makes an order "making" rather than "taking" (or vice versa) is its price relative to that of the existing orders in the order book. Consider the book shown in Fig. 3. In it, a bid to buy shares at $29.49 would be called by market participants a "liquidity-providing," "adding," or "passive" order; as already noted, in what follows, I'll call it a "making" order, because the most systematic use of orders of this kind is in "market-making." The matching engine (the exchange system that maintains the order book and executes a trade if it finds a bid and offer at the same price) cannot execute a bid at $29.49 immediately, because there are no offers to sell at that price to match it with. Instead, the matching engine would simply add it to the order book's list of bids at $29.49. It remains in the book—available for other market participants to execute against (hence the terminology of "market-making," "liquidity-providing," "adding," or "passive")—until it is executed or canceled. In most exchanges, the list is a queue, a time-priority list: a new bid to buy at $29.49 will be executed only when all earlier bids at that price have been executed or canceled. In contrast, a bid to buy at $29.50 would be a "taking" order. (Other terms for it would be "aggressive" or "removal.") The matching engine can execute a bid at $29.50 straightaway (perhaps only partially if it is large), because there are offers to sell at that price. Doing so removes these existing orders from the book: hence the terminology of "taking" or "removal."

In an exchange of the sort in which HFT is prevalent, most "making" and most "taking" orders are placed by algorithms. The core mechanism of direct interaction

BIDS TO BUY

$29.49	100	100	200	
$29.48	50	30		
$29.47	100			
$29.46	50	100	100	100
$29.45	200			

OFFERS TO SELL

$29.54	100	200	
$29.53	50		
$29.52	40	50	
$29.51	50	50	200
$29.50	100	100	100

Fig. 3 An order book. The bids or offers at a given price form a time-priority queue, with the earliest received shown here on the left (*source*: author's interviews and observations of trading)

among trading algorithms is the straightforward process of matching these orders one with the other: "taking" orders are matched with (executed against) "making" orders at the same price. That mechanism is so simple that it is hard to imagine that much pivots on it, and the difference between the prices of a "making" and a "taking" order on a modern electronic exchange is normally very small (the one-cent difference in Fig. 3 between a bid that "makes" and a bid that "takes" is 0.03 percent of the market price, and differences of roughly that relative size are typical). Yet, this simple mechanism and tiny difference resonate strongly. The material exigencies of this mechanism of interaction in large part generate the huge infrastructure of datacenters and communication links sketched above, and—at least for some participants and to some extent—"making" and "taking" can be experienced as having a moral, affective weight.

That moral weight is worth exploring briefly before turning in more detail to the material practices of algorithmic "making" and "taking." The words chosen by interviewee BQ are unusually strong, but the underlying sentiment is not unique: "I tend to want to work at [HFT] companies that are 'makers' because I see the inherent evil in the 'takers'." This moral preference for "making" rests on the legitimacy of its most systematic form: "market-making." In the order book in Fig. 3, for example, the first-in-the-queue bid to buy 100 shares at $29.49 and the first-in-the-queue offer to sell 100 shares at $29.50 might both have been entered by the same market-making algorithm. Although the goal is economic (to earn the one-cent difference between those two prices, along with any "rebates" or other payments the exchange may make to incentivize market-making), algorithmic

"making" inherits the legitimacy of a traditional human role: that of the "market-maker," who constantly stood ready both to buy the financial instrument being traded and to sell it (at a higher price, the difference in the past usually being proportionately much larger than the one-cent difference in Fig. 3).[10]

Because other market participants' bids and offers arrive only sporadically, market-makers—whether human or algorithmic—provide a service to market participants who want to transact immediately. Interviewee OH highlights the potency of that source of legitimacy when she recounts an episode—at the height of the global financial crisis of 2007–2008—in the algorithmic market-making firm for which she then worked (a firm I have visited myself several times to speak to others in it). A software developer had left the firm saying, "I couldn't look my grandmother in the face anymore and say I worked in finance." The firm's chief executive called a meeting of all its employees in the large, open-plan trading room at the center of its offices and (as my interviewee recollects) told them, "I'm going to explain to you why you should be able to look your grandmother in the face: because we're market-makers and we provide liquidity."

Specialists in "taking" reject this "moralization" of making (as indeed do some specialists in the latter). For example, interviewees BY and CV, both from the same "taking" firm, cite taking's central role in what economists call "price discovery," its role (via arbitrage: see below) in keeping prices in different markets aligned, the "service" (interviewee CV) it provides to those who wish to trade using "making" orders (which, other things being equal, is cheaper than "taking"), and the plain fact that without "taking," an exchange would have no trading. Furthermore, invocations of market-making's legitimacy can sometimes smooth over what one might call "rough edges" in the actual practice of automated trading. At a trader's conference in the Netherlands in 2019, I listened to a senior member of the country's financial regulatory body praise "market-makers formerly known as prop[rietary] traders or high-frequency traders." Just the previous day, though, I had sat in a café with an experienced Amsterdam trader, interviewee CS, who had told me how arbitrage (which involves exploiting fleeting price discrepancies by buying and selling as close as possible to simultaneously and thus typically involves "taking") was interwoven with market-making, both in what he individually did and in the wider trading of the city's HFT firms. "[B]asically any money you can see lying around you will pick up," he said.

Legitimacy—being able to "look your grandmother in the face"—is, furthermore, seldom a day-to-day concern of market-makers. More pressing is the often precarious economics of the activity. As AG puts it, "you make a little bit of money," from your algorithm repeatedly selling at a price higher than that at which it buys, but "you periodically get run over": your market-making algorithm buys when prices are about to fall or sells when prices are about to rise. No market-maker reported being

---

[10]For example, Nasdaq dealers' avoidance of odd-eighth price quotes, at a time in which US share prices were denominated in eighths of dollars, notoriously made this difference usually no less than 25 cents as recently as the early 1990s (Christie & Schultz, 1994).

able entirely to avoid being run over, but all try as hard as they can (often success-fully) to ensure that the resultant losses do not fully cancel out the repeated small gains. To avoid this, a market-making algorithm has to keep its "inventory" (its aggregate trading position) reasonably small. "[Y]ou have to actively control your inventory," says AE: if inventory starts to rise, a market-making algorithm will "shade" its bids or offers so as to reduce it. If, for example, too many of its bids have been executed, it will reduce the price of its offers so as to make those more attractive. If that fails, it may begin to "take": in this example, reducing its inventory by executing against existing bids in the order book. (That occasional need for a market-making algorithm to "take" is one of the ways in which the divide between "making" and "taking" is not absolute.)

In addition, almost all market-making algorithms use "signals" such as "futures lead" to make predictions of near-term price movements and use those predictions to minimize the risk of being "run over." As BL says, "markets move and you need to know when they're going to move because [otherwise] you'll be inventorying at a terrible price." These signals need to be "squashed," as BM puts it: to inform how an algorithm trades, they need to be reduced to a single indicator. Although a variety of mathematical forms of "squashing" are in use, HFT interviewees consistently report that by far, the most common is for algorithms to combine signals via what is essentially a linear regression equation, in which a set of predictor variables (here, signals) are each "weighted" so that in combination, they best predict the value of a single "dependent variable." Interviewees used a variety of expressions to refer to this dependent variable—such as "fair value" (AF and AQ), "fair price" (AE), "microprice" (AN), and even "perfect price" (AM)—but it was most commonly called a "theoretical value."[11] Neither of those two words should, however, be overinterpreted: what the term refers to in HFT is simply a near-term prediction of the price of the financial instrument being traded, "the price you can reasonably expect to transact at the near future" (AF).

Price prediction is helpful in market-making for at least two related reasons. One is discussed in the next section: the need to cancel bids or offers as quickly as possible if prices are about to move in such a way as to render them "stale," so avoiding "taking" algorithms being able to "pick them off" (make a profit by executing against them). The other reason is to help ensure that a market-making algorithm's bids or offers are as close as possible to the front of the "queue" for execution. If prices change (in Fig. 3, e.g., if all the offers to sell at $29.50 are executed against or canceled), then market-making algorithms will race to populate a new "level," as market participants would put it (e.g., race to bid at $29.50). Success in the race is vital to the mundane economics of market-making: a bid or offer that is not at or near the head of the queue is likely to be executed, if at all, only in adverse circumstance, such as when the "side" of the order book (bids or offers) it is on is "crumbling" (emptying out) and prices are about to move adversely.

---

[11] Not all HFT firms seek to estimate theoretical value. Thus, interviewees AI, AQ, AU, and BP reported that their algorithms' predictive efforts did not take this form.

8   Taking: Picking Off Stale Quotes

Algorithmic "taking" is a heterogeneous activity. Within HFT, its most highly regarded form is price prediction that either is mathematically more sophisticated than that engaged in by market-making algorithms or draws on more diverse data than they do. (The background constraint on market-making algorithms is the need for speed, both to achieve favorable queue position and for the reason we are about to discuss.)

That quantitatively sophisticated "taking" is, however, in many cases, what interviewee BW calls an "aspiration." In practice, a great deal of "taking" has a much simpler form: the picking off of stale quotes.[12] This involves the arrival of a "signal" of a kind that is widely understood within HFT as being of predictive value: in the case of share trading, for example, a move in the market for share-index futures or a substantial change in the order book for the same shares in a different datacenter. A signal of this kind immediately renders many market-making algorithms' bids or offers out-of-date ("stale"). If that happens, a "taking" algorithm has no need for quantitative sophistication: if it is fast enough, it can simply "pick off"—profit by trading against—knowably stale bids or offers.

The possibility of "picking off" creates a speed race that is at least as important as and probably more important than the race among market-making algorithms to be at the head of the queue. "Making" algorithms race to cancel stale bids or offers, and "taking" algorithms race to execute against those stale quotes: the pressing nature of this race is amply testified to by my interviewees, and it is modeled insightfully by Budish et al. (2015). This race forces HFT algorithms to operate at nanosecond speeds, and its exigencies are central to the mundane economics of HFT: they create, for example, the incentive—and in many cases the need—to pay to build or use the fastest possible links among datacenters.

Nontrivial sums of money are at stake between "making" and "taking" algorithms. Budish et al. (2019, p. 40) estimate that in 2015, total annual "arbitrage rents" (their term) in US share trading—the money at stake in the "making"/"taking" race—were between $3.1 and $3.7 billion (ibid.). This figure, though, is mostly likely larger than the aggregate profits earned by HFT firms from trading US shares. As is clear from interviews, these "rents" (which in the model of Budish et al., 2015 are paid—via bid-offer spreads—by the end investors in the stock market such as pension and mutual funds) are shared between HFT firms; exchanges; and the suppliers of technologies, communication links, and other services to HFT firms. For example, Budish et al. (2019, p. 40) estimate that the "exchange-specific speed technology revenues" of the three main "families" of US equities exchanges—what those exchanges earn from, e.g., selling fast data—were between $675 and $790 million in 2015 and between $874 and $1024 million in 2018 (ibid.).

---

[12]For brevity, I will not attempt to discuss forms of taking other than quantitatively sophisticated price prediction and "picking off": see MacKenzie (2018a).

9   Intervening in Making-Taking Interaction

The involvement of exchanges in the interaction among HFT algorithms is not restricted to the sale of data and associated technologies: at least in some cases, exchanges intervene materially in that interaction, usually to encourage "making."[13] There is no evidence in my interviews that this is because they regard "making" as more moral than "taking." Their fear, rather, is of "empty screens" (interviewee GI)—order books devoid of bids and offers—which make an exchange fatally unattractive to traders and institutional investors.

Anyone who knows the actor-network theorist Bruno Latour's famous example of material politics, the traffic-slowing "sleeping policeman" (Latour, 1999, pp. 187–188), cannot help but be delighted by the fact that the currently most salient interventions in "making"/"taking" interaction are "speed bumps." The best known device (highlighted in Michael Lewis's, 2014 bestseller *Flash Boys*) is a 60-kilometer coil of fiber-optic cable installed in one of the share-trading datacenters (NY5) by the new US stock exchange, Investors Exchange (IEX), through which all incoming orders to IEX (and all market data from IEX) have to pass, slowing them down by 350 microseconds. The coil, however, is a less than decisive intervention in the interaction between "making" and "taking" algorithms, because it slows down both categories equally.

At the time of writing (August 2020), the current focus of controversy is *asymmetric* speed bumps, which impose delays on "taking" algorithms while not delaying cancellations of orders by "making" algorithms, thus protecting the latter from being "picked off." Despite several proposals by exchanges to install such speed bumps, the regulator of US share trading, the Securities and Exchange Commission, has so far rejected all these proposals, on the grounds that they unfairly discriminate among market participants. The domain in which intervention of this kind is most common has therefore been foreign exchange, with its very different political economy. Because the trading of foreign exchange is inherently transnational in a world in which—the European Union aside—financial regulation is still mainly national, it is regulated only lightly, and trading venues have no need to gain regulators' permission for changes in trading systems. In addition, banks, with their sometimes slow technical systems, are still important foreign exchange market-makers and retain considerable influence and resources, having the capacity—interviewees report—to put pressure on trading venues in which they are losing money to HFT algorithms.

The various material interventions in the interaction of algorithms in foreign exchange trading are too numerous to discuss in detail, so let me simply sketch the most sophisticated of them, a module, described to me by interviewee GS, that was added to the Thomson Reuters trading system in 2016. The module examines incoming buy and sell orders for each of the currency pairs being traded, classifies

---

[13]They also intervene economically in US share trading, most importantly by paying "rebates" to those who place "making" orders that subsequently are executed against.

them as either "taking" or "making," and adds them to the corresponding buffer.[14] The first order to enter an empty buffer starts a "timer" that runs for 3 milliseconds (thousandths of a second), at which point the buffer is emptied by sending the bids or offers it contains to the matching engine in a random order.[15] Randomization stops, e.g., the fastest market-making algorithm always getting to the head of queues, and crucially, *cancellations* of orders are not placed in a buffer, but sent to the matching engine immediately. This gives substantial protection to market-making algorithms: if the market moves, they have 3 milliseconds (a long time, by HFT standards) to cancel their stale quotes before they are picked off.

Asymmetric speed bumps are an issue that divides market participants sharply. One of my interviewees told me that the head of a major trading firm had phoned an exchange that was proposing to install such a speed bump to warn it that the firm would cease trading on any of the exchanges owned by the company if the speed bump was installed. That is hearsay, and I have no independent evidence of the phone call, but the fact that my interviewee believed that it had taken place is indicative of the sensitivity of the issue.

10   Conclusion

Much more will be needed than what has been possible in this chapter to develop further the treatment in the social studies of finance of issues such as power, the state, and politics. Nevertheless, I hope that the idea of "material political economy"—and the emphasis on *all* of the three words—is a useful contribution. I do not, of course, intend that idea to replace other broadly anthropological/sociological approaches to economic life (such as the "cultural economy" of du Gay & Pryke, 2002), but as suggested above, it is one way of seeking to bring together the ethnographic strength of the best research in the social studies of finance with the emphasis on politics and power in other versions of political economy.

It must of course be acknowledged that the "Einsteinian" materiality of HFT is unusual: I know of no other economic activity in which time delays measured in the billionths of seconds are consequential. Nevertheless, in a world in which the digital economy is becoming ever more salient, it is always worth remembering that, however "virtual" digital processes appear, they all have material underpinnings (as Dourish (2017) emphasizes). That is very much the case, for example, for cryptocurrencies such as bitcoin, with the staggering quantities of electricity needed

---

[14]There are separate buffers for "making" orders at different prices.

[15]The randomization is by firm: all the orders in the buffer from the same firm are grouped, and only one of them is implemented before the Reuters algorithm moves to an order submitted by the next firm in the randomly ordered list of firms. For more detail on the module and its rationale, see Melton (2017).

for the "mining" activities that are so central to them (for the material political economy of which, see MacKenzie, 2019b).

Even bitcoin, though, is a little esoteric. What is far from esoteric are the big social media, electronic commerce, and search platforms that now saturate the everyday life of so many people: Facebook, for example, reports that on average, 2.1 billion people—more than a quarter of Earth's population—use one of its services every day.[16] Although, when using many such systems, human beings can tolerate delays of the order of 1 second (a million times longer than the magnitude of delay that can be fatal in HFT), the scale at which those systems need to operate makes the necessary computation—especially memory accesses—materially extraordinarily demanding. It hardly needs to be said, in addition, that there is a mundane political economy to most such systems: that of digital advertising. Controversy swirls around these systems, but little of it so far has concerned their material underpinnings, except for the issue of electricity consumption and environmental impact. "Material political economy" is, however, not just a perspective but a hypothesis: that, just as in the case of HFT, if we can begin to dig deeper into the materiality of the giant systems of the digital economy, we will find that there is also a material politics of these systems and that it is economically consequential.

References

Bell, B., & Van Reenen, J. (2013). Bankers and their bonuses. *Economic Journal, 124*(February), F1–F21.

Borch, C., & Lange, A.-C. (2017). High-frequency trader subjectivity: Emotional attachment and discipline in an era of algorithms. *Socio-Economic Review, 15*(2), 283–306.

Bourdieu, P. (1997). Le champ économique. *Actes de la Recherche en Sciences Sociales, 119*(Sept), 48–66.

Budish, E., Cramton, P., & Shim, J. J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *Quarterly Journal of Economics, 130*(4), 1547–1621.

Budish, E., Lee, R.S., & Shim, J.J. (2019). A theory of stock exchange competition and innovation: Will the market fix the market?. *NBER working paper* no. 25855. National Bureau of Economic Research.

Callon, M. (Ed.). (1998). *The laws of the markets*. Blackwell.

Christie, W. G., & Schultz, P. H. (1994). Why do NASDAQ market makers avoid odd-eighth quotes? *Journal of Finance, 49*(5), 1813–1840.

Cowan, R. S. (1983). *More work for mother: The ironies of household technology from the open hearth to the microwave*. Basic Books.

Dourish, P. (2017). *The stuff of bits: An essay on the materialities of information*. MIT Press.

---

[16]See https://newsroom.fb.com/company-info/ (accessed October 6, 2019).

du Gay, P., & Pryke, M. (Eds.). (2002). *Cultural economy: Cultural analysis and commercial life*. Sage.

Fligstein, N., & Dauter, L. (2007). The sociology of markets. *Annual Review of Sociology, 33*, 105–128.

Fligstein, N., & McAdam, D. (2012). *A theory of fields*. Oxford University Press.

Godechot, O. (2012). Is finance responsible for the rise in wage inequality in France? *Socio-Economic Review, 10*, 447–470.

Hobson, J. M., & Seabrooke, L. (2007). Everyday IPE: Revealing everyday forms of change in the world economy. In J. M. Hobson & L. Seabrooke (Eds.), *Everyday politics of the world economy* (pp. 1–23). Cambridge University Press.

Kluttz, D. N., & Fligstein, N. (2016). Varieties of sociological field theory. In S. Abrutyn (Ed.), *Handbook of contemporary sociological theory* (pp. 185–204). Springer.

Lange, A.-C. (2016). Organizational ignorance: An ethnographic study of high-frequency trading. *Economy and Society, 45*(2), 230–250.

Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Harvard University Press.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Laumonier, A. (2019). *4*. Zones sensibles.

Law, J., & Mol, A. (2008). Globalisation in practice: On the politics of boiling pigswill. *Geoforum, 39*, 133–143.

Levy, J. I. (2006). Contemplating delivery: Futures trading and the problem of commodity exchange in the United States, 1875–1905. *American Historical Review, 111*(2), 307–335.

Lewis, M. (2014). *Flash boys: Cracking the money code*. Penguin.

MacKenzie, D. (2016). Must do better. *London Review of Books, 38*(9), 29–30.

MacKenzie, D. (2018a). "Making", "taking" and the material political economy of algorithmic trading. *Economy and Society, 47*(4), 501–523.

MacKenzie, D. (2018b). Material signals: A historical sociology of high-frequency trading. *American Journal of Sociology, 123*(6), 1635–1683.

MacKenzie, D. (2019a). Market devices and structural dependency: The origins and development of "dark pools". *Finance and Society, 5*(1), 1–19.

MacKenzie, D. (2019b). Pick a nonce and try a hash. *London Review of Books, 41*(8), 35–38.

MacKenzie, D. (2021). *Trading at the speed of light: How ultrafast algorithms are transforming financial markets*. Princeton University Press.

Melton, H. (2017). Market mechanism refinement on a continuous limit order book venue: A case study. *ACM SIGecom Exchanges, 16*(1), 74–79.

Menkveld, A. J. (2016). The economics of high-frequency trading: Taking stock. *Annual Review of Financial Economics*, 81–24.

Mirowski, P., & Nik-Khah, E. (2007). Markets made flesh: Performativity, and a problem in science studies, augmented with consideration of the FCC auctions. In D. MacKenzie, F. Muniesa, & L. Siu (Eds.), *Do economists make markets? On the performativity of economics* (pp. 190–224). Princeton University Press.

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press.

Philippon, T. (2015). Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. *American Economic Review, 105*(4), 1408–1438.

Philippon, T. (2019). *The great reversal: How America gave up on free markets*.

Philippon, T., & Reshef, A. (2012). Wages and human capital in the U.S. finance industry: 1909–2006. *Quarterly Journal of Economics, 127*(4), 1551–1609.

Seyfert, R. (2016). Bugs, predations or manipulations? Incompatible epistemic regimes of high-frequency trading. *Economy and Society, 45*(2), 251–277.

Part III
Lessons from History and Great Economists

Marx and Hayek on "Real" Versus "Less Real" Explanations for the Fragility of Capitalism

Arie Arnon

For all except true believers in the gospel according to Neo Classical Theory it is clear that the study of capitalist economies requires an analysis of how monetary and financial institutions affect the macroeconomic and microeconomic characteristics of the economy. One striking aspect of the history of the one hundred years since the publication of the third volume [of *Capital* in 1894], the 111 years since the death of Marx, has been the increasing complexity of the financial structure.[1]

  Minsky (1994). "Marxian Economics: A Centenary Appraisal." *Hyman P. Minsky Archive*. Paper 170.

1 Introduction and Outline of the Argument

The tensions and links between "real" and "monetary" (or more generally "financial") facets in economic analysis are known to be confusing and even misleading. Almost all scholars, including the two addressed in this paper, analyzed non-barter economies where money and banking (and more sophisticated financial arrangements) were present. Political economists in the nineteenth century and in the first half of the twentieth century were alert to the phenomena of cycles and periodic crises in the economic system and addressed it often. As we know, the community of

---

[1] http://digitalcommons.bard.edu/hm_archive/170 (p. 2). Minsky's comments were made in a conference held in 1994 "Marxian Economics: A Centenary Appraisal—Karl Marx's Third Volume of *Capital*: 1894–1994." The written comments were never published in the conference volume; they are held in the Bard College Digital Commons, where the Minsky Archive is deposited since he was a distinguished scholar of the Levy Institute. The comments are Minsky's discussion of Foley's and Sardoni's papers presented at the conference. The volume of the papers and some of the discussions appeared in a book with the conference title in 1998, edited by Riccardo Bellofiore.

A. Arnon (✉)
Department of Economics, Ben-Gurion University of the Negev, Beersheba, Israel
e-mail: arnona@bgu.ac.il

economists tended to underestimate the importance of instabilities by the turn of the millennium. However, as Desai (2014) and many others wrote, the 2008 threats to the economic system made modern economists more aware of the systemic instabilities in the capitalistic system.[2]

Karl Marx's and Friedrich von Hayek's thoughts on the fragility of capitalism—a system characterized in their analyses by the competitive production of goods and services made for sale on the markets—are essential to understanding their works. Production and selling, or the monetary circulation of commodities and their production, are interconnected. Thus, as the above quote from Minsky states, production cannot be understood separately from money (and finance), an observation that was significant for both Marx and Hayek. However, I will argue with many, but certainly not all, interpretations that Marx had been rightly perceived as emphasizing more the "real" rather than the "monetary" (or "financial") causes in explaining the fragility of capitalism in the books he published on economics (1859, 1867).[3] One could describe the latter sphere as the "less real" in terms of its importance in determining the processes in capitalism. That is, at least for Marx in his main texts, the "real" causes were more fundamental and, in many ways, unescapable, leading to inevitable consequences.

Moreover, in his writings, Marx did not accept the view that there existed any systemic intervention—any policy—that could prevent the collapse of capitalism: neither a policy concerning the "real" forces at work nor one that would be implemented in the "less real" spheres. For Hayek, the fragility was primarily the result of "less real" factors, definitely not causes in the production process itself, and he believed that the system could continue functioning properly if societies would only listen to him and accept his (changing) analysis and recommended policies. In other words, Hayek believed that if society would accept his recommendations as to what had to be done to avoid the fragility in the system, the undesirable consequences of that fragility would vanish. Let us elaborate on Marx's and Hayek's views.

2 Karl Marx (1818–1883)

Marx's analysis of the fragility of capitalism in the two books he published during his lifetime, *A Contribution to the Critique of Political Economy* (1859) and *Capital, Vol. I* (1867), addressed a system where money exists and production was done under competitive conditions by capitalists who were selling goods on the markets.

---

[2] See also Desai (2006).

[3] For a recent discussion linking Marx's and Minsky's ideas concerning the role of finance in crises, see De Grandi and Tutin (2020). In the paper, the authors explore "Marx's writings on money, finance, credit and crises, [which] although they mostly consist in unfinished notes and drafts, contain the lineaments of a coherent theory of financial instability, which can be better understood in the light of Minsky's hypothesis of financial instability" (p. 854).

In the subsequent two volumes of *Capital*, posthumously published in 1885 and 1894 after Friedrich Engels' edited drafts written mainly in the 1860s, Marx referred to crises while addressing money, banking, and finance more broadly than in his two early books. Many of the debates on the roles of "real" versus "less real" factors in the analysis were due to the gaps between what Marx published himself and what Engels edited and published under Marx's name after his death. Several drafts written by Marx, which were never previously published but are now available, could help in better understanding Marx's views of capitalism and crises.

In both the writings published during Marx's lifetime as well as in the posthumously published texts, the value category takes center stage. Value, shorthand for exchange value—not use value—is set apart in Marx's analysis from market prices and production prices. Value was the foundation for Marx's (and the Classicals') analysis, while "surplus value" was a key concept in his thinking about crises and fragility. "Surplus value" is embedded in the process of production under capitalism, where capital hires labor power and miraculously manages to create more value in the production process. Nowhere else can surplus be born but in production; the discovery of the origin of "surplus value" was, in Marx's view, his most important and novel discovery. The unique relation of capital to labor in the production process, the famous role of the unpaid fraction of labor's contributions, explains the miracle behind "surplus value" and was also crucial to Marx's understanding of the fragility of capitalism.

In Marx's writings, three major explanations for crises emerge; importantly, the fragility of the system would uphold in all three also under barter, when money does not exist and ignoring finance. The three famous elucidations, known briefly as (1) underconsumptionism, (2) the tendency of the rate of profit to fall, and (3) "disproportions," the possible chaos concerning production in decentralized economies, would all result in crises (see Foley, 1986, Chapter "Desperation by Consent: Inequality and Financial Crises", and 2010). The famous (or infamous) prediction that capitalism could not escape its collapse was proposed by Marx based largely on his value analysis, since the first two of the three key explanations for crises suggested by Marx were based on his value and "surplus value" concepts. Hence, for Marx, crises were fundamentally the result of "real" processes located in the production sphere and were not the consequences of mainly "monetary" or "financial" causes. The monetary and financial considerations could aggravate the various fragilities,[4] but for him, these were more secondary causes, not the primary ones. De Grandi and Tutin (2020, p. 877) concluded in a recent paper:

> For Marx any crisis starts with a "Minsky moment" of liquidity crisis. [...] he gives a rather clear account of the mechanisms leading to a fragile financial structure. His notion of "sensitivity" or fragility of the productive structure to financial shocks has striking similarities with Minsky's financial instability hypothesis.

---

[4] See *Capital, Vol. III* and the *Theories of Surplus Value* volumes (Marx, 1969), known sometimes as *Capital, Vol. IV*.

It is important to note that for Marx crises are imminent and manifest the fragility of capitalism. Of course, they are not the only manifestation of the fragility of capitalism. The conflict between capital and labor, the class society, also produces fragilities that are manifested in other facets beyond economic crises. Monetary and financial causes contribute to the fragilities and could aggravate them and hence affect crises. However, one should note that no policy could prevent crises in the "pure" capitalistic system, a conclusion that is contrary to the common belief of most modern economists who followed Keynesian thinking of whatever variety. Crises could be postponed, as Lenin and Luxemburg argued, by expanding the system, using noncapitalistic areas to solve its contradictions through new demands, military expenditures, and so on. However, once the world was fully under the capitalistic system, they argued that those solutions would stop functioning. Moreover, under the above conditions, the fragilities would appear both in a more competitive capitalist system and in a less competitive one, whether authorities intervened or not.

In Marx's first book on economics in 1859, *A Contribution to the Critique of Political Economy*, published after a series of crises occurred during the first half of the nineteenth century in several countries, Marx argued—criticizing both David Hume and David Ricardo—that those crises were the results:

> [...] of big storms on the world market, in which the antagonism of all elements in the bourgeois process of production explodes; the origin of these storms and the means of defence against them were sought within the sphere of currency, the most superficial and abstract sphere of this process. The theoretical assumption which actually serves the school of economic weather experts as their point of departure is the dogma that Ricardo had discovered the laws governing purely metallic currency. It was thus left to them to subsume the circulation of credit money or bank-notes under these laws (Marx, 1907a, b, p. 182).

Thus, these experts tried to explain crises simply by addressing "credit money" under the same laws governing "purely metallic currency." Marx chose another path, following the Banking School and Thomas Tooke rather than David Ricardo and the Currency School (see Arnon, 1984, 1991, 2011).

In recent years, with the publication of previously unpublished segments of Marx's writings in the new MEGA volumes (MEGA^2), the materials expanded the research into Marx's economic thinking on crises and the fragility of the capitalist system (see van der Linden & Hubmann, 2018; Mori, 2017, 2018a, b, 2019; and Clarke, 1994). The *Books of Crisis*, written by Marx from November 1857 to February 1858, was published in MEGA^2 in 2017 and is of great interest. It was written around the time Marx prepared the *Grundrisse* (Rough Draft), which also remained unpublished for many years. The obvious motivation for writing the texts was the very serious world crisis in 1857–1858. Kenji Mori, the editor of the MEGA^2 volume where *The Books of Crisis* was published, explained the enormous interest in the events: "The 1857 crisis and the business cycle leading to it were unique [...]. It was the first global economic crisis in history. The crisis was, just like the 2008 crisis, triggered by a financial crisis in New York and spread like a wildfire

across all countries involved in the world market" (2019, p. 81).[5] In his text, Marx emphasized, elaborated, and clarified the key role of production in explaining crises, supporting the claim that in his view, "real" causes were largely responsible for the fragility of the capitalist system. Mori stated that during the few years before 1857, as Marx was anticipating another crisis to arrive, he believed that he knew what type of crisis to expect:

> Marx was sure that the next crisis would be a "double crisis" in the sense of a simultaneous and interconnected collapse of *industrial* and *produce markets, caused* by *over-production* on the industrial market and *under-production* on the produce market (Mori, 2019, pp. 82–83).

Moreover, Mori claims that "Marx worked on the *Books of Crisis* under the influence of volumes five and six of *A History of Prices*" published in 1857 by Thomas Tooke (1774–1858) and William Newmarch (Mori, 2018a, p. 913) and that "one may say that both volumes served Marx as models for empirical research" (p. 915). However, Mori ignores Tooke's and the Banking School's vital influence on Marx's thinking concerning the subjects of money, banking, and finance, the "less real" sphere. In fact, Marx read not only the volumes of *A History of Prices* as Mori wrote but also, what Mori seemed to ignore, the less empirical and more analytical *An Inquiry into the Currency Principle* written by Tooke in 1844, where he presented his Banking School views, as well as other texts which significantly influenced Marx's thinking on money and banking. Thus, the well-established influence of Thomas Tooke and the Banking School more generally on Marx's ideas concerning the "less real" sphere which received much attention in the secondary literature was ignored by Mori. Additionally, Marx read those texts much earlier than Mori indicated.[6]

In explaining the 1857–1858 crisis in *The Books of Crisis*, Marx indeed used the "double crisis" argument; it suited the conditions in England but was not relevant as a general explanation (see Mori, 2018a and 2019, pp. 83–84). The main general cause for crises was the situation which Marx termed a "general overproduction." This crucial concept was developed in the *Grundrisse*, the unpublished Rough Draft written in 1857–1858 which was published many years later. In this text, Marx defined a "general overproduction," using what Mori calls an "input-output system."

However, the fundamental cause and the primary explanation for overproduction are to be found in the "peculiarities" of fixed capital, which is characterized by a

---

[5]Mori (2019, pp. 81–82) continues with two more reasons for the uniqueness of the 1857 crisis: "2) It is recognized by modern researchers of economic history to be the first crisis in which the capitalist economy was sufficiently developed to make modern business cycle models applicable. According to Hughes (1956), for example, the Hicksian trade cycle theory can be suitable to explaining the 1857 crisis. We will see later that Marx's understanding of crises at that time turns out not to be very far from it. 3) It was the crisis Marx investigated most intensively in his life time, i.e. he reported as an economic journalist (namely for the newspaper New York Daily Tribune) in real time, the whole business cycle of ten years beginning in 1848 and ending with the crisis in 1857 (which continued until early 1858)."

[6]For details, see Arnon (1984 and 2011, Chap. 16); see also Betancourt and Manigat's (2018) discussion of James Stuart's role in shaping Marx's monetary theory.

"long utilization period" as well as by a "long construction period" (Mori, 2019, p. 93):

> While the earlier concept of "double crisis" had consisted in the underproduction of raw materials and overproduction of industrial products, Marx now developed it to a dynamic concept of two successive periods of fixed capital. Although further explication of economic crises based on the peculiarities of fixed capital has been never provided by Marx himself, his particular reasoning of the causality obviously entitles him to a position in a line of theories that we call here *the production theory of crisis* (Mori, 2019, pp. 94–95; my emphasis).

In formulating his production theory of crisis, Marx was influenced by another Banking School scholar, this time James Wilson (1805–1860), the founder and first editor of *The Economist*.[7] He was the first in a long line of future advocates of such theories[8] which Mori (2019, p. 95) summarized:

> The overall characteristic of this type of theories is often labeled as "non-monetary over-investment theories" (Haberler, 1958, p. 72). More specifically, however, it can be characterized by attaching decisive importance to fixed capital, especially its periodicity of construction and utilization period.

Thus, Mori's argument suggests that among the important clarifications concerning the development of Marx's thinking on crises are ideas which can be found in the drafts which were kept until recently in the archives. The *Book of Crisis* written by Marx while facing a rolling real crisis only emerged in 2017 in volume IV/14 of MEGA[2] (see Mori, 2017). In the three notebooks, one can find the data collected by Marx and his commentaries on the crisis as it was progressing in France, the United Kingdom, and elsewhere, covering various "real" and "less real" markets. Marx founded his explanation of crisis on "general overproduction" and focused his attention on fixed capital and its "peculiarities." As noted above, he focused on the long utilization and long construction of fixed capital.[9] Interestingly, Hayek, to whom we will turn soon, was also influenced by Wilson.

Paula et al. (2013) delineated Marx endeavors in studying the next crisis after 1857–1858, that of 1866. They focused largely on another unpublished notebook

---

[7] "Marx read Wilson's book and several articles in The Economist and excerpted them even in detail in his so-called London Notebooks, so Wilson's influence on the formation of Marx's theory cannot be denied (see MEGA IV/7, pp. 74, 440–60)" (Mori, 2019, p. 95); the book is by Wilson (1847).

[8] Mori lists: "Albert Aftalion, Dennis Robertson, Gustav Cassel, Frederick Lavington, Arthur Spiethoff, Mentor Bouniatian, Arthur Cecil Pigou, and up to John Hicks" (p. 95)

[9] "In his process of empirical studies on the 1848–58 business cycle in real time, including the *Books of Crisis*, Marx acquired an insightful intuition that there is a material basis for business cycle, and that in particular the occurrence of crisis and its periodicity is closely related to peculiarities of fixed capital, especially the 'transformation' of circulating capital into fixed capital and 'continuity of production' based on the distinction between construction and utilization periods. By representing this very idea and trying to theorize the intuition, Marx ought to be affiliated to a stream of crisis theories that stress the importance of fixed capital, reaching from Wilson through Aftalion up to Hicks" (Mori, 2019, p. 109).

which Marx prepared in 1868–1869 (marked in the archives as B113) and concluded that:

> By analyzing the contents of Notebook B113 and of the source materials which were used in its composition, our aim is to suggest the relevance of this manuscript source for an adequate understanding of the prominent place occupied by monetary and financial issues amidst Marx's investigations in the late 1860s (Paula et al., 2013, p. 166).[10]

Furthermore, an important point for the historiography of Marx's analysis of the role of finance in crises was the claim the authors made that "it seems reasonable to suppose that Notebook B113 was originally prepared as raw material for a deep revision of" (Paula et al., 2013, p. 180) Marx's writings on finance and crises as they appeared in *Capital, Vol. III* (in part five). The well-known debates concerning this volume, particularly how accurately they reflected Marx's views about finance and crises, are troubling to many scholars since the volume was published in 1894 after being edited by Engels after Marx's death. Thus, it seems safe to speculate that the notebooks containing Marx's studies, which were lying for many years in the archives—the *Books of Crisis* now in MEGA^2 (IV/14), Notebook B113 as well as other notebooks—will contribute to future research attempting to clarify Marx's ideas on finance and crises in capitalism. Paula et al.'s (2016) paper consciously contributed to such an endeavor: "The notebooks of 1868–1869 provided a further step in his [Marx's] understanding of capitalist crises, and might have thus constituted essential material for a revision of Volume III, making them essential for a proper understanding of its contents" (p. 196). Based on Paula et al.'s, 2016 paper, which circulated in various versions in the years before its publication, De Deus et al. (2016, p. 474) provided a summary of the general lessons the authors of these texts learned from the notebooks:

> Between 1857 and 1866, Marx established a clear agenda, not only for his research but also for the actual writing of *Capital*. In the middle of this trajectory, reality would pose some questions to the object he logically built: the panic of 1866 provided new elements about credit, stock markets, finance and the money market. In fact, the panic presented specific features, if compared with the previous crises and panics of the 19th century.

In the late 1860s, Marx's analysis of crises in capitalism relied on the concept of surplus value, articulated in the "law of the tendency of the rate of profit to fall." However, by the time of the 1866 crisis, Marx had a firmer grasp of the growing roles of money, credit, and finance in theorizing crises than he did after 1857. His understanding in Notebook 113 was in this sense clearly different from that in the *Books of Crises* relating to the 1857–1858 crisis. De Deus et al. (2016, p. 476) follow the differences and emphasize:

> At that time [after 1857–8], one can say that Marx knew what he was looking for. As was the case in 1857, Marx was concerned with a complete picture of events and with the most important interpretations presented by newspapers and other authors. Nevertheless, in 1868, Marx had already written about the law of the tendency of the rate of profit to fall and the

---

[10]For more on Notebook B113 and the various sources Marx used for studying the financial system, see pp. 163–166.

counteracting influences, chief among them the credit system and the stock market. The notebooks of 1868–69 thus presented many notes about the transformations under way in capitalism and in its financial institutions. Most important, some of Marx's main concerns were the role of the state in the crisis, the suspension of the Bank Act of 1844, and the intense debates in the British Parliament, which would eventually produce reform and improvement in the legislation.

Thus, there was a process of change in Marx's thinking concerning the fragility in capitalism, a process that deserves more analysis based on the new available sources.

3   Friedrich A. von Hayek (1899–1992)

Friedrich A. von Hayek was quite influential and well known in the 1930s. After he had published his PhD dissertation, *Monetary Theory of the Trade Cycle*, first in German in 1929, he gave a celebrated series of lectures in the London School of Economics (LSE) in 1931 which were published as *Prices and Production.*[11] The former book was written before the Great Depression, while the latter was written after it began. Hayek's messages in both are clear: the fragilities in the capitalist system are due to occurrences that originated within the "less real" sector. The banks expanded the supply of money and credit and were responsible for the disorders that were behind the booms and indispensable busts in the economy. Had the banks refrained from those undesirable actions, the fragility would have disappeared. The interruptions they introduced caused the system to deviate from the optimal levels of interest rates, distorted relative prices, and produced the instability and fragility. If the banks and central banks (and other intermediaries) would have stopped destabilizing the equilibrium rates which existed when they did not interfere, the capitalist system would not have shown signs of fragility.

Hayek's position on policy in the 1930s was famously on the side of fewer interventions in the economy. His paper "Economics and Knowledge" (1937) started his more methodological inquiry, leading him toward a very skeptical view of what we could know and hence about what we should do: not much, in his view. This turn to methodology reinforced his conclusions: since the knowledge about the economy was restricted and policy could likely worsen the situation, not improve it, refrain from interferences in the economy. Thus, the problem of missing information in the economy led Hayek to strengthen his objections to the rising Keynesian tendency.

It is important not to forget that not only in those years, the years of the Great Depression, but also during the years of the hegemonic Keynesian era, Hayek did not propose the elimination of the power in the hands of the banks or other financial institutions to influence the economy and hence affect interest rates and relative prices. Throughout the years, he accepted the common monetary architecture while warning its controllers to be cautious. It was only starting in the mid-1970s, after

---

[11] *Monetary Theory of the Trade Cycle* was published in English in 1933; *Prices and Production* second, expanded, edition, was published in 1935.

Hayek received the Nobel Prize at the age of 75, that he outlined his very different "free banking" doctrine, a kind of "End the Fed" position.

In *Monetary Theory and the Trade Cycle*, written before the Great Depression, Hayek made the case that an abstract theory, described by him as "a monetary explanation of the trade cycle," was preferable to its alternatives, identified as "real" explanations. Furthermore, he argued that the theories proposed by Knut Wicksell and Ludwig von Mises, among the many monetary theories he had examined, were the closest to "true" explanations, though neither was perfect ([1929b] 1933, pp. 47–48).

In the second chapter of the book entitled "Non-Monetary Theories of the Trade Cycle," Hayek surveyed many of the "best known" real theories, among them those by Wesley C. Mitchell and Alvin H. Hansen. In his view, they failed to explain why the economy moved away from "equilibrium." These various theories, known as "disproportionality theories," began by assuming one major cause for a slump, a cause that is rooted in the boom that precedes it: production capabilities grow by more than what consumption warrants (ibid., p. 56). Hence, prices of production goods rise relative to those of consumption goods, a phenomenon familiar to those researching the business cycle at the time.

He identified three types of "real" explanations:

a) Those which claim that the increases in demand for the consumption of goods affect the production of goods from different orders differently
b) Those which describe the same phenomenon but argue that it results from special circumstances concerning savings and investment
c) Those based on "psychological variants," as Hayek called them (ibid., pp. 60–61).

Hayek argued that none of these "real" explanations provided a convincing answer to a simple question: "Why do the forces tending to restore equilibrium become temporarily ineffective and why do they only come into action again when it is too late?" (ibid., p. 65). In other words, "real" theories do not offer a satisfactory explanation for why the price system malfunctions (ibid., pp. 70–71).

What is different and distinctive about "monetary" explanation as opposed to "real" ones? The crucial answer focused on the effects that money and credit have on the economy that "real" factors do not have. These effects are necessary for a satisfactory explanation of trade cycles:

> Every change in the volume of means of circulation is, in fact, an event to be distinguished from all other real causes, for the purpose of theoretical reasoning; for, unlike all others, it implies a loosening of the inter-relationships of equilibrium. No change in "real" factors, whether in the amount of available means of production, in consumers' preferences, or elsewhere, can do away with that final identity of total demand and total supply on which every conception of economic equilibrium is based. *A change in the volume of money, on the other hand, represents as it were a one-sided change in demand, which is not counterbalanced by an equivalent change in supply* (ibid., pp. 92–93; my emphasis).

The uniqueness of the monetary explanation under what Hayek called "static equilibrium theory" was that the "finality and 'closedness' of the system" dissipate

because of a monetary change. Importantly, the characteristics of equilibrium in the system (finality and closedness) do not change because of "real" factors, i.e., in such circumstances the equilibrium remains the same. Logically, this is the most important point that emerges from *Monetary Theory and the Trade Cycle*. As a result of monetary changes, one can have different equilibrium conditions that are in line with cyclical phenomena. Hence, the cycle is not reflecting disequilibrium in the economy, but rather, it is reflecting changing equilibrium positions. For Hayek, the former approach—cycles as the result of disequilibrium—was not a satisfactory theoretical explanation. Anyone who understands that cycles exist in reality and are the result of fundamental forces in the system would have to explain them as the outcome of equilibrium considerations. In other words, cycles should be explained as the result of what Hayek termed "a priori" competitive forces. A monetary explanation does that; real explanations do not.

The monetary explanations contradict, of course, the idea of a well-functioning system where there is no money. However, the trade cycle is part and parcel of a monetary economy. Hence, monetary theory should not just study the value of money but "should also study those phenomena which distinguish the money economy from the equilibrium inter-relationships of barter economy which must always be assumed by 'pure economics'" (ibid., p. 104). Hayek compares a barter economy to a monetary one and introduces the rate of interest into the analysis: "[I]n a barter economy, interest forms a sufficient regulator for the proportional development of the production of capital goods and consumption goods." When money is absent, the rate of interest is capable of preserving equilibrium and circumventing cycles by preventing "excessive extension of the production of production goods" beyond savings created in the economy. When money appears on the scene, the mechanism that both guarantees equilibrium and prevents cycles in the economy does not work.

Hayek briefly reviews the history of monetary explanations in *Monetary Theory and the Trade Cycle*. He expressed appreciation for some of the predecessors he reviewed but insisted that none of them reached the right conclusions. Among those whom he mentions approvingly is Henry Thornton, who continued to attract Hayek's attention in the coming years, as did David Ricardo, Alfred Marshall, and others. But it was the early studies by Knut Wicksell and Ludwig von Mises that clearly influenced Hayek and which he found revealing and enlightening. Hayek claimed that the focus of most researchers on the general price level, also known as the purchasing power of money, was the fundamental error responsible for the failure to reach a satisfactory theoretical explanation for the trade cycle. Wicksell, whose distinction between the "money rate of interest" and the "natural rate" (1898) influenced all later discussions, emphasized the two rates' impact on the price level. But the economic variables affected by interest rates were more fundamental to the working of the economy and to its equilibrium than was the price level alone. In particular, Hayek assessed the changes in the supply of savings and the demand, or production, of capital goods that necessarily occur in response to changes in interest rates.

In a chapter entitled "The Fundamental Cause of Cyclical Fluctuations," Hayek elaborated on the "element whose presence forms the 'necessary and sufficient' condition for the emergence of the Trade Cycle" (Hayek, 1929b 1933, p. 140). This new element was the "elasticity" in the volume of money in the economic system, which was responsible for "deviations of the money rate of interest from the equilibrium rate [...] which must be regarded as the cause of the periodically recurring disproportionalities in the structure of production" (ibid., p. 139). In order to understand the elasticity in the monetary system of banking, one has to distinguish between the analysis of a single bank and of the banking system as a whole and to appreciate the similarity between "paper circulation," that is, notes, and the "balances in the banks," that is, the banks' liabilities. The first authors to contribute to an understanding of this elasticity were Henry Thornton, James Pennington, and the members of the Banking School.[12] What is crucial to understand, argued Hayek, was that the banking system could create more credits than the sums deposited with it (ibid., pp. 152–163).

According to Hayek, when banks are faced with more demand for credit due to better opportunities for investors (rising natural rate) as opposed to demand due to lower bank interest rates, they are able to supply that demand. The reason is that the banks determine their own cash reserves, and under the new conditions, based on their self-interest, they will decide whether to extend loans. The credit expansion, motivated first by the demand for bank credit from businesses whose rate of profit is higher than the rate they pay to the banks, is further extended by the banks. The latter, calculating their own interests, support extending credit. Prices in the economy change accordingly in a predictable and well-documented fashion, which is at the core of Hayek's explanation: prices of capital goods increase more than those of consumption goods.

Thus, the gap between the money rate and the natural rate is behind both the boom and the bust (ibid., pp. 173–176). The additional credit, created as a result of both the investors and the banks acting according to their own best interests, allows more investment than savings warrant:

> By creating additional credits in response to an increased demand, and thus opening up new possibilities of improving and extending production, the banks ensure that impulses towards expansion of the productive apparatus shall not be so immediately and insuperably balked by a rise of interest rates as they would be if progress were limited by the slow increase in the flow of savings. *But this same policy stultifies the automatic mechanism of adjustment which keeps the various parts of the system in equilibrium, and makes possible disproportionate developments which must, sooner or later, bring about reaction* (ibid., pp. 177–178, my emphasis).

Hayek was quick to point out that elasticity in the supply of credit had many advantages for the economy, although he identified some disadvantages as well. Once a "single price has been fixed at a different level from that which it would have formed in a barter economy," which happens when we have such elasticity, "a shift

---

[12] See Hayek ([1929b] 1933, note on pp. 152–153); see also Hayek (1991, Chap. 11).

in the whole structure of production is inevitable" (ibid., p. 179). The failure of the automatic mechanism of adjustment under credit creation was fundamental to Hayek's theoretical construct. When there is no elasticity in the supply of credit, the rate of interest in the market and the natural rate take care of the adjustment, which cannot turn stultified, that is, cannot become ineffective:

> The immediate consequence of an adjustment of the volume of money to the "requirements" of industry is the failure of the "interest brake" to operate as promptly as it would in an economy operating without credit. This means, however, that the new adjustments are undertaken on a larger scale than can be completed; a boom is thus made possible, with the inevitable recurring "crisis." *The determining cause of the cyclical fluctuation is, therefore, the fact that on account of the elasticity of the volume of currency media the rate of interest demanded by the banks is not necessarily always equal to the equilibrium rate, but is, in the short run, determined by considerations of banking liquidity* (ibid., pp. 179–180; emphasis in the original).

Hayek emphasized that the monetary explanation of the trade cycle does not assume that the initial cause is always monetary; the initial cause could, in fact, take many different forms, real or monetary.[13] It does assume, however, that the economic system will respond to the "initial change" by beginning a cycle, instead of "reacting [. . .] with an immediate 'adjustment'" that would form a new equilibrium. The trade cycle is explained by "monetary factors," hence the term "monetary theory of the trade cycles." The monetary explanation fundamentally rests on a process in which additional credit is the driving force; the trade cycle theory cannot be understood without such additional credit. As a result of credit creation, the pricing process departs from the "course deduced in static theory."

The many attempts to arrive at a modified system that would secure a stable price level are all missing the point since they are not capable of undoing the trade cycle and are looking in vain for stability. The only proper answer to "cyclical fluctuations" is to do away with the "elasticity" in credit, for example, to "keep the total amount of bank deposits entirely stable," as Hawtrey appeared to suggest in *Monetary Reconstruction* (1926) (see Hayek, 1929b 1933, pp. 181, 190). Hayek, however, hesitated, suggesting that "this seems to us purely Utopian" since banks would have to assume "the role of brokers, trading in savings" and since it would "necessitate the complete abolition of all bank-money—i.e. notes and cheques." Even if practical, the price of such a change in terms of secular growth in the economy would cause many to reject it. A better alternative, Hayek told his readers, was to accept the fact that fluctuations were here to stay and that one can only minimize their damage. This conclusion had far-reaching policy consequences, of course, as Hayek proposed an explanation for the trade cycle, not a cure.

---

[13] "It must be emphasized first and foremost that there is no necessary reason why the initiating change, the original disturbance eliciting a cyclical fluctuation in a stationary economy, should be of monetary origin. Nor, in practice, is this even generally the case. The initial change need have no specific character at all, it may be any one among a thousand different factors which may at any time increase the profitability of any group of enterprises" (Hayek, 1929b 1933, pp. 182–183).

The theoretical arguments underlining Hayek's discussions in *Monetary Theory and the Trade Cycle* were not always unambiguous and were open to various interpretations. He seemed not fully satisfied with what he wrote in 1929, as one can sense from the comments he added to the English translation of the book. In the translation, he quite often directed his readers to consult the theoretical arguments as he presented them in *Prices and Production*, the four celebrated lectures he delivered in January 1931 at the LSE. However, while following Hayek's advice, one should not forget that in the time between these two studies, the Great Depression had begun. As we now know, that shattering experience changed the context of the debate dramatically. From then on, the debate was no longer focused on the theoretical explanation of the trade cycle, but on policies to counter a major crisis. These policies, or, more accurately, the recommendation to avoid active policies, were at the core of what made Hayek's views so well known and controversial during the 1930s.

Hayek on Policy in the Early 1930s

In *Prices and Production*, Hayek pledged to deal with "the problems connected with a monetary policy suitable for the prevention of crises" ([1931] 1935, p. 99). As argued above, the general price level was the wrong target for policy. Even if monetary policy was to achieve a stable general price level, Hayek described in detail how money would not be neutral and the "real" economy would still respond to "money" in many instances. For example, Hayek argued that during the "upward swing of the cycle," the total circulating medium would increase. Not only should the banks refrain from expanding it, but also they should contract credit. This action, observed Hayek, was "probably entirely Utopian" since the "general opinion" was that the central banks should support trade (ibid., pp. 116–117):

> I am strongly convinced that, if we want to prevent the periodic misdirections of production caused by additional credit, something very similar to the policy outlined above, absurd as it may seem to those accustomed to present-day practice, would be necessary. I do not delude myself that, in the near future, there will be any opportunity of experimenting with such a policy (ibid., pp. 117–118).

Thus, bankers should not follow the levels of production or of trade as guidelines for their policy. He argued that such attempts would only be appropriate for "a [central] monetary authority for the whole world: action on the part of a single country would be doomed to disaster" (ibid., p. 125).

It was not until the mid-1970s, after Hayek received the Nobel Prize and started practicing economics again, that he systematically readdressed policies. Then 75 years old, he adopted new positions on the appropriate monetary regime and on monetary policies that were only partially the outcome of economic analysis but more in line with his general position on markets and governments. Between 1975 and 1978, Hayek outlined an argument for "free banking" for the first time (see

Hayek, 1978). Hayek's mature views raise interesting questions about issues concerning his methodology and his economics: Was the new research program he promoted in the 1970s on free banking the (late) result of his post 1937 methodology?

4   Synopsis: Marx's and Hayek's Reading of Nineteenth-Century British Monetary Theory

Keynes's *General Theory* (1936) associated the fragility of the capitalist system with the problems which arose from the lack of correspondence between the actual aggregate demand and what is necessary for achieving full employment. Monetary and financial aspects play a role, but the crucial failure in the system is on the "real" side though not necessarily on its production segment. I would argue that it resembles, to some degree, Marx's approach. However, contrary to Marx, Keynes proposed a solution to the fragility in the form of managing and directing aggregate demand. The "socialization of investment" was his more radical form of intervention in the market process. Money, credit, and finance were important in Keynes's thinking, of course, although, instead of just adding another cause for the fragility in the system, they also provided more tools for interventions that could save the system from its destructive internal forces.

Hayek, like Adam Smith, believed that on the "real" side, there were no foundations for fragility. In his view, the causes for fragility in the system were in the money, banking, and finance spheres where excess supplies created instabilities through their influence on the rates of interest, triggering inappropriate relative prices. If the institutions behind the departures of interest rates from their natural levels stopped pushing the economy away from equilibrium—which would be the case if they didn't interfere—the system would remain stable, and fragilities would disappear. Moreover, even in the imaginary case—a system with no interventions and no instabilities, as in Hayek's mature view—there would be no room for correcting policy since no information would exist for making such policy work.

Intriguingly, Hayek's early studies of the history of monetary theory, before the mid-1970s, pointed to very different conclusions from those he reached later. During his years in Vienna, after his visit to New York in 1924–1925 and before he left for London in 1931, Hayek had planned to write a book on the history of monetary theory, for which he had prepared four chapters by 1929.[14] The most compelling

---

[14] (1) "Genesis of the Gold Standard in Response to the English Coinage Policy in the 17th and 18th Centuries;" (2) "First Paper Money in eighteenth Century France"; (3) "The Period of the Restriction, 1797–1821, and the Bullion Debate in England;" and (4) "The Dispute Between the Currency School and the Banking School, 1821–1848." These chapters, written in German, were not published until 1991, when they were translated and appeared as Chapters 9–12 in *The Trend of Economic Thinking*, volume 3 of *The Collected Works of F.A. Hayek*. Hayek intended to revise the chapter, but according to Bartley, the editor of the volume, "Hayek did not find the opportunity to

theoretician to emerge from Hayek's studies of past monetary thinking, the one whose ideas attracted him most, was Henry Thornton. What impressed Hayek about Thornton was the fact that he understood—and Hayek thought he was the first to do so—the complex process by which exchange rates are determined, an issue that scholars in the 1920s struggled with as well. Although Hayek had some minor disagreements with Thornton, as he wrote in his 1929 chapter "The Period of the Restriction, 1797–1821, and the Bullion Debate in England," he concurred with Thornton on several major issues, including his criticism of Smith concerning the latter's famous Real Bills Doctrine. Hayek argued that Thornton showed "how fear of progressive devaluation of money accelerates [money] velocity, but he is even more concerned about the fact that conversely, in times of crisis, when everyone attempts to hoard cash, the velocity of circulation is reduced and may trigger a serious shortage of money" ([1929a] 1991, p. 194). This led to Hayek's rather surprising observation about Thornton's position concerning interventions in the monetary system, i.e., monetary policy, in his seminal *Paper Credit*.[15]

Analyzing the "correct response" to "internal drains"—the money shortages that resulted from a public demand for cash, which was characteristic in times of crises—Hayek recognized Thornton's "major contribution to economic science: the correct response for a note-issuing bank is to maintain the level of circulating money rather than decrease it, if a panic is to be avoided" (p. 194; my emphasis). Prior to Thornton, scholars had not discussed the responsibility of the note-issuing Bank of England for the proper functioning of the economy. Its acceptance by the monetary orthodoxy, as described by Fetter in his classic 1965 work, was a process that took many decades (see Arnon, 2011, 2017). By 1929, Hayek was already fully aware of the pioneering status of Thornton in this context and described that achievement as "Thornton's final and perhaps most significant contribution," adding dramatically that it "has been nearly neglected until now [1929]." This contribution, he added, "concerns the foundations of the discount policy pursued by note-issuing banks." Hayek expounded on the historical perspective of this major theoretical breakthrough:

> Thornton raises the question whether there exists a natural tendency to keep note circulation within limits that exclude a dangerous devaluation of notes. In answering this question he first demonstrates that neither reliance on the wealth of the borrower nor limitation to genuine commodity-backed notes can offer assurance against this danger. Even if these two aspects are taken into account, it would not prevent an unreasonable increase in the number of borrowers nor an unwarranted proliferation of commodity-based notes. Thornton reiterates emphatically that every time the prevailing profit rate in business exceeded the interest rate of the bank, there would be a tendency to over-issue notes. The bank should therefore attempt to adjust its interest rate to the market rate. Thornton thereby not only resolved in advance a controversy that was to rage in the middle of the nineteenth century

---

revise" them (1991, p. 128, note 3). Bartley explains that the book contract was cancelled after Hayek moved to London in 1931 and Hitler rose to power in 1933.

[15]The full title: *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*, published originally in 1802; republished in 1939 with an introduction by F.A. von Hayek

between the currency and the banking schools,[16] but even anticipated by nearly a century K. Wicksell's theory about the significance of the "money interest rate" falling behind the "natural" interest rate (Hayek, 1929a 1991, pp. 194–195).

Thus, the right quantity of notes in circulation could *not* be left to a "natural tendency." One should take careful note of Hayek's description of the two mechanisms that *do not guarantee* such a tendency: one cannot trust either "the wealth of the borrower" or the use of "genuine commodity-backed notes" as assurance against too many notes. Thus, Thornton rejected two well-known mechanisms, with Hayek in seeming agreement; one was Smith's Real Bills Doctrine, and the other was the convertibility of bank notes. The latter approach was famously the foundation of the Currency School, while Thornton would have important influence on the Banking School and, indirectly at least, on Marx.

Hayek's continued fascination with Thornton resulted in the republication of *Paper Credit* in 1939, reestablishing the book's significant place in the history of monetary theory. This endeavor included Hayek's long and important "Introduction" to the volume where he expressed the same admiration for Thornton as he had shown in 1929: "It is not too much to say that the appearance of the *Paper Credit* in 1802 marks the beginning of a new epoch in the development of monetary theory" (Hayek, 1939, p. 36). Importantly, he also added to the list of Thornton's contributions the discovery of "forced saving," an issue which Hayek himself addressed in the early 1930s. Hayek told his readers that Thornton wrote how the "expansion of credit will in the first instance lead to the employment of 'antecedently idle persons'" but added that, as these were limited in number, the increased credit "will set to work labourers, of whom a part will be drawn from other, and perhaps, not less useful occupations" (1939, p. 49; Hayek noted Thornton 1802, p. 236). "This leads him," wrote Hayek, "to one of the earliest expositions of what has become known as the doctrine of 'forced saving.'"[17]

The increase in capital that may result from "excessive issue of paper" is due to the fact that the laborer "may be forced" to consume less. Hayek also lists among Thornton's achievements—as proof of the "height of his intellectual power"—his discussion of the role of the rate of interest on which Thornton "breaks entirely new ground." Hayek concluded by saying that J.S. Mill was the last author to do "anything like justice to Henry Thornton," drawing attention to the strange story of the disappearance of Thornton from the known economic literature by the 1870s.

Thus, Thornton's perception that money and credit have considerable influence on the "real" economy clearly influenced Hayek's early, pre-"free banking," views. It also indirectly influenced Marx, through the Banking School on which Thornton clearly had a significant impact. Although, as argued above, for Marx, the

---

[16]Hayek's note: "Chapter. "Inequality, Economic Policy and Household Credit in the US: The Roots of Unsustainable Finance", especially pp. 283–290 of the original edition and pp. 399–410 of the German translation," which appear on pp. 251–256 in Hayek's, 1939 edition

[17]Hayek developed the argument concerning "forced savings" in one of his better-known papers in 1932.

fundamental source of fragility was to be found primarily in the "real" sphere of the economy, even though throughout his writings one can find important traces of the "less real" factors as well.

References

Arnon, A. (1984). Marx's theory of money: The formative years. *History of Political Economy, 16* (4), 555–575.

Arnon, A. (1991). *Thomas Tooke: Pioneer of monetary theory*. Michigan University Press.

Arnon, A. (2011). *Monetary theory and policy from Hume and Smith to Wicksell: Money, credit and the economy*. Cambridge University Press.

Arnon, A. (2017). *Hayek's early studies of past monetary theories: The intriguing attitude to Henry Thornton*. Paper presented at the History of Economics Society 44th annual conference, Toronto, Canada, 22–26 June.

Bellofiore, R. (Ed.). (1998). *Marxian economics: A reappraisal essays on Volume III of capital, Volume 1: Method, value and money*. Macmillan.

Betancourt, R. G., & Manigat, M. P. (2018). James Steuart and the making of Karl Marx's monetary thought. *European Journal of the History of Economic Thought, 25*(5), 1022–1051.

Clarke, S. (1994). *Marx's theory of crisis*. Macmillan Press.

De Deus, L. G., Paula, J. A., Gama Cerqueira, H. E. A., Eduardo, H., & Suprinyak, C. E. (2016). A theory in the making: Marx's drafts of *Capital* and the Notebooks on the crisis of 1866. *Science & Society, 80*(4), 468–480.

De Grandi, A., & Tutin, C. (2020). Marx and the "Minsky moment", liquidity crises and reproduction crises in *Das Kapital*. *European Journal of the History of Economic Thought, 27*(6), 853–880.

Desai, M. (2006). Hayek and Marx. In E. Feser (Ed.), *The Cambridge companion to Hayek* (pp. 67–81). Cambridge University Press.

Desai, M. (2014). Marx, Keynes and Hayek and the Great Recession of 2008. In R. Bellofiore & G. Vertova (Eds.), *The Great Recession and the contradictions of contemporary capitalism* (pp. 50–64). Elgar.

Fetter, F. W. (1965). *Development of British monetary orthodoxy, 1797–1875*. Cambridge University Press.

Foley, D. K. (1986). *Understanding Capital: Marx's economic theory*. Harvard University Press.

Foley, D. K. (2010). Lineages of crisis economics from the 1930s: Keynes, Hayek, and Schumpeter. *Eastern Economic Journal, 36*(4), 413–422.

Haberler, G. (1958). *Prosperity and depression. A theoretical analysis of cyclical movements* (4th ed.). George Allen and Unwin.

Hawtrey, R. G. (1926). *Monetary reconstruction* (2nd ed.). Longmans, Green.

Hayek, F.A. ([1929a] 1991). The period of restriction, 1797–1821, and the Bullion Debate in England. In *The collected works of F. A. Hayek, Vol. 3, The trend of economic thinking: Essays on political economists and economic history* (pp. 177–215). Ed. by W.W. Bartley III and S. Kresge. The University of Chicago Press.

Hayek, F.A. ([1929b] 1933). *Monetary theory and the trade cycle*. Translated from the German by N. Kaldor and H. M. Croome. Kelley. Reprint 1966.

Hayek, F.A. ([1931] 1935), *Prices and production*. Second revised and enlarged edition, with a Foreword by L. Robbins. Kelley. Reprint 1967.

Hayek, F. A. (1932). A note on the development of the doctrine of "forced saving". *Quarterly Journal of Economics, 47*(1), 122–133.

Hayek, F. A. (1937). Economics and knowledge. *Economica, 4*(13), 33–54.

Hayek, F.A. (1939). Introduction. In H. Thornton ([1802] 1939), *An enquiry into the nature and effects of the paper credit of Great Britain* (pp. 11–63). A.M. Kelley, 1978; also in Hayek (1991), *The collected works of F. A. Hayek, Vol. 3, The trend of economic thinking: Essays on political economists and economic history* (pp. 295–344). Ed. by W.W. Bartley III and S. Kresge. The University of Chicago Press.

Hayek, F.A. (1978). *The denationalization of money: An analysis of the theory and practice of concurrent currencies*. Second revised and enlarged edition. Institute of Economic Affairs; also in Hayek (1999), *The collected works of F. A. Hayek, Vol. 6, Good money. Part II: The standard* (pp. 128–229). The University of Chicago Press.

Hayek, F.A. (1991). *The collected works of F. A. Hayek, Vol. 3, The trend of economic thinking: Essays on political economists and economic history*. Ed. by W.W. Bartley III and S. Kresge. The University of Chicago Press.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

van der Linden, M., & Hubmann, G. (Eds.). (2018). *Marx's Capital: An unfinishable project?* Brill.

Marx, K. ([1857–1858] 1973). *Grundrisse: Foundations of a critique of political economy*. Penguin.

Marx, K. ([1857–1858] 2017), *Marx-Engels-Gesamtausgabe IV,* vol. 14, *Karl Marx: Manuskripte zum zweiten Buch des "Kapitals" 1868 bis 1881* [MEGA^2, IV/14]. Akademie Verlag.

Marx, K. ([1859] 1970). *A contribution to the critique of political economy*. With an introduction by M. Dobb. Progress Publishers.

Marx, K. ([1867] 1907a). *Capital, Volume I*. C.H. Kerr.

Marx, K. ([1885] 1907b). *Capital, Volume II*. C.H. Kerr.

Marx, K. ([1894] 1909). *Capital, Volume III*. C.H. Kerr.

Marx, K. ([1861–63] 1969). *Theories of surplus value*. [Volume IV of *Capital* in three parts]. Lawrence & Wishart. Reprint 1972.

Minsky, H.P. (1994). Marxian economics: A centenary appraisal (12-15-1994). *Hyman P. Minsky archive*, Paper 170. http://digitalcommons.bard.edu/hm_archive/170

Mori, K. (2017). Einführung zu [Introduction to] *Marx-Engels-Gesamtausgabe IV*, vol. 14 [MEGA^2 IV/14] (pp. 511–559). Akademie Verlag.

Mori, K. (2018a). Karl Marx's *Books of crisis* and the concept of double crisis: A Ricardian legacy. In M. van der Linden & G. Hubmann (Eds.), *Marx's Capital: An unfinishable project?* (pp. 206–227). Brill.

Mori, K. (2018b). The *Books of crisis* and Tooke–Newmarch excerpts: A new aspect of Marx's crisis theory in MEGA. *European Journal of the History of Economic Thought, 25*(5), 912–925.

Mori, K. (2019). Karl Marx's *Books of crisis* and the production theory of crisis. In H.-M. Trautwein (Ed.), *Studien zur Entwicklung der ökonomischen Theorie, Vol. XXXIV, Neue Perspektiven auf die politische Ökonomie von Karl Marx und Friedrich Engels: Die Marx-Engels-Gesamtausgabe (MEGA)* (pp. 81–111). Duncker & Humblot.

Paula, J. A., Cerqueira, H., Cunha, A., Suprinyak, C., De Deus, L., & Albuquerque, E. (2013). Notes on a crisis: The *Exzerpthefte* and Marx's method of research and composition. *Review of Radical Political Economics, 45*(2), 162–182.

Paula, J. A., Cerqueira, H., Cunha, A., Suprinyak, C., De Deus, L., & Albuquerque, E. (2016). Investigating financial innovation and stock exchanges: Marx's Notebooks on the crisis of 1866 and structural changes in capitalism. In R. Hecker, R. Sperl, & C.-E. Vollgraf (Eds.), *Beiträge zur Marx-Engels-Forschung, Neue Folge 2014/2015. Zu den Studienmaterialien von Marx und Engels* (pp. 194–217). Hamburg.

Thornton, H. ([1802] 1939). *An enquiry into the nature and effects of the paper credit of Great Britain.* With an introduction by F.A. von Hayek. A.M. Kelley. Reprint 1978.

Wicksell, K. ([1898] 1936). *Interest and prices: A study of the causes regulating the value of money.* Translated by R.F. Kahn, introduction by B. Ohlin. A.M. Kelley. Reprint 1965.

Wilson, J. (1847). *Capital, currency and banking*. The Economist.

Financial Instability and Crises in Keynes's Monetary Thought

Jörg Bibow

1 Introduction

Following the Rome conference theme, "Financial instability, market disruptions and macroeconomics: lessons from economic history and the history of economic thought," I will attempt to shed light on these issues by revisiting Keynes's writings from *Indian Currency and Finance* (1913) to *The General Theory* (1936). The analysis reveals Keynes's astute concerns about the stability/fragility of the banking system, especially under deflationary conditions.

The analysis proceeds as follows. Section 2 investigates Keynes's early monetary work, *Indian Currency and Finance*, which shows Keynes's keen concerns for financial instabilities and his appreciation of the important role of central banks as lender of last resort. Section 3 revisits Keynes's internationally influential attack on the Versailles Treaty, *The Economic Consequences of the Peace* (1919), in which Keynes analyzes the challenges posed by debt overhangs in the aftermath of World War I and the deflationary strategy for their resolution that is the core of the infamous "peace" treaty. Section 4 examines Keynes's *A Tract on Monetary Reform* (1923) and *A Treatise on Money* (1930). The background to the former work is the price-level instability experienced in Britain and elsewhere in the aftermath of WWI. The background to the latter work is Britain's stagnation in the second half of the 1920s following the country's return to gold at sterling's prewar parity. Britain's specific struggles with its chosen internal devaluation in the context of a lively world economy in the second half of the 1920s were very different from the worldwide deflationary environment of the Great Depression that formed the background to *The General Theory*.

J. Bibow (✉)
Department of Economics, Skidmore College, New York, USA
e-mail: jbibow@skidmore.edu

Section 5 takes a close look at Keynes's assessments and writings during the Great Depression, revealing his deep understanding of the havoc wreaked by deflation in banking systems—an issue pushed aside by assumption in *The General Theory*. Section 6 concludes.

2   Early Concerns Regarding Banking Instabilities

Keynes's *Indian Currency and Finance*, published in 1913, concerns both India's currency and exchange rate arrangements and its financial structure. Keynes describes India's banking system as a dual one, featuring both Western-style banks and "native" financial institutions (Shroffs, Marwaris, and other private bankers and moneylenders), with both strands of the system operating in parallel and without full arbitrage of financial conditions between them. While Keynes assesses India's exchange rate arrangements, with its gold exchange standard linked to sterling and the City of London, as suitable and progressive, he identifies various vulnerabilities of its evolving banking system compared to the situation in Britain and elsewhere.

According to Keynes, the contemporary banking situation in England had developed into a stable "checking currency" system in which checks facilitated payments and gold did not circulate but only served as a reserve for external purposes. Bank runs had become infrequent as depositors' confidence in the system had grown sufficiently, while the Bank of England had perfected its toolset for managing sterling's gold anchor (Keynes, [1913] 1971, p. 13).

By contrast, while chiefly satisfied with its gold exchange standard, Keynes was concerned about the stability of India's banking system and the availability of finance for facilitating the country's development.

Externally, the situation was characterized by a good measure of stability at the time, in principle. Nevertheless, Keynes was worried that stability might breed instability and that short-term foreign indebtedness was inherently risky. He observes that:

> the business of financing Indian trade, so far as it is carried out by banks with their seat in London, is in the hands of a very small number of banks. They stand, broadly speaking, in an exceedingly strong financial position supported by large reserve funds. In this matter India is now enjoying the fruit of past disasters and of conditions in which the struggle for existence was too keen to allow any but the fittest to survive. If the present spell of prosperity lasts too long, she will no doubt lose it (Keynes, [1913] 1971, p. 147).

Keynes adds that "There is, prima facie, some danger to the stability of the Indian financial system in the fact that its money market is largely financed by funds raised, not permanently but for short periods, in a far-distant foreign center" (Keynes, [1913] 1971, pp. 149–150).

Internal financing arrangements were far more fragile in Keynes's view. The Indian rupee traditionally circulated in the form of silver coins. The note issue in circulation had grown significantly but remained wholly disassociated from banking.

As a result, India's currency was internally "absolutely inelastic" (ibid., p. 40), which was a serious problem since the demand for currency was strongly seasonal.

Banking deposits, too, had grown very strongly in recent times, but not so with banks' cash reserves. Paid-up bank capital ratios had also sharply declined. Keynes (ibid., p. 53) saw this as a serious source of fragility in view of India's history of confidence crises, bank runs, and deep-seated hoarding habits (featuring the "barren accumulation of gold").

Keynes discusses the role of banking regulation (minimum capital and liquidity requirements and lending restrictions) and identifies the sharp deposit expansion of native moneylenders, with their "hopelessly inadequate" cash reserves, as particularly dangerous:

> it is hard to doubt that in the next bad times they will go down like ninepins. If such a catastrophe occurs, the damage inflicted on India will be far greater than the direct loss falling on the depositors. The growth of banking habits in India is, of course, of the utmost importance to the country's economic development. A startling series of failures will do much to retard it (Keynes, [1913] 1971, p. 159).

The ultimate problem was that India was lacking a central bank. While the government was de facto fulfilling some central banking functions, Keynes considered this feature as the key source of weakness and potential instability (Keynes, [1913] 1971, p. 166). The absence of a proper central bank meant that India was lacking an effective lender of last resort in case of banking crises. Keynes states that he:

> would emphatically apply to India the well-known doctrine which the powerful advocacy of Mr. Bagehot raised in England many years ago to an impregnable position in the unwritten constitution of this country—the doctrine, namely, that in a time of panic the reserves of the Bank of England must, at a suitably high rate, be placed at the disposal of the public without stint and without delay (Keynes, [1913] 1971, p. 115).

Keynes's prominent role on the "Royal Commission on Indian Finance and Currency" that was set up in May 1913 allowed him to address the issue of establishing a central bank in India in an annex of the committee's report—which lent its support to the matter.

When Britain came off gold for good in September 1931, India and the remainder of the British Empire (except for South Africa) stayed aligned with sterling rather than gold on that occasion. De Cecco (1974) highlights India's critical role within the empire of earning a dollar surplus. "Sterling balances," short-term debts accumulated by Britain largely during World War II within the empire, became a major challenge in the Anglo-American negotiations concerning the postwar settlement and involving Keynes—when America insisted on abolishing trade preferences and restoring sterling convertibility (too) early, which ushered in the Sterling Crisis of 1947.

By the time of Keynes's death in 1946, Britain was no longer in the favorable creditor position Keynes had emphasized in 1913, identifying an asymmetry in creditor-debtor relations that became central in his thinking about international relations:

The position of a country which is preponderantly a creditor in the international short-loan market is quite different from that of a country which is preponderantly a debtor. In the former case, which is that of Great Britain, it is a question of reducing the amount lent; in the latter case it is a question of increasing the amount borrowed (Keynes, [1913] 1971, p. 13).

3 Early Concerns Regarding Creditor-Debtor State Relationships

Like India, Germany provided another recurrent theme in Keynes's theorizing and policy advising on world affairs, starting with *The Economic Consequences of the Peace*, published in 1919. Central to Keynes's critique of the Treaty of Versailles were excessive debt obligations, government debt obligations arising from the war, and imposed war reparations.

Keynes begins the work with describing the high degree of economic integration that the world and specifically Europe had reached prior to WWI (also known as the era of globalization, mark 1). He argues that the victors should practice magnanimity and solidarity rather than vengeance. Reviving prosperity in war-ridden Europe—rather than deliberately starving and disintegrating the continent—should feature Germany as an integral part (Keynes, [1919] 1971, pp. 186–187). It was going to take another world war to finally see world leaders approach the matter in the spirit proposed by Keynes in 1919 in the aftermath of WWII.

Instead, the Treaty of Versailles imposed on Germany an external debt that exceeded the indemnity France paid to Germany in 1871 by a factor of 13. The treaty meant that Germany lost significant parts of its territory and population, as well as its merchandise fleet, colonies, and other foreign assets. Yet, it was obliged to deliver trade surpluses for decades to come while facing new trade restrictions. Germany would have needed to go through a massive internal devaluation to deliver the goods and pay off the debts.

But colossal adjustment would not befall Germany alone. The recipient countries of German transfers would need to adjust, too, to enable Germany to run huge and persistent trade surpluses. Their demonstrated unwillingness to do so underlined that Versailles provided the script for general deflation, which would hardly make the overshadowing debt problem go away—but quite the opposite.

But the deeper underlying problem was that it was not only Germany that was made to owe huge debts to its victors. Rather, Keynes observes in the final chapter of the book:

The war has ended with everyone owing everyone else immense sums of money. Germany owes a large sum to the Allies; the Allies owe a large sum to Great Britain; and Great Britain owes a large sum to the United States. The holders of war loan in every country are owed a large sum by the state; and the state in its turn is owed a large sum by these and other taxpayers. The whole position is in the highest degree artificial, misleading, and vexatious. We shall never be able to move again, unless we can free our limbs from these paper shackles. [. . .] As regards internal debt, I am one of those who believe that a capital levy for the extinction of debt is an absolute prerequisite of sound finance in every one of the

European belligerent countries. But the continuance on a huge scale of indebtedness between governments has special dangers of its own (Keynes, [1919] 1971, pp. 177–178).

The general (debt) deflation implanted in the Treaty of Versailles was temporarily avoided by two factors. The Wall Street lending machine not only fired up asset prices in America but also enabled a boom even in Germany in the second half of the 1920s[1]—until a sudden stop/reversal in capital flows turned boom into bust and saw Germany achieve trade surpluses when Chancellor Heinrich Brüning finally imposed on Germany the deflation provided for at Versailles 10 years earlier. Germany's reparation obligations were reduced in several renegotiations until they finally ended in 1932—when the Great Depression was ravaging.

4   Internal Equilibrium in Theory and British Practice: From the *Tract* to the *Treatise*

Keynes's *A Tract on Monetary Reform* analyzes the violent disturbances to the price level that occurred in Britain and other countries during WWI and its aftermath. Keynes's key policy advice is that monetary policy should primarily aim at maintaining internal equilibrium rather than be tied to some external commitment such as the gold standard. Acknowledging important changes in global balance and power, Keynes argues that the latter would mean having British monetary policy largely determined by the US Federal Reserve Board. Instead, the Bank of England should apply its monetary policy foremost to stabilizing the credit cycle and price level. In addition, some degree of exchange rate stability would still be attainable through central bank cooperation, Keynes argues.

Of greatest relevance in view of the conference theme is the first chapter of the book in which Keynes investigates the consequences of inflation and deflation on the distribution of wealth and income and economic activity. Keynes argues that the economic organization of modern societies is based on money contracts and presupposes for its proper and fair functioning that the value of money is held fairly stable.

Historically, price stability has not been the norm though. Even the nineteenth-century gold standard era saw periods of significant inflation and deflation. But the experience of relative stability over the period as a whole may have encouraged the illusion of its permanence. Considering long periods, the experience has been one of inflation and progressive deterioration in the value of money, driven by "the impecuniosity of governments and the superior political influence of the debtor class" (Keynes, [1923] 1971, pp. 8–9). What has changed under the "phase of capitalism, as developed during the nineteenth century, [is that] many arrangements

---

[1] "The United States lends money to Germany, Germany transfers its equivalent to the Allies, the Allies pay it back to the U.S. government. Nothing really passes—no one is a penny the worse" (Keynes, [1926a] 1978, p. 281).

were devised for separating the management of property from its ownership" and the emergence of an elaborate "investment system" (read: financial system) based on money contracts and featuring the "payment of fixed sums of money over a long period of time" (ibid., p. 4).

Keynes divides society into three classes: the "investing class" (read: wealth owners), the business class, and the (wage) earner. Concerning the distribution impact, Keynes, ([1923] 1971, p. 29) concludes that those who "in good faith have committed their savings to titles to money rather than to things" bear the brunt of inflations that may initially benefit business men but that once the "atmosphere of confidence" is destroyed, the real damage arises from the curtailment of investment and capital accumulation. His analysis then elaborates further on the diminution of the production of wealth that inflations and deflations are likely to cause, distinguishing price-level changes, expectations of price-level changes, and the degree of confidence in which such expectations are held. Keynes concludes that:

> inflation is unjust and deflation is inexpedient. Of the two perhaps deflation is, if we rule out exaggerated inflations such as that of Germany, the worse; because it is worse, in an impoverished world, to provoke unemployment than to disappoint the *rentier*. But it is not necessary that we should weigh one evil against the other. It is easier to agree that both are evils to be shunned. The individualistic capitalism of today, precisely because it entrusts saving to the individual investor and production to the individual employer, presumes a stable measuring-rod of value, and cannot be efficient—perhaps cannot survive—without one (Keynes, [1923] 1971, p. 36).

What is surprising in Keynes's analysis is that, while he highlighted the importance of the existence of the financial system, he does not consider any impacts of price-level changes on financial institutions but focuses on wealth owners and debtors only. Supposedly borrowers protect themselves from deflation through default, possible repercussions of which Keynes fails to discuss here except for the generally depressing effect of deflation on economic activity and employment as such, which will likely feature defaults on debts as well.

Monetary policy should stabilize the internal price level and avoid cyclical fluctuations in "business, prices, and employment" (Keynes, [1923] 1971, p. 138). Controlling banks' credit creation is a central part of monetary policy mainly operating through "bank rate" policy, but the central bank must also be ready to adjust the liquidity in the system by varying its own investments. He advises the central bank coordinates with the treasury, since short-term government debts are close substitutes for central bank deposits (ibid., pp. 141–142). But Keynes's analysis of the operational side of monetary policy remains rudimentary, and financial stability concerns are not part of his investigation in the *Tract* either (ibid., p. 148).

Against Keynes's advice in the *Tract*, Britain returned to gold at the prewar parity in 1925, only to depart from gold again in 1931. The widespread nostalgia for gold was based on illusions, as Keynes explained in the *Tract*: "most important of all—in the modern world of paper currency and bank credit there is no escape from a 'managed' currency, whether we wish it or not; convertibility into gold will not alter

the fact that the value of gold itself depends on the policy of the central banks" (ibid., p. 136).

The Economic Consequences of Mr. Churchill, published in 1925 in response to Britain's return to gold at the prewar parity, provides an important link between the *Tract* and the *Treatise*, and it captures the main event that was shaping Britain's economic performance during that time. Returning to gold at the prewar parity meant leaving traded-goods industries uncompetitive. Restoring equilibrium would involve deflation and unemployment (Keynes, [1925] 1972).

Keynes had diagnosed inflations and deflations as economically and socially harmful in the *Tract*. That some prices adjust faster than others and that economies may get stuck in disequilibrium for longer periods of time were the central themes of the *Treatise*. Deflation as a cumulative process, featuring stagnation and protracted unemployment, was the peculiar British experience at the time. The background and experience of the 1930s were a different one: it was no longer merely Britain's need to deflate in order to restore equilibrium (internal devaluation) that was the issue; now, it was deflation all round—leading to deeper deflation and depression all round. It was the theoretical framework of the *A Treatise on Money* that provided the theoretical lens through which Keynes viewed the early stages of the Great Depression.

Keynes's *Treatise* is mainly an investigation into the causes of, and ways to control, the business cycle (or credit cycle). Echoing his frustration as expressed in his earlier *Tract* about the limited practical usefulness of the quantity theory of money, Keynes sets out to devise a dynamic theoretical framework suitable for analyzing states of transition and disequilibrium. His so-called fundamental equations, featuring distinct concepts of investment and saving, provide the core of his theoretical apparatus. The key idea is that disequilibria between investment and saving (resembling disequilibria between spending in the economy or business receipts on the one hand and business costs on the other) are the drivers of "profit inflation/deflations" which, in turn, constitute the primary force behind the business cycle, in Keynes's view.

Keynes uses Wicksell's notion of the "natural rate" of interest and disparities between the natural rate and "market rate" of interest, as an alternative way of depicting dynamic cumulative processes. The alternative mode of presentation more easily lends itself to capturing Keynes's foremost public policy concern: managing the banking system in such a way as to stabilize credit and the economy by keeping the market rate in line with the natural (i.e., equilibrium) rate of interest. Since "booms and slumps are simply the expression of the results of an oscillation of the terms of credit about their equilibrium position" (Keynes, [1930a] 1971, p. 165), where:

> credit is the pavement along which production travels; and the bankers, if they knew their duty, would provide that transport facilities to just the extent that is required in order that the productive powers of the community can be employed to their full capacity. It has been a principal object of this treatise to give a clear answer to these perplexities. What is the true criterion of a creation of credit which shall be non-inflationary (free, that is to say, from the taint of profit inflation—income inflation is a different matter)? We have found the answer to

lie in the preservation of a balance between the rate of saving and the value of new investment (Keynes, [1930b] 1971, p. 197).

Despite all the attention paid to institutional and behavioral banking matters, the *Treatise* is quiet on financial instability. Keynes investigates the credit cycle, not financial crises. There are rare hints sprinkled here and there about the Wall Street crash of 1929 and its immediate aftermath. But there is only a brief section titled "The Slump of 1930" at the end of the second volume in which Keynes offers his account of how developments over the 1920s led to said slump that marked the early stage of the Great Depression.

In Keynes's narrative of events, the natural rate, which was temporarily elevated in the aftermath of WWI, declined markedly in the second half of the 1920s, while the market rate remained stuck at inappropriately high levels. The latter phenomenon was due to both misguided monetary policies and special market factors. He singles out borrowings by governments under treaty obligations (distress borrowers) and attempts to restore the gold standard as a factor raising central banks' demand for gold, accompanied by long-term borrowings by governments and banks for the purpose of building up liquid reserves (in terms of dollar and sterling balances), as factors leading to tight credit conditions facing genuine business borrowers. As a third class of "artificial" borrowers, becoming a formidable force in 1928–1929, Keynes identifies speculative borrowers participating in the stock market boom on Wall Street (and elsewhere). The divergence between the market rate of interest and the natural rate set off deflationary forces which were reinforced by the "slump psychology in the minds of entrepreneurs" (Keynes, [1930b] 1971, pp. 341–342). He warns that profit deflation will push entrepreneurs to seek recourse to an "assault on the money incomes of the factors of production [. . .] a dangerous enterprise in a society which is both capitalist and democratic" (ibid., p. 346). He urges to apply "bank-rate policy and open-market operations à outrance" (ibid., pp. 346–347) in order to effect a proper adjustment of the market rate of interest. But the *Treatise* is mainly about the stagnation experience of the 1920s rather than the world crisis of the 1930s.

5 Keynes as an Investor and Economic Commentator During the Great Depression

One finds Keynes's interpretation and explanation of unfolding events beginning with the Wall Street crash of 1929 in his contemporary writings and commentaries. His theoretical framework of the *Treatise* informed his views of developments at the early stages of the Great Depression. But given the speed at which he moved on from the *Treatise* to *The General Theory*, his perspective at the peak of the crisis in 1932–1933 was already closer to his later work. His other informative writings are partly related to Keynes's own financial investment activities and partly to his journalistic activities, speeches, and correspondence.

While the Wall Street crash of October 1929—just before Irving Fisher (1929) had famously declared that "stock prices have reached what looks like a permanently high plateau"—is widely seen as the starting shot of the Great Depression, there was some controversy in the years leading up to the precrash market peak in 1929 about stock market overvaluation, inflation risks, and Federal Reserve policy that Keynes participated in.

An early example is Keynes's memo of 14 April 1927, prepared for a National Mutual Board meeting, in which he disagrees with a bearish account on the American stock market's position by his long-time investment partner O.T. Falk. In Keynes's view, "none of the usual symptoms preceding serious malaise exist at the present time" (Keynes, 1927).[2] However, the expectation of a declining rate of interest appears to be a central part of his reasoning: "a steady downward tendency of the rate of interest is to be anticipated [. . .] I do not see much risk of a movement of interest upwards" (Keynes, 1927).

In his memo "Is There Inflation in the United States" (dated 1 September 1928), Keynes returns to the US position, now applying his *Treatise* concepts of industrial circulation and financial circulation. Keynes attests that stock prices were "prima facie, judged by dividend yield and past records, [. . .] extremely dear" (Keynes, [1928] 1973, p. 57). His key concern however was that aggressive tightening by the Federal Reserve would cause a slump in (real) investment, choking the economy. Keynes argues that rising stock prices as such do not represent inflation and that any effects on the economy and inflation would only arise indirectly by (over)stimulating investment. The impact of bank lending on asset prices is complex, as banks may either boost prices by buying assets themselves or lending to players who do or merely be facilitating "difference of opinion" (contrasting views within the markets) by lending to bullish investors and simultaneously providing term deposits that satisfy the liquidity preference of bearish stock sellers. He judges that short-term loans to the stock market are always a factor but rarely a decisive one and that the "existence of a powerful opposition to the bull market is a moderating influence on investment." He warns that expansion of finance-related dealings (financial circulation) can even have a deflationary effect on the economy—unless accommodated by the Federal Reserve—namely, if financing Wall Street curtails business loans (industrial circulation). He advises against attacking the rising stock market by monetary tightening: "Continued monetary stringency might easily bring about [. . .] a depression, or at least the anticipation of one. But it can hardly be the intention of the [Federal Reserve Board] to bring about a business depression" (ibid., p. 58).

Ten months later, Keynes once again responded to Falk's outlook for US stocks, which had turned bullish again in the meantime. In a memo dated 2 July 1929, Keynes observes the following: "I seem to be destined to disagree with Mr. Falk about the prospects of the American market. [. . .] Taking everything into account,

---

[2] All unpublished Keynes sources quoted in this paper are held in the King's College Archive Centre, Cambridge, The Papers of John Maynard Keynes, JMK, quoted with permission.

the market now seems to me a much more dangerous one for Bulls than I thought it [ten months ago]." He concludes that:

> to enter the market at this moment is to enter it at a time when prices are discounting the future to an unexampled extent, yet when the present is so good that the economic body will surely have to take a rest before it will have the strength for yet a further leap forward. That the United States often does the unexpected is undoubted. But the market seems to me a dangerous one, relatively to the prospects of large further gains, for newcomers to enter at this state (Keynes, 1929).

Wall Street crashed 3 months later in October 1929, and the US economy was to take a sharp turn for the worse in due course—with global repercussions that became known as the Great Depression.

Just around that time, in late October 1929, the Dutch company Philips invited Keynes to contribute to the Economic Intelligence Service (EIS), reporting mainly but not exclusively on the British Empire situation. Starting in January 1930, Keynes wrote monthly reports for them until the end of 1931, followed by quarterly reporting in the years 1932–1934. These reports offer valuable insights into his contemporary assessments of the evolving situation during the Great Depression.

Keynes's first EIS letter of January 1930 titled "The Position in Great Britain" attests that the world economy would require "a prolonged period of *very* cheap money" and expresses particular concern about "the state of the main overseas markets, which is almost wholly due to the reduced volume of loans issued on the international investment markets and to falling commodity prices. Both these troubles are traceable to the almost unprecedentedly dear money prevailing throughout the world in the middle of 1929, coming on top of a situation which was already none too good owing to the previous fall of commodity prices" (Keynes, 1930c). Keynes added that he was particularly disturbed about the outlook in Germany for 1930.

His May 1930 EIS letter starts:

> Up to Easter it appeared as if the situation might be stabilizing itself. But since Easter there has been a serious setback in every respect. We have to adjust our opinions to the fact that we are undoubtedly in the middle of a major depression at least as serious as the major depressions of the pre-war period, such for example as 1907. In fact, the extent and rapidity of the fall of international wholesale prices has been greater than on any occasion, apart from the 1921–22 slump, in the last seventy years (Keynes, 1930d).

He adds that, in contrast to Britain, in the United States, observers were:

> not pessimistic enough because they are inclined to regard [what is happening] as one of the several minor recessions which have been experienced since the war, and not as a major recession. Indeed, the over-optimism of the United States seems to me to be a real danger, since there may be a revulsion of feeling as it becomes clearer that a rapid recovery is scarcely to be hoped for (Keynes, 1930d).

The global situation deteriorated further in the second half of 1930. In September 1930, Keynes declared that "we are now in the middle of one of the major slumps of recent economic history. Indeed, if we measure the magnitude of events by the fall of prices from the average of the preceding three or four years, the extent and rapidity

of the fall are perhaps the most severe that have ever been experienced in modern times" (Keynes, [1930e] 1981, p. 390). By now, Keynes was especially alarmed about the situation in Germany and the United States. But he expected things to get worse in general. In a letter to Walter Case dated 16 December 1930, Keynes observes the following:

> My own up-to-date feeling as to the general position is one of extreme pessimism. Sometimes I think that we are only approaching the acute phase of the depression and that there may still be some shattering disillusions to come. However this may be, I cannot perceive the least reason in the world for expecting an early recovery (Keynes, 1930f).

Keynes elaborates on the disastrous consequences of deflation in his two-piece article titled "Great Slump of 1930." Stressing that deflation increases the burden of the debt, he focuses on "bonded debt" and international debt relationships rather than banking. He questions "whether the necessary adjustments could be made in time to prevent a series of bankruptcies, defaults, and repudiations which would shake the capitalist order to its foundations [and warns:] Here would be a fertile soil for agitation, seditions, and revolution" (Keynes, [1930g] 1972). He advises that internationally coordinated and courageous central bank action by the three key creditor countries (the United Kingdom, the United States, France) was needed to stop the slump (ibid., pp. 145–146).

But his outlook was quickly turning more gloomy. His February 1931 EIS letter states the following:

> previous letters were pessimistic but not pessimistic enough. [. . .] The prospect of long series of defaults during 1931 is not to be excluded. [. . .] I do not see the slightest signs of the foundations of recovery being laid anywhere. It is at least possible that the position may get worse, or if not that a *long* interval may elapse before it gets really better (Keynes, 1931a).

In his May 1931 EIS letter, Keynes observes the following:

> I have been forecasting for some time that the American market will suffer at least one further disillusion, and that this might have a serious effect on prices. This disillusionment has now duly matured. The foolish and quite unfounded hopes of an early recovery have now been abandoned by almost everyone. There is indeed, no rift in the clouds. I still see no reason whatever for any real recovery in world prosperity in the near future. It would be safer to base all plans on this assumption than on more optimistic hopes (Keynes, 1931b).

His June 1931 EIS letter, his last one before his visit to the United States, highlights the renewed plunge in US and global stock markets as expressions of "panic":

> In many cases the decline of prices has gone much further than on previous occasions of slumping values in the last year or two; and for the first time it is possible to say that what can only be described as panic prices, having little or no relation to intrinsic values, are prevailing. This suggests that we are now entering the crisis, or panic, phase of the slump. I am inclined to think that when we look back on this particular slump we shall feel that this phase has been reached in the summer months of 1931, rather than at any earlier date. That is to say, I should expect the climax of the slump to be reached in the course of the next ensuing months. It may very well be that the worst point of the crisis will be precipitated by a rapid deterioration of the German position. The difficulties of the Credit-Anstalt in Vienna have produced a great impression on high financial circles in London and America, and have given them a great shock (Keynes, 1931c).

The American Visit, 1931

Keynes developed a much clearer picture of the severity of the crisis and the centrality of banking in it when he visited the United States in June–July 1931, departing from Britain on 30 May 1931 (Skidelsky, 1992). But the situation in Europe was also fast deteriorating following the collapse of Vienna's Credit Anstalt Bank earlier in May, which saw banking problems spreading across Austria, Germany, and Central Europe. Even before Keynes's return to Britain in mid-July, market volatility had reached London, with sterling coming under pressure on the foreign exchanges and money market rates rising as the Bank of England was losing gold.

In the United States, Keynes gave a number of lectures; participated in discussion groups; and met with economists, central bankers, bankers, government officials, and other informed observers. He quickly became acutely aware that the US banking system was in a critical shape. The fragility of US banks arose from their risky and illiquid balance sheet positions—paired with a rising preference for safety and liquidity on the public's part. The risk was that deflation would produce more stress in the domestic US banking system, just as debtor countries might be driven into default owing to falling prices of the products they export.

Keynes gave two lectures at the New School for Social Research in New York City on 15 and 18 June. In the first lecture, titled "Do We Want Prices to Rise?," he discusses the option of deflation as expediting "liquidation" versus reflation of prices. He firmly rejects the expediency of the former option, describing liquidation as a "polite phrase for general bankruptcy . . . national debts, war debts, obligations between the creditor and debtor nations, farm mortgages, real estate mortgages;—all this financial structure would be deranged by the adoption of Dr. Sprague's proposal [of deflation].[3] A widespread bankruptcy, default and repudiation of bonds would necessarily ensue. Banks would be in jeopardy" (Keynes, [1931d] 1981, p. 547). In his second lecture, titled "What Can We Do to Make Prices Rise?," he advises to "increase the quantity and reduce the cost of banking credit" (Keynes, [1931e] 1981, p. 550) but acknowledges that "the present depression may prove to be one of the longest on record [because] the rate of interest may need to be driven down a great way before we recover" (ibid., p. 553).

---

[3] Oliver Mitchell Wentworth Sprague was a Harvard economist and banking expert who had close ties with the Federal Reserve and from 1930 to 1933 also advised Bank of England Governor Montagu Norman. Keynes refers here to a speech Sprague had delivered in London in May 1933 in which he contrasted a "monetary school" and an "industrial and economic equilibrium" school which held opposing views on the causes of the crisis and policy prescriptions. While the former, of which Keynes in his speech identifies as a proponent, wanted central banks to boost their liquidity creation to counter financial fragility and stop and reverse deflation, the latter, including Sprague, argued that a general deflation of wages and prices had to run its course for the economy to equilibrate again. According to Sprague, the latter school represented the consensus view at the three leading (American, British, and French) central banks (see also Meltzer, 2003).

His subsequent Harris Foundation lecture series in Chicago provides more elaborate analyses of the investment slump and surge in unemployment. Excessively high interest rates feature prominently. According to Keynes, interest rates needed to decline as the expansion progressed but instead increased—for which he blames the Federal Reserve (Keynes, [1931f] 1973, p. 350).

In his policy advice on what in his view is "essentially a technical banking problem [. . .] [and] pre-eminently the business of the central banker," Keynes, ([1931g] 1973, p. 363) emphasizes the need to restore confidence and to reduce long-term interest rates. Additionally, he suggests that debt-financed public construction programs should be part of the solution.

In private correspondence with Hubert Henderson on 22 June 1931, Keynes appears to be far less optimistic that his suggested credit-easing advice could actually work because the banks are already in very poor shape, with a substantial proportion being insolvent, a situation that he had most underestimated before his visit. Their precarious position stems from their investments in high-risk bonds that have plunged in value, on the one hand, and "unrest amongst depositors," ready to run at any moment, on the other:

> This means that the banks in their turn are extraordinarily nervous, even those which are perfectly solvent, since they never know when they may have to support a run from their depositors. Accordingly they have an absolute mania for liquidity. They put pressure on their customers to repay loans, since loans and advances are non-liquid in an emergency. Generally speaking, they turn all the assets they can into a fairly liquid form and in some cases keep an abnormally large amount of till money. As long as this mentality exists on the part of depositors and banks, and it is obvious that in the circumstances it is entirely intelligible, since many banks are in fact not safe, whilst the members of the general public cannot tell which the dangerous ones are, it overshadows the whole situation. It is a large part of the explanation of the failure of the bond market to make more progress. Whenever the less saleable bonds improve a little in price and become more saleable, some bank takes the opportunity to get more liquid. It is indeed a vicious circle. The anxiety of the banks to get liquid keeps the bond market weak (I mean the bond market for second-grade bonds) and so long as the bond market is weak the position of the banks remains precarious. . . . It is the weakness of the banking system all over the country which primarily stands in the way of the usual remedy, cheap and abundant credit, failing to take effect (Keynes, [1931h] 1981, pp. 556–557).

In his speeches in Chicago, Keynes addresses the "Hoover Moratorium," which was declared on 20 June 1931 and effectively ended the Versailles reparations saga, deploying his *Treatise* model to explain the ongoing slump. His remarks at a round table on 1 July 1931 at the Harris Foundation Institute shed more light on why he considers banking fragility as particularly critical, highlighting that deflation will be even more damaging when bank intermediation rather than direct debt financing is involved:

> The more I hear about the situation here, the more I attach importance to [banking fragility]. I think one of the occasions of the excessive lending by banks imprudently against real estate is the fact that the volume of their deposits is so large in relation to the quantity of short-term sound assets that there are to go around that they are driven to the investment not entirely suitable for a banking system. Also, this makes the whole community much more vulnerable to a change in the value of money, because you have a whole lot of people owing a whole lot

of other people very large sums of money, and then holding assets against those, instead of there being a direct investment by the public in the assets; if you got the direct assets themselves, changes in the value of money would not be so dangerous, but if the ultimate owners of the wealth have it in terms of money, and somebody owes them that money and buys assets against it, then any change in the value of money sets up an appalling strain and it is a very dangerous position (Keynes, [1931i] 1981, p. 539).

Keynes returned to England on 18 July 1931, a few days after the publication of the Macmillan Report (13 July 1931) and Germany's default (on 15 July 1931), events that added to the pressures on London with its large net short-term debtor position. During his Atlantic crossing, Keynes prepared a memorandum for the Economic Advisory Council that he was a member of on economic conditions in the United States. Keynes expresses the belief that the Federal Reserve leadership leaned decidedly in favor of raising prices (the "monetary school") as opposed to letting deflation running its course (the "equilibriumists") but senses that they may be less confident than Keynes in their powers to accomplish their aim. Keynes views the US domestic situation primarily as a construction slump and again emphasizes the pivotal role of banking fragility: "The banks' ratio of capital and reserves to liabilities is often small, so that even moderate losses wipe it out" (Keynes, [1931j] 1981, p. 568). The trouble is that banking losses in the United States were far worse than moderate:

> The truth is that the financial structure of the United States is no more able than that of the rest of the world to support so terrific a change in the value of money. The vast growth of bank deposits and of bonded indebtedness in that country interposes a money contract between the real asset on the one hand and the ultimate owner of wealth on the other. A depreciation in the money value of the real asset, sufficient to cause margins to run off, necessarily tends to burst up the whole structure of money contract, particularly those short-term contracts represented by bank deposits. I think it would be true to say that the first preoccupation of Governor Meyer is to restore solvency and liquidity to his member banks and that the objective of restoring the level of output to normal is at present remote. [...] After observing this background, I understand much better than I did part of what lies behind Dr. Sprague's ["equilibriumist"] attitude, his feeling that no good can come until insolvencies have progressed much further and a large amount of monetary indebtedness wiped out. But if we are to proceed along these lines, how much of the financial structure would be left standing when we were done, I do not know. To counsel this way out seems to me to be, even for a philosopher, a counsel of despair; whilst for one attached to a central bank, it is suicidal, because it means sacrificing what the central banks exist to safeguard. But whilst we ought, in my judgement, to bend every effort to move the course of events in the opposite direction to that which Dr. Sprague recommends, it is not wise to underestimate the extraordinary difficulties in doing so (Keynes, [1931j] 1981, pp. 571–572).

Keynes's spontaneous reactions in speeches and correspondence to the unfolding US banking crisis were also captured in "The Consequences to the Banks of the Collapse of Money Values," published in August 1931. Keynes describes how expanding banking systems and bond and mortgage markets have established a "veil of money between the real asset and the wealth owner." While observers are generally "familiar with the idea that changes in the value of money can gravely upset the relative positions of those who possess claims to money and those who owe money," causing wealth transfers between creditors and debtors, Keynes draws attention to

another development that comes to the fore when there is a very large change in the value of money and a hefty decline in the money value of assets. Keynes describes that banks will normally leave a sizeable margin when lending against collateral (i.e., not make loans to the full value of the asset), and this margin of safety protects them against asset price declines. He then observes that for the first time in modern history, a "world-wide collapse over almost the whole field of the money values of real assets" has created a situation where "the 'margins' have run off" (Keynes, [1931k] 1972, p. 173).

Keynes argues that widespread deflation and collapse in asset prices left the economy in the grip of a credit crunch that risks further choking economic activity, even if general panic has not yet struck. Almost every asset class is part of the banks' problem. In many countries, business (working capital) loans are in the worst condition as "in present circumstances for many classes of producers of raw materials, of farmers and of manufacturers, there are no profits and every prospect of insolvencies, if matters do not soon take a turn for the better" (ibid., p. 175).

Keynes seems to put the blame squarely at the door of the monetary authorities for doing "too little, too late." His outlook seems grim when he attests that, by the summer of 1931, developments had reached the point where avoiding the worst had become "extraordinarily difficult":

> I believe that, if today a really conservative valuation where made of all doubtful assets, white a significant proportion of the banks of the world would be found to be insolvent; and with the further progress of Deflation this proportion will grow rapidly. Fortunately our own domestic British Banks are probably at present—for various reasons—among the strongest. But there is a degree of Deflation which no bank can stand. And over a great part of the world, and not least in the United States, the position of the banks, though partly concealed from the public eye, may be in fact the weakest element in the whole situation. It is obvious that the present trend of events cannot go much further without something breaking. If nothing is done, it will be amongst the world's banks that the really critical breakages will occur. ... The present signs suggest that the bankers of the world are bent on suicide. At every stage they have been unwilling to adopt a sufficiently drastic remedy. And by now matters have been allowed to go so far that it has become extraordinarily difficult to find any way out (Keynes, [1931k] 1972, pp. 176–178).

Keynes's July 1931 EIS letter sums up his by now very gloomy global outlook based on his US visit: "A somewhat intensive sight of present conditions in the United States has served to convince me that quite apart from European complications, it is quite out of the question that there should be anything which could be called a true recovery of trade at any time within, say, the next nine months. The necessary foundations for such a recovery simply do not exist" (Keynes, 1931l).

While his American visit had brought banking troubles to the center of his attention, upon returning home, Keynes's interest naturally shifted back to British economic conditions in general and the country's balance of payments and currency situation in particular. He writes to Richard Kahn on 13 August 1931, prophesying sterling's departure from gold "within a month unless heroic measures are taken" (Keynes, [1931m] 1981, pp. 594–595).

Sterling's departure from gold on 21 September 1931 enlarged Britain's policy space, and Keynes was going to call on the Bank of England to engage in actively

lowering longer-term interest rates in due course. Keynes was acutely aware that Britain had only bought itself some breathing space while remaining entrapped in the global depression, mentioning in a letter to Walter Case dated 2 November 1931 the following:

> I am sure that British optimism is at present carried too far by enthusiasm over the changed local situation and is paying too little attention to the continuance of depressed and dangerous conditions abroad. [...] So I expect the world depression to continue some time yet and one day Great Britain will wake up to the fact that there are strict limits to the degree of recovery which a country can gain from what is nothing but a change in tis domestic circumstances and in its relationship to the rest of the world (Keynes, [1931n] 1981, pp. 11–12).

The German Visit, 1932

Keynes's visit to Germany in January 1932, which included a lecture in Hamburg, titled "The Economic Prospects 1932," and a meeting with Chancellor Heinrich Brüning in Berlin, afforded him new insights into the depth of the crisis in that country. In an article titled "An End of Reparations?," published a week after his visit, he observes that Germany was "in the grip of the most terrible deflation that any nation had experienced" (Keynes, [1932a] 1978, p. 366). In his Hamburg lecture, Keynes remarked that it was surprising that Germany has not yet suffered a "collapse of her political and social organization." He also looks back and recalls his *The Economic Consequences of the Peace*, describing the Versailles Treaty as "one of the greatest errors of international statesmanship ever committed" (Keynes, [1932b] 1982, p. 46).

His Hamburg lecture confirms that at this point, Keynes saw the Great Depression foremost as a financial crisis rather than a severe business downturn. In fact, he now saw the risk that the financial structure might fully collapse. His assessment of the situation features an idea that appears here for the first time: a "competitive panic to get liquid" driving the general collapse in prices that, according to Keynes, "feeds on itself":

> We are now in the phase where the risk of carrying assets with borrowed money is so great that there is a competitive panic to get liquid. And each individual who succeeds in getting more liquid forces down the price of assets in the process of getting liquid, with the result that the margins of other individuals are impaired and their courage undermined. And so the process continues. It is, perhaps, in the United States that it has proceeded to the most incredible lengths. But that country only offers an example, extreme owing to the psychology of its people, of a state of affairs which exists in some degree almost everywhere. The competitive struggle for liquidity has now extended beyond individuals and institutions to nations and to governments, each of which tries to make its international balance sheet more liquid by restricting imports and stimulating exports by every possible means, the success of each on in this direction meaning the defeat of someone else. Moreover every country tries to stop capital development within its own borders for fear of the effect on its international balance. Yet it will only be successful in its object in so far as its progress towards negation is greater than that of its neighbors. Where and how is this ghastly internecine struggle to stop?

> [...] Through lack of foresight and constructive imagination the financial and political authorities of the world have lacked the courage or the conviction at each state of the decline to apply the available remedies in sufficiently drastic doses; and by now they have allowed the collapse to reach a point where the whole system may have lost its resiliency and its capacity for a rebound (Keynes, [1932b] 1982, pp. 39–41).

While the self-defeating "competitive panic to get liquid" risks collapsing the financial structure, Keynes was hopeful that Britain's abandonment of the gold standard would unleash positive consequence in Britain and beyond. Apart from stopping the general decline in prices in Britain and the sterling area, it meant focusing pressures on key countries occupying net creditor positions:

> The rest of the world owes them money. They will not take payments in goods; they will not take it in bonds; they have already received all the gold there is. The puzzle which they have set to the rest of the world admits logically of only solution, namely that the rest of us should find some way of doing without their exports. The expedient of continually reducing world prices failed; for prices were dragged down equally everywhere. But the expedient of exchange depreciation relatively to gold will succeed. Thus a process has been set moving which may relieve in the end the deflationary pressure. The question is whether this will have time to happen before financial organization and the system of international credit break under the strain. If it does, then the way will be cleared for a concerted policy of capital expansion and price raising—which one can call inflation for short—throughout the world. For the only alternative solution which I can envisage is one of the general default of debts and the disappearance of the existing credit system, followed by a rebuilding on quite new foundations (Keynes, [1932b] 1982, pp. 44–45).

Keynes subsequently reworked his Hamburg lecture and delivered it twice in Cambridge in February 1932. In the updated version, he elaborated on the "competitive struggle for liquidity," featuring in the financial panic by contrasting microeconomic and macroeconomic reasoning and providing more global macroeconomic context as well. Keynes observes the following:

> We have here an extreme example of the *disharmony* of general and particular interest. Each nation, in an effort to improve its relative position, takes measures injurious to the absolute prosperity of its neighbors; and since its example is not confined to itself, it suffers more from similar action by its neighbors than it gains by such action itself. Practically all the remedies popularly advocated today are o this internecine character. Competitive wage reductions, competitive tariffs, competitive liquidation of foreign assets, competitive currency deflations, competitive economy campaigns, competitive contractions of new development—all are of this beggar-my-neighbor description. The modern capitalist is a fair-weather sailor. As soon as a storm rises he abandons the duties of navigation and even sinks the boats which might carry him to safety by his haste to push his neighbors off and himself in. [...] For one man's expenditure is another man's income. Thus whenever we refrain from expenditure, whilst we undoubtedly increase our own margin, we diminish that of someone else; and if the practice is universally followed, everyone will be worse off. An individual may be forced by his private circumstances to curtail his normal expenditure, and no one can blame him. But let no one suppose that he is performing a public duty in behaving in such a way. An individual or an institution or a public body, which voluntarily and unnecessarily curtails or postpones expenditure which is admittedly useful, is performing an anti-social act (Keynes, [1932c] 1982, pp. 52–53).

In his final EIS letter of December 1934, Keynes declares that he was "inclined to be a little more optimistic about the possibilities of the first half of 1935" (Keynes,

1934). Keynes put his money where his mouth is. He had taken heavy losses on his private financial investments in the run-up to the Wall Street crash and Great Depression, primarily on commodity speculations that went awry. Keynes's private wealth peaked in 1936, the year *The General Theory* was published, owing primarily to concentrated investments in British equities.

Despite Keynes's deep understanding of what happened in the banking and global financial systems during the Great Depression, these issues are not developed or hardly addressed in *The General Theory*. Banking calamities are set aside by the assumption of a constant money stock that allows Keynes to focus on the liquidity preference of the general public and offer a general theory of employment rather than a special theory of great depressions (Bibow, 2009, 2020).

6   Concluding Remarks

This contribution examines Keynes's writings from 1913 to 1936 to distill his evolving views on financial instability.

Keynes showed keen concerns for financial instabilities in his early monetary work, *Indian Currency and Finance*. Regarding Indian banking, he emphasized its early stage of development, regulation, and the absence of a central bank as potential sources of vulnerability. From early on, Keynes appreciated the importance of a central bank and firmly upheld the Bagehot principle (Keynes, [1926b] 1981). In *Economic Consequences of the Peace*, international debtor-creditor relationships (and global imbalances) then took center stage. While his *Tract on Monetary Reform* emphasized price stability as vitally important and investigated the wreckage caused by inflations and deflations, the book is quiet on financial structure. Similarly, his *A Treatise on Money* is strong on details of financial structure and bank behavior, but the book is a work on (normal) business cycles rather than crisis economics. In many ways, *The General Theory* is even more down to basics, focusing squarely on what determines the level of employment at any time, while financial institutional detail and bank behavior are largely left working quietly behind the scenes, and financial instabilities are barely mentioned.

It is his writings from the early 1930s, especially those related to his American visit in 1931 and German visit in 1932, that provide clear evidence of Keynes's deep understanding of the banking and financial troubles existing at the time. Based on his explorations into bank money and bank behavior in the *Treatise*, Keynes exposed a clear grasp of how debt, banking, and deflation interact in a (global) "competitive panic for liquidity" that can wreck the financial structure and much else besides.

These calamities were part of the background that informed Keynes's theorizing as he was traveling toward his final major monetary work: *The General Theory*. That all around deflation is very unlikely to stabilize economies equipped with a banking system was made all too obvious to miss at the time by unfolding events. That actions that can be rational and effective when practiced by individual actors or small

units may turn viciously destabilizing when followed universally informs the new macroeconomics of *The General Theory*.

But in that work, the scenario of a "competitive panic for liquidity" represents only an extreme possibility in liquidity preference theory. The theory concerns the general insight that financial conditions as determined in the financial system under the guidance of the monetary authority are shaping economic realities—rather than any real forces providing uniquely determined anchors for the system, with finance as a mere reflection of that economic reality.

References

Bibow, J. (2009). *Keynes on monetary policy, finance and uncertainty: Liquidity preference theory and the global financial crisis*. Routledge.

Bibow, J. (2020). *The general theory* as "Depression Economics"? Financial instability and crises in Keynes's monetary thought. *Levy Economics Institute Working Paper* no. 974, October. Levy Economics Institute of Bard College.

De Cecco, M. (1974). *Money and empire: The international gold standard, 1890–1931*. Basil Blackwell.

Fisher, I. (1929). Comment in a speech at the monthly dinner of the Purchasing Agents Association at the Builders Exchange Club, New York. *New York Times*, 16 October.

Keynes, J. M. ([1913] 1971). *Indian currency and finance*. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 1. Macmillan.

Keynes, J. M. ([1919] 1971). *The economic consequences of the peace*. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 2. Macmillan.

Keynes, J. M. ([1923] 1971). *A tract on monetary reform*. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 4. Macmillan.

Keynes, J. M. ([1925] 1972). *The economic consequences of Mr. Churchill*. The Hogarth Press. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 9: *Essays in Persuasion* (pp. 207–230). Macmillan.

Keynes, J. M. ([1926a] 1978). The progress of the dawes scheme. *The Nation and Athenaeum*, 11 September. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 18: *Activities 1922–1932: The end of reparations* (pp. 277–282). Macmillan.

Keynes, J. M. ([1926b] 1981). Bagehot's Lombard Street. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 19(2): *Activities 1922–1929: The return to gold and industrial policy. Part 2* (pp. 465–472). Macmillan.

Keynes, J. M. (1927). Memorandum by Mr. J.M. Keynes for Board Meeting. 20 April. Keynes Papers: NM/2/1/154–8.

Keynes, J. M. ([1928] 1973). Is there inflation in the United States? As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 13: *The general theory and after: Part 1, Preparation* (pp. 52–59). Macmillan.

Keynes, J. M. (1929). Untitled memorandum. 2 July Keynes papers: NM/2/3/14–16.

Keynes, J. M. ([1930a] 1971). *A treatise on money: The pure theory of money*. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 5. London: Macmillan.

Keynes, J. M. ([1930b] 1971). *A treatise on money: The applied theory of money*. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 6. Macmillan.

Keynes, J. M. (1930c). The position in Great Britain. Economic intelligence service, 7 January. Keynes Papers: BM/2/5–11.

Keynes, J. M. (1930d). Untitled letter. Economic intelligence service, 6 may. Keynes papers: BM/2/5/24–26.

Keynes, J.M. ([1930e] 1981). The future of the rate of interest: Prospects of the bond market. *The index*, Sept., 178–186. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 390–399). Macmillan.

Keynes, J. M. (1930f). Letter to Walter Case. 16 December. Keynes papers: BM/2/39.

Keynes, J. M. ([1930g] 1972). The great slump of 1930. *The Nation and Athenaeum*, *48*, Part 1: 20 December, 402; Part 2: 27 December, 427–428. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 9: *Essays in Persuasion* (pp. 126–134). Macmillan.

Keynes, J. M. (1931a). February letter. Economic intelligence service, 11 February. Keynes papers: BM/5/74–76.

Keynes, J. M. (1931b). May letter. Economic intelligence service, 5 may. Keynes papers: BM/5/90–94.

Keynes, J. M. (1931c). June letter. Economic intelligence service, 29 may. Keynes papers: BM/5/95–97.

Keynes, J. M. ([1931d] 1981). Do we want prices to rise? Lecture at the new school for social research, New York, 15 June. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 544–553). Macmillan.

Keynes, J. M. ([1931e] 1981). What can we do to make prices rise? Lecture delivered at the new school for social research, New York, 18 June. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 549–553). Macmillan.

Keynes, J. M. ([1931f] 1973). The originating causes of world-unemployment. Lecture 1, delivered at the Harris Foundation, Chicago, June. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard* Keynes, Vol. 13: *The general theory and after: Part 1, Preparation* (pp. 343–352). Macmillan.

Keynes, J. M. ([1931g] 1973). The road to recovery. Lecture 3, delivered at the Harris Foundation, Chicago, June. As reprinted in D. Moggridge (Ed.), *The Collected Writings of John Maynard Keynes*, Vol. 13: *The General Theory and After: Part 1, Preparation* (pp. 358–367). Macmillan.

Keynes, J. M. ([1931h] 1981). Letter to Hubert Henderson, 22 June. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 555–557). Macmillan.

Keynes, J. M. ([1931i] 1981). Discussion leader at Harris Foundation roundtable, 1 July. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 529–544). Macmillan.

Keynes, J.M. ([1931j] 1981). Economic Advisory Council: A note on economic conditions in the United States. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 561–588). Macmillan.

Keynes, J.M. ([1931k] 1972). The consequences to the banks of the collapse of money values. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 9: *Essays in Persuasion* (pp. 150–158). Macmillan.

Keynes, J.M. (1931l). July 1931 letter. Economic intelligence service, undated. Keynes papers: BM/5/98–104.

Keynes, J.M. ([1931m] 1981). Letter to R.F. Kahn, 13 August. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 594–595). Macmillan.

Keynes, J. M. ([1931n] 1981). Letter to O.T. Falk, 18 September. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 20: *Activities 1929–1931: Rethinking employment and unemployment policies* (pp. 611–612). Macmillan.

Keynes, J. M. ([1932a] 1978). An end of reparations? *The New Statesman and Nation*, 16 January, 57–58. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 18: *Activities 1922–1932: The end of reparations* (pp. 366–369). Macmillan.

Keynes, J. M. ([1932b] 1982). The economic prospects 1932. Speech delivered in Hamburg, 8 January. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 21: *Activities 1931–1939: World crises and policies in Britain and America* (pp. 39–48). Macmillan.

Keynes, J.M. ([1932c] 1982). The world's economic crisis and the way of escape. Lecture delivered in Cambridge on 4 February and 18. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 21: *Activities 1931–1939: World crises and policies in Britain and America* (pp. 50–62). Macmillan.

Keynes, J. M. (1934). Notes on the economic situation. Economic intelligence service, 16 December. Keynes papers: BM 1/184–87.

Keynes, J. M. ([1936] 1973). *The general theory of employment, interest and money*. As reprinted in D. Moggridge (Ed.), *The collected writings of John Maynard Keynes*, Vol. 7. Macmillan.

Meltzer, A. H. (2003). *The history of the Federal Reserve, volume 1: 1913–1951*. University of Chicago Press.

Skidelsky, R. (1992). *John Maynard Keynes: The economist as saviour, 1920–1937*. Macmillan.

Debt and Debt Management: Reflections on a Fable by Kalecki

Jan Toporowski

In 2015, I published a reconstruction of a Kalecki fable on credit and debt (Toporowski, 2015). At the time, I thought that it provided an intriguing insight into the working of the monetary transmission mechanism in the financial system. Although my discussion of that working remains, I think, fundamentally correct, I have since been able to think through more carefully the implications of Kalecki's fable for debt and debt management. These are essentially new implications and have radical consequences for policies of debt management. They remain "lessons from history" because Kalecki's economic analysis, and the function of debt in that analysis, was drawn from the historic period in which he came to his mature understanding of capitalism. This was a period in which governments were oppressed by debt, and this is referred to in Kalecki's fable. I present those implications and consequences in this chapter, along with the version of the fable that I published in 2015.

Kalecki would occasionally tell stories illustrating the paradoxical, even irrational, aspects of capitalism. Perhaps best known is his tale of two railway lines in competition with each other and suffering from excess capacity, whose only long-term solution is the construction of successively further railway lines along the same route to ensure full capacity utilization on railway lines already built (Kalecki, 1990).

In the course of an exchange of views on the Greek crisis with Kalecki's student, Kazimierz Łaski, I passed on to him a fable of private debts in a Greek village being cleared through the circulation of a €100 note brought to the village by a German tourist and returned to him at the end of the story. Łaski pointed out that the story was much older and had in fact been originally told to him by Kalecki when the two men were working together in the 1960s. I was able to reconstruct Kalecki's story from the fragments remembered by Łaski, adding to it my own explanatory remarks,

---

J. Toporowski (✉)
SOAS University of London, London, UK
e-mail: jt29@soas.ac.uk

to make the story comprehensible to the twenty-first century reader. The story is presented in the next section. A humorous, Mexican version of the story is available on YouTube.[1] Literary versions of these debt processes are given in my original paper.

1 Kalecki's Fable

Following the death, in 1935, of the Polish military dictator Józef Piłsudski, his regime continued under a group of his military cronies, known as the "colonels," who increasingly modelled their regime on that of Mussolini in Italy. One of the colonels, who was responsible for economic development, wanted to understand the economic principles behind government policies of economic recovery. Kazimierz Łaski thinks it may have been Roman Górecki (1889–1946), a lawyer who had joined Piłsudski's "Legion," a militia formed to fight alongside the Austrian army in 1914. Piłsudski, himself a man of limited military training, was assiduous in promoting his supporters. By 1926, just before Piłsudski's military coup against the civilian government, Górecki had been advanced to the rank of general. Shortly after the coup, he was appointed head of the National Economic Bank (*Bank Gospodarstwa Krajowego*). Górecki went on to write a short book that was translated and published in London, detailing the economic success of the new Polish state. The book, complete with hagiographic descriptions of Piłsudski and his chosen head of state Ignacy Mościcki, concludes its introduction with the following summary of Polish economic policy (Górecki, 1935, p. 17):

> [...] the rational economic policy of the Government is based on three fundamental principles: (1) the maintenance of a stable currency based upon a balanced national budget and favourable trade balance. (2) The creation of conditions favourable to the process of internal capitalization (i.e., investment JT). (3) The adaptation of all elements of economic life to the new conditions brought about by the crisis.
>
> Experiments of every kind have been avoided, particularly the experiment of artificial "manipulation of business conditions," in the belief that only by means of simple, classic, orthodox methods can positive results be achieved.

Górecki was minister of trade and industry in the Polish government from October 1935 to May 1936. As reported by Łaski, he called in Kalecki's colleague from the Institute for the Study of Business Cycles and Prices (*Instytut Badań Konjunktur Gospodarczych i Cen*), Ludwik Landau, to explain the principles behind the "new economics." Landau had just been fired from the institute and was then working at the national statistical office, Główny Urząd Statystyczny (see Landau, 1957; Kalecki, 1997).

Landau explained at length the principles of effective demand and credit cycles determining the levels of output and employment in a market capitalist economy at

---

[1] At https://www.youtube.com/watch?v=Aw6zsJBVMCM&t=42s

any one time. The colonel (or "General" Górecki) had evident difficulty in grasping this. In a final effort to break through the colonel's incomprehension, Landau told the following story:

> In Eastern Poland, far from any electricity connection, to an impoverished Jewish shtetl, whose residents were mired in debt and living on credit, a wealthy and pious Jew arrived and checked into the local inn, taking care to pay his hotel bill in advance. On Friday, to avoid breaking the Sabbath injunction against carrying money, he handed over to the inn-keeper for safe-keeping a $100 note. The denomination of the banknote is significant because the Polish currency, the złoty, had been placed on a dollar standard in 1934 at a rate of zł.5.26 to the US$. Efforts to maintain this parity included a restriction on the issue of banknotes. Hence the widespread use of informal credit and debt.[2] Early on Sunday, the wealthy and pious Jew left the inn before the inn-keeper had a chance to return the banknote.

> After a few days, the inn-keeper decided that the wealthy Jew was not going to return. So he took the $100 note and used it to clear his debt with the local butcher. The butcher was delighted and gave the note for safe-keeping to his wife. She used it to clear her debts with a local seamstress who made up dresses for her. The seamstress who herself was in debt, was pleased to take the money to repay her rent arrears with her landlord. The landlord was happy to get his rent at last and gave the money to pay his mistress, who had been giving him her favours without any return for far too long. The mistress was pleased because she could now use the note to clear off her debt at the local inn where she occasionally rented rooms.

> So it was that the bank-note finally returned to the inn-keeper. Although no new trade or production had occurred, nor any income been created, the debts in the shtetl had been cleared, and everyone looked forward to the future with renewed optimism.

> A couple of weeks later, the wealthy and pious Jew returned to the inn, and the inn-keeper was able to return to him his $100 note. To his amazement and dismay, the wealthy Jew took the note, set fire to it at the paraffin lamp that, in the absence of electricity, illuminated the table, and used the flame to light his cigarette. On seeing the inn-keeper's dismay the wealthy Jew laughed and told him that the banknote was forged anyway.

Landau finished his story and waited for understanding to seize the colonel. Beads of sweat appeared on the colonel's forehead, from the intellectual effort at comprehension. Finally, the colonel's face lit up with comprehension. "Really! I knew from the very beginning that there was something wrong with that Jew. Of course, the money was forged!"

Kalecki's conclusion from this was that unfortunately, too many people think like the colonel, and very few people understand the story as we do.

---

[2]The Polish writer Stefan Themerson related to the author how, faced with a shortage of banknotes and bank credit, Themerson's publisher in the 1930s, Gebethner and Wolff, had offered Themerson a wagonload of coal in payment for royalties. Themerson refused this payment, and they finally agreed that Themerson would receive as payment a stay in a hotel in the Polish mountain resort of Zakopane, where the publisher had credit.

2   Debt Payment Processes

This section interprets Kalecki's debt fable as demonstrating how debts may be serviced purely by financial transfers, without any income from the real economy through the autonomous circulation of money in the financial system. This interpretation challenges the classical theory of interest according to which the rate of interest is derived from the rate of profit in the economy, because debt obligations have to be serviced from income generated in the real (nonfinancial) economy (Toporowski, 2020a).

To demonstrate the autonomy of debt payments with respect to the real economy, it is convenient to start from a situation in which there is debt but no real economic activity. Consider a group of mutually indebted agents (MIAs) who have no income from any productive economic activity. These may be rentiers (with assets, in the form of debt claims against each other, but no income, apart from the interest on those claims) or retired persons holding such debt claims and owing money but without any pensions. How can they make debt payments to each other? The answer, as the above fable shows, is very simple. Those who *receive* interest or debt payments can make payments from the interest or payments on debts that they have received. The rest can *sell* financial assets, in the form of the debt claims that they hold, in order to make their own debt payments. Obviously, for such payments to be made, the system requires some minimum holdings of monetary assets (bank deposits), say $\acute{M}$.

Strictly speaking, the money or liquidity required to settle debt payment in this situation depends on three factors. First of all, there is the actual sum of debts outstanding (in money terms). Secondly, there is the distribution of debts and debt claims among the MIAs. Total debts obviously equal to total debt claims, but not for every agent, so that there will be some *net* debtors and some *net* creditors. Thirdly, there is the sequence over time of contracted debt payment (interest and principal repayment) commitments of the MIAs. In turn, these depend on the term (short or long term) of the debts and whether the interest payments are due annually, quarterly, or weekly.

If actual holdings of money or bank deposits, M, are less than $\acute{M}$, the MIAs have to depend on net revenue from real activities (surplus). Only in this sense is interest an actual deduction from a surplus generated in the real (nonfinancial) economy, as affirmed by the theory of interest that prevailed in classical political economy that the rate of interest is ultimately determined by the rate of profit in the economy (Toporowski, 2020a). The widely held view that interest must somehow be paid out of a surplus generated in the real economy is derived from this classical view

This link to current surplus is the foundation of the Wicksellian theory that the money rate of interest may deviate from the rate of profit but that the rate of profit places a ceiling on the interest that may be paid in an economy. Kalecki's fable challenges this determination of interest by the rate of profit by considering a situation in which the mutually indebted characters in the story had no income but

were able to pay off their debts with a token money (the detail of its forgery is incidental to the financial circulation of money).

To test this supposed link between interest and income from nonfinancial activities, we may consider a situation in which the circle of mutually indebted agents not only have no other income, except the interest that they may receive on their debt claims on each other, but also one in which their holdings of money, M, are less than the amount Ṁ required to allow payments to be made to each other. In this situation, interest payments cannot be made in full. Where would our MIAs find the money to pay what they owe to each other? The obvious answer is for them to sell the debt claims that they hold in exchange for money and use that money to make the payments required on their debts. In this situation, the velocity of circulation of money among the MIAs would increase, because it is now not only being used to pay interest. The money in financial circulation is now also being used to buy and sell debt claims.

What happens if the MIAs jointly become net sellers of their debt claims? In that situation, the price of their debt claims will fall in relation to the interest that comes with ownership of the debt claim. This price will continue falling until, at some point, the yield on the debt claims will attract into the market for debt claim holders of money who will wish to obtain a generous income from holding debt claims rather than money. This will continue until the market for debt claims has absorbed enough money for the indebted to be able to service their debts.

In a credit economy, where our mutually indebted agents hold their money as bank deposits, commercial banks may ease the shortage of money by offering credit against the security of debt claims held by the borrower. With bank deposits acceptable as means of payment of interest, bank lending has the capacity to ease the shortage of money but at the expense of increased borrowing by our MIAs, who now not only owe money directly to each other but also to banks. However, although the total debts of the MIAs have increased, the monetary resources, in the form of bank deposits of MIAs, have increased by the same amount. Their net debt has not increased, and the interest that they pay to their bankers is offset by the interest that our MIAs may now receive on their bank deposits. Thus, the contribution of commercial banks to debt management is the creation of new debt claims (lending or "advances"). With each new bank debt claim, an equivalent deposit is created (and hence an addition to M).

However, this mechanism requires a commercial banking system willing to grant, on demand, new credit against the value of debt claims. Even with such a facility, the creation of new credit against existing debt claims or the attraction of money into the market for debt claims may not be reliable ways of raising the money stocks required to clear debt payments. Holders of money stocks may be deterred from buying debt claims by the prospect that their purchase may not be sufficient to stem the fall in the prices of those debts, and banks may be deterred from making sufficient advances by the fall in the value of the security that the debt claims offer. The markets for debts are the hunting grounds of speculators who buy debts in anticipation of a rise in their value. This, after all, is the basis of Keynes's famous "beauty contest" theory of financial speculation (Keynes, 1936, "Some Reflections on Financial Instability in

Macro Agents-Based Models: Genealogy and Objectives"). If the mechanism of selling off debt claims breaks down or banks cease to make adequate advances against debt claims, then defaults on debt payments may arise.

Keynes's great rival, Ralph Hawtrey, called this failure of liquidity in debt markets "credit deadlock," in reference to the situation after the First World War, when the foreign debts of the Western powers were supposed to be paid from reparations paid by the government of Germany, which had been stripped at Versailles of its ability to pay those reparations.[3] He recommended that "a credit deadlock which is impervious to cheap money may (thus) yield to treatment of through open market purchases of securities" by the central bank. In Hawtrey's view, such operations would make banks more liquid and therefore more inclined to lend (Hawtrey, 1938, p. 256). But the same logic applies to direct purchases of debt claims in the event of a debt market breakdown, if net selling continues. Net selling by indebted agents requires a net buyer, and if commercial banks will not do this on behalf of their depositors, the central bank can perform this operation through its open market operations. We are familiar with this from the quantitative easing programs that have been pursued by central banks in the United States, Europe, and Japan, since the crisis of 2008.

This now introduces a fourth factor affecting the minimum money stocks or liquidity necessary to ensure debt payments, namely, the class of debt and from which agent debt claims are purchased by the central bank or the class of debt accepted as security against loans by commercial banks and in turn the sequence by which the additional liquidity meets the contracted debt obligations of MIAs. This is not a random process, subject to shocks. At any one time, these factors are given by the structure and distribution, among agents, of debts inherited from the past. The difference between the minimum monetary stock required for clearing payments and the actual money stocks held by the agents ($\acute{M}$-M) is therefore determinate at any one time, even though the process of making debt claims liquid (selling them to the central bank or to commercial banks or borrowing against them from commercial banks) will modify the structures of debt and debt claims.

In the analysis, we have reached the point where the debts of mutually indebted agents with no income have been used to expand borrowing from commercial banks, and the liquidity of those banks is maintained by the central bank through buying debt claims from commercial banks. These are the conditions for keeping the portfolios of MIAs sufficiently liquid for them to make payments on their debts to each other, directly or through commercial bank intermediation, or to the central bank. Over time, more interest is diverted among MIAs, through commercial banks, or paid to the central bank as opposed to other MIAs, as their debts pass into the

---

[3]This contradiction in the financial policies of the Western powers had been noted by Keynes in his critique of the Versailles settlements, *The Economic Consequences of the Peace* (Keynes, 1920, Chapters "High-Frequency Trading and the Material Political Economy of Finance" and "Marx and Hayek on "Real" Versus "Less Real": Explanations for the Fragility of Capitalism"; see also Keynes, 1922, "Marx and Hayek on "Real" Versus "Less Real": Explanations for the Fragility of Capitalism").

ownership of the central bank. "In the long run we are all dead" (Keynes, 1923, p. 80), and on decease, our remaining debt claims pass into the possession of commercial banks or the central bank to help write off debt claims against the deceased held by those commercial banks or the central bank. The net debt of the group of mutually indebted agents is not changed, until the last of them inherits the last debt claims against himself.

3   Some Implications for Debt Management

Four features of this process, of maintaining payments on debts without drawing on income from the real economy, warrant comment. First of all, there is the growing intermediation by banks to maintain a level of money in the economy sufficient to allow payments to be made on the debts of individuals with no income other than the interest that comes with their ownership of debt claims. Agents who started off borrowing money from each other end up owing larger amounts of money apparently to banks but, indirectly, to other agents who hold bank deposits. These deposits are backed in bank balance sheets by debts owed to banks by agents who took out loans against the security of their initial debt claims against each other.

A second feature is the market in debt claims. Debt claims can only be used as collateral for additional bank loans if the debt claims can be sold into a market. Transforming debts into marketable securities is a procedure known as "securitization." Hyman Minsky, who observed the rise of securitization in the United States in the 1980s, was correct to point out that the function of securitization is to make portfolios of assets (in this case debt claims) more liquid (Minsky, 2008). Without such securitization, the system of debt would have collapsed if all mutually indebted agents had insufficient money to service their debts. Securitization allowed them to obtain such money, either by selling their debt claims to each other or by obtaining credit from a bank secured against an asset or debt claim that the bank could sell if the borrower defaulted. Finally, securitization allows banks to maintain the liquidity of their balance sheets by selling the debt claims in their possession to another commercial bank or a central bank in exchange for reserves that can be used to settle payments between banks.

A third feature of the process is that it explains how financial crisis, in the form of default on debt payments, arises in an economy because commercial banks stop lending and central banks stop buying. Both of these breakdowns in the system were central causes of the 2008 crisis, even if they were not widely recognized as such. To this crisis, we can apply the "lessons of history" that Kalecki taught us. The conventional account of the crisis has it that financial deregulation and the indebting of poorer households had infected collateralized debt obligations with bad debt, and this aroused suspicion of securitization among banks and in securities markets. This was undoubtedly the sentiment that accompanied the filing for bankruptcy of Lehman Brothers in September 2008, now widely taken as the start of the crisis. In fact, the crisis had very little to do with the dressing up of bad debt as good

securities by unscrupulous bankers and everything to do with long-term institutional trends in banking and capital markets.

On the banking side, the key long-term institutional development has been the migration of the inter-bank money markets from the official markets for unsecured short-term borrowing to the foreign exchange swaps market. There are various reasons behind this migration. But key among them is the lower cost of lending that is effected through the foreign exchange swaps market because borrowing in that market is secured on a future repurchase agreement. Under the Basel regulations since 1988, less capital has to be set aside for this, than for unsecured lending. However, the overnight and other short-term rates of interest in the market for unsecured lending and borrowing remained the litmus test that central banks applied to their official interest rates: As long as these short-term rates were close to the official policy rates of the central banks, their monetary policy committees could assure themselves that the "monetary transmission mechanism" from those policy rates to the economy at large, via the inter-bank market, was functioning correctly. In fact, the mechanism was only appearing to function correctly because, as the money markets migrated to the foreign exchange swaps market, central banks found themselves increasingly taking their local market for unsecured short-term funds onto the central banks' balance sheets where, naturally, the rates for those funds conformed to policy rates but only because the rates for unsecured funds were now more or less administered by central banks (Shabani et al., 2021).

On the capital market side of the financial system, where long-term rates of interest are determined, liquidity in those markets was drying up from the 1990s, principally because of the maturing of pension funds in the United States and the United Kingdom. The first sign of this difficulty was the bursting of the dot-com bubble in 2001. The contribution of this illiquidity to the 2008 crisis was that it became increasingly difficult for banks to float off their loan books as collateralized debt obligations into the capital market.[4] Banks took to "warehousing" their unsold bonds' off-balance sheet in subsidiaries (special-purpose vehicles) financed with unsecured short-term borrowing (Nesvetailova, 2010). As the inter-bank market became less active, it became more difficult to roll over that borrowing. When rolling over proved impossible, the special-purpose vehicles went into default.

It can be argued that much the same was also true of the nonfinancial corporate sector, where a merger and takeover boom since the 1990s, financed with short-term borrowing, had loaded up large industrial corporations with (unsecured) short-term borrowing. This too, like the collateralized debt obligations of banks, was supposed to be "funded" or refinanced by stock issues into the capital market. But this proved impossible, and when the corporations could no longer rollover their short-term debt, they too fell into default. But this was not before they had reduced severely their fixed capital investment programs. Since it is private sector investment that

---

[4] See Toporowski (2020b). The thinning liquidity of the capital market was not apparent from the usual indicator of such liquidity, the yield on US government bonds, which has fallen since the turn of the century with the growing use of such bonds as security in swap transactions.

determines the business cycle in capitalism, it has been the fall in fixed investment that transmitted the financial crisis to the real economy (Toporowski, 2016).

On the commercial banking side, therefore, the story of the 2008 crisis runs in accordance with the Kalecki fable and the failure of banks to inject into financial circulation the credit necessary for debt management. On the central banking side, there was a similar failure. From the 1990s onward, the ruling policy doctrine has been the neo-Wicksellian view that the way to regulate inflation or the business cycle is through the short-term rate of interest, guided by the policy rates of interest of central banks. Under the assumption of arbitrage with perfect liquidity, modelling showed that short-term rates of interest could be kept in line with policy rates by setting rates for the deposits and borrowing of reserves at the central bank. This appeared to work. But it was because the inter-bank market for unsecured borrowing was being increasingly brought onto the balance sheet of the central bank, where rates were administered by that bank, rather than because a condition of perfect liquidity prevailed among banks (Toporowski, 2019). In effect, central banks gave up the provision of liquidity to securities markets out of a misguided belief that they should not interfere with the effective "pricing of risk" that was supposed to be the function of those markets. Lender of last resort facilities was constrained by bureaucratic procedures: In Britain, when Northern Rock came knocking at the door of the Bank of England, in July 2007, rescue efforts had to be coordinated with the government's treasury department and a separate Financial Services Authority.

Liquidity provision and open market operations returned in the guise of "quantitative easing," ostensibly because the reduction of policy rates of interest almost down to zero left little room for the operation of the neo-Wicksellian solution of reductions in those rates. Liquidity provision quickly stabilized financial systems and even caused stock market booms, except in the European Monetary Union, where the neo-Wicksellianism had been embedded into the statutes of the European Central Bank. The misconstruction of the 2008 crisis, as being due to overvalued subprime mortgages backing collateralized debt obligations, was discreetly contradicted by the profits that the US Federal Reserve and the US Treasury made on the securitized mortgages that they bought up from US banks under the Troubled Assets Recovery Program. The crisis was not one of insolvency and fraud, but of illiquidity. Too many commentators have followed the colonel in Kalecki's fable, interpreting a crisis of liquidity as a fraud because they do not understand the way in which insufficient liquidity in debt markets precipitates default on debt payments.

The analysis leads directly to a paradoxical conclusion, namely, that the answer to a debt crisis is more debt. This is paradoxical because the first two features discussed above, namely, growing financial intermediation and securitization, are commonly associated with financialization, which is widely held to contribute to financial crisis and leads the editor of Minsky's essay on securitization into a wrong interpretation of that essay (Minsky, 2008). It is an irony of the literature on financialization that it seeks to constrain, or even eliminate, in the case of securitization, precisely those mechanisms that facilitate the avoidance of financial crisis. To a great extent, this is because of a misreading of the 2008 crisis, which was indeed preceded by rising financial intermediation and securitization over an extended period. However, to

focus merely on this trend is to overlook the failures in financial intermediation, in particular in the capital markets, and the difficulties of securitization that started in the 1990s, well *before* the crisis itself (Toporowski, 2016).

A fourth, and perhaps least apparent, feature of Kalecki's fable of debt management is that, in a situation of mutual indebtedness, it treats the problem of payments on debt as essentially a transfer problem, i.e., how to ensure that there is enough liquidity in the portfolios of indebted parties to ensure that mutual debt obligations can be settled. This involves swapping assets (debt claims) for money among the indebted parties or borrowing from banks (the resulting deposits being backed by banks' loans to indebted parties) or purchase of debt claims by a central bank. In any case, the financial assets that are created by debt (debt claims) remain within the system as a kind of "inside" debt.

This is not the case when debt is owed to someone who is "outside" the system, i.e., when the creditor does not owe money to anyone else within the system so that, on receipt of debt payments, the "outside" creditor merely takes the money and does not use it to make payments to anyone else in the system. To some degree, this is the case with the "central bank" in the system described in the previous section. But that central bank husbanded its resources for the greater good of the financial system, using the reserves that it can today create in order to buy debt claims from MIAs (and when it did not, as demonstrated in 2008, it led to crisis). Of course, the more of these debt claims that it buys, the more monetary resources it will take out of the system in the future, with payments on the debt claims that the central bank has purchased. This is why central banks need to maintain continuous open market operations to regulate the liquidity of the financial system.

A much more obvious case of "outside debt" was the international gold standard, a history from which, unfortunately, too many lessons were learnt. In that system, gold was the ultimate monetary resource but brought no return to its possessor, since the price of gold was constant. In that situation, the rate of interest was the only incentive for parting with gold. A central bank, mindful of its responsibility to issue only banknotes backed by gold, had only a limited capacity to buy in debt claims, and private creditors had even less incentive to plow back payments received into buying debt claims. There is a sense in which the gold standard made all debt into "outside" debt, inhibiting the financial circulation of money. This was the background to the "classical" theory of interest that tied interest to the rate of profit. This system, with its periodic banking crises when banks ran out of reserves, is the one that the more ardent critics of "financialization" would have restored.

It was only with the development of fractional reserve banking and the widespread use of bank deposits as means of payment that a credit system emerged in which commercial banks could lend to create deposits, and central banks buy more freely (see Hobson, 1913, pp. 89–92; De Cecco, 1984; Chick, 1992). By facilitating interest payments through financial circulation, the credit system finally released interest and debt payments from their dependence on surplus or income generated in the nonfinancial economy. Yet, the understanding that prevails today, of banking and financial crisis, has learned the wrong lessons from history. The literature is almost wholly based on the idea that financial crisis arises because financial

intermediaries have short-term liabilities and hold illiquid debt claims, as banks did under classic capitalism of the early years of the gold standard. But that was before the development of capital markets opened new possibilities of making debt claims liquid (Toporowski, 2020b). As Kalecki's fable illustrates, once debt becomes "inside debt," interest and payments on debt need only financial circulation for their prompt settlement. For the system as a whole, debts are truly "fictitious capital" that may be settled with a forged banknote.

References

Chick, V. ([1986] 1992). The evolution of the banking system and the theory of saving, investment and interest. In S.C. Dow and P. Arestis (Eds), *On money, method and Keynes: Selected essays* (pp. 193–205). Palgrave Macmillan.

De Cecco, M. (1984). *Money and empire: The international gold standard 1890–1914*. St. Martin's Press.

Górecki, R. (1935). *Poland and her economic development*. George Allen & Unwin.

Hawtrey, R. G. (1938). *A century of bank rate*. Longmans Green and Co.

Hobson, J. A. (1913). *Gold, prices and wages with an examination of the quantity theory*. Methuen.

Kalecki, M. ([1933] 1990). Stimulating the business upswing. In J. Osiatyński (Ed.), *Collected works of Michał Kalecki*, Vol. I: *Capitalism: Business cycles and full employment*. Clarendon Press.

Kalecki, M. ([1964] 1997). Ludwik Landau: Economist and statistician. In J. Osiatyński (Ed.), *Collected works of Michał Kalecki*, Vol. VII: *Studies in applied economics 1940–1967, miscellanea*. Clarendon Press.

Keynes, J. M. (1920). *The economic consequences of the peace*. Macmillan.

Keynes, J. M. (1922). *A revision of the treaty being a sequel to the economic consequences of the peace*. Macmillan.

Keynes, J. M. (1923). *A tract on monetary reform*. Macmillan.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

Landau, L. (1957). *Wybór Pism*. Państwowe Wydawnictwo Naukowe.

Minsky, H. P. (2008). Securitization. *Levy Economics Institute policy note* no. 2008/2. The Levy Economics Institute of Bard College.

Nesvetailova, A. (2010). *Financial alchemy in crisis: The great liquidity illusion*. Pluto Press.

Shabani, M., Stenfors, A., & Toporowski, J. (2021). Japanese banks in the international money markets. In A. Stenfors & J. Toporowski (Eds.), *Unconventional monetary policy and financial stability: The case of Japan*. Routledge.

Toporowski, J. (2015). A Kalecki fable on debt and the monetary transmission mechanism. *LSE Financial Markets Group special paper* no. 239, August. London School of Economics.

Toporowski, J. (2016). The crisis of finance and the crisis of accumulation: It was not a Lehman brothers moment. In T. Subasat (Ed.), *The great financial meltdown: Systemic, conjunctural or policy created?* (pp. 236–247). Edward Elgar.

Toporowski, J. (2019). Open market operations. In D. G. Mayes, P. L. Siklos, & J.-E. Sturm (Eds.), *The Oxford handbook of the economics of central banking* (pp. 436–453). Oxford University Press.

Toporowski, J. (2020a). Anwar Shaikh and the classical theory of interest: A critical note. *Cambridge Journal of Economics, 44*(2), 465–474.

Toporowski, J. (2020b). Financialisation and the periodisation of capitalism: Appearances and processes. *Review of Evolutionary Political Economy, 1*(2), 149–160.

Part IV
Income Distribution and the Social Roots of Economic Crisis

Desperation by Consent: Inequality and Financial Crises

Amos Witztum

1   Introduction

The rise in income inequality which preceded the recent financial crisis has given the occasion for a greater interest in the possible relationship which may exist between inequality and the development of severe economic fluctuations triggered by financial activities.[1] Apart from the obvious question regarding the way in which inequality may trigger financial instability, there is also the issue of the economic role of inequality itself. The reason why this may matter is because the ability of the government to address the possible effects of inequality on the financial system will be constrained by the economic significance of inequality. Namely, if there are efficiency considerations associated with inequality, eradicating it for the sake of stability may prove to be too costly. On the other hand, this will raise questions about the meaning of efficiency altogether if it appears to be so closely linked with instability (on top of sanctioning inequality).

On the face of it, to ponder at all, the economic role of inequality may sound like an odd question to ask given the pillars of the two Welfare Theorems; after all, efficient competitive equilibria (i.e. allocations where individuals solve their own—private—economic problem) can, in principle, accommodate any distribution of

---

[1] See, for instance, Stiglitz (2009), Milanovic (2009), Wade (2009), Fitoussi and Saraceno (2010), Rajan (2010), Reich (2010), Kumhof et al. (2012, 2015), Galbraith (2012), Palley (2012), and Stockhammer (2015). There are, of course, dissenting voices like Bordo and Meissner (2012) who claim otherwise, but it is important to emphasize that their analysis is based only on the share of the top 1%. This means that they ignore those at the bottom end of the distribution whose behaviour could have been significant in facilitating the crisis when lines of credit become available to them.

A. Witztum (✉)
Centre for the Philosophy of Natural and Social Sciences (CPNSS), London School of Economics, London, UK
e-mail: A.Witztum@lse.ac.uk

income. Hence, if inequality triggers financial instability, then we should "simply" aim at achieving the efficient outcome commensurate with a more equal distribution. In other words, this is all a matter of the right *corrective* policy. However, if we bear in mind that in the presence of missing markets, efficient outcomes are limited by specific ownership (hence, corporate) structures (and subsequently, a more limited span of income distributions), and if these structures generate an instability trigger- ing distribution of income, then the problem becomes more endemic and cannot be "simply" resolved by policy alone. In turn, this will lead to the question of whether income distribution is just the afterthought of efficiency—as is embedded in the modern economics' psyche—or, as was the case in classical thinking, one of its building blocks.

To begin with, however, one must wonder about the possible connection that may exist between income distribution and financial crises. If we accept that what seems common to most financial crises is strongly associated with a tendency for excessive risk-taking behaviour, then we would expect to find some correlation between inequality and risk-taking. Indeed, there is some evidence to this effect. Falk et al. (2015), who surveyed 80,000 subjects in 76 countries, find that there is indeed a positive correlation between inequality and risk-taking preferences. But the fact that there is a greater tendency to risk-taking when there is inequality does not in itself suggest a crisis. For this to follow, we would expect greater risk to be taken by those who are most likely to default: the poor.

From the perspective of modern economics, this is difficult to explain for two main reasons. Firstly, all individuals are rational (regardless of their wealth or income) which means that unless they make consistent mistakes, they are unlikely to act in a way which will expose them to excessive risk or lead them to default on their debt (which is, of course, costly to them too). Secondly, given that according to modern theory greater risk-taking (less risk aversion) is usually correlated with income or wealth,[2] then those who take greater risk are also those who, in general, would be less likely to default. For these two reasons, at some fundamental level, inequality as a cause of a crisis seems inconsistent with modern analysis.

Of course, attempts have been made to offer some explanations as to why the poor would take greater risk than theory predicts. A great deal of them come from game theory and experimental economics where the validity of the findings and the nature of explanations are somewhat limited.[3] Notwithstanding, one thing which

---

[2]See, for instance, a discussion on both theory and evidence in Courbage et al. (2018).

[3]See, for instance, Genakos and Pagliero (2012), Dijk et al. (2014), Bosch-Domènech and Silvestre (2006), Payne et al. (2017), and Hopkins (2018). Payne et al. (2017) show that, in general, participants in experiments would take higher risks when there is inequality in rewards. This still seems insufficient to justify excessive risk taking by the poor. Hopkins (2018), on the other hand, provides a more theoretical argument where he finds that poorer participants (in terms of wealth) would take greater risks than anticipated when the inequality in rewards increases, but at the same time, this would decrease with an increase in inequality in wealth. This explanation is conditioned on the validity of the tournament metaphor for economic life and the generality of the utility function employed. Bosch-Domènech and Silvestre (2006) find that the predictions of the theory are

seems common to all is indeed the presumption of the rationality of the agents which suggests that poorer household may indeed engage in greater risk than theory predicts but not necessarily excessive risk sufficient to trigger a financial crisis.

A less ad hoc type of explanation can be found in the claim that inequality means that as the income of the rich increases, more funds would be directed towards the provision of loans for the poor.[4] Ostensibly, the greater availability of loans would reduce the cost of borrowing and thus increase the number of riskier projects adopted. However, even though overall debt increases with inequality, the fact is that the real interest rate in the USA, for instance, rose dramatically in the years before the crisis.[5] Therefore, it is not enough to know that there are more funds available; we need to know more about who it is that is doing the borrowing. If it is indeed the poor then given their lack of wealth, they are more prone to bankruptcies, and a crisis would indeed ensue, but with the increase in the cost of borrowing, it would suggest that the poor would be the least involved in borrowing[6] even though the line of reasoning here suggests a transfer of funds from the rich to the poor.

Perhaps a less aggregate approach to credit may give more credence to the presumption that funds have been made available to the poor. Brei et al. (2018) find that from a relatively low threshold of the availability of credit (as a percentage of GDP), it appears that the expansion of market credit is correlated with an increase in inequality while bank credit is less so. While overall causality here is a problem, it may not be entirely unreasonable to suppose that had the poor relied on banks alone for their loans, they might not have received the means to fund their excessive risk-taking. Therefore, if the availability of funds to the poor is associated mainly with market credit and if the poor are indeed excessive risk-takers,[7] then the correlation between market credit and inequality may be considered as a potential cause of a financial crisis.

Notwithstanding the availability of funds, the question remains, why would the poor become excessive risk-taker? To make the poor behave in a manner which evidently seems irrational, we would need to justify a departure from standard behaviour. Not in the sense that individuals are always irrational but rather in the

---

actually correct when small gambles are involved but not so when gambles involve large sums of money. That is, when the rewards are small, wealthier people are less risk-averse than their poorer fellows in society. But, when gambles involve large rewards, the risk taking of the poor exceeds that of the wealthy. Here again, we are at the mercy of the significance of experimental results.

[4] See, for instance, Kumhof et al. (2015).

[5] From 1.605% in 2004 to 5.223% in 2007 (according to the World Bank Data). In the UK, the increase was much less pronounced but nonetheless an increase (from 1.866 in 2004 to 2.892 in 2007).

[6] Coibion et al. (2016) show that within the USA, in areas of high inequality, the low-income population did not borrow more. This is not a definitive argument against the possibility that the poor borrow more as the paper is not looking at different societies but rather different zip code addresses.

[7] Dick and Lehnert (2007) show that deregulation in the US market which led to the increased availability of credit also led to a parallel increase in private bankruptcies (measured by individuals filing for Chap. 7).

sense where people are driven to a point of resorting to irrational actions. Such a departure cannot be explained simply by changes in the constraint individuals face (or the lure of potential rewards) but rather through either an altogether change in motivation or a breakdown in behaviour due to deeply frustrated expectations. Modern economics does not offer such an explanation at all.

In principle, to expect individuals to change their behaviour so dramatically as to act outside their generally attributed motivation, there needs to be a trigger. As we suggested, such a trigger could be associated with a sense of frustration about the normal way of behaviour. If individuals behave in a particular manner, they will have certain expectations about the outcomes of such a behaviour. If these expectations are repeatedly frustrated at the time where the accepted view is that the system works well, an agent is more likely to respond by abandoning the way they normally behave and opt for a radical act which they may hope will break their continuous inability to move closer toward their desired aims. Such an act cannot be judged by rational considerations as it would be an expression of desperation.[8]

But while modern economics does not account for changes in behaviour attributed to income distribution, it does implicitly recognize the presence of potential continuous frustrated expectations. We will show how within the model of general equilibrium, one can identify domains where the expectations of some agents, could be deemed to have been frustrated even though on the face of it, they are all rational and receive from the system that which they have *reason to expect*. At the same time, however, we will show how this problem which may appear to have a trivial solution through the reallocation of initial endowments (i.e. the second welfare theorem) runs into immense difficulties given the constraint on permissible distributions which arises from the problem of incompleteness in a world of uncertainty.

2   Frustrated Expectations

Why would a rational agent who finds himself at the bottom of the income distribution behave in a way that is different from the one we would normally expect him to behave and which may jeopardize him even more? In modern economics this is, of course, impossible. As we said earlier, all agents are equally rational and act within

---

[8]A similar line of reasoning has been quite prevalent following the last financial crisis. See, for instance, Rajan (2010) and Reich (2010) who argue that the crisis arose due to the attempt by the poor and lower-middle-class families to keep up with consumption when the growth in their earnings stalls. While the term desperation is not explicitly used, there is a question why would rational beings desire to consume beyond their means? As we mentioned earlier, Kumhof et al. (2015) seem to suggest an alternative, non-behavioural explanation in terms of the increased bargaining power of the rich through their channelling of a greater share of their increased income in the form of loans to the poor. Nevertheless, this does not alter the fact that for this to hold, the poor must indeed feel desperate to make use of these loans (and increase their indebtedness) in order to catch up with consumption.

that which they have reason to expect. Namely, given their endowment and market conditions, they will choose the best means to maximize their wellbeing. From this perspective, the term excessive is inconceivable. Yes, they may take greater risk than normally expected when the circumstances are such that they may consider the extra risk as useful in solving their economic problem or even in better positioning them within society (if such a position is part of their wellbeing). But they are going to approach this in a rational manner which means that they are unlikely to generate bad debt unless they make consistent mistakes which, in terms of modern analysis, would deem them as irrational.

In fact, even this last claim does not sit well with standard theory and raises the question of the ad hoc nature of the circumstances in which one can argue that it would be rational for the poor to become greater risk-taker than is normally supposed. Consider an agent with income Y who faces a gamble in which he, or she, may win h with probability p or lose h with probability $(1 - p)$. One way of gauging the level of this agent's risk aversion (or willingness to take risks) would be to examine the probability of winning for which he, or she, will be indifferent between their existing income (or wealth) and taking the gamble. We define this probability as p^*.[9] This condition, containing the Arrow-Pratt measure of risk aversion, is given as follows assuming that the agent's VNM utility function is $v(y)$:

$$p^* = \frac{1}{2} - \frac{v''(y)}{v'(y)} \cdot \frac{h}{4}.$$

The effects of the level of income (y) from which the agent launches his, or her, gamble is somewhat ambiguous as $\frac{\partial p*}{\partial y} = -\left(\frac{v'''(y) \cdot v'(y) - (v''(y))^2}{(v'(y))^2}\right) \cdot \frac{h}{4}$, but it is easy to see that if we suppose most agents to be risk-averse in principle (i.e. a concave VNM function), then using the most common form of $v(y)$ which represents such risk aversion ($v(y) = \alpha \cdot lny$) will make this expression negative. Hence, as income increases, the less risk-averse people become. The poor, therefore, would be more risk-averse than the rich and, therefore, are unlikely to be the origin of greater systemic risk-taking. But more importantly, we can see that the greater dispersion of return—which is the way some of the attempts to explain the rational behaviour of the poor interpret the presence of inequality—would not yield the required outcome. It is evident that the increase in h would lead to greater risk aversion (higher p^*) at any level of income. Namely, it is not even consistent with the standard theory that the poor would respond to apparent opportunities by taking greater risk than they would normally do. Therefore, it is difficult to pin the rise in risk-taking among those

---

[9]If p^* is the probability where an agent will be indifferent between taking or not taking a gamble (and then remaining with his, or her, current income or wealth), then it means that if the actual probability of winning is lower, the individual will not take the gamble. If an individual is risk-neutral, she/he will be indifferent when the probability of winning is 0.5. This means that they will not take the gamble if the actual probability is $p < p^* = 0.5$. The more risk averse an agent is the higher will have to be the probability of winning for him, or her, to agree to take the gamble.

Fig. 1 General equilibrium
with subsistence domain
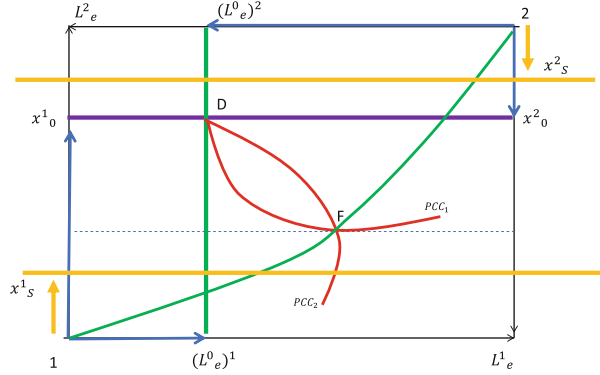
who are more likely to default, on the kind of rational considerations which modern
economics attributes to individuals.

So, what other possible reason could there be for agents to depart from what may
be considered as rational behaviour and embark on behaviour which may be reckless
both to them and society? The answer, as we suggested earlier, must therefore be an
act of desperation.

Consider an economy with two agents (or types of agents) who are engaged in the
production of a single commodity x which is the product of their labour
$(\bar{x} = \alpha \cdot (L_1 + L_2))$[10] while each one of them owns a share of the asset used in the
production process (say, land). To simplify the story, we assume the amount to be
produced as given and that both agents have the same productivity.[11] The initial
endowment reflects a certain distribution of labour and produce between the two
individuals where the latter represents initial ownership structures. In other words,
the total output of x is given ($\bar{x}$), and it is distributed between the individuals
according to their share of ownership (in, say, the land). So, in the diagram below,
the distribution of ownership is depicted by the purple line (which defines how $\bar{x}$ is
divided between them $\bar{x} = x^1 + x^2$). The distribution of the burden between them
(how much each one of them contributes to the production of $\bar{x}$) is given by the green
line which denotes the distribution of leisure that will result from a given distribution
of burden (work). We assume these structures (of ownership and burden) to have no
effect on productivity. Naturally, the agents—who derive their wellbeing from
consumption and leisure—can trade ownership (which is the source of consumption)
for leisure and thus improve their condition to the best which is possible for them
given the initial endowment and without making anyone worse off. We can depict
this story with the traditional Edgeworth box (see Fig. 1).

---

[10] L_1, L_2 represent the labour input of 1 and 2, respectively. Note that if the natural length of the day
is $\bar{L}$, then their leisure is defined to be: $L^i_e = \bar{L} - L_i$.

[11] This means that there is always the same output and the only thing the individual trade is their
share in labour and their share in output. One could have allowed the output to vary, but this would
have made the diagrammatic exposition impossible. There will be no qualitative difference in the
conclusion from this exposition.

The horizontal axis depicts the leisure each agent gets while the vertical axis depicts the quantity of the good they are allotted. Point D represents the initial endowment which means that given a certain ownership structure, which gives each agent claim over the produce derived from his, or her, property rights, and the historical division of labour between them (so that their leisure is given by $(L^0_e)^1$, $(L^0_e)^2$ respectively), agent 1 would command x^1_0 of the total output while agent 2 would get x^2_0. Naturally, competition will allow both of them to improve their position by trading ownership shares for leisure and thus move to point F where agent 1 forgoes some of his claim for the output in return for more leisure and agent 2 the other way around. The two red lines (PCC) mark the price-consumption curve that will lead us to the competitive equilibrium that is, of course, Pareto efficient. Each point on these lines captures a point of maximum utility for a given price (rational choices).

Here, at the beginning of the story, agent 1 was rich but had to work hard while agent 2 was poor but worked a little. In equilibrium there seems to be greater equality between them. This, of course, does not mean that competitive outcomes are necessarily consistent with either equality or morally acceptable allocations but in our case, point F does represent a more equitable allocation than D[12] as both individuals have better access to output while sharing more equitably their inputs. Consider now the two heavy orange lines. These lines represent what we may call subsistence. What we mean by subsistence is not necessarily physical subsistence but rather a socially constructed minimum of social acceptability in consumption. To be a valued member of society, one would need to be able to participate in certain social norms which may require a certain level of income (here captured by x). For simplicity sake we assume leisure not to be part of the minimum social bundle. This means that any allocation which would bring one of them below this line would mean that this individual cannot socially subsist.[13] He, or she, could well subsist privately, but their aspirations to be respected members of society would not be fulfilled. Having said this, the argument could have also been framed in terms of physical subsistence. The problem in such a case would be that individuals' preferences would then more likely be of a Stone-Geary type which would slightly change the argument. We will nevertheless discuss both:

In the left-hand diagram (in Fig. 2), we have the case where we are referring to social subsistence. Here, individuals' preferences cover the entire domain even though there may be social implications to being in part of the domain. If the initial endowment reflected a different ownership structure and work history, point D would now be at a place where agent 2 commands a greater share of the output while contributing very little labour. This, perhaps, is more representative of the normal state of affairs where property rights give someone almost full command over the output and the other has to rely mainly on his, or her, labour. Still, the

---

[12] There is a discussion of this issue in Witztum (2019, Chap. 4).

[13] In modern literature this seems to have been a forgotten episode where a discussion took place about what can be called a survival equilibrium (see Coles & Hammond, 1995).
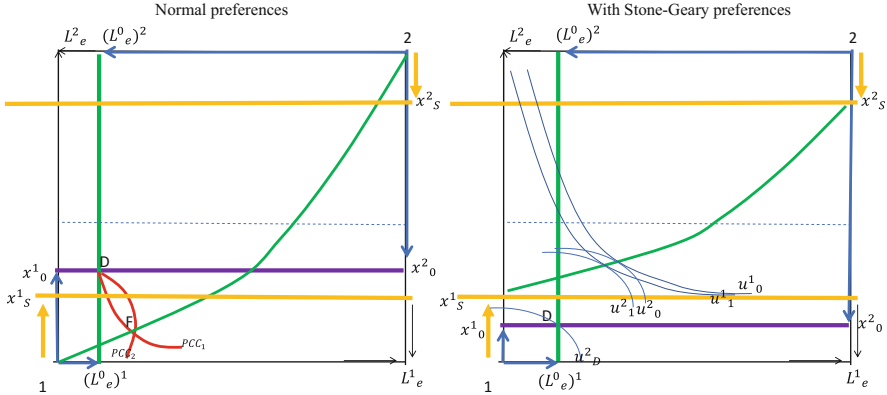
Fig. 2 General equilibrium with subsistence domain: desperation by consent

wonders of the market suggest that even in the face of this initial position, all agents can improve their position; they can rationally choose the best means to maximizes the wellbeing which they have reason to expect (determined by their initial endowment). This, of course, will take them to point F where agent 1 will trade some of this already meagre consumption of *x* for more leisure.

The difficulty of this situation is obvious. By moving to point F where, according to economic theory, they are both better off and the outcome is efficient, individual 1 is now consuming the good below the level required for social subsistence. The meaning of it is that the personal preferences of the individuals are such that he feels that he cannot go on working so much and is willing to forgo some of the material wellbeing to have more leisure so that he can physically recover. The fact that he is willing to pay this price realizing that he is losing his position in society does not mean that this is something the individual is happy to do. Being without any leisure at all has become a private impossibility and while the individual needs the social association, he has been pushed into a corner where he has no choice but gives greater weight to his personal need over his social need. The move from D to F appear perfectly rational, but this would be a distortion of the idea of choice. The individual is indeed free to choose between D and F, but he has lost his sovereignty in the sense that he would have rather not be in this position in the first place. Why must he abandon his need to be socially connected only to fulfil some basic private needs while the others do not face such a dilemma at all? Clearly, when the individual trades his way to point F, he must be deeply frustrated and may resort to actions that will allow him to move out of D but not all the way to F. The market as it stands does not offer this mechanism, and therefore, the actions the agent may take could be deemed as excessive risk-taking.

Had the subsistence level of consumption represented physical levels of subsistence, individuals' preferences are more likely to be represented by Stone-Geary type of preferences where indifference curves are asymptotic to the minimum level of consumption. This is depicted in the diagram (Fig. 2) on the right. Clearly, in such

a case, the position of point D in the left-hand diagram would not have led to point F. However, equally possible would be the situation where the endowments of the agents would be at point D which itself is below the physical survival line for at least someone. In such a case, in the absence of non-market mechanisms, the position of the agent in the market is the worst one possible. He will desperately wish to climb out of the predicament of physical decline by offering all his remaining leisure (rejuvenating time) for the sake of additional material goods. It could also push him towards acts of desperation.

The case of the Stone-Geary preferences and the physical survival is, in some sense, simpler as the desperation of the agent is apparent. The more worrying situation is the one we mentioned earlier where agents, led by their pressing personal concerns, slip into a domain where they realize that other expectations which they may have had can no longer be fulfilled. The desperation may be less obvious but nonetheless present and become the engine that may lead to acts outside the bounds of standard rationality. These, however, can only lead the agents to point F in the left-hand diagram. Such an almost unnoticed slip into the desperate domain is indeed a potential characteristic of a system with uncertainty and missing markets. It could explain the cause of desperation as well as the inability of modern economics to reconcile stability with efficiency.

What all of this suggests is that it is conceivable, even within modern economics, to imagine situations where the distribution of income (as reflected in the initial position of the two agents in our story) can become a source of desperate actions. However, it is also possible to see how easy it is for society to resolve these problems. In the end, within the framework of analysis which we employed, there is the second welfare theorem which suggests that if we did not like point F because it may push someone to desperation and irrational acts, there is a simple remedy: reallocate the initial endowments. Simply by changing the ownership structure (even for a given division of labour, i.e. along the green vertical lines), we can make sure that competitive outcomes never push anyone beyond the subsistence line and into the zone of desperation.

Based on this, we should have reached the conclusion that even if income inequality had been a trigger for excessive risk-taking, a simple correction of income distribution should have resolved the problem. The ideas of competition and financial markets need not be re-examined as one can find an allocation which will yield both efficiency and stability (in the sense of pushing no one into the desperate zone). However, the problem with our exposition so far is that there is no uncertainty and there are no financial markets. What we will try to show now is that when we move into this domain serious problems may arise which will make such a corrective policy, futile.

3   Stability Versus Efficiency

To see how in spite of the ability which society may have to change the initial endowment, the competitive outcome could still lead to an allocation where some agents may become severely frustrated to a degree that will change the way in which they behave, we must examine the effects of uncertainty on the promise embedded in modern economics. In a nutshell, we will show how uncertainty with missing markets (and incomplete contracts) severely limits the freedom which society may have in managing income distributions in a manner that will serve both efficiency and stability at the same time. To remind the reader, the problem of incomplete (or missing) markets brings to the forefront of economic analysis the allocation of property rights. In the case of certainty, this restriction did not pose a problem for the validity of the second-welfare theorem which allows all possible ownership structures (distribution of property rights)—and hence, all possible income distributions—to be associated with efficiency and the solution of the economic problem. But in uncertainty, incompleteness also means incomplete contracts which give owners of assets residual rights. These, in turn, condition productive efficiency on certain distributions of assets and, thus, diminish the domain of admissible distributions of income that can support productive efficiency which is, of course, the necessary condition for allocative efficiency. Consequently, the question that arises is whether the distribution of ownership which is necessary for productive efficiency will not force on the economy a distribution of income where some will find their expectations severely frustrated. If this is the case, the efficiency of the system becomes the cause of instability and one will not be able to have both.

Consider a world with two individuals (or two types of individuals), R (for Raskolnikov) and G (for Goriot), and one type of commodity: corn. R owns a plot of land which is located at the top of a mountain while individual G owns a plot located at the bottom of a valley. The plots are not of the same size, and we may suppose that the plot owned by R is larger than that of G. Suppose too that there are two states of nature. It will be either sunny (state 1) or stormy (state 2). If it is stormy (state 2), the crops on R's plot (on top of the hill) is decimated, but G's plot (at the protected bottom of the valley) will yield its maximum yield ($\bar{c}_2$). If state 1 materializes, the sunny state, R's plot on top of the mountain will flourish and yield its maximum ($\bar{c}_1$), but the plot of G, at the bottom of the valley, will suffer devastating frost and will therefore yield nothing. The initial endowments of the two individuals are given by $\omega_G = (0, \bar{c}_2)$ and $\omega_R = (\bar{c}_1, 0)$, respectively (where the index represents the state of the world).

Clearly, neither R nor G are happy to face a state of the world in which they will have absolutely nothing. They may starve to death in such a situation. Hence, they will seek insurance against this by trying to strike a deal with each other. Raskolnikov will offer to sell Goriot corn at an agreed price (p_1), if state 1 materializes (this is when he has a bountiful yield). But Raskolnikov is not interested in the money; he would still need corn to survive. Therefore, he will offer this deal to Goriot in return for a similar offer, according to which, Goriot will agree to sell corn

to Raskolnikov at an agreed price (p_2), if state 2 materializes (the stormy one) when Goriot has bountiful yields. Naturally, each one of them will only offer a fraction of their yield as each one of them would need to retain some of the corn for themselves. A case of full insurance would be a case where the amount of corn they keep for themselves and the corn they buy if they have none will be the same. It means that they will consume the same amount of corn whichever turns out to be the state of nature. The agreed price will be determined in the markets for these claims as many people will seek to buy corn in state 1 and sell in state 2 as well as many who would wish to sell corn in state 1 and buy in state 2.

The economic raison d'etre of financial markets is indeed this one: to allow the individuals who would otherwise be stuck with their endowments ($\omega_G = (0, \bar{c}_2)$ and $\omega_R = (\bar{c}_1, 0)$) to shift through trade to a point of consumption like this ($c^0_{1, R}, c^0_{2, R}$) for R and ($c^0_{1, G}, c^0_{2, G}$) for G (which we will call, point E, see fig. 3 below). Though we mentioned full insurance, this is unlikely to be the case given the asymmetry in the sizes of the plots. However, the markets for state-contingent claims is only imaginary. In reality, as these are promises, there is no sufficiently widespread trade like this. Hence, the alternative to this would be for the individuals to trade in their assets in order to reach the point they would have liked to reach had they been able to trade in state-contingent claims.

How can this happen? Well, in our case it is easy enough. If Raskolnikov sold a fraction of his land, say θ, to Goriot, in exchange for a fraction (φ) of Goriot's land, the distribution of corn between the two states of the world will now become $((1 - \theta) \cdot \bar{c}_1, \varphi \cdot \bar{c}_2)$ for R and $(\theta \cdot \bar{c}_1, (1 - \varphi) \cdot \bar{c}_2)$ for G. If $c^0_{1, R} = (1 - \theta) \cdot \bar{c}_1$; $c^0_{2,R} = \varphi \cdot \bar{c}_2$ and in the same manner for Goriot, the trade-in assets successfully produced the outcome they would have reached had they been able to trade in state-contingent claims, namely, point E. Now, neither of them faces a possible starvation. The actual economic role of financial markets, in this respect, is to emulate through trade-in assets the outcome that would have emerged in a world of state-contingent claims.

What we saw here is that in a world of what we call complete markets, there is evidently no problem for financial institutions to deliver any allocation people would have liked to have reached if they substituted trade in state-contingent claims with the trade in assets. This means that competitive decentralization fully serves the objectives for which the two agents have engaged in market interactions. However, we know that in the case of missing markets (in the sense of the required correspondence between the number of assets and the number of states of the world), none of this will be true and financial markets will always fail in reaching points which people would have liked to have reach if they could trade in state-contingent claims. But this is too easy, instead, let us suppose that there are enough assets and that the chosen allocation could indeed be reached through such a trade. However, the aspect of missing markets which we will preserve is its effect on contracts. Even if the number of assets and states of nature are the same, we still have the problem of an incomplete contract. While some may question that financial markets are indeed incomplete in terms of the number of assets, there can be no question with regard to the incompleteness of contracts.

In such a case, ownership of assets comes with what is known as residual claims. This means that owners of assets can always cream off some of the return and leave a smaller fraction of it for the terms of the contract. As a result, efficiency implies that it is usually better for an agent who operates an asset also to own it. This point, explored in great details in Hart (1995) but has long before been explored by J.S. Mill, makes the ownership structure a question of productive efficiency.[14] As such, it creates a serious hurdle for the ability of economics to free efficiency from specific ownership structures which may produce unsavoury distributions of income both in terms of their moral acceptability but more to the point, in terms of generating frustrated expectations.

In our case, if G and R trade in their land but not their labour,[15] it means that G works on a plot only a part of which is owned by him. The rest would be owned by the other agent R. A similar situation will now face R too. Evidently, G and R will also have contracts (incomplete) with each other as employees in that part of the asset which is owned by the other person. Consequently, there will be some loss of output due to the reduced incentives each one of them faces when working the part of land belonging to the other. Therefore, it is evident that R and G will not be able to reach the allocation they aimed at $(c^0_{1,R}, c^0_{2,R}) = ((1 - \theta) \cdot \bar{c}_1, \varphi \cdot \bar{c}_2)$ for R and $(c^0_{1,G}, c^0_{2,G} = (\theta \cdot \bar{c}_1, (1 - \varphi) \cdot \bar{c}_2)$ for G. Instead, because of the lack of incentives triggered by the fact that some of the produce of one's effort will go to someone else, the total yield of Raskolnikov will now be $\gamma \cdot \bar{c}_1$ and the total yield of Goriot $\delta \cdot \bar{c}_2$ where we assume here that $\gamma < \delta$. This means the loss in output in the larger plot will be greater than in the smaller one. Hence, the new outcome will now become $(c^1_{1,R}, c^1_{2,R}) = ((1 - \theta) \cdot \gamma \cdot \bar{c}_1, \varphi \cdot \delta \cdot \bar{c}_2)$ for Raskolnikov and $(c^1_{1,G}, c^1_{2,G} = (\theta \cdot \gamma \cdot \bar{c}_1, (1 - \varphi) \cdot \delta \cdot \bar{c}_2)$ for Goriot (assuming that neither of them can separate their productivity in their part of their plot from their productivity in the part belonging to the other). This outcome is clearly not the same as point E for which they were aiming when they started trading their assets. Instead, they will find themselves at point which we may call E'.

This is clearly not an outcome which the individuals have expected. It is important to bear in mind that when they struck the deal, they had not known anything about either γ or δ which were unobservable. So, when they determined the split in ownership, they were aiming to reach allocation E which, to remind you, is the one they would have wanted to reach if they could have traded in state-contingent claim. This whole story can now be seen in the diagram in Fig. 3.

Initially, they were at point ω. Had they been able to trade in state-contingent claims, they would have ended at the equilibrium point E. The red line from the origin depicts the equilibrium price of the state-contingent claim (which will also be the intersection of the price-consumption curves of both agents). In the case of complete markets and contracts, they could reach point E by exchanging ownership. This will happen when, as we said above, $(c^0_{1,R}, c^0_{2,R}) = ((1 - \theta) \cdot \bar{c}_1, \varphi \cdot \bar{c}_2)$ for R

---

[14] See a discussion in Witztum (2012).

[15] We can imagine that each one has special expertise in working on their specific plots.
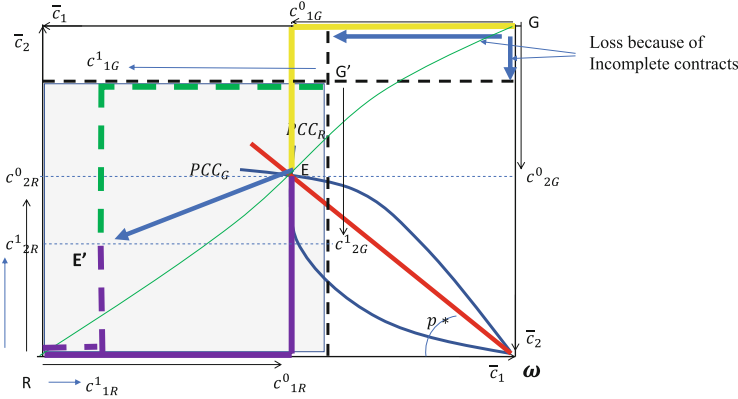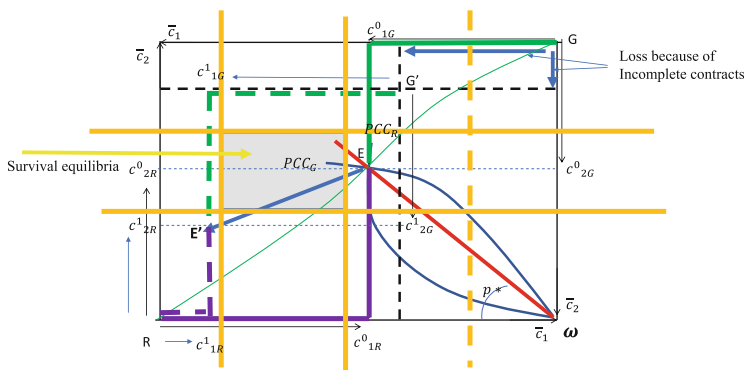
Fig. 3 General equilibrium: trade in assets with incomplete contracts

and $(c^0_{1,G}, c^0_{2,G} = (\theta \cdot \bar{c}_1, (1 - \varphi) \cdot \bar{c}_2)$ for G. This is depicted by the heavy yellow and purple lines. The heavy horizontal yellow line represents the share of corn in state 1 that belongs to Goriot due to his share in the ownership of the plot on the hill. The heavy vertical yellow line (to point E) represents the share Goriot retains in his own plot which produces corn in state 2. Correspondingly, the horizontal heavy purple line represents the share of corn in state 1 which will reflect the retained ownership on plot 1 by Raskolnikov. The heavy purple line rising to E represents the share of R in the corn yield of plot 2 which is due to his share in the ownership of plot 2. As you can see, adding the horizontal yellow line with the horizontal purple line produces the entire output of plot 1.

Now, in the presence of incomplete contracts, as they exchange the assets (the ownership of the plots), they are not aware that because of the incompleteness of the contract, the Edgeworth box shrinks (due to decline in incentives) from point G to G' or to the shaded area in the diagram in Fig. 3. This means that their respective share in the produce in each plot will be represented by the corresponding broken lines. As you can see, this will lead to a distribution of corn according to point E'.

What is interesting about point E' is that it is well beyond the core of the original expected equilibrium at point E. This means that someone, if not both will have their expectations frustrated. As we assumed that plot 1 (on top of the hill) is larger than the one in the valley, the losses of output in it due to the incentive problem emanating from the residual claims derived from the new ownership structure will be greater (this is depicted by the blue arrows from point G leading to G' in the diagram in Fig. 3).

So, what will be the meaning of it? Well, firstly, it will not be an efficient outcome, and financial markets would have failed their economic purpose. Secondly, there will be a lot of frustrated expectations. Notice that in the diagram in Fig. 3, the real disappointment must be that of Raskolnikov who at point E' has very little corn in either state of the world. Goriot, in this scenario, is clearly better off than

Fig. 4 General equilibrium: trade in assets with incomplete contracts and survival restrictions

Raskolnikov but not necessarily better off than he would have been had the markets worked well and led him to point E.

Notice however that the frustrations we mention here are not due just to the fact that the new equilibrium (at E') is outside the original core. It is more associated with what is sometimes called "survival equilibria" (in particular, in development literature) and which is in our case, the area where both agents can socially subsist. What we mean by this, as we said before, is that individuals are capable of performing private as well as social activities commensurate with the standards of social behaviour in their society. In our case, this would simply mean that they have enough corn in each state of the world to be able to interact according to basic social standards. To see this, let us add the survival lines to our diagram (Fig. 4):

The heavy orange lines, as before, represent the lower boundaries below which lies the domain where the agent received too little corn, in either states of the world, to be able to fulfil what he, or she, deems to be appropriate to their social standing (or sense of association). But as we said before, it could also be physical subsistence. The grey box in the diagram in Fig. 4 contains all those allocations where both agents can socially subsist or "survive".[16] Everything in the grey box is socially acceptable even though not all allocations there are efficient. It is easy to see that the allocation to which the individual aspired and the one that would have been the equilibrium had they been able to trade in state-contingent claims (point E) would have been within the survival box of the problem they could have faced had ownership stayed as it were at the beginning (delineated by the broken vertical orange line). This would have included the allocations within the core of the original problem but a lot of other allocations which may indeed be outside this core. In other words, those allocations within the survival box are the allocations which would fit the general social expectations which the agents have had before they engaged in trade.

---

[16]The broken vertical orange line gives a sense where the survival box would have been had there been no loss of corn due to an incomplete contract.

When trade leads them to an outcome which is outside the domain of what is socially acceptable, some agents—those for whom the outcome is adverse—may wish to break the rules or at least, resort to desperate actions in order to help them move towards the survival zone. As economics offers no rational way in which individuals can reach the survival box from point E', agents' actions may not always correspond to choosing the best means to an end. The attraction of greater risk-taking is quite strong, and individuals are likely to become less risk-averse even though standard theory expects them to be more risk-averse than other members of society.

As far as the question of the availability of fund is concerned, we must bear in mind that when some find themselves in the domain of desperation, others have done quite well. This could suggest that for those who benefited from the situation, the outcome in either states of nature would be better than they initially expected. This means that they are likely to offer the frustrated ones the necessary funds for them to execute their greater risk-taking desire. In some sense, one can say that the rich provide the poor with the rope by which to hang themselves. Had the funds not become available for the poor, it is not unlikely that the desire for greater risk-taking activity may translate into social disobedience and riots. In this respect, financial crises, paradoxically, are a way of avoiding social upheavals.

Either way, this story reveals the dilemma facing modern economics because of its fundamental belief that competitive decentralization works well in delivering an allocation which is both efficient (solves the economic problem) and socially desirable. Firstly, instead of attributing all these glorious properties to a single competitive equilibrium, modern economics faces three competing allocations that cannot really be reconciled. These are the socially desirable one; the one that is needed for financial markets to achieve their economic objectives; and the one that supports productive efficiency in the case of an incomplete contracts. In the Walrasian scheme, all three can be one. In fact, even if we assume—which is far from obvious—that the first two can become one, the tension with the third may become detrimental. In our example, point E represents the potential amalgamation of all three. There is no obvious reason to suppose that this (more or less equal) allocation is not the socially desirable one. It is also the one in which financial markets succeed in solving the economic problem and the one—in case of complete contracts—which is also productive efficient. However, the fact that under incomplete contracts E will always mean E' means that the three allocations can never become the market solution to the economic problem. Secondly, financial crises demonstrate that even in a society where there is no danger of physical survival, individuals may still behave in an unexpected way because the expectations they have from society are not only material and the ability of the markets to fulfil these expectations—given the inherent inefficiency of financial markets (always giving E' when we seek E)—will always be frustrated.

4   Conclusion

There can be little doubt that the recent financial crisis revealed the importance of income distributions to the question of financial stability. It was preceded by a sharp and consistent increase in inequality. This raised two questions: the first, whether inequality could have been the trigger of financial instability, and the second, depended on the answer to the first question, is whether the remedy should be to focus on correcting income distributions.

The most obvious explanation as to why inequality may trigger a crisis rests with both the increased availability of market funds and excessive risk-taking by those who cannot really afford it who then, inevitably, default and bring down the entire system. The former is fairly straightforward, and the evidence seems to support it too. The latter is more problematic in the sense that it does not sit well with the tenets underlying the economic system in which financial markets have an integral role. After all, the rationality of agents upon which we identify the merits of competitive decentralization does not easily lend itself to the claim that the poor—those who cannot afford it—would resort to behaviour characterized by irresponsible and excessive risk-taking. With rational agents, inequality could not have triggered a crisis, in particular, as the cost of borrowing seemed to have increased as well.

Therefore, there is a need to explain why it is at all possible, within the modern paradigm, for such behaviour to emerge even though we may generally wish to hold on to the view that agents are rational in the way depicted by modern economics (i.e. seeking the best means to an end in a consistent manner). One such explanation can be given by identifying two general types of competitive equilibria, those which lie in the core or are not too far from it, and those which are at the edges when one group of individuals are in a position where they may be better off than they were before trade, but the outcome nevertheless frustrates something fundamental about their expectations from the system. These frustrated expectations, then, can give a reason why even in a world of rational agents, there are conditions under which they break and resort to more extreme forms of behaviour which may indeed include excessive risk-taking.

The problem that is created by the existence of such domains where competitive equilibria are a trigger for frustrated expectation is that it limits the set of admissible distributions of income. In a world of certainty, even with missing markets, this does not create a difficulty as society can simply reallocate initial endowments to achieve the appropriate distributional outcome and thus preserve efficiency and stability. However, in a world of uncertainty, the problem of incomplete markets (or contracts) restricts the domain of ownership structures (and subsequently, distributions of income) for which efficiency can be maintained. We show how this can lead to an inherent contradiction between those distributions commensurate with efficiency and those where expectations are not frustrated.

References

Bordo, M. D., & Meissner, C. M. (2012). Does inequality lead to a financial crisis? *Journal of International Money and Finance, 31*(8), 2147–2161.

Bosch-Domènech, A., & Silvestre, J. (2006). Do the wealthy risk more money? An experimental comparison. In C. Schultz & K. Vind (Eds.), *Institutions, equilibria and efficiency. Essays in honor of Birgit Grodal* (pp. 95–116). Springer.

Brei, M., Ferri, G., & Gambacorta, L. (2018). Financial structure and income inequality. *BIS working papers* no. 756. Bank for International Settlements.

Coibion, O., Gorodnichenko, Y., Kudlyak, M., & Mondragon, J. (2016). Does greater inequality lead to more household borrowing? New evidence from household data. *Federal Reserve Bank of San Francisco working paper series* no. 2016–20. Federal Reserve Bank of San Francisco.

Coles, J. L., & Hammond, P. J. (1995). Walrasian equilibrium without survival: Equilibrium, efficiency, and remedial policy. In K. Basu, P. K. Pattanaik, & K. Suzumura (Eds.), *Choice, welfare and development: A festschrift in honour of Amartya K. Sen (pp. 32–64)*. Oxford University Press.

Courbage C., Montoliu-Montes, G., & Rey, B. (2018). How vulnerable is risk aversion to wealth, health and other risks? An empirical analysis for europe. *GATE working paper* no 1827. Université Lumière Lyon 2, Groupe d'Analyse et de Théorie Economique.

Dick, A., & Lehnert, A. (2007). Personal bankruptcy and credit market competition. *Federal Reserve Bank of New York staff reports* no. 272. Federal Reserve Bank of New York.

Dijk, O., Holmen, M., & Kirchler, M. (2014). Rank matters–the impact of social competition on portfolio choice. *European Economic Review, 66*, 97–110.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. B., & Sunde, U. (2015). The nature and predictive power of preferences: Global evidence. *IZA discussion paper* no. 9504. IZA Institute of Labor Economics.

Fitoussi, J.-P., & Saraceno, F. (2010). Inequality and macroeconomic performance. *Documents de travail de l'OFCE* no. 2010–13. Sciences-Po, Observatoire Francais des Conjonctures Economiques.

Galbraith, J. K. (2012). *Inequality and instability: A study of the world economy just before the great crisis*. Oxford University Press.

Genakos, C., & Pagliero, M. (2012). Interim rank, risk taking, and performance in dynamic tournaments. *Journal of Political Economy, 120*(4), 782–813.

Hart, O. (1995). *Firms, contracts and financial structure*. Clarendon Press.

Hopkins, E. (2018). Inequality and risk-taking behaviour. *Games and Economic Behavior, 107*, 316–328.

Kumhof, M., Lebarz, C., Rancière, R., Richter, A., & Throckmorton, N. (2012). Income inequality and current account imbalances. *IMF working papers* no. WP/12/08. International Monetary Fund.

Kumhof, M., Rancière, R., & Winant, P. (2015). Inequality, leverage and crises. *American Economic Review, 105*(3), 1217–1245.

Milanovic, B. (2009). Global inequality and the global inequality extraction ratio: The story of the past two centuries. *Policy research working paper series* no. 5044. The World Bank.

Palley, T. I. (2012). *From financial crisis to stagnation: The destruction of shared prosperity and the role of economics*. Cambridge University Press.

Payne, B. K., Brown-Iannuzzi, J. L., & Hannay, J. W. (2017). Economic inequality increases risk taking. *Proceedings of the National Academy of Sciences, 114*(18), 4643–4648.

Rajan, R. G. (2010). *Fault lines: How hidden fractures still threaten the world economy*. Princeton University Press.

Reich, R. B. (2010). *Aftershock: The next economy and America's future*. Random House.

Stiglitz, J. (2009). The rocky road to recovery. *Project Syndicate*, 6 January. Accessed July 3, 2021, from http://www.project-syndicate.org/commentary/the-rocky-road-to-recovery

Stockhammer, E. (2015). Rising inequality as a cause of the present crisis. *Cambridge Journal of Economics, 39*(3), 935–958.

Wade, R. (2009). From global imbalances to global reorganisations. *Cambridge Journal of Economics, 33*(4), 539–562.

Witztum, A. (2012). The firm, property rights and methodological individualism: Some lessons from J.S. Mill. *Journal of Economic Methodology, 19*(4), 339–355.

Witztum, A. (2019). *The betrayal of Liberal economics* (Vol. 2). Palgrave Macmillan.

Inequality, Economic Policy, and Household Credit in the USA: The Roots of Unsustainable Finance

Orsola Costantini

The debt-to-income ratio of American families started its unprecedented ascent back in the 1980s. Already at the time, the phenomenon attracted the attention of economists, generating reactions which can be roughly divided into two categories. On one side, some scholars developed interpretations based on a life-cycle savings approach with rational expectations, updating the models by Modigliani and Muth, and supported the optimistic view that households increase their indebtedness because they expect an increase in cash flows and want to smooth their consumption accordingly. On the other side, a number of economists explored the links between the rise in household debt, the fall in the labor share of income, and the financialization of the economy, building on the work of Kaldor, Kalecki, and Keynes on inequality and radical uncertainty. Their view of the risks associated with financial deregulation and the suppression of wages as a stable source of demand was much more somber.

The life-cycle approach prevailed at the time and was foundational to a consensus view in macroeconomics which supported the widespread financial deregulation and active monetary policy of the 1980s, 1990s, and 2000s. The decades were dubbed as the Great Moderation: a prolonged period of reduced output volatility.

Eventually, however, the financial crash in 2007 and the ensuing crisis awoke the world to the fragilities that had been building up under the surface. While shocking details emerged about how decades of wild deregulation had transformed the credit industry, inequality had its comeback moment: a plethora of studies by scholars from different schools of thought pointed to the dramatic increase in the debt of the lowest income groups and to the predatory practices of credit institutions that targeted them.

However, evidence has emerged that some of the most influential research works written shortly after the crisis overestimated the role of consumer spending and of

O. Costantini (✉)
United Nations Conference on Trade and Development, Geneva, Switzerland
e-mail: orsola.costantini@un.org

the financial decisions of low-income households both in the debt boom and in the wave of defaults. The new findings have crucially exposed simplistic readings of the link between income inequality and financial fragility, such as those put forward by Mian and Sufi and coauthors (Mian et al., 2013; Mian & Sufi, 2015).

Based on the new evidence, alternative interpretations suggested that household debt and financial fragility are essentially an upper-middle class story, related to housing rather than to consumption, and basically unaffected by poverty and income inequality (Jordà et al., 2016; Mason, 2018). This new literature has accordingly moved away from income inequality and turned to a portfolio approach, to understand the source of financial fragility and crises in a system prone to speculative boom-bust cycles, attributing systemic risk to the exposure of upper-middle class families to asset price fluctuations (Bartscher et al., 2020).

In this chapter, I provide a different interpretation of the relationship between debt, inequality, and crises. Crucially, I refuse to identify the macroanalysis of household indebtedness with a mere study of the quantitative contributions to its accumulation over time or of the immediate causes of the latest housing boom and its crash in 2008. On the contrary, I consider the overindebtedness of American families as a long-term phenomenon, whose macroeconomic impact goes beyond the burst of one financial crisis and which is situated within a specific historical context characterized by intense financialization, globalization, and economic policy activism.

First, I present evidence from the Fed's Survey of Consumer Finances[1] (from 1989 to 2019), showing that the financial conditions of the bottom of the personal income distribution have changed dramatically over the past decades, bearing signs of distress associated with the compression of available incomes. This is particularly evident when we take into account the burden of necessary expenditures and debt servicing. The overindebtedness of the low and middle-income households, I argue, is a crucial new factor that characterizes household finances since the 1980s compared to the previous period. This is true even though it may not have been the immediate cause of the housing and mortgage boom of the 2000s, which had a more typical speculative genesis that spanned over 6 years.

Then, I discuss how this increase of the indebtedness of the American households has been a key factor in the dynamics and perpetuation of a fragile and unstable, and yet exceptionally resilient, economic framework. In fact, the unprecedented possibility for households to access credit has represented the emergence of a significant "external" source of profit for firms, which can be associated with what Michał Kalecki dubbed "domestic exports" (Kalecki, 1990, pp. 164–173). With that term, the Polish economist referred to public net spending as a domestic source of demand that did not generate from within the capitalist productive process and, together with net exports, could provide an additional and often critical support to economic

---

[1] The Survey of Consumer Finances (SCF) is a triennial cross-sectional survey of US families of the US Federal Reserve. The survey data include information on families' balance sheets, pensions, income, and demographic characteristics.

reproduction and growth.[2] In recent decades, households' net borrowing played precisely that role, in a context of wage repression and reduction of public deficit spending. However, as opposed to public debt, the sustainability of household credit accumulation is subject to inherent uncertainty: the incomes, which households need to pay back the amounts borrowed, derive from the decisions of a different sector, that of firms, which are at best indirectly influenced by household spending when they determine the size of employment and production. Indeed, in a context of secular fall in the rate of investment, of real wage compression and of reduction in net public spending in the USA (and the other developed countries), household "autonomous spending" has sustained firms and banks earnings but ultimately resulted in an unprecedented accumulation of debt and financial fragility for the sector.

Nevertheless, in the face of this fragility, debt kept growing for many decades and even during a crisis, in 2001. Only in 2008 there was a deflagration and a sharp deleveraging, but shortly after the end of the 2008–2009 crisis, households went on to accumulate new debt at a slower but still alarming pace. What explains this extraordinary persistence, and what are its consequences? The answer lies precisely in the ability of the other domestic "external" channel, government and central bank, to intervene.

This interaction between two sources of "domestic exports" is what lies at the root of the economic dynamics of the past 40 years. The alternate surges and falls in household net borrowing and public net spending have subsidized the US economy but have also accompanied it into a phase of low growth and stagnant productivity, punctuated by financial excesses.

1 The Distribution of Household Debt

The households in the upper-income quintiles hold the large majority of the total dollar value of debt, most of it in mortgages. Bartscher et al. (2020), for example, show that the mortgage debt of the households between the 50th and 90th income percentiles explains a large part of the increase in the sector's indebtedness since the 1950s. This increase reflects a higher dollar value of mortgages (intensive margin), more than a greater number of mortgage originations (extensive margin). Also Adelino et al. (2016, p. 1636) find that "the majority of new mortgages by value were originated to middle-class and high-income segments of the population even at the peak of the boom."

Similarly, high-credit score borrowers have held the majority of mortgage originations in the pre-crisis decade (Ferreira & Gyourko, 2015; Adelino et al., 2016), while the share of originations to subprime borrowers (those with a credit score

---

[2] As I shall explain later, the importance of the distinction can be understood best when put in the context of Kalecki's reading of the work of Rosa Luxemburg on the accumulation of capital.

below 660) relative to high credit score borrowers remained stable across the pre-crisis period. In direct contrast with the results by Mian et al. (2013), it appears that "although the pace of origination rose in low-income ZIP codes, this increase did not translate into significant changes in the overall distribution of credit, given that it started from a low base (borrowers in low-income and subprime ZIP codes obtain fewer and significantly smaller mortgages on average)" (Adelino et al., 2016, p. 1636).

Moreover, Amromin and Paulson (2009) find that holders of prime mortgages originated in 2006 and 2007 were extremely sensitive to home price changes, more so than in the past and customarily predicted. Haughwout et al. (2011) show that the effect was largely driven by the behavior of a small group of real estate investors, who were prime borrowers and fell in the 50th–90th income percentile group. For them, there was little to lose from a default and quickly reacted to the changed conditions in the market and in the cost of credit. Indeed, speculative activity certainly rests at the heart of the housing crisis, but it is not representative of the large majority of the debt holders. In fact, if we exclude speculative business activity, most households will try to keep their home for as long as possible, regardless of the changes in its price and depending exclusively on their available income.

Indeed, we know that, if the largest share of the dollar value of total delinquent mortgages belonged to prime borrowers, the default rate remained significantly higher for subprime borrowers (Amromin & Paulson, 2009; Haughwout et al., 2019).[3]

Moreover, while it is unsurprising that mortgage debt appears as the first contributor to household debt growth, evidence of a strong increase in the indebtedness of lower-income classes and lower credit score individuals is similarly uncontroversial. Figure 1 shows how the debt to pre-tax income ratio of three groups in the equivalent income distribution ladder,[4] that is, the income of the household, adjusted by the number of its members, has changed from 1989 to 2019. The choice to look at the distribution of debt by equivalent income rather than total income groups reflects the obvious observation that available income and spending decisions depend on the number of members of the household: a single person household and family of three earning the same total income face very different financial conditions and are likely to make different decisions.

The bottom 50% of the equivalent income distribution saw an increase of more than 35 percentage points in the course of the 1990s and again in the following decade, before the global financial crisis. The rate of growth is of course much larger, given the low starting point: more than 50% in the first time interval. The figure would be higher if, as is customary, I included in the estimate of the total debt the

---

[3] In addition, most prime mortgages for amounts at or below $417,000 are guaranteed through the government-sponsored enterprises (Amromin & Paulson, 2009).

[4] I use the customary OECD formula, which assigns weight 1 to the head, 0.75 to any additional adult, and 0.5 to each child (OECD Project on Income Distribution and Poverty, via www.oecd.org/social/inequality.htm accessed in August 2018).
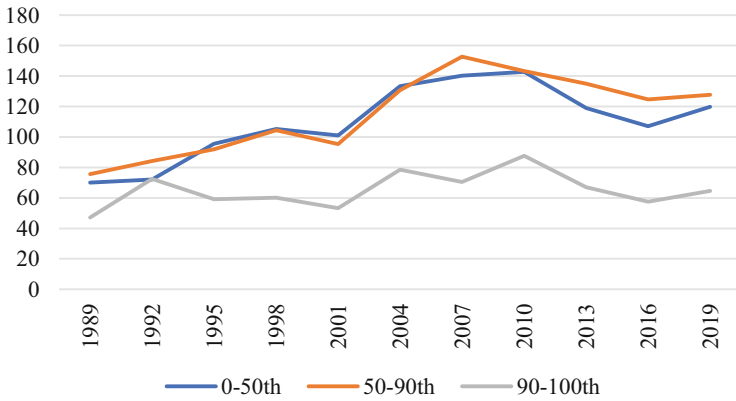
Fig. 1 . Debt to income ratio by percentile groups of equivalent income (*source*: own calculation from Survey of Consumer Finances (debt does not include education loans whose repayment is not yet due)

education loans whose repayment is not yet due—that is, during school and in the following grace period. However, I decided to ignore them to avoid considering cases of households with large debt and no or low incomes, but for which the outstanding liabilities do not yet influence current financial strain.

The debt-to-income ratio of the group between the 50th and 90th percentile increased by 48 percentage points (a record jump) in the 1998–2007 period, which corresponds to the housing boom. The "excess" growth from the previous period, however, was more than entirely canceled out in the following decade, when there was a − 25 percentage points fall. In the postcrisis period, the debt-to-income ratio of the lowest equivalent income group did not fall immediately but ultimately reverted, in 2016, to its 1998 level (keep in mind that I am not including the education loans of individuals still in school). Hence, in the housing boom period, the size of the indebtedness of the bottom group of the equivalent income distribution (as a group) changed significantly relative to both before and after—at least until 2016. However, data from the 2019 survey show that the bottom group's indebtedness has picked up again, so that ultimately the 1989–2019 rate of growth of the debt-to-income ratio is very similar across the group of households below the top 10% of the distribution.

But the sectoral, or even the household-level, debt-to-income ratio per se is not very informative of the financial conditions of families. In fact, financial strain should be considered in relation with the expenditures faced by individuals. Households not only must pay off debt coupons and interests but also cover for the cost of essential goods and services, as well as maintain a buffer of precautionary saving for future unexpected occurrences, for many of which social security has ceased to provide (Pressman & Scott, 2009). Researchers have found, for instance, that poor health contributes to financial distress (Lyons & Yilmazer, 2005) and an increase in

Table 1 Median debt payments to income ratio (%) by equivalent income group (*source*: own calculation from Survey of Consumer Finances)

	0–50th	50–90th	90–100th
1989	18.3	15.6	11.0
1992	16.0	16.7	13.3
1995	18.4	16.7	11.2
1998	19.9	18.4	11.7
2001	19.6	16.9	10.9
2004	20.0	18.4	12.7
2007	20.9	19.6	12.3
2010	19.7	19.4	12.7
2013	17.1	17.5	10.9
2016	15.8	15.8	10.8
2019	17.2	16.1	9.9

Table 2 Median equivalent income by equivalent income group (*source*: own calculation from Survey of Consumer Finances)

	0–50th	50–90th	90–100th
1989	14,653	43,527	120,891
1992	18,322	49,576	116,396
1995	15,541	42,738	105,991
1998	17,004	46,885	112,153
2001	17,824	51,987	138,633
2004	18,100	52,490	132,115
2007	17,887	52,270	145,610
2010	16,909	49,105	133,155
2013	16,727	48,540	135,124
2016	17,744	52,423	155,481
2019	19,437	56,895	152,718

the health insurance coverage of the population positively affects household finances and credit scores (Mazumder & Miller, 2016).

Table 1 shows the median debt payments to income ratio of the different equivalent income groups and Table 2 the median equivalent income in 2019 dollars. A quick comparison suggests that the propensity of households to spend out of available income has much increased. Households with a large share of credit card or other installment consumer debt, typically located at the bottom of the income distribution, are particularly exposed (for contractual reasons) to an increase in the servicing costs, which occurred in spite of low official rates (Mora, 2014; Costantini, 2020). This also suggests that, even when individuals did not originally get into debt for necessitous spending, they may have remained locked into a situation of overindebtedness that reduced their more liquid savings and capacity to face income, interest rates, and other costs shocks.

This has systemic consequences. Generally speaking, if disposable income shrinks, it is more difficult to reduce the share of spending. But because of the stretch of past financial decisions, expenditures become increasingly rigid to downward income fluctuations and less dependent on a current evaluation of income availability and consumption or investment desires. Rather than a stickiness of

habits, there is a *stickiness of costs*, which generates a smooth flow of transfers to the rentier sector and sustains household propensity to consume out of available income (Costantini & Seccareccia, 2020). Of course, this becomes particularly worrisome as the main family asset—its house—devalues dramatically, as it occurred in 2007. But trouble occurs even absent such an extreme event: the first house is a very illiquid asset for low and middle-income households, who will sell it only as a last resort remedy. Hence, as long as they are still paying off their mortgage, they are not financially distinguishable from renters.

The contribution of debt servicing to the macro trend of household debt growth has been significant, as highlighted by Mason and Jayadev (2014). The two authors showed that, if in the run-up to the crisis, from the late 1990s to 2007, household spending (net new borrowing) excluding interest played an important role in the growth of the debt-to-income ratio, in the long run interest rate charges and as well as subdued inflation have been the prime contributors to the growth in the aggregate. Adopting the terminology proposed by Minsky (1992), we can say that households as a group were speculative borrowers; that is to say, they could not save enough to service the debt and stabilize it relative to income.

Indeed, households' propensity to save follows a downward trend since 1981. Cynamon and Fazzari (2017) calculate that the savings of the American families, minus their debt payments, fell below zero as soon as 1987. Throughout the period, the real per capita personal outlays outpaced personal disposable income and grew more smoothly than in the past, notably during the crisis of 2000 and following a recession. Bibow (2010) further documents that the saving rate of the top 5% of the income distribution fell significantly in the years of the dot-com bubble, while the propensity to save of the rest of the households—which had remained somewhat stable in the course of the 1990s—started falling immediately after the bubble burst and throughout the global financial crisis.

In conclusion, the involvement of households in the lower and middle class in the debt boom is clear in the data. With it, came an unprecedented increase in their financial distress and fragility, which was however compatible with an increase in the propensity to spend (including debt servicing costs), thus providing a smooth source of liquidity to the rentier and nonfinancial corporate sectors.

2 The Long-Term Macroeconomic Consequences of Household Overindebtedness

Households' net financial position started to worsen in the mid-1980s, with a through at the turn of the century. It improved slightly after the global financial crisis, before reverting to the previous trend. A comparison of households net borrowing with that of the other sector of the economy, in Fig. 2, clearly suggests that, in four decades, the former has become a central feature and source of liquidity for the American economy. Most interestingly, it appears to mirror almost exactly
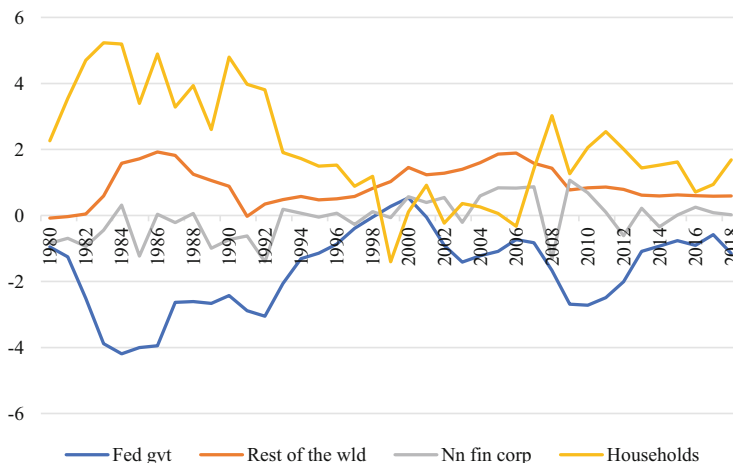
Fig. 2 Net lending/borrowing of the federal government and households to GDP (%) (source: FRED, Federal Reserve Economic Data Series: FGLBAFQ027S; RWLBACQ027S; NCBLAFQ027S; HHNONLA027N)

the net flow of the government, while private nonfinancial corporations circled around a balanced position.

From this optic, and using a terminology advanced by Michal Kalecki, household net borrowing (spent domestically) can be interpreted as a source of "domestic exports," that is, a domestic source of demand that does not derive from incomes generated within the capitalist productive process (Kalecki, 1990, pp. 164–173; Costantini, 2015). With that expression, originally, Kalecki referred to the government deficit which, together with actual net exports from abroad, "enable the capitalists to make profits over and above their own purchases of goods and services" (Kalecki, 1991a, p. 246).

Using the adjective "external," Kalecki acknowledged the influence exerted on his analysis by Rosa Luxemburg who, in her 1913 book *The Accumulation of Capital*, developed the idea that external demand is essential to extended reproduction (Luxemburg, 2003; Bellofiore, 2009).

Luxemburg, like Kalecki, used Marx's schemes of reproduction as the starting point of her analysis. But "both used Marxian [schemes] to search for the limits of capitalist accumulation. Using more modern words, they treated capitalism as a system, limited by effective demand, sharply distinguishing the production of commodities from their realization" (Kowalik, 2009, p. 111, cited in Harcourt & Kriesler, 2012). In Luxemburg's view, the accumulation of capital was only plausible as long as capitalists believed that their surpluses could be monetized. That outcome, however, was only possible—for the capitalists as a class—if "external" sources of demand intervened, effectively subsidizing the capitalist system.

This "aspect of accumulation of capital [does not depend on what happens inside the plants] but concerns the relations between capitalism and the non-capitalist

modes of production which start making their appearance on the international stage. Its predominant methods are colonial policy, an international loan system—a policy of spheres of interest—and war" (Luxemburg, 1960, p. 432). Imperialism, in her view, was best described as a specific mode of accumulation. But also the government contributed with its demand, when not funded by taxation.

Albeit considering Luxemburg's view as extreme and admitting the theoretical possibility of extended reproduction in a closed capitalist system, Kalecki recognized the historical relevance of external markets "[w]ithout [which], profits are conditioned by the ability of capitalists to consume or to undertake capital investment" (Kalecki, 1991a, p. 246; Kalecki, 1991b), an activity, the latter, which he believed to be filled with uncertainty, which forces firms to engage in competition, innovation, and distributive conflict.

While Kalecki especially emphasized the role of the government, the sectoral flows in Fig. 2 suggest that the reduction of household net lending has represented a primary new and additional source of liquidity, in a context of reduction in public deficit spending. Such inflow has appeased competition among firms and between firms and workers and increased profitability, in spite of wage repression and fall in employment rates (Storm, 2017; Taylor & Öner, 2020).

A similar framing of the phenomenon of household debt has been advanced by Seccareccia (2011) who expressed the problem in terms of the theory of the circuit.[5] In addition, more recently, Fiebiger (2018) Fiebiger and Lavoie (2019) have presented empirical exercises that show that household residential investment (as semiautonomous demand) has contributed significantly to the American business cycle.[6]

My reading, however, points also to the fact that household debt accumulation is itself the result of a targeted political strategy (Costantini, 2015, 2020). Not only financial deregulation and direct incentives have in no small measure boosted the increase in household indebtedness, but the fall in net savings is also clearly the reflection of a contextual, programmed, retrenchment of net public spending and, more specifically, of welfare spending, coupled with the privatization of many social services. In this context, firms and government have relied on household indebtedness in order to avoid the (short-term) consequences of inequality and wage compression on revenues and political stability. However, a clarification here is necessary: I do not ascribe to the idea advanced by Rajan (2010) that politicians in power "cynically" favored easy credit and related policies, for political calculus, that is to temporarily please the impoverished middle class. I rather believe that it was the successful lobbying of the financial industry which led to the outcome (Ferguson et al., 2017).

---

[5]On Rosa Luxemburg as a forerunner of the circuit theory, see Bellofiore (2009) and Bellofiore and Passarella (2009).

[6]Other authors in the tradition of the Sraffian Supermultiplier have also contributed to the empirical debate, see for instance Pérez-Montiel and Pariboni (2021).

Either way, however, this strategy has a limit. The net borrowing of government (and firms) as sectors, in fact, creates the conditions (revenues) for itself to be repaid. But household debt, in most cases, does not generate directly any revenue for those who bear it. Indirectly, it may stimulate spending thus generating positive profit expectations for firms, who then may start new investment plans and increase employment.

But, as we have seen, firms have done just the opposite: They have reduced the domestic effort to create (good) jobs and repressed wages. On the contrary, they have used an unprecedented portion of their earnings to buy their own equities and to allow their executives to cash in (Lazonick, 2016). That is why, to allow the system to persist, governments and central banks have learned to deal with emergencies, providing timely stimulus and prompt transfers to the too big to fail institutions, thus filling the gap left by the private borrowers.

Once again, the lesson taught by Luxemburg ([1913] Luxemburg, 1960, p. 491, my translation) comes to mind:

> the actual historical process of the capitalist dynamics [cannot be grasped] if we transcend from all the conditions that determine its historical reality. The accumulation of capital as a historical process rolls out, from beginning to end, within an environment with many pre-capitalist formations, and opens its way in a constant political struggle, and in a perennial game of actions and reactions with them.

Hence, while "theoretically" unsustainable, a system that relies on household autonomous spending can persist and show remarkable resilience, thanks to the policy decisions that, first, encouraged the increase in borrowing and then compensated its fall. Most importantly, however, the *timely and short-lived* nature of such stimulus does not provide the appropriate context for a democratic discussion over a fair allocation of resources. In fact, the system just described is compatible with the idea of "economic alarmism" as a political strategy to capture the state (Caffè, 1976): it is most convenient for elites to reduce prudential economic interventions in order to take advantage of the emergency to apply measures that do not command democratic support by depicting them as necessary. In the case of 2008, governments lavished the financial sector and the financialized nonfinancial corporate sector with public money and continue to do so, under the provisions of the Dodd-Frank Act (Stiglitz, 2010; Kregel, 2012; Kane, 2020).[7]

While there is no ultimate and all foreseeing Architect, the ideological nature of this institutional and political setting emerges clearly from the economic literature commissioned to support it. The contributions go from the life-cycle models with rational expectations and asymmetric information, which supported the deregulation of the credit market and the low-interest rates policy of Alan Greenspan (Hall, 1978; Hall & Mishkin, 1980; Zeldes, 1989; Krueger & Perri, 2006), to the new neoclassical synthesis that prescribed, first, monetary activism (Eggertsson & Woodford, 2004;

---

[7]Several other studies have explored the private capture of the neoliberal state in general terms (see, for instance, Crouch, 2009), and Thomas Ferguson has studied the phenomenon in the American case in a series of articles and books (e.g., Ferguson & Johnson, 2013)

Tcherneva, 2008) and then its coordination with timely and short-lived deviations from the government budget balance (Bernanke, 2008).[8]

3   Conclusions

In contrast with what recent literature argues, I contend that the overindebtedness of the household at the bottom of the income distribution remains relevant for macro-economic stability. On the one hand, the upper-middle class holds the largest share of debt and contributed the most to the debt and mortgage boom of the 2000s. On the other hand, data from the Survey of Consumer Finances confirm that the families in the lower 50% of the equivalent income distribution saw their debt-to-income ratio dramatically increase over the past decades together with financial strain and debt service payments.

Household spending, in this context, provided a smooth flow of revenues to the corporate and rentier sectors. In fact, since the 1990s, net household borrowing (minus so financed imports) can be described as a source of internal exports: an inflow of liquidity and source of revenue from within the country but external to the productive system and its distributive framework. Such inflow appeases competition among firms and between firms and workers and increases profitability.

But the unsustainable nature of this autonomous spending from the household sector forces governments and central banks to deal with all emergencies, tempo-rarily injecting new liquidity into the system: a more typical form of domestic exports. While this dynamic has so far allowed big oligopolistic corporations to avoid the short-term consequences of income inequality on their balance sheets, it has actually supported and directly subsidized the financialization of the economy and thus financial instability.

In contrast with recent literature that takes a portfolio (and wealth inequality) approach to household financial instability, I conclude that a policy attempting to reduce fragility cannot avoid to tackle the structure of aggregate demand and the sources of income inequality.

References

Adelino, M., Schoar, A., & Severino, F. (2016). Loan originations and defaults in the mortgage crisis: The role of the middle class. *The Review of Financial Studies, 29*(7), 1635–1670.

Amromin, G., & Paulson, A. L. (2009). Comparing patterns of default among prime and subprime mortgages. *Economic Perspectives, 33*(2), 18–37.

---

[8]For a better discussion of those, I refer to Roncaglia (2019).

Bartscher, A.K., Kuhn, M., Schularick, M., & Steins, U. (2020). Modigliani meets Minsky: Inequality, debt, and financial fragility in America, 1950–2016. *FRBNY staff report* no. 924. Federal Reserve Bank of New York.

Bellofiore, R. (2009). The monetary circuit of capital in the Anti-Critique. In R. Bellofiore (Ed.), *Rosa Luxemburg and the critique of political economy* (pp. 53–63). Routledge.

Bellofiore, R., & Passarella, M. (2009). Finance and the realization problem in Rosa Luxemburg: A "Circuitist" reappraisal. In J.-F. Ponsot & S. Rossi (Eds.), *The political economy of monetary circuits* (pp. 98–115). Palgrave Macmillan.

Bernanke, B. (2008). *Testimony before the committee on the budget*. U.S. House of Representatives.

Bibow, J. (2010). Financialization of the U.S. household sector: The "subprime mortgage crisis" in U.S. and global perspective. *IMK study* no. 3/2010. Macroeconomic Policy Institute (IMK) at the Hans Boeckler Foundation.

Caffè, F. (1976). *Un'economia in ritardo: contributi alla critica della recente politica economica italiana*. Bollati Boringhieri.

Costantini, O. (2015). A burning debt. The influence of household debt on investment, production and growth. *MEQ working paper* no. 1/2015. University of Bergamo, Department of Management, Economics and Quantitative Methods.

Costantini, O. (2020). Household debt and the rentier share of income. In H. Bougrine & L. P. Rochon (Eds.), *Economic growth and macroeconomic stabilization policies in post-Keynesian economics* (pp. 68–89). Edward Elgar Publishing.

Costantini, O., & Seccareccia, M. (2020). Income distribution, household debt and growth in modern financialized economies. *Journal of Economic Issues, 54*(2), 444–453.

Crouch, C. (2009). Privatised Keynesianism: An unacknowledged policy regime. *The British Journal of Politics and International Relations, 11*(3), 382–399.

Cynamon, B. Z., & Fazzari, S. M. (2017). Household income, demand, and saving: Deriving macro data with micro data concepts. *Review of Income and Wealth, 63*(1), 53–69.

Eggertsson, G.B. & Woodford, M. (2004). Optimal monetary and fiscal policy in a liquidity trap. *NBER working paper* no. 10840. National Bureau of Economic Research..

Ferguson, T., & Johnson, R. (2013). When wolves cry "wolf": Systemic financial crises and the myth of the danaid jar. In L.-P. Rochon & M. Seccareccia (Eds.), *Monetary economies of production banking and financial circuits and the role of the state* (pp. 73–98). Edward Elgar Publishing.

Ferguson, T., Jorgensen, P., & Chen, J. (2017). *Fifty shades of green: High finance, political money, and the U.S. Congress*. Report for the Roosevelt Institute.

Ferreira, F. & Gyourko, J. (2015). A new look at the U.S. foreclosure crisis: Panel data evidence of prime and subprime borrowers from 1997 to 2012. *NBER working paper* no. 21261. National Bureau of Economic Research.

Fiebiger, B. (2018). Semi-autonomous household expenditures as the causa causans of postwar US business cycles: The stability and instability of Luxemburg-type external markets. *Cambridge Journal of Economics, 42*(1), 155–175.

Fiebiger, B., & Lavoie, M. (2019). Trend and business cycles with external markets: Non-capacity generating semi-autonomous expenditures and effective demand. *Metroeconomica, 70*(2), 247–262.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy, 86*(6), 971–987.

Hall, R. E. & Mishkin, F. S. (1980). The sensitivity of consumption to transitory income: Estimates from panel data on households. *NBER working paper* no. 505. National Bureau of Economic Research.

Harcourt G. C. & Kriesler, P. (2012). Michał Kalecki and Rosa Luxemburg on Marx's schemes of reproduction: Two incisive interpreters of capitalism. *UNSW Australian School of Business Research paper* no. 2012–34. USNW Business School.

Haughwout, A., Tracy, J., & van der Klaauw, W. (2011). Real estate investors, the leverage cycle, and the housing market crisis. *FRBNY staff report* no. 514. Federal Reserve Bank of New York.

Haughwout, A., Lee, D., Scally, J., Thomas, L., & van der Klaauw, W. (2019). Handbook of the consumer chapter: Trends in household debt and credit. In A. Haughwout & B. Mandel (Eds.), *Handbook of US consumer economics* (pp. 23–51). Elsevier.

Jordà, Ò., Schularick, M., & Taylor, A. M. (2016). The great mortgaging: Housing finance, crises and business cycles. *Economic Policy, 31*(85), 107–152.

Kalecki, M. ([1933] 1990). On foreign trade and "domestic exports". As reprinted in Osiatynski, J. (ed.), *Collected works of Michał Kalecki,* Vol. I: *Capitalism: Business cycles and full employment* (pp. 165–173). Oxford University Press.

Kalecki, M. ([1954] 1991a). Theory of economic dynamics: An essay on cyclical and long-run changes in capitalist economy. As reprinted in Osiatynski, J. (Ed.), *Collected works of Michał Kalecki*, Vol. II: *Capitalism*: *Economic dynamics* (pp. 203–348). Oxford University Press.

Kalecki, M. ([1967] 1991b). The problem of effective demand with Tugan-Baranovsky and Rosa Luxemburg. As reprinted in Osiatynski, J. (Ed.), *Collected works of Michał Kalecki*, Vol. II: *Capitalism: Economic dynamics* (pp. 451–458). Oxford University Press.

Kane, E. (2020). Immaculate deception: How (and why) bankers still enjoy a global rescue network. *INET working paper* no. 130. Institute for New Economic Thinking.

Kowalik, T. (2009). Luxemburg's and Kalecki's theories and visions of capitalist dynamics. In R. Bellofiore (Ed.), *Rosa Luxemburg and the critique of political economy* (pp. 102–116). Routledge.

Kregel, J. A. (2012). Minsky and the narrow banking proposal: No solution for financial reform. *Public policy brief* no. 125. Levy Economics Institute of Bard College.

Krueger, D., & Perri, F. (2006). Does income inequality Lead to consumption inequality? Evidence and theory. *Review of Economic Studies, 73*(1), 163–193.

Lazonick, W. (2016). The value-extracting CEO: How executive stock-based pay undermines investment in productive capabilities. *INET working paper* no. 58. Institute for New Economic Thinking.

Luxemburg, R. ([1913] 1960). *L'Accumulazione del capitale. Contributo alla spiegazione economica dell'imperialismo e ciò che gli epigoni hanno fatto della teoria marxista. Una anticritica.* Einaudi.

Luxemburg, R. ([1913] 2003). *The accumulation of capital.* Routledge.

Lyons, A. C., & Yilmazer, T. (2005). Health and financial strain: Evidence from the survey of consumer finances. *Southern Economic Journal, 71*(4), 873–890.

Mason, J. W. (2018). Income distribution, household debt, and aggregate demand: A critical assessment. *Levy Economics Institute working paper* no. 901. Levy Economics Institute of Bard College.

Mason, J. W., & Jayadev, A. (2014). "Fisher dynamics" in US household debt, 1929–2011. *American Economic Journal: Macroeconomics, 6*(3), 214–234.

Mazumder, B., & Miller, S. (2016). The effects of the Massachusetts health reform on household financial distress. *American Economic Journal: Economic Policy, 8*(3), 284–313.

Mian, A., & Sufi, A. (2015). Fraudulent income overstatement on mortgage applications during the credit expansion of 2002 to 2005. *NBER working paper* no. 20947. National Bureau of Economic Research.

Mian, A., Rao, K., & Sufi, A. (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics, 128*(4), 1687–1726.

Minsky, H. (1992). The financial instability hypothesis. *Levy Economics Institute working paper* no. 74. Levy Economics Institute of Bard College.

Mora, N. (2014). The weakened transmission of monetary policy to consumer loan rates. *Federal Reserve Bank of Kansas City Economic Review*, Q I, 93–117.

Pérez-Montiel, J. A., & Pariboni, R. (2021). Housing is NOT ONLY the business cycle: A Luxemburg-Kalecki external market empirical investigation for the United States. *Review of Political Economy*. https://doi.org/10.1080/09538259.2020.1859718.

Pressman, S., & Scott, R. (2009). Consumer debt and the measurement of poverty. *Review of Social Economy, 67*(2), 127–146.

Rajan, R. (2010). *Fault lines: How hidden fractures still threaten the world economy*. Princeton University Press.

Roncaglia, A. (2019). *The age of fragmentation: A history of contemporary economic thought*. Cambridge University Press.

Seccareccia, M. (2011). Financialization and the transformation of commercial banking: Understanding the recent Canadian experience before and during the international financial crisis. *Journal of Post Keynesian Economics, 35*(2), 277–300.

Stiglitz, J. (2010). *Freefall. America, free markets, and the sinking of the world economy*. Norton and.

Storm, S. (2017). The new normal: Demand, secular stagnation, and the vanishing middle class. *International Journal of Political Economy, 46*(4), 169–210.

Taylor, L., & Öner, Ö. (2020). *Macroeconomic inequality from Reagan to Trump: Market power, wage repression, asset price inflation, and industrial decline*. Cambridge University Press.

Tcherneva, P. R. (2008). The return of fiscal policy: Can the new developments in the new economic consensus be reconciled with the post-Keynesian view? *Levy Economics Institute working paper* no. 539. Levy Economics Institute of Bard College.

Zeldes, S. P. (1989). Consumption and liquidity constraints: An empirical investigation. *Journal of Political Economy, 97*(2), 305–346.

Profitability, Stimulus Policy, and Finance

Anwar Shaikh

1 Introduction

> The engine which drives Enterprise is [...] Profit (Keynes, [1930] 1976, *A Treatise on Money*, p. 148).
>
> Sales without profits are meaningless (Braham, 2001, *Business Week*).

The classical theory posits that the rate of growth of capital is driven by the *expected* net rate of profit (i.e., by the difference between the expected rate of profit and the interest rate). Keynes's theory of effective demand relies on the same proposition. In both schools, other things being equal, an increase in the expected profit rate will have a positive, and a rise in the rate of interest a negative, impact on the net rate of profit and hence on the rate of accumulation of capital.

But there are also central differences between the two approaches. In the classical tradition, the expected rate of profit is linked to the actual rate of profit in the manner of Soros's theory of reflexivity, whereas in Keynes's general theory, the expected rate of profit is left "hanging in the air." In addition, in classical theory the interest rate is linked to the profit rate, whereas in Keynesian theory it is tied to liquidity preference. Finally, we know that Keynes grounded his macroeconomic argument in the notion of "atomistic competition" and that even Kalecki's initial formulation of his theory of effective demand assumed "free competition." The classical analogue is the theory of "real competition," which I would argue that is the appropriate foundation for Keynesian macroeconomics. In real competition, firms set prices, are demand-conscious, and struggle for market share in the context of intraindustrial price competition and interindustrial investment flows motivated by profit rate differences.

A. Shaikh (✉)
Department of Economics, New School for Social Research, New York, USA
e-mail: shaikh@newschool.edu

Massive deficit-financed growth in Hitler's Germany and Roosevelt's World War II US economy succeeded in greatly expanding output and employment without raising prices and interest rates and lowering the profit rate. This was precisely because the normal market linkages between the positive and negative effects were blocked by price, wage, and interest rate controls. In the postwar period, all these markets returned to normal. Then much more modest stimuli in advanced countries eventually induced inflation alongside stagnation. The same problem appeared more recently in Lula's Brazil in which strongly positive stimulus and wage increasing effects eventually gave way to inflation and stagnation.

2 Effective Demand and Competition

Keynes grounded his argument in the notion of "atomistic competition" and even Kalecki's initial formulation of his theory of effective demand assumed "free competition" (Leijonhufvud, 1967, p. 403; Davidson, 2000, p. 11; Kriesler, 2002, pp. 624–625). Keynes begins his treatment of effective demand by noting that firms must produce on the basis of expected proceeds because production takes time. In his analysis of aggregate demand, he assumes that individual firms take the demand for their own product into consideration, which as Asimakopulos remarks is "a relation that does not hold for [. . .] [a perfectly] competitive firm." In articles published after the *General Theory*, Keynes explicitly states that (Asimakopulos, 1991, pp. 40–41, 43, 48):

> entrepreneurs have to endeavour to forecast demand [. . .] they endeavour to approximate to the true position by a method of trial and error. Contracting where they find that they are overshooting their market, expanding where the opposite occurs. It corresponds precisely to the higgling of the market by means of which buyers and sellers endeavour to discover the true equilibrium position of supply and demand.

This vision of competition is what I call real competition: Firms set prices, are demand conscious, and struggle for market share while subject to price competition and undertake investment across industries motivated by profit rate differences. Equilibration is turbulent, and balances are achieved in and through errors. There is no notion of equilibrium as a state of rest.

I would argue that the theory of real competition provides a natural foundation for the theory of effective demand. Expected profitability is turbulently regulated by actual profitability. The normal capacity rate of profit is determined in the classical-Sraffian manner by the wage share under given production conditions and over time also by changes in capital intensity arising from technical change. The interest rate is determined by profit rate equalization forces and/or monetary policy, the market interest is further affected by supply and demand, and a rise in the interest rate will lower net profitability, other things being equal (Shaikh, 2016, pp. 615–637). In this sense, my work seeks to return the theory of effective demand to a classical ground.

3 The Rate of Profit

In the classical tradition, the rate of profit is *the* central variable, based on the existence of surplus labor and a surplus product in any given technological and working conditions (Sraffa, 1960, pp. 3–11).[1] Here, the causation goes from a socially determined real wage and length and intensity of the working day, and a historically established technology, to the profit rate.

Capitalism and its characteristic mechanization process have enormously increased the productivity of labor. From the side of capital, labor is just another "input" into a profit-based production process and, like any other input, should be used as "efficiently" as is feasible (Foner, 1955, pp. 14–15). This means making the working day as long and hard as possible. In Great Britain in the eighteenth and nineteenth centuries, working days of 14, 16, and even 18 hours were not uncommon, and the intensity of labor rose along with the length of the working day (Kuczynski, 1972, pp. 46–48). In the United States even in the late nineteenth century, working days varied between 10 and 15 hours a day, in many cases for 7 days a week (Barger & Schurr, 1944, pp. 73–74; Foner, 1955, p. 22).

On the other side, workers have constantly fought against these conditions, through resistance and sabotage, through unionization, and through political struggles culminating in legislation. A roughly 8-hour working day at a regulated pace is the achieved norm in advanced countries, although immigrations and undocumented workers often work much harder and more intensively (Nomani et al., 1995). And in developing countries, 60–70-hour workweeks of intense labor under poor working conditions are quite common.

The capitalist development of technology is also social-historical. Adam Smith explains why an extended division of labor can greatly reduce capitalist costs of production, and Marx does the same for machinery (Marx, 1967a, Part II). Thus, at any given moment of time, the observed production coefficients depend on labor conditions, the degree of utilization of capital, and existing prices and costs. Lurking behind this is history, the role of the state, and the influence of the conjunctural factors such as wars and economic crises.

---

[1] Sir James Steuart argues, and Marx concurs, that there are really two sources of aggregate profit: profit on transfer of wealth, which Marx calls profit on alienation, and profit on production, based on the creation of a surplus product. Marx begins with the latter, because it is the principal mode of profit in industrial capitalism. He reserves the former for more concrete discussions involving "the distribution of surplus value among classes in the form of profit, interest and rent," the analysis of financial capital and the analysis of profit in pre-industrial (merchant) capitalism (Marx, 1963, p. 42; 1967a, p. 163; Shaikh, 2016, pp. 208–212).

4   The Rate of Interest

Keynes has a well-defined liquidity preference theory of the interest rate, while Marx's famously unfinished volumes II–III of *Capital* seemed to offer more than one approach (Shaikh, 2016, pp. 475–480). However, we can trace out a consistent classical line of argument that links the interest rate to the profit rate. Smith estimates that in Great Britain the long-run interest rate is roughly half of the profit rate (Smith, 1937, p. 200). Ricardo also says that the rate of interest is ultimately "regulated by the rate of profits" (Ricardo, 1951, p. 363). In this regard, at one point Marx (1971, Addenda, pp. 461, 505, 509) says that, "[a] general rate of interest corresponds naturally to the general rate of profit" and that for "Smith, Ricardo and all other economists worth mentioning, [. . .] the average rate of interest is determined by the average rate of profit." Itoh and Lapavitsas (1999, pp. 70, 95–96) note that according to Marx, banking capital participates in the equalization of profit rate. Panico (1988, pp. 184–186) develops this line by integrating banking into a Sraffian model of prices of production. I extend this through a formal treatment of the normal run interest rate as the price of production of finance, determined in the same manner as any other price of production through the equalization of profit rates.

The defining characteristic of a financial intermediary such as a bank is that its "output" is a money quantity, which in the case of banks is the total quantity of (new) loans (LN). Hence, the sales revenue of a bank is the total nominal interest receipts on loans $(i \cdot LN)$ and the price per unit loan is simply the interest rate. Let $ucr =$ the real unit average costs of banking, which includes processing of loans, deposits, withdrawals, etc., and let $\kappa r =$ real unit capital costs of banking, where capital includes fixed capital and bank reserves. Let $p =$ the price level, and assuming for simplicity that the nominal counterparts of banking operating costs and capital costs are $p \cdot ucr$ and $p \cdot \kappa r$, respectively, we can write the price of production of bank as the interest rate corresponding to the normal rate of profit (r) determined in the classical manner, subject to the constraint that $r < i$ (i.e., that the normal return on capital be greater than the cost of capital).

$$i = p \cdot ucr + r \cdot p \cdot \kappa r = p(ucr + r \cdot \kappa r) \text{ subject to the constraint that } r < i \quad (1)$$

Three things follow. First, the nominal interest i rises with the price level at any given real cost structure. In their *History of Prices*, Tooke & Newmarch, 1928) observed that the interest rate and the price level tend to move together (Panico, 1988, pp. 323, 439; Itoh & Lapavitsas, 1999, p. 29; Shaikh, 2016, p. 452). Marx remarks on this finding in his commentary on Tooke on interest rates (Marx 1967b, Chap. 23, p. 370). Gibson (1923) rediscovered the same pattern almost a century after Tooke. Keynes calls this "one of the most completely established empirical facts in the whole field of quantitative economics" (Keynes, 1976, vol. 2, p. 198).

Second, the *real* interest rate $\frac{i}{p} = ucr + r \cdot \kappa r$ depends on real costs and the general rate of profit. In this definition, the real interest rate is a relative price just like the relative price of corn or steel, and it falls or rises with its own relative real costs. Note

how different this production-based real interest rate $\frac{i}{p}$ is from the Fisherian consumption-based real interest rate $i - \pi$ where $\pi = \frac{\dot{p}}{p}$ is the rate of inflation[2] (Shaikh, 2016, pp. 466–467). The Fisherian argument implies that the nominal interest rate varies with the rate of change of prices, rather than with the price level as the classical argument. This is why Gibson's (re)discovery is labeled Gibson's Paradox: It is empirically true but contradicts the orthodox theory, standing in the Pantheon alongside Leontief's paradox, the equity premium paradox, and so on.

Third, allowing for shocks and policies, the market interest rate fluctuates around the normal rate. Monetary policies that have been in force in the United States since the early 1980s to systematically hold down the nominal interest rate have three contradictory consequences: They raise the net rate of profit $(r - i)$ *relative* to the path of the profit rate itself (which is itself changing) and thereby raise the growth rate above that dictated by the profit rate alone; they lower the profits of financial institutions and encourage them to turn to risker high return financial schemes; and they encourage businesses and households to borrow more, thereby increasing debt loads in the private nonfinancial sector.

The collapse of the subprime mortgage sector in 2007 triggered a global general crisis that spread rapidly across an already fragile global economy (Shaikh, 2016, p. 736).

5 Expectations and Macroeconomic Outcomes: Instability as the Means to Stability

The treatment of expectations in modern economics has always been a weak point in economic analysis. Three formulations tend to appear in macroeconomic models. Adaptive expectations used in econometric models of the 1960s and 1970s rely on past observations of some variable to project its future values. Rational expectations in use since the 1980s assume that expectations of agents are "model consistent" in the sense that agents in a model arrive at forecasts that are stochastically equivalent to those of the model itself. And VAR models of expectations use small vector autoregression (VAR) models with equations for a few key economic measures to create forecasts.[3]

Both Marx and Keynes begin by noting that the capitalist investment is undertaken on the highly uncertain basis of *expected* net profitability, the excess of the expected rate of profit over the interest rate that is its benchmark. In Marx, a business cycle boom may be spurred by a rise in the expected net rate of profit, but as the

---

[2] Fisher claimed that the expected real rate of interest, the difference between the money rate and the expected rate of inflation, would equal the rate of return in the real (i.e., nonfinancial) sector (McCulloch, 1982, pp. 47–49; Ciocca & Nardozzi, 1996, p. 34).

[3] Federal Reserve, https://www.federalreserve.gov/pubs/bulletin/1997/199704lead.pdf

boom proceeds, it raises prices, and the market interest rate tightens the labor market and raises the wage share, lowers the profit rate, and lowers the net rate profit rate even further due to the rise in the interest rate. At some point the downward path of the actual net profit rate asserts itself over the expected rate, and the upturn will give way to a downturn. Hence "we observe the cycles in which modern industry moves—state of inactivity, mounting revival, prosperity, over-production, crisis, stagnation, state of inactivity, etc." (Marx 1967b, Chap. 22, p. 360). Note that in Marx's argument, expectations change *actual* variables: a rise in the expected net rate of profit invokes more investment, higher growth, a rise in the interest rate, a rise in the wage share, and a reduction in the actual profit rate and in the actual net rate of profit.

Keynes's says something similar: A rise in the expected return on investment (marginal efficiency of capital) sparks a boom, which eventually leads to rising interest rates and rising costs. At some point "some catalyst, often minor in itself, causes market sentiment to shift, and precipitates a downward movement [...] [in which there is a] sharp decline in the marginal efficiency of capital." However, in Keynes' theory the relation between the actual and expected rate of profit is outside the short run, and his attempt to link the two remains unclear (Asimakopulos, 1991, pp. 70–84, 132).

Marx and Keynes share the notion that expected outcomes affect actual outcomes and that actual outcomes in turn affect expected ones. But their remarks are not formally developed nor integrated into their frameworks. This is where George Soros comes into the picture.

Drawing on his training in philosophy and on his deep practical knowledge of financial and commodity markets, Soros proposes a general theory of reflexivity applicable to many different phenomena. I will focus on its economic applications. He advances three general theses: expectations affect actual outcomes; actual outcomes can affect fundamentals; expectations are in turn influenced by the relation of actuals to fundamentals. The result is a process in which actual variables oscillate turbulently around their gravitational values.

Consider the stock market. An expectation of higher stock prices will induce investors to buy more stocks, so actual stock prices will rise. But as actual prices continue to rise beyond fundamental stock prices, the growing gap undermines expectations to the point that investors begin to sell off. The boom turns into a bust. Soros tells us that his goal has always been to exit the market before the bust, but he (wisely) does not tell us how he measures the fundamentals nor how he determines the appropriate point of departure. At a practical level, the price-earnings ratio is widely used by stock market analysts, and potential turning points are judged by their distance from a level deemed to be sustainable (fundamental). There are several definitions of the sustainable levels, and when the actual price-earnings ratio is consistently above these fundamentals, the chorus of warnings rises sharply. Robert Shiller's CAPE (cyclically adjusted price-to-earnings) ratio, created as the
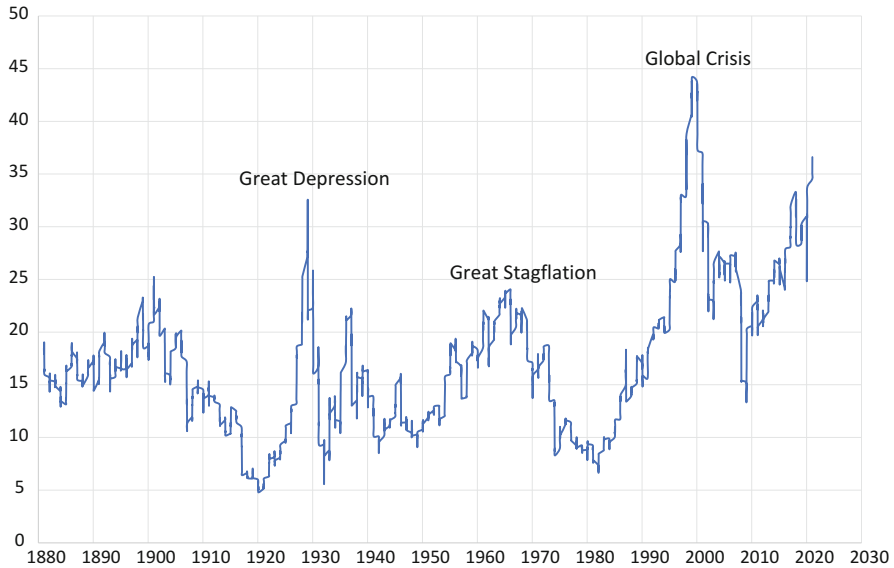
Fig. 1 Shiller CAPE (cyclically adjusted price-earnings) ratio. *Source*: http://www.econ.yale.edu/ ~shiller/data.htm

ratio of stock prices to an inflation-adjusted 10-year average of company earnings,[4] is widely used.

Figure 1 displays the CAPE from 1860–2021. Two things are immediately evident. First, stock market behavior is characterized by recurring sharp fluctuations. These are its normal instabilities. Second, the normal fluctuations sometimes ride much longer waves culminating in great peaks and subsequent crashes. Over the long run, the CAPE measure averages around 17. Its highest levels have preceded major economic crises: The Great Depression beginning in 1929, the Great Stagflation (inflation with stagnation) of 1968–1992, and the Global Crisis beginning in 2007. As of this writing, CAPE stands at 36.61, higher than any previous peak except that of the Global Crisis.

According to Soros, the interactions between expected, actual, and fundamental variables lead to the actual variables gravitating around their fundamental values. All three are moving in a process in which the actual price first overshoots, and then undershoots, the fundamentals. There is never an equilibrium moment, only alternating phases of instabilities (Soros, 2009, pp. 50–75, 105–106). The gravitational centers are path-dependent (Arthur, 1994; David, 2001)[5] because expectations can affect fundamentals; the system is non-ergodic because the future is not a stochastic

---

[4] http://www.econ.yale.edu/~shiller/data.htm

[5] Path dependence implies that a variable's gravitational center is itself dependent on a particular historical path taken by the variable.

reflection of the past (Davidson, 1991),[6] the existence of extended disequilibrium phases invalidate the efficient market hypothesis, and the reactions of fundamentals to actual outcomes invalidate the notion of rational expectations (Soros, 2009, pp. 58, 216–222). I have shown that the theory of reflexivity can be formalized in a simple and general manner with testable propositions (Shaikh, 2010).

In a classical context, the three equivalent variables are the expected, actual, and normal net rates of profit. An expectation of higher net profits sparks an increase in investment and employment. Increased wage and property income raise consumption demand, which in combination with the increased investment demand further raises aggregate demand. This may continue for an extended period. But the rise in real wages due to tighter labor markets erodes the normal actual rate of profit, and the rise in interest rates erodes the actual net rate of profit even more. As the boom sails on, the gap between expected and actual net profits grows, until at some point expectations reverse course and the boom turns into a bust. This is turbulent gravitation, not equilibrium repose.

Fiscal stimulus raises aggregate demand, while monetary stimulus that lowers the interest rate raises the net profit rate. Both can mitigate a bust. But they can also stretch out a boom, perhaps to highly dangerous levels.

6 Stimulus from Increased Purchasing Power

Ricardo famously insisted that an exogenous infusion of aggregate demand would have no impact on supply, since supply was already determined by a fully utilized stock of capital. Hence, an increase in the quantity of money would only increase the price level without having any real effects (Ricardo, 1952, 434–436).

Marx argues that capital is never fully employed in the Ricardian sense. Some part of aggregate capital is always in money form looking for new commodities to invest in; some in machines, materials, and labor-power looking to be engaged in production; some in production itself; and some in finished products looking for sale. Normal production also involves the economic utilization of the existing stock of capital, so that one-shift or even two-shift output with an 8-hour working day at normal intensity may be very far from engineering capacity involving three shifts and long intense working days (Shaikh, 2016, pp. 120–164). In this light, he argues that the flood of new gold flowing out of California mines in the 1840s enhanced global purchasing power and raised global output as it spread from the New World to the Old (Rist, 1966; Marx, 1973, p. 623). As he puts it, "[I]f it were not in the nature of capital to be never completely occupied [. . .] then no stimuli could drive it greater

---

[6] An ergodic stochastic process is one in which "averages calculated from past observations cannot be persistently different from the time average of future outcomes." Samuelson (1969) "made the acceptance of the 'ergodic hypothesis' the sine qua non of the scientific method in economics" (Davidson, 1991, pp. 132–133).

production." This same responsiveness of production to effective demand is central to Keynes' theory.

7   Inflation and Stagflation

The monetarist (quantity theory) of inflation argues that capitalism tends to fully employ all those who wish to work. Then the level of real output at any moment of time is determined by the production conditions (production function) and the supply of labor. It follows that the growth of this full-employment level of output is determined by population growth and the growth of output per unit labor (labor productivity growth) due to technical change. A growth of money supply which is greater than the growth of full-employment output generates inflation (Friedman, 1970). The Keynesian theory takes a different tack. A rise in effective demand raises the level of nominal output. If unemployed labor is available, real output and hence employment will absorb the increase in effective demand. But once there is near full employment, further stimulus will lead to price increases. Finally, post-Keynesian (Kaleckian) theory relies on the notion that firms maintain fixed profit margins (markups) over labor and input costs. In a closed economy, input costs are themselves the contemporaneous costs of their own inputs and labor costs and so on. Thus, prices can be resolved into direct and indirect labor costs and the corresponding (fixed) direct and indirect profit margins. Then inflation is the result of increases in domestic wages and in import prices (Shaikh, 2016, pp. 235, 589).

In the postwar period, the Keynesian theory had the upper hand in policy: If there was unemployment, stimulate the economy, recognizing that inflation might arise as unemployment entered some lower region (the Phillips curve tradeoff). This seemed to work until the late 1960s. But then unemployment began to rise, and stimulus policies began to generate inflation even though the economy was nowhere near full employment. Real output growth slowed down (stagnation), unemployment rose, and yet inflation accelerated: stagflation had arrived. Keynesian theory was in turmoil. Friedman stepped into this breach with the audacious proposition that the rise in measured unemployment was due to increasing numbers of people choosing not to work in the face of the unemployment and income support policies of the welfare state. The economy was always at effective full employment, he said, which is why repeated Keynesian stimulus policies gave rise to inflation. This notion of capitalism as a system of automatic full employment was carried over into all subsequent orthodox models (Shaikh, 2016, pp. 566–584).

From a classical perspective, full employment is a temporary phenomenon. If the labor market is tight, real wages rise, and profitability rises. Labor market tightness and the increased cost of labor *accelerate* ongoing labor-saving technical change, increase the incentive to induce non-working portions of the population to enter the labor force (increase the labor participation rate) and strengthen the incentive to import labor from other regions or countries. This increases the rate of growth of the labor supply. At the same time, the decline in profitability tends to slow down output

growth and decrease the rate of growth of the demand for labor. The pool of the unemployed fills back up (1967a, Chap. 25).

So if labor supply is not the long-run limit to growth, what is? The question takes us back to Ricardo, Marx, and von Neumann. If we treat aggregate output as a single commodity (corn), as is done in most macroeconomic models, then inputs of corn seed and labor give rise to an output of corn. Subtract the corn inputs and the corn equivalent of workers' wages, and you get a corn surplus product. The ratio of the corn surplus to the corn stock of capital is the profit rate. On the other hand, the rate of growth of the system depends on how much of the surplus corn is reinvested into further seed inputs and further wage payments. The theoretical maximum physical growth rate is when all the surplus is reinvested in growth: The maximum growth rate is the ratio of corn surplus to corn capital, which is the same as the profit rate. This Ricardian proposition is implicit in Marx's three-sector schemes of reproduction whose maximum balanced growth rate arises when all of the surplus value is reinvested, and it is explicit in the von Neumann multi-sector model of maximum growth (Pasinetti, 1974, fn. 1 p. 104; Kurz & Salvadori, 1995, pp. 383–384). This upper limit is not attainable in practice even if the economy is stimulated in that direction because bottlenecks in output growth will increasing crop up as the economy's growth rate approaches its maximum (Erlich, 1967, pp. 609–610; Pasinetti, 1977, pp. 208–216). This is similar to the Keynesian notion that labor supply bottlenecks will be increasingly common as the economy heads towards full employment. In the Keynesian case, the ratio of employed labor to the maximally available labor, i.e., labor utilization (employment) rate, is a measure of the degree of output slack. The classical equivalent is the ratio of the actual growth rate to the theoretical maximum, which I call the growth utilization rate. This approach is shown to have strong empirical support in a variety of countries and a variety of time periods (Handfas, 2012; Shaikh, 2016, p. 712).

8 Historical Stimulus Policies

Under normal conditions, continued stimulus policies initially raise output and employment growth. But as the product market tightens, prices and interest rates will tend to rise, and as the labor market tightens, real wages will rise. The rise in real wages will lower the rate of profit and the rise in interest rates will further lower *net* rate of profit. If the stimulus is pushed too far, the automatic feedback mechanism of relatively free markets will lead to inflation and stagnation. It follows that if one wants to push stimulus policies to their limits, there must be controls on wage, price, and interest rates—the free market must be curtailed.

In 1930s Germany, in a period of what Kalecki calls "military Keynesian" (Toporowski, 2016), interest rates were kept low, and large budget deficits were used to greatly expanded output and eliminate massive unemployment. Prices, wages, and even general business practices were directly controlled to such an extent that from 1933 to 1938, despite this unprecedented expansion in output and

employment, German real wages *fell* by roughly 25% even as productivity greatly increased. Hence, the wage share fell even more, and the profit rate rose fourfold from −7% in 1931 to 15% by 1939.

In the United States during World War II, Federal spending rose sixfold, the public debt relative to GDP rose from 50% to 120%, national output shot up, and 17 million new civilian jobs were created. Here too, interest rates were kept low, and regulations on prices and wages kept real wages from rising faster than productivity: real wages in manufacturing rose by 50 percent, while industrial productivity increased by 96 percent, so the wage share fell, and the after-tax US corporate profits doubled. As in the German case, the State was able to suppress the normal feedback loops between massive stimuli and increases in the wage share and in the interest rate, i.e., it was able to prevent decreases in the net rate of return on investment.

In the postwar era, governments all over the developed world expressed a strong commitment to maintaining a high level of employment and rising levels of incomes. The Keynesian theory seemed the ideal foundation, and the Phillips curve offered a clear-cut means of assessing the tradeoff between lower unemployment rates and higher inflation rates. But now the reaction to stimuli was left to the market. By the 1970s the Phillips curve had fallen apart, and both unemployment and inflation rose together—something quite unexpected within the theory. The Great Stagflation had begun. What had gone relatively unnoticed was that in the United States between 1948 and 1972, the wage share *rose* from 53 to 59%, while the corporate profit rate fell 18.3 to 11%. The Reagan-Thatcher reaction reversed these trends: from 1982 to 2007, real wages were held in check while productivity continued to grow, so the wage share fell, and the profit rate stabilized. At the same time, monetary policy was employed to reduce the interest rate in an unprecedented manner, from 10.7 to a mere 4.4%. A stabilized profit rate and a falling interest rate turned out to be good for economic growth: *unemployment fell from about 10 to 4.6%*. But inequality also increased dramatically, financial activities were deregulated, and financial capital poured over the whole globe. The resulting financial and speculative bubble finally burst in 2008 (Shaikh, 2016, pp. 724–745).

There are more recent examples of the profitability-based limits to stimulus policies. In Brazil, two successive Lula governments from 2003 to 2010 focused on the expansion of mass consumption by enhancing the incomes of poorer families, increasing the minimum wage, and financing public and private investment in social infrastructure through increased access to credit and subsidized interest rates. Poverty fell, unemployment fell, growth averaged a robust 4%, *the wage share rose*, and the profit rate fell. From 2011 onward, growth fell by half to 2.14% over 2011–2014 and went sharply negative to—3.8% in 2015 (Carvalho & Rugitsky, 2015).

9   The Lesson from Stimulus Policies

There is no question that modern economies can be stimulated by an injection of purchasing power, i.e., by new private, public, or by foreign credit directed towards commodity expenditures. And if the stimulus is deficit-financed, the resulting increase in national debt can be accommodated within broad limits. In some notable cases, such as the United States in World War II and Germany in the 1930s, massive stimuli led to large increases in output and the virtual elimination of unemployment. Yet in the heyday of the Keynesian policy during the 1970s, much more modest stimuli eventually led to rising unemployment *and* rising prices and interest rates: stagflation. In more recent times beginning with the early 2000s, the initial effectiveness of stimulus policies in Brazil has given way to economic decline and inflation.

The aim of this chapter has been to show that the differences in the two types of outcomes had principally to do with the feedback between stimulus and net profitability. At the heart of this lies the behavior of the wage share, i.e., of real wages *relative* to productivity and of interest rates. If the wage share rises, the profit share falls, which lowers the rate of profit, lowers the rate of growth, and raises unemployment. If stimulus policies are increased to maintain employment in the face of slowing growth, prices will begin to rise. Since the rate of interest is linked to the price level, it too will rise, which will further reduce the net rate of profit and growth. If monetary policy is used to reduce the real interest rate by making the nominal interest rate fall relative to the price level, this may mitigate the fall in the net profit rate at the expense of increasing reliance on debt finance and concomitant increases in debt burden. The boom may be sustained for some time, until the limits of debt-fueled growth assert themselves.

The lesson is that a sustainable stimulus policy must not only attend to demand and interest rates but also to the relation between real wages to productivity: the wage share is the key variable. It is not a matter of wage-led growth vs. profit-led decline as independent outcomes but rather of the need to thread the narrow path between the two.

References

Arthur, W. B. (1994). *Increasing returns and path dependence in the economy*. University of Michigan Press.

Asimakopulos, A. (1991). *Keynes' general theory and accumulation*. Cambridge University Press.

Barger, H., & Schurr, S. H. (1944). *The mining industries, 1899–1939: A study of output, employment and productivity*. National Bureau of Economic Research.

Braham, L. (2001). The business week 50. *BusinessWeek*, (23 March).

Carvalho, L. & Rugitsky, F. (2015). Growth and distribution in Brazil the 21st century: Revisiting the wage-led *versus* profit-led debate. *Department of Economics, FEA-USP working paper* no. 2015–25. University of San Paulo.

Ciocca, P., & Nardozzi, G. (1996). *The high price of money: An interpretation of world interest rates*. Clarendon Press.

David, P. A. (2001). Path dependence, its critics and the quest for "historical economics". In P. Garrouste & S. Ioannides (Eds.), *Evolution and path dependence in economic ideas: Past and present* (pp. 15–40). Edward Elgar.

Davidson, P. (1991). Is probability theory relevant for uncertainty? A post Keynesian perspective. *The Journal of Economic Perspectives, 5*(1), 129–143.

Davidson, P. (2000). There are major differences between Kalecki's theory of employment and Keynes's general theory of employment interest and money. *Journal of Post Keynesian Economics, 23*(1), 3–25.

Erlich, A. (1967). Notes on the Marxian model of capital accumulation. *American Economic Review, 57*(2), 599–615.

Foner, P. S. (1955). *History of the labor movement in the United States, volume II: From the founding of the American Federation of Labor to the emergence of American imperialism*. International Publishers.

Friedman, M. (1970). The counter-revolution in monetary theory. In *First Wincott memorial lecture delivered at the Senate House, University of London, 16 September*. Institute of Economic Affairs for the Wincott Foundation.

Gibson, A. H. (1923). The future course of high class investment values. *Banker's Magazine (London), 1*(15), 15–34.

Handfas, A. (2012). Demand pull and supply resistance in a classical inflation model. PhD diss, New School for Social Research.

Itoh, M., & Lapavitsas, C. (1999). *Political economy of finance*. Macmillan.

Keynes, J. M. ([1930] 1976). *A treatise on money*. Harcourt, Brace and Company.

Kriesler, P. (2002). Was Kalecki an "imperfectionist"? Davidson on Kalecki. *Journal of Post Keynesian Economics, 24*(4), 623–630.

Kuczynski, J. (1972). *A short history of labour conditions under industrial capitalism in Great Britain and the Empire*. Barnes and Noble Books.

Kurz, H. D., & Salvadori, N. (1995). *Theory of production: A long-period analysis*. Cambridge University Press.

Leijonhufvud, A. (1967). Keynes and the Keynesians: A suggested interpretation. *American Economic Review, 57*(2), 401–410.

Marx, K. (1963). *Theories of surplus value. Part I*. Progress Publishers.

Marx, K. (1967a). *Capital* (Vol. I). International Publishers.

Marx, K. (1967b). *Capital* (Vol. III). International Publishers.

Marx, K. (1971). *Theories of surplus value. Part III*. Progress Publishers.

Marx, K. (1973). *Grundrisse*. Vintage.

McCulloch, J. H. (1982). *Money & inflation: A monetarist approach*. Academic Press.

Nomani, A.Q., Rose, F., & Ortega, B. (1995). Legal beat: Labor department asks $5 million for alleged worker enslavement. *Wall Street Journal* (Eastern edition), 16 August.

Panico, C. (1988). *Interest and profit in the theories of value and distribution*. St. Martin's Press.

Pasinetti, L. L. (1974). *Growth and income distribution. Essays in economic theory*. Cambridge University Press.

Pasinetti, L. L. (1977). *Lectures on the theory of production*. Columbia University Press.

Ricardo, D. (1951). *The works and correspondence of David Ricardo, Vol. I: On the principles of political economy and taxation*. Ed. by P. Sraffa with the collaboration of M.H. Dobb. Cambridge University Press.

Ricardo, D. (1952). *The works and correspondence of David Ricardo, Vol. V: Speeches and evidence 1815–1823*. Ed. by P. Sraffa with the collaboration of M.H. Dobb. Cambridge University Press.

Rist, C. (1966). *History of monetary and credit theory from John Law to the present day*. Augustus M. Kelley.

Samuelson, P. A. (1969). Classical and neoclassical theory. In R. W. Clower (Ed.), *Monetary theory* (pp. 1–15). Penguin.

Shaikh, A. (2010). Reflexivity, path-dependence and disequilibrium dynamics. *Journal of Post Keynesian Economics, 33*(1), 3–16.

Shaikh, A. (2016). *Capitalism: Competition, conflict, crises*. Oxford University Press.

Smith, A. (1937). *The wealth of nations*. Modern Library.

Soros, G. (2009). *The crash of 2008 and what it means*. Public Affairs.

Sraffa, P. (1960). *Production of commodities by means of commodities*. Cambridge University Press.

Tooke, T., & Newmarch, W. (1928). *A history of prices and the state of circulation from 1793 to 1837*. Adelphi Company.

Toporowski, J. (2016). Multilateralism and military Keynesianism: Completing the analysis. *Journal of Post Keynesian Economics, 39*(4), 437–443.

Part V
Today's Macroeconomics Confronts Economic Crises

Some Reflections on Financial Instability in Macro Agent-Based Models: Genealogy and Objectives

Muriel Dal Pont Legrand

1 Introduction

The 2008 financial crisis has greatly increased criticism of the capacity of modern macroeconomic models to deal with large-scale crises and financial instability. Dynamic stochastic general equilibrium (DSGE) models are criticized for their core assumptions which assume all coordination issues as resolved and enable an exclusive focus on the behavior of optimal dynamics (cf. Colander et al., 2008) but hinder our understanding of how relatively small shocks can generate a large scale crisis. In the years following the 2008 crisis, macroeconomic agent-based models (MABMs)[1] came to be seen by many as directly challenging DSGE models in relation to this issue (cf. Delli Gatti (2017)). This challenge takes both direct and indirect forms, and the arguments culminated in 2018 with the publication of two special issues in the *Oxford Review of Economic Policy* and the *Journal of Economic Perspectives*.[2] The present paper sets out not to discuss the relative merits of DSGE models and MABM in general or to analyze the capacity of DSGE models to deal with financial issues[3]; rather it focuses on the contribution made by MABMs to our

---

[1] Agent-based models (ABMs) is the term used by several different disciplines. Agent-based computational economics (ACE) models are a specific application of an ABM. ACEs include a variety of models, essentially FABM (financial agent-based models) and MABM (macro agent-based models). All of them stem from the complex economics contribution(s) initiated for the most part by the Santa Fe Institute (cf. Fontana, 2010). In this paper, we focus on MABM.

[2] Respectively noted Symposia (2018a) and (2018b)

[3] See Claessens and Kose (2017) for a survey of financial frictions, Benes et al. (2014) for an examination of DSGE and financial crises, and Trautwein (2021) who shows how DSGE models investigate the role of financial frictions more intensively and the conclusions reached by these authors.

M. Dal Pont Legrand (✉)
Université Côte d'Azur, CNRS GREDEG, Nice, France
e-mail: Muriel.dal-pont-legrand@univ-cotedazur.fr

understanding of financial instability (and subsequent macroeconomic volatility). Thus, following a discussion of the different MABMs communities, we identify the various analytical (historical) roots[4] underlying the MABMs' research program and the lines of research explored to deal with large-scale crises and financial instability. We conclude with some remarks about the possible consequences of those developments on the future of macroeconomics.

The paper is organized as follows. Section 2 briefly describes the theoretical context in which MABMs emerged. Section 3 identifies previous research programs which to a greater or lesser extent have influenced the current MABMs' research agenda. Section 4 examines how financial instability is addressed and emphasizes the specificity of the different MABMs' modeling strategies. Section 5 concludes with some initial reflections on the status of those models and the degree to which they are considered a serious challenge to DSGE models in economic policy.

2   A Brief History of the Emergence of Agent-Based Macroeconomics

It is recognized that MABMs may in part be an expression of the resistance against the diffusion of DSGE models[5]; however, this is not to reduce the importance of their own research agenda. In this section, we briefly describe the theoretical underpinnings of the recent MABM microfoundational program.

Traditional Critiques in a New Favorable Context

General equilibrium (GE) approaches[6] support the idea that markets are inherently stable, a dynamic property which is linked strongly to their being built on "sound microeconomic foundations."[7] This view was propounded in the early 1970s by

---

[4]Dosi and Roventini (2017) examine the reasons for the rapid diffusion in Italy of the MABM approach. In doing so, they provide some interesting links between current Italian MABM groups and various Italian political economy traditions which might explain why Italy seems to be such fertile ground for the introduction of these new approaches. However, they do not provide a systematic analysis of these models' various analytical roots which is what we aim to do in the present paper.

[5]For a detailed discussion of the possible mutual influence between DSGE and MABMs, see Dal Pont Legrand et al. (2021).

[6]Here, we are thinking of the equilibrium macroeconomics developed along new classical school lines.

[7]The relationship between macro- and microeconomics has a long history; however, today's "microeconomic foundations" refer strictly to methodological individualism (cf. Duarte & Tadeu Lima, 2012).

Robert Lucas who later defined his goal as aimed at reincorporating "aggregative problems such as inflation and the business cycles within the general framework of 'microeconomic' theory" (Lucas Jr., 1987, p. 107). The main justification at the time for the "hyper-rational, self-interested agent typically used in standard macro models was that it was consistent with the characterization used in micro theorizing" (Colander et al., 2008, p. 236). According to this view, there are no macroeconomic phenomena which cannot be modeled (and then explained) by microeconomics and, consequently, aggregate macroeconomic variable dynamics are identical to those related to a single "standard"[8] microeconomic agent. Clearly, this removes the need for any distinction between micro and macro (Lucas Jr., 1987, p. 107–108). Although this methodological turn was triggered by Lucas's seminal contribution, it remains a pillar of even the most recent macroeconomic models, i.e., DSGE models.[9]

This (specific) conception of macroeconomic microfoundations overlooks the various ways in which the micro- and macroeconomic levels are intertwined[10] and has been attacked regularly (Kirman, 2011, p. 19):

> it is of no use looking at some "representative individual" in order to understand what will happen at the aggregate level. You would not imagine looking at the behaviour of a representative ant if you wanted to predict the evolution of the activity of the nest. In this view aggregate activity is not a blown-up version of individual behaviour. The passage from micro to macro is more complex than a simple adding up of independent individuals. [...] In economics, if we are interested in macroeconomic relations concerning the reactions to changes in various aggregate variables, we should not start at the level of the isolated rational individual.

Indeed, if we cease to consider individuals in isolation then if their interactions matter, *emergence* happens,[11] and logically the "standard" aggregation method becomes inappropriate. This is an old question whose proper modeling (and simulation) was made possible only when economics started to borrow tools from complex systems analysis.[12] This research program was launched largely by the efforts of the Santa Fe Institute in the 1980s (cf. Arthur, 2014). Economists who

---

[8]"By 'standard' microeconomics, we mean microeconomics founded on rational individuals i.e., based on a selfish *Homo Oeconomicus* who makes axiomatic-defined rational calculations aimed at maximizing a context-independent utility function" (Delli Gatti et al., 2011, p. 2).

[9]It is commonly assumed that DSGE models represent the current consensus which mixes the core new classical school approach with New Keynesian elements. For more details of this mix, see de Vroey (2016) or Duarte (2015).

[10]Cf., for instance, Hoover (2012) for a detailed exposition of the three main *microfoundational programs* or Backhouse and Boianovsky (2012) for a history of disequilibrium microeconomic foundations.

[11]Emergence occurs when an entity is observed to have properties which its individual parts on their own do not have. Also, these properties emerge only through the interactions among these parts. For a more detailed discussion, see the introduction in Delli Gatti et al. (2008a).

[12]We understand complex systems analysis as related mainly to two domains: nonlinear dynamic systems and complex adaptative systems, both used to produce endogenous dynamics, networks models, etc.

adopted this approach considered that both the empirical evidence and the experimental results show that aggregation can generate regularities: "quite simple and not hyper-rational individual rules when aggregated become well shaped: regularities emerge from individual 'chaos'" (Delli Gatti et al. 2008a, p. 1). This led to the idea that rather than focusing on "what individuals think and do," an alternative position would be to develop "micro foundations (that) might well mean how the macro-structure of the beehive influences the distribution of the behaviors of the bees. A sort of macrofoundation of the micro" (Dosi & Roventini, 2019, p. 3). This constitutes a foundational argument against methodological individualism: "Any meaningful model of the macro economy must analyze not only the characteristics of the individuals *but also the structure of their interactions*" (Colander et al., 2008, p. 237, emphasis added).

Gradually, larger numbers of economists began to explore this alternative modeling strategy on the basis that it might be shown to be a better guide for policy makers,[13] a growing tendency after the failure of DSGE models to explain the 2008 crisis and the subsequent recession. There were calls (Colander et al., 2008) for an "empirically based macroeconomic model" i.e., for a model which although a reduced form of the observed economy would nevertheless be able to accommodate some of its essential elements and characteristics:

> What must be refused is not reductionism per se, that is the idea that to understand a complex system we need an adequate description of the individual characteristics and of the network of interactions and its constituents, but methodological reductionism in its strongest form, according to which "the whole is simply the sum of the parts" [reference here to Dawkins, 1976]. On the contrary, in a complex system, the whole constitutes something which is *more and different* than the mere linear combination of its constitutive parts (Delli Gatti et al., 2011, p. 6, original emphasis).

In the context of the 2008 crisis, one of the most salient failures of DSGE models was their inability to address the issue of instability. More generally, what was criticized was their contribution to the idea that "free markets" are fundamentally stable in *every circumstance* such that a single unique model can be applied to all possible situations that was attacked[14]: in other words, "Much that is true when the economy is stable ceases to be true when it is not" (Leijonhufvud, 2014, p. 763). Nevertheless, the growing influence of DSGE models buoyed by the context of the *Great Moderation* led economists, bankers, regulators, and policy makers to

---

[13]The speech delivered by Jean-Claude Trichet to the European Central Bank Conference (Frankfurt, 18 November 2010) is an example of the sudden interest of policy makers in MABMs: "The atomistic, optimizing agents underlying existing models do not capture behaviour during a crisis period. We need to deal better with heterogeneity across agents and the interaction among those heterogeneous agents. [. . .] Agent-based modelling dispenses with the optimization assumption and allows for more complex interactions between agents. Such approaches are worthy of our attention."

[14]The new classical school developments have resulted in most economists losing interest in studying stability properties. However, it has been demonstrated that under reasonable informational assumptions, there is no adjustment process able to guarantee the existence of a convergence to equilibrium (Kirman, 2016).

disregard the possibility of serious systemic instability. The pre-crisis DSGE models say very little about agent heterogeneity,[15] and nothing about scaling effects or coordination issues, all elements that many economists consider likely played a major role in the 2008 economic crisis and the subsequent recession. When the crisis forced economists to reconsider the soundness of their modeling strategy, MABMs found new support, and in recent years, their results and conclusions have become increasingly widespread.[16]

A Research Program

Based on agents' heterogeneity and decentralized coordination mechanism(s), MABMs can deal explicitly with coordination issues and disequilibria dynamics. These elements naturally predisposed the MABM literature to focus on (financial) instability issue(s) to try to *explain* (cf. Dal Pont Legrand & Hagemann, 2019) (and not just to *mimic*) large-scale crises, i.e., the essence being to understand how relatively small shocks are at the origin of deep downturns, in order to understand in turn the salient disproportion between cause(s) and effect(s), between shocks and propagation.[17]

The aggregate properties of such an economy are obtained by summing the microeconomic dynamics of heterogeneous agents, i.e., macroeconomics from the bottom-up.[18] The heterogeneity applies to several agent dimensions (individual behaviors). It is generally assumed that agents are linked through their social interactions, through markets or via networks. Their interactions tend to follow simple behavioral rules which reflect bounded rather than perfect individual rationality[19] and are based on local information. Agents use decision-making heuristics which enable learning and adaptive behaviors. A system is considered *complex* if the highest level is not the result of an aggregation process but rather emerges from the set of interactions among multiple agents. As a result of these interactions, (M)ABMs can be characterized by externalities, nonlinearities, and dynamic processes with positive feedbacks. Logically, MABMs have come to represent an opportunity for heterodox theories to benefit from explicit microeconomic foundations. Nevertheless, it should be remembered that ABM are only "tools": They can

---

[15] Despite several previous attempts, the incorporation of heterogeneity in DSGE models became a major objective mainly after the crisis.

[16] However, the increased number of papers does not indicate cross-fertilization with DSGE models. This is a different issue which is examined in Dal Pont Legrand et al. (2021).

[17] Cf. Stiglitz (2015) for a detailed discussion.

[18] This is the title of Delli Gatti et al.'s, 2011 book.

[19] Some (M)ABM models are based on a rational expectations hypothesis.

co-construct theories but "by definition" are not core elements of a single specific approach.[20]

3 Understanding (Macro-)Financial Instability: Old Roots for New Microeconomic Foundations

As emphasized above, MABMs are founded on renewed microeconomic foundations. As this research program became more substantial, it attracted many economists including a few who were already working on instability issues within different theoretical traditions so that the development of (M)ABMs provided an opportunity to revive older ideas and concepts. For some MAB modelers, the search for the microeconomic foundations of macroeconomic financial instability was the main objective[21]; for others it was one of the useful ingredients which allowed a better understanding of economic dynamics.[22] Eight MABM families can be identified (Dawid & Delli Gatti, 2018); however, in this paper, we focus exclusively on two that directly or indirectly address the instability issue in relation to the functioning of the financial sphere, namely, K&S (Keynes & Schumpeter) and CATS (computational adaptive system).[23] Before discussing the specificity of their respective contributions to financial instability, we would emphasize their common roots in Axel Leijonhufvud's great influence and commitment to heterogeneous interacting agents.

Leijonhufvud's contribution to the search for disequilibrium microfoundations is well-known and strongly associated to the work of Clower (cf. Backhouse & Boianovsky, 2012 or De Vroey, 2016). However, his contribution to the introduction of complexity in economics has been so far rather neglected. He was a member of the UCLA Center for Computable Economics and later was a prominent figure at the Trento University[24] where a group of economists began exploring the first ABMs. While Leijonhufvud never developed a specific computational model, he had a

---

[20]For instance, see Heise (2017) for a characterization of mainstream complexity economics.

[21]Cf. the work initiated by Domenico Delli Gatti and Mauro Gallegati, a research program clearly inspired by Minsky.

[22]Cf. the Keynes and Schumpeter (K&S) models.

[23]It should be noted that among the various macroeconomic research programs based on ABM, the so-called AGH (Asharf, Gershman and Howitt) research program is not without connection with the literature we examine. Indeed, this program was strongly influenced by Clower and Leijonhufvud so that, similar to as other MABMs, the approach is linked to several of Leijonhufvud's contributions (1977, 1986, 1993, 1997, 2014). Moreover, Howitt is a member of the scientific board of the Trento Summer School created by Leijonhufvud. However, those models (Ashraf et al. 2016, 2017; and Howitt, 2012) do not emphasize financial instability but concentrate more on money and coordination.

[24]Axel Leijonhufvud was appointed Professor of Monetary Theory and Policy at the University of Trento in 1995 and later founded the Trento Summer School in Adaptative Economic Dynamics.

definite influence and tried to unite MABM-sympathizers to work on the same agenda.[25] His approach to computational economics was that he saw it as a possibility to explain how individuals, for whom the equilibrium price vector initially was unknown, could coordinate in order to identify it: "[t]he economy should be looked at as a machine that has to compute the equilibrium" (Leijonhufvud, 1993, p. 7).[26] So clearly, Leijonhufvud[27] paved the way to new microeconomic foundations[28] which could explain how decentralized (or poorly coordinated) economies work.

Heterogeneous interacting agents are clearly at the origin of endogenous business cycles dynamics. In recognition of their Schumpeterian inspiration, CATS and K&S models build on the concept of aggregation which explains business cycles as resulting from "the complex interactions of firms and industries (a procedure reminiscent of Schumpeter, 1939) in which small shocks and endogenous elements coexist" (Delli Gatti et al. 2008a, p. 2). Business cycle economists generally felt challenged by the recurrence of upturns and downturns in aggregate output, and it is possible to distinguish two approaches to these problems. The first is equilibrium-based and considers that cycles analysis should be interpreted in the context of the decomposition between impulse and propagation, a research program that goes back to Slutsky, 1937) and Frisch (1933) and extends to present-day DSGE models. The second approach is based on disequilibrium and/or nonlinearities (along the lines of Kaldor or Goodwin) and considers regular oscillations as endogenous phenomena. While clearly anchored in the impulse/propagation tradition, DSGE models are nevertheless at the origin of an increased overemphasis on shocks.[29] As emphasized by Stiglitz (2015), this neglect of a propagation mechanism hindered their ability to explain how small shocks can produce large fluctuations, a point on which they are clearly challenged by MABM.

Indeed, proponents of MABMs consider that if all these approaches lead to deadlock, this is because none of them is appropriate to analyze the "interaction between statements at the microeconomic level in terms of behavioral rules and aggregate categories, like income, expenditures or savings" (Delli Gatti et al. 2008a, p. 3). Although these researchers do not doubt that individual behaviors are at the origin of fluctuations, the behavior of the whole system is different from the behavior of any one of its constitutive elements: heterogeneity and interactions were considered the two key elements of the CATS and K&S research programs. Since heterogeneity has been proven empirically to be a remarkable source of financial fragility, CATS extended Minsky's research program, building on heterogeneity and agents'

---

[25]This point was described by Domenico Delli Gatti and Hans Michael Trautwein (who both occasionally participated to the so-called Trento Group) in separate interviews held in 2020.

[26]On this point, he refers to Goodwin (1951, pp. 1–2) who argues that it now seems permissible "to regard the motion of an economy as a process of computing answers to the problems posed to it."

[27]See Leijonhufvud (1977, 1986, 1993, 2014) and Colander et al. (2008).

[28]See Dosi's tribute to Leijonhufvud in *Trento Summer School. 19 Years with a Passionate Teacher* (p. 45) at https://event.unitn.it/aed-summerschool/libro_trento_summer_school.pdf.

[29]See Dal Pont Legrand and Hagemann (2019) for a discussion of the relative weights of shocks and propagation over the evolution of business cycles theories.

interactions. The K&S program uses this heterogeneity in various ways: for instance, firm as well as expectations heterogeneity are at the origin of financial fragility and, this way, introduce a Minskyan dimension.

Foundations of K&S (Keynes and Schumpeter)

The K&S research program originated in Santa Anna School of Economics in Pisa, initiated by Giovanni Dosi and driven by various combinations of co-authors (Dosi et al., 2008, 2010, 2013, 2015, 2020; Dosi & Roventini, 2017, 2019). Its Schumpeterian influences are clear certainly partly due to Dosi's personal research trajectory. The models are based on two key elements identified as Keynesian demand/supply side interactions and Schumpeterian theories of technology-fueled economic growth. A financial amplification via credit cycles was introduced into this combination of elements.

Interconnection Between the Demand and Supply Sides The main objective was to provide support for the role of long run (not just short run) demand dynamics which, in turn, prompted the reemergence of the need for countercyclical policies. More precisely, K&S develops a sort of Schumpeterian endogenous growth model enabling examination of both long-run growth and short-run fluctuations:

> Business cycles are endogenous and have a genuine Keynesian origin. The production and investment choices of firms can lead to coordination failures in the goods markets which in turn affect aggregate output and unemployment dynamics (Dosi & Roventini, 2017, p. 9).

Depending on the favorable coincidence between on the one side innovative exploration of new technologies and on the other side (sufficient) demand, Dosi and Roventini identified two possible distinct growth regimes characterized by different short-run fluctuations and unemployment levels.

The Minsky Connection Although avoiding an overwhelming focus on the Minsky connection (Minsky, 1986), K&S clearly introduces Minskyian elements (see Dosi et al., 2015, 2020), i.e., a credit cycle hypothesis. Banks' activities are procyclical: They support firms' development and allow leveraged activities. Conversely, during recessions, they drastically reduce the availability of credit at a time when firms most need it. The survival probability of those firms then decreases and default-loan losses outweigh banks' net worth. The spread of this scenario results in a banking crisis and governments generally being forced to offer bailouts. However, as the 2008 crisis shows, this does not avoid a "credit crunch," and the large public deficits can lead to a sovereign debt crisis.

Schumpeter Meeting Keynes? A Growth Cycle Dimension Although the modeling strategy is different, this research program has much in common with Stiglitz (1993). Stiglitz built a sequential rather than an ABM such that as demand decreases (recession), revenue and bank credit also decrease. This forces firms to reduce the

amounts of resources they allocate to R&D which in turn reduces the expected growth rate. Joint analysis of growth cycles means that this approach can be used to examine the impact not only of monetary or fiscal policies but also of technologies and industrial policies and, in each case, enables investigation of the short- and long-run consequences. All economic policies (fiscal, industrial, monetary, etc.) that dampen fluctuations have long-run impacts: Dampening fluctuations smooths investment and production over the business cycles and fosters R&D and eventually growth. Thus, to some extent Keynesian countercyclical policies are compatible with Schumpeterian policies.[30] This approach underlines the inefficiency of austerity policies which deter growth and are self-defeating since they do not work to stabilize public finances.[31]

Roots for CATS (Computational Adaptive System)[32]

Complexity Economics In the pre-CATS era, the tendency was to refer to "complex dynamics" with implicit references to chaos dynamics. The pioneers of this (vast) program on (endogenous) complex economic dynamics include Goodwin and Day. They proposed mainly macroeconomic models with no microeconomic foundations, which were highly parameter sensitive and difficult to validate empirically. Those models did not easily allow theoreticians to draw economic policy conclusions, and despite their merits, they were mostly ignored by policy makers. However, in the 1980s, promoted in part by the Santa Fe Institute, and especially in the 1990s, agent-based computational economic (ACE) models were developed and shifted the emphasis from the macro to the micro level of analysis.[33] At this time, Delli Gatti and Gallegati were developing models based on network analysis and nonlinearities[34] and were in contact with the Santa Fe Institute.[35] There are various

---

[30]Concerning economic policy, business cycles, and growth in Schumpeterian analysis, see Dal Pont Legrand and Hagemann (2017a). In this paper, Keynesian and Schumpeterian countercyclical or stabilizing economic policy are clearly distinguished. In relation to Schumpeter's notion of the "recuperative powers of capitalism" and its differences with the modern notion of "productive recessions," see Dal Pont Legrand and Hagemann (2017b).

[31]See Dawid and Delli Gatti (2018) and Dosi and Roventini (2017) on this point, and for a more detailed accounting of the economic policy implications of the K&S approach in general, see Fagiolo and Roventini (2012).

[32]We refer here to Battiston et al. (2009) and to a series of papers (with different combinations of co-authors) Delli Gatti et al. (2005, 2006, 2007, 2008a, b, 2009, 2010a, 2011) and Assenza et al. (2015) and Assenza and Delli Gatti (2019).

[33]Delli Gatti et al. (2008a, p. 8) use the same argument.

[34]Despite some evident merits, those tools were suffering from limitations: Network modeling tools (at least in economics) were considered as rather "mechanical," while nonlinear dynamics were difficult to validate empirically.

[35]Cf. Interview with Delli Gatti in January 2018 when he confirmed that both he and Gallegati visited the Santa Fe Institute in 2004 where they also met Farmer. At this time, almost certainly due

reasons for this rather sudden diffusion of these new models. It is undeniable that ABMs benefited from the development of new computational tools which were accessible to nonspecialists unfamiliar with programming languages.[36] However, although necessary these technical conditions are not sufficient to explain the sudden emergence and development of ACE models. Our intuition is that these developments were driven by the parallel evolution of the economics of information which increased interest in research into agents' interactions: see for instance Gallegati et al. (2003) or Delli Gatti et al. (2008a), papers which among others refer explicitly to the connection with the economics of information.

Hyman Minsky and Financial Fragility[37] Hyman Minsky had a direct influence on the CATS program proposed by Domenico Delli Gatti and Mauro Gallegati.[38] Specifically, these latter were seeking an appropriate way to reintroduce Minsky's financial fragility concept within a model. Intuitively, the mechanism leading to financial fragility is that during a phase of economic prosperity, leverage builds progressively, exposure to (credit) risk increases, and the system becomes extremely fragile meaning that a relatively small shock can have huge consequences. These ideas were proposed and developed by Hyman Minsky in the 1970s and 1980s but could be even more relevant in today's context of high(er) connectivity within the banking and financial network, which causes the significance of exposure to risk to increase. In the current economy, there is almost no chance that a shock will be confined to a limited part of the network. It should be noted that, in order to capture financial fragility, the CATS program investigated different mechanisms with the result that it uses a set of different models (although they have some common fundamental elements) while so far, the K&S research program is clearly identified as an approach based on one core model which is applied (and adapted) to different issues.

Keynesian Roots Demand is an important element of the CATS model. This introduces a Keynesian flavor, but this is not the most salient feature of the model which is focused on source(s) of financial instability and their macroeconomic impact. However, it is clear that these models and some post-Keynesian models apply common agent behavioral rules under uncertainty so that to some extent, the (new-)Keynesian connection, apart from the already mentioned Leijonhufvud's influence, is related to the treatment of information.

---

to his investment in econophysics, Mauro Gallegati was convinced by the applicability of complexity and ABM tools to his and Delli Gatti's joint research program.

[36] Sophisticate programming languages are still used but software such as NetLogo requires no specific ex ante knowledge which boosted the popularity of this modeling tool.

[37] Cf. interview with Domenico Delli Gatti conducted by Muriel Dal Pont (January 2018, Milano).

[38] Mauro Gallegati was awarded his PhD degree in 1989 from Marche Polytechnic University; his supervisor was Hyman Minsky.

Indeed, the economics of information owes much to the contributions of Greenwald and Stiglitz (GS)[39] with Joseph Stiglitz a regular co-author on MABM papers.[40] We can identify two distinct channels of influence. First, following GS (1987, 1990, 1993), Delli Gatti and Gallegati (DGG)[41] envisaged a GS-style financial accelerator model to explore the consequences of the degree (and distribution) of agent heterogeneity for their GS-type financial fragility. Since agents possess "local" information, imperfect information is a crucial element in their model. Also, since prices cannot reflect all the relevant information, agents are led to interact outside the price system. As in GS (1990), DGG modeled a leveraged aggregate supply, and again as in GS (1993), this meant that the probability of bankruptcy was incorporated in the firm's profit function. Thus, the model takes account of risk-uncertainty, and also the producer's realization that its future depends not only on its actions but also on the actions of others. Second, building on Stiglitz and Greenwald (SG) (2003), the emerging CATS team published three NBER working papers co-authored with GS: Battiston et al. (2009) and Delli Gatti et al. (2007, 2008b). Later, we can see the influence of SG (2003) mainly through the way they analyze the link between money and credit and the subsequent capacity of credit to stabilize the economy. Their 2003 book explores the functioning of a credit-based economy. They emphasize how credit differs from other commodities in being based explicitly on information and default risk. They analyze the consequences in terms of economic policy objectives and efficiency. This led SG (2003) to model other types of agent financial linkages in addition to the traditional lender/borrower links. In these models, relatively small disturbances can weaken large populations, add to the increasing financial fragility of the entire system, and amplify shocks to the system. SG's (2003) monetary and fiscal policy conclusions are both innovative and revive forgotten views. However, at this stage their work did not constitute a "complete" network analysis: the linkages among agents were given so the model was mainly static and lacked economic motives. Because the microeconomic foundations of the SG framework were based on heterogeneous interacting agents, it became a candidate for ABM applications. The first paper coauthored by individuals from these two communities was published in 2006 (Delli Gatti et al., 2006).

This section has examined how ABM benefited from long-established research programs. This supports the idea that ABM are primarily and mainly tools that can be exploited by various research agendas. This is not to minimize their theoretical contribution but rather to show that it may depend on the context in which they are mobilized. Along those lines, we next examine how the modeling strategies adopted

---

[39] We would not want to suggest that Akerlof made no contribution to the theory of information, but the CATS program does not refer explicitly to his work.

[40] Stiglitz is a regular co-author of CATS papers. He met Mauro Gallegati in 1989 (cf. interview of DDG January 2018) and contributed to business cycle models and network financial analysis. Recently he has published on K&S (see Dosi et al., 2020).

[41] DGG refers to the work produced by these two authors independent of the larger CATS team.

by the two main MABMs communities determined the nature of their contribution to macroeconomics.

4   Modeling Strategies and Research Agenda

While the two programs have a great deal of common analytical grounding, they follow distinct modeling strategies which in fine introduce distinctive elements in their respective research agendas. It is our objective in this section to survey their actual contributions and how they defined their research agendas. We first examine the two strategies adopted by the CATS' team and then the strategy adopted by the K&S research program.

In exploring the CATS modeling strategy, it is clear that the objective was to explore financial mechanisms and instability to enable a better explanation of the observed fluctuations.[42] This led to the development of different models exploring various microeconomic interactions involving heterogeneous agents. While the team became increasingly concerned with empirical validation, their work like early work on complexity economics is marked by a clear normative dimension. They mainly produced financial business cycles models and financial network dynamic analyses.

Financial Business Cycles

Financial business cycle research was aimed at improving the capacity of business cycle models to reproduce stylized facts and observed regularities, and this was the focus of the CATS research program. Its contribution centered on introducing heterogeneity in a financial accelerator model. In 2003, Gallegati et al. showed that in models where firms are characterized by two sources of heterogeneity based on their balance sheets and their size, idiosyncratic shocks can generate large fluctuations. The 2005 business cycle model which was based on a scaling approach was a crucial advance. It modeled heterogeneous agents' behavior explicitly in a decentralized economy. It introduced a leveraged aggregate supply in line with GS (Greenwald & Stiglitz, 1990, 1993) in a MABM model to examine the interactions among different financially fragile firms and the banking sector. The distribution of heterogeneity[43] proved an essential factor explaining the amplified fluctuations. The team provided further developments of and extensions to this initial

---

[42] Delli Gatti and Gallegati were in close contacts with Hyman Minsky long before they started with ABM. They always had the objective to understand financial fragility microeconomic foundations.

[43] In this model, the authors introduce heterogeneity in firm' size and firm growth rates. The initial distribution of this heterogeneity matters, and the moments of distribution by financial position vary with the business cycle.

model, gradually building a general approach which revealed the pillars of this new macroeconomics paradigm, a project described very clearly in Delli Gatti et al. (2011).

Assenza et al. (2015) paper was an important turning point. Their model analyzed the connection between capital and credit, and its influence at the macroeconomic level. It introduced a distinction between capital and consumption goods and the idea of "a stylized supply chain where upstream firms [...] supply a durable and sticky input [capital] to the downstream firms, who produce consumption goods" (Assenza et al., 2015, p. 5). Both types of firms are reliant on bank loans to fill their finance gap. This model combined with the tensions and frictions related to labor (and goods) markets explains how distortions in respective financial conditions can lead to crises. Finally, Assenza and Delli Gatti (2019) proposed a hybrid macroeconomics and agent-based model (M&ABM) which examines a (financial) market based on an ABM. This ABM generates artificial data represented by an aggregate variable which then is introduced into the macroeconomic (i.e., DSGE-type) model. Ultimately, these macroeconomic model variables have impacts at the ABM micro (individual) level which result in two feedback effects. In this model, firms' financial conditions (net worth) are heterogeneous which has direct consequences for the different external financial premiums that exist because financial frictions are considered. In this context, agents determine their optimal investment levels: When the interest rate increases, the firm's profits and consequently net worth decrease. So, interest rate changes affect the distribution of firm heterogeneity. The main findings of this model are (i) that the diffusion of shocks depends on the degree of firm heterogeneity, and (ii) that the distribution of firm heterogeneity is affected by the shocks. Not only does their paper show that MABM allows the emergence of the observed aggregate fluctuations; it provides a direct comparison to traditional macroeconomic models.

Credit Network Matters

Superimposing a network structure on a (M)ABM is another modeling strategy. Credit relationships by their nature are complex (Delli Gatti, Gaffeo, & Gallegati, 2010b). First, they involve both heterogeneous agents and different categories of agents (banks, financial markets, pension funds, etc.). This results in different types of credit: (i) inside credit which is credit between the same class of agents (e.g., inter-firm credit) and (ii) outside credit which is credit between different classes of agents (e.g., households and banks). Second, the credit network is not given but is continuously evolving: Some relationships are interrupted, and some new relationships are forged. Third, credit networks are fragile since if one agent's net worth is sufficiently affected by a shock, its resulting financial condition will affect the financial conditions of its links (in the same or a different class). In the case of bankruptcy, there may be a domino effect such that a relatively small shock could generate huge fluctuations. An interesting finding in this literature is that although a larger variety

of credit sources should diversify lender risk, the existence of large lenders involving many borrowers propagates financial distress. This idea emerged in SG (2003) and was later examined through an ABM lens by Delli Gatti et al. (2006) and Battiston et al. (2009)[44]; the size effect ("too big to fail") proposed in the face of a crisis could be dominated by a *connection effect*; i.e., the more connected the institutions/agents, the more important it is to bail them out ("too connected to fail").[45] In 2012, Battiston et al. employed this framework to explore the effect of risk diversification on systemic risk. They analyze default cascades in financial networks and identify two potential external effects which contribute to the ongoing discussion on the stabilizing effect (i.e., the impact or lack of it on the systemic risk) of risk diversification strategies.

The K&S strategy looks different. Although the model calibration came later, from the very first model they produced, the aim was to compete directly with DSGE models and to provide a general model, i.e., to explain growth, cycles, and unemployment dynamics.

Long-Run Dynamics and Financial Factors

The K&S program pays attention to both long-run dynamics and fluctuations but considers them as the joint product of capital accumulation. Dosi and Roventini (2017) proposed a Schumpeterian growth model based on complex interacting heterogeneous agents in which demand (Keynes) and credit cycle (Minsky) elements affect short-run decisions which in turn determine long-run growth. The core of these dynamics lies in the coincidence (or not) between innovation (emergence and diffusion) and (sufficient) demand. Banks can amplify but are not the source of fluctuations. Finally, the model can reproduce persistent fluctuations and match the business-cycle properties concerning productivity, price, inflation, and markups. The modularity of MABMs, in general, and the strong emphasis K&S puts on the articulation between demand and supply forces prompted many similar papers investigating economic and, especially, monetary and fiscal policies (Fagiolo & Roventini, 2012).

---

[44] Stiglitz was a co-author of most of the papers produced by the CATS' team dealing with credit network, so Delli Gatti et al. (2006, 2007, 2008c, 2009, 2010a) and Battiston et al. (2009 and 2012).

[45] This literature has similarities to work on stress tests (Aymanns et al., 2018).

5   Some Preliminary Conclusions

This literature provided the microeconomic foundations for the emergence of macroeconomic financial instability. Rather than introducing market frictions, it uses local interactions among heterogeneous agents. Indeed, the essence of (M)ABM is the decentralized nature of the economic system,[46] a framework which renews macroeconomists' interest in coordination issues. Because these models benefit from computational techniques, they are able to overcome the traditional aggregation constraint and represent an opportunity to introduce agent heterogeneity, social interactions, and bounded rationality.[47]

Introducing interactions adds some degree of the social dimension compared to consideration of individuals in isolation in DSGE models. Epstein (2006) clearly distinguishes between the MAB approach which refers to a social science model of the economy and standard economics based on a natural sciences model. He considers that natural science models treat the social world as a natural world, i.e., with invariant laws, whereas social science models can influence the laws that govern the world since they assume interactions between the development of a science and changes in behaviors based on the agents' knowledge about these developments. Thus, there is a clear ontological divide over the nature of economic agents (cf. Davis, 2018). Social science models conceive economic agents as elements of a larger network. They see these economic agents as of particular relevance to the analysis of financial instability which depends not only on the agents' characteristics and decisions but also on the decisions of other agents, thus on their different distribution and on the network morphology. Because those models describe the economy as a "complex, adaptive, dynamic system," learning and expectations[48] matter, and the adaptivity presumes that agents are backward-looking and learn from past events and can extrapolate.[49]

In the development of an empirical validation method, (M)ABM gradually abandoned the pure theoretical (normative) position and became more applied in order to provide alternative economic policy guidance. In addition, although initially those models suffered from the diversity of their modeling strategies and consequent unstable economic policy conclusions, they have improved in this area. Currently, in both the MABM and DSGE literatures, there is (i) a clear epistemic community of MABM authors, and (ii) within that community there are clear groups organized around the same core models and concentrated on specific questions and (iii) a

---

[46]Models with decentralized multi-market transactions.

[47]There are also mainstream complexity economics which means that the ACE need not reject rational expectations or equilibrium (cf. Heise, 2017).

[48]Leijonhufvud (1986, p. 4) argues that the dependence on the state of expectations is "the main reason why macroeconomists do not compare favorably with natural scientist when it comes to predictions."

[49]Foresight is possible but over finite horizons and sequentially modified in light of realized outcomes (cf. Delli Gatti et al. 2008c).

definite (and increasing) overlap among the topics addressed and some convergence in their economic policy conclusions which is reinforcing their credibility and (iv) greater convergence in terms of empirical validation methods (Lux & Zwinkels, 2018) which is enabling more systematic comparison. Finally, DSGE modelers initially refused to investigate this approach arguing that, in the absence of any guarantee of *control*,[50] they preferred *discipline* (standard microeconomic foundations); however, current MABMs enable greater control and are able to compete on that front with DSGE models (Chatelain & Ralf, 2018).

It is undeniable that MABMs have passed an important theoretical and empirical milestone. We have shown that they have new microeconomic foundations for financial fragility which allow interesting economic policy conclusions (Fagiolo & Roventini, 2012). Are these aspects sufficient for policy makers to consider them an alternative to DSGE models? Indeed, while today MABMs have gained reputation and have been introduced in various central banks (Haldane & Turrell, 2018), they are still not fully integrated in the policy maker toolkit (Plassard, 2020). Perhaps a hybrid model[51] as proposed by Assenza and Delli Gatti (2019) might have a better chance of achieving this—at least in the short run. Whether this potentially would increase the appreciation of a "genuine" MABM is difficult to estimate.

References

Arthur, B. W. (2014). Complexity economics: A different framework for economic thought. In *Complexity and the economy*. Oxford University Press.

Ashraf, Q., Gershman, B., & Howitt, P. (2016). How inflation affects macroeconomic performance: An agent-based computational investigation. *Macroeconomic Dynamics, 20*(2), 558–581.

Ashraf, Q., Gershman, B., & Howitt, P. (2017). Banks, market organization, and macroeconomic performance: An agent-based computational analysis. *Journal of Economic Behavior & Organization, 135*, 143–180.

Assenza, T., & Delli Gatti, D. (2019). The financial transmission of shocks in a simple hybrid macroeconomic agent-based model. *Journal of Evolutionary Economics, 29*, 265–297.

Assenza, T., Delli Gatti, D., & Grazzini, J. (2015). Emergent dynamics of a macroeconomic agent-based model with capital and credit. *Journal of Economic Dynamic & Control, 50*, 5–28.

[50]The capacity to "control" models is understood here as the decision to stick to traditional microeconomic foundations and, this way, to stabilize the results and as well as economic policy conclusions. Still today, DSGE modelers consider that any possible alternative microeconomic foundations strategy is excluded because in addition to not easily being justified, it increases the range of possible economic policy conclusions.

[51]"Hybrid models" are recent attempts to incorporate insights from agent-based computational economics into DSGE models.

Aymanns, C., Farmer, J. D., Kleinnijenhuis, A. M., & Wetzer, T. (2018). Models of financial stability and their application in stress tests. In *Handbook of computational economics, volume 4: Heterogeneous agent modeling* (pp. 329–391). Elsevier.

Backhouse, R. E., & Boianovsky, M. (2012). *Transforming modern macroeconomics: Exploring disequilibrium microfoundations, 1956–2003*. Cambridge University Press.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C., & Stiglitz, J. E. (2009). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *NBER working paper* no. 15611. National Bureau of Economic Research.

Battiston, S., Delli Gatti, D., Gallegati, M., Greenwald, B. C., & Stiglitz, J. E. (2012). Default cascades: When does risk diversification increase stability? *Journal of Financial Stability, 8*(3), 138–149.

Benes, J., Kumhof, M., & Laxton, D. (2014). Financial crises in DSGE models: A prototype model. *IMF working paper* no. WP 14/57. International Monetary Fund.

Chatelain, J.-B., & Ralf, K. (2018). Publish and perish: Creative destruction and macroeconomic theory. *History of Economic Ideas, 26*(2), 65–101.

Claessens, S., & Kose, M. A. (2017). Macroeconomic implications of financial imperfections: A survey. *BIS working papers* no. 677. Bank for International Settlements.

Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., & Mehrling, P. (2008). Complexity and dynamics in macroeconomics: Alternatives to the DS GE models. *American Economic Review: Papers & Proceedings, 98*(2), 236–240.

Dal Pont Legrand, M., & Hagemann, H. (2017a). Business cycles, growth and economic policy: Schumpeter and the great depression. *Journal of History of Economic Thought, 39*(1), 19–33.

Dal Pont Legrand, M., & Hagemann, H. (2017b). Schumpeter's analysis of the cleansing effect re-considered. *Journal of Economic Perspectives, 31*(1), 245–256.

Dal Pont Legrand, M., & Hagemann, H. (2019). Propagation and impulses in equilibrium business cycles theories: From interwar debates to DSGE consensus. *GREDEG working paper* no. 2019–1. Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis.

Dal Pont Legrand, M., Baccini, A., Cioni, M., & Petrovich, E. (2021, 5 January 2021). Is cross-fertilization possible in macroeconomics? DSGE confronted to MABM models. *Seminarios del GIDE, FCEA*, Universidad de la Republica, Uruguayad, 4 March 2021. Presented also at the ASSA annual meeting, HES session, *Macro agent-based vs. DSGE modeling: A short history of two competing approaches to macroeconomics*.

Davis, J. (2018). Agent-based modeling's open methodology approach: Simulation, reflexivity, and abduction. *Oeconomia History, Methodology, Philosophy, 8*(4), 509–529.

Dawid, H., & Delli Gatti, D. (2018). Agent-based macroeconomics. In *Handbook of computational economics, volume 4: Heterogeneous agent modeling* (pp. 63–156). Elsevier.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Delli Gatti, D. (2017). The crisis of economic theory and the complexity view: A note. *Italian Economic Journal, 3*(3), 419–422.

Delli Gatti, D., Di Guilmi, C., Gaffeo, E., Gallegati, M., Giulioni, G., & Palestrini, A. (2005). A new approach to business fluctuations: Heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic Behaviour and Organization, 56*(4), 489–512.

Delli Gatti, D., Gallegati, M., Greenwald, B. C., Russo, A., & Stiglitz, J. E. (2006). Business fluctuations in a credit-network economy. *Physica A: Statistical Mechanics and its Applications, 370*(1), 68–74.

Delli Gatti, D., Gallegati, M., Greenwald, B. C., & Stiglitz, J. E. (2007). Net worth, exchange rates, and monetary policy: The effects of a devaluation in a financially fragile environment. *NBER working paper* no. 13244. National Bureau of Economic Research.

Delli Gatti, D., Gaffeo, E., Gallegati, M., Giulioni, G., & Palestrini, A. (2008a). *Emergent macroeconomics*. New Economic Windows.

Delli Gatti, D., Gaffeo, E., Gallegati, M., & Desiderio, S. (2008b). Adaptive microfoundations for emergent macroeconomics. *Eastern Economic Journal, 34*, 441–463.

Delli Gatti, D., Gallegati, M., Russo, A., Greenwald, B. C., & Stiglitz, J. E. (2008c). Financially constrained fluctuations in an evolving network economy. *NBER working paper* no. 14112. National Bureau of Economic Research.

Delli Gatti, D., Russo, A., Greenwald, B. C., & Stiglitz, J. E. (2009). Business fluctuations and bankruptcy avalanches in an evolving network economy. *Journal of Economic Interaction and Coordination, 4*(2), 195–212.

Delli Gatti, D., Russo, A., Greenwald, B. C., & Stiglitz, J. E. (2010a). The financial accelerator in an evolving credit network. *Journal of Economic Dynamics and Control, 34*(6), 1627–1650.

Delli Gatti, D., Gaffeo, E., & Gallegati, M. (2010b). Complex agent-based macroeconomics: A manifesto for a new paradigm. *Journal of Economic Interaction and Coordination, 5*(2), 111–135.

Delli Gatti, D., Desiderio, S., Gaffeo, E., Cirillo, P., & Gallegati, M. (2011). *Macroeconomics from the bottom-up*. Springer Verlag.

De Vroey, M. (2016). *A history of macroeconomics from Keynes to Lucas and beyond*. Cambridge University Press.

Dosi, G., & Roventini, A. (2017). Agent-based macroeconomics and classical political economy: Some Italian roots. *Italian Economic Journal, 3*(3), 261–283.

Dosi, G., & Roventini, A. (2019). More is different... and complex! The case for agent-based macroeconomics. *LEM papers series* no. 2019/01. Laboratory of Economics and Management, Sant'Anna School of Advanced Studies.

Dosi, G., Fagiolo, G., & Roventini, A. (2008). Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *LEM working paper series* no. 2008/21. Laboratory of Economics and Management, Sant'Anna School of Advanced Studies.

Dosi, G., Fagiolo, G., & Roventini, A. (2010). Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics & Control, 34*(9), 1748–1767.

Dosi, G., Fagiolo, G., Napoletano, M., & Roventini, A. (2013). Income distribution, credit and fiscal policies in an agent-based Keynesian model. *Journal of Economic Dynamics & Control, 37*(8), 1598–1625.

Dosi, G., Fagiolo, G., Napoletano, M., Roventini, A., & Treibich, T. (2015). Fiscal and monetary policies in complex evolving economies. *Journal of Economic Dynamics & Control, 52*, 166–189.

Dosi, G., Napoletano, M., Roventini, A., Stiglitz, J. E., & Treibich, T. (2020). Rational heuristics? Expectations and behaviors in evolving economies with heterogeneous interacting agents. *Economic Inquiry, 58*(3), 1487–1516.

Duarte, P. G. (2015). From real business cycle and new Keynesian to DSGE macroeconomics: Facts and models in the emergence of a consensus. *FEA/USP working paper series* no. 2015–05. University of San Paulo, Department of Economics.

Duarte, P. G., & Tadeu Lima, G. (2012). *Microfoundations reconsidered: The relationship of micro and macroeconomics in historical perspective*. Edward Elgar.

Epstein, J. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.

Fagiolo, G., & Roventini, A. (2012). Macroeconomic policy in DSGE and agent-based models. *Revue de l'OFCE, 124*(5), 67–116.

Fontana, M. (2010). The Santa Fe perspective on economics: Emerging patterns in the science of complexity. *History of Economic Ideas, 18*(2), 167–196.

Frisch, R. (1933). Propagation problems and impulse problems in dynamic economics. In K. Koch (Ed.), *Economic essays in honour of Gustav Cassel* (pp. 171–205). Allen & Unwin.

Gallegati, M., Giulioni, G., & Kichiji, N. (2003). Complex dynamics and financial fragility in an agent-based model. *Advances in Complex Systems, 6*(3), 267–282.

Goodwin, R. (1951). Iteration, automatic computers, and economic dynamics. *Metroeconomica, 3*(1), 1–7.

Greenwald, B. C., & Stiglitz, J. E. (1987). Money, imperfect information and economic fluctuations. *NBER working paper* no. 2188. National Bureau of Economic Research.

Greenwald, B. C., & Stiglitz, J. E. (1990). Macroeconomic models with equity and credit rationing. In R. G. Hubbard (Ed.), *Asymmetric information, corporate finance, and investment* (pp. 15–42). Chicago University Press.

Greenwald, B. C., & Stiglitz, J. E. (1993). Financial market imperfections and business cycles. *Quarterly Journal of Economics, 108*(1), 77–114.

Haldane, A. G., & Turrell, A. E. (2018). An interdisciplinary model for macroeconomics. *Oxford Review of Economic Policy, 34*(1–2), 219–251.

Heise, A. (2017). Whither economic complexity? A new heterodox economic paradigm or just another variation within the mainstream? *International Journal of Pluralism and Economics Education, 8*(2), 115–129.

Hoover, K. D. (2012). Microfoundational programs. In P. G. Duarte & G. Tadeu Lima (Eds.), *Microfoundations reconsidered: The relationship of micro and macroeconomics in historical perspective* (pp. 19–61). Edward Elgar.

Howitt, P. (2012). What have central bankers learned from modern macroeconomic theory? *Journal of Macroeconomics, 34*(1), 11–22.

Kirman, A. (2011). *Complex economics. Individual and collective rationality*. Routledge.

Kirman, A. (2016). Ants and monoptimal self-organization. Lessons for macroeconomics. *Macroeconomic Dynamics, 20*, 601–621.

Leijonhufvud, A. (1977). Costs and consequences of inflation. In G. C. Harcourt (Ed.), *The microeconomic foundations of macroeconomics* (pp. 265–312). Macmillan.

Leijonhufvud, A. (1986). Whatever happened to Keynesian economics? *UCLA economics working paper* no. 419. University of California Los Angeles, Department of Economics.

Leijonhufvud, A. (1993). Towards a not-too-rational macroeconomics. *Southern Economic Journal, 60*(1), 1–13.

Leijonhufvud, A. (1997). Macroeconomics and complexity: Inflation theory. In W. B. Arthur, S. Durlauf, & D. A. Lane (Eds.), *The economy as an evolving complex system II* (pp. 321–336). Santa Fe Institute and Wesley Addison.

Leijonhufvud, A. (2014). Economics of the crisis and the crisis. *European Journal of History of Economic Thought, 21*(5), 760–774.

Lucas, R. E., Jr. (1987). *Models of business cycles*. Basil Blackwell.

Lux, T., & Zwinkels, R. C. J. (2018). Empirical validation of agent-based models. In *Handbook of computational economics, volume 4: Heterogeneous agent modeling* (pp. 437–488). Elsevier.

Minsky, H. (1986). *Stabilizing an unstable economy*. Yale University Press.

Plassard, R. (2020). The incorporation of agent-based model into the Bank of England's toolkit. *GREDEG working paper* no. 2020–30. Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), Université Côte d'Azur.

Schumpeter, J. A. (1939). *Business cycles. A theoretical, historical and statistical analysis of the capitalist process, 2 vols*. McGraw-Hill.

Slutsky, E. ([1927] 1937). The summation of random causes as the source of cyclic processes (in Russian). In *Problems of Economic Conditions*, The Conjuncture Institute, Moscow, 3(1); revised English version in *Econometrica*, *5*, 105–146.

Stiglitz, J. E. (1993). Endogenous growth and cycles. *NBER working paper* no. 4286. National Bureau of Economic Research.

Stiglitz, J. E., & Greenwald, B. C. (2003). *Towards a new paradigm in monetary economics*. Cambridge University Press.

Stiglitz, J. E. (2015). Towards a general theory of deep downturns. *NBER working paper* no. 21444. National Bureau of Economic Research.

Symposia. (2018a). Rebuilding macroeconomic theory. *Oxford Review of Economic Policy, 34* (1–2).

Symposia. (2018b). Macroeconomics a decade after the great recession. *Journal of Economic Perspectives, 32*(3).

Trautwein, H.M. (2021). Financial instability and frictions: Can DSGE models finally address the critical issues?. *In this volume*.

Trichet, J. (2010). *Reflections on the nature of monetary policy non-standard measures and finance theory*. Speech by Jean-Claude Trichet, President of the ECB, Opening address at the 6th ECB Central Banking Conference Frankfurt, 18 November 2010. Available at: https://www.ecb. europa.eu/press/key/date/2010/html/sp101118.en.html

Financial Instability and Frictions: Can DSGE Models Finally Address the Critical Issues?

Hans-Michael Trautwein

1 Introduction

The global financial crisis of 2007–2009 was also a crisis for economics, or so it has been widely perceived. Before long, the royal question "why did nobody see it coming?", posed by Queen Elizabeth II at the London School of Economics in November 2008, had become part of the folklore in critical comments on the state of macroeconomics. The answer given by many commentators was that dynamic stochastic general equilibrium (DSGE), the dominant modeling technology, "is hopelessly inadequate for dealing with financial crises and their aftermaths" (Leijonhufvud, 2011, p. 6). The critics pointed out that the standard models—as they were run in terms of DSGE under the interchangeable labels of new Keynesian economics and new neoclassical synthesis—failed to predict the financial crisis because their assumptions about intertemporal optimization and financial market efficiency excluded the occurrence of such crises by definition. Prior to the crisis, the imperative use of DSGE had led to a neglect of earlier insights about financial instability that could be found in the works of Irving Fisher, John Maynard Keynes, Hyman Minsky, and others. At the time of the crisis, when financial markets had their "Minsky moment" full of bankruptcies and other disruption, the opinion spread quickly that—to put it in the words of the representative critic Axel L.—"dynamic stochastic general equilibrium theory has shown itself an intellectually bankrupt enterprise" (Leijonhufvud, 2008, p. 6).

Given such verdicts of hopeless inadequacy, it might have been expected that mainstream macroeconomics would turn away from the DSGE approach. Not long after the outbreak of the crisis, however, V.V. Chari boldly asserted in a testimony before the Committee on Science and Technology in the US House of

H.-M. Trautwein (✉)
Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
e-mail: michael.trautwein@uol.de

Representatives that "[a] useful aphorism in macroeconomics is: 'If you have an interesting and coherent story to tell, you can tell it in a DSGE model. If you cannot, your story is incoherent'" (Chari, 2010, p. 35). At that time, such a statement sounded like the old quip "If reality does not conform to theory, then so much the worse for reality." Yet, in the following years, the DSGE industry saw a boom in the literature on "financial frictions" which has managed to produce stories about financial instability and crises in terms of dynamic stochastic general equilibrium. Some papers even refer explicitly to Fisher, Keynes, Minsky, or other older economists.

Have the critics been proved wrong? Can DSGE models finally address the critical issues of financial instability? To which extent? These are the questions to be answered in the following. In order to assess progress in DSGE, relevant points of criticism will be specified, in Sect. 2, with regard to the state of macroeconomics at the outbreak of the global financial crisis. The canonical three-equation model developed by Michael Woodford (2003) will serve as a reference target of criticism, since it had set a standard for DSGE modeling, accompanied by claims to represent an all-encompassing new synthesis of twentieth-century macroeconomics (Woodford, 2009). This "consensus view" revolved around "frictions" in terms of nominal rigidities which cause deviations from the optimal general equilibrium, accommodating both monetarist and Keynesian views that had prevailed in the heyday of the old neoclassical synthesis. The reduction of macroeconomic theorizing to a frictions view has, however, been criticized as precluding the understanding of fundamental problems of intertemporal coordination or saving-investment imbalances, which are at the root of financial instability and which had been in focus along different lines of thinking in the Wicksell–Keynes tradition.

Section 3 recapitulates criticism that has been directed at Woodfordian DSGE to the Wicksell–Keynes tradition. The wider distinction between a "Wicksell connection" and a "Fisher connection," originally introduced by Leijonhufvud (1981), is used to outline the criteria by which financial frictions DSGE will be assessed. Leijonhufvud's *macroeconomics* family tree is modified to include earlier theories of financial instability and updated to demonstrate how the current literature is connected to earlier stages of new Keynesian economics.

Section 4 outlines the main developments in financial frictions DSGE and takes a closer look at a representative model (by Gertler et al., 2016) that reflects on the state of the art by telling a DSGE story about shadow banking and the recent financial crisis. Section 5 provides an assessment of what has been achieved by those reformulations of the DSGE approach and which inadequacies remain in the light of earlier thinking about financial instability.

2 Pre-crisis DSGE

In January 2008, a few months before the panic that followed the collapse of Lehman Brothers, Michael Woodford presented a paper on "Convergence in Macroeconomics: Elements of the New Synthesis" at the annual meeting of the American Economic Association. He argued that it "is now widely agreed that macroeconomic analysis should employ models with coherent intertemporal general-equilibrium foundations" (Woodford, 2009, p. 269). These models require "that all equations of the model be derived from mutually consistent foundations, and that the specified behavior of each economic unit make sense given the environment created by the behavior of the others" (ibid., p. 270). Woodford (ibid., p. 269) pointed out that:

> [i]n this respect, the methodological stance of the New Classical school and the real business cycle theorists has become the mainstream. But this does not mean that the Keynesian goal of structural modeling of short-run aggregate dynamics has been abandoned. Instead, it is now understood how one can construct and analyze dynamic general equilibrium models that incorporate a variety of types of adjustment frictions that allow these models to provide fairly realistic representations of both short-run and long-run responses to economic disturbances.

The new synthesis was thus based on twisting RBC (real business cycle) theory into new Keynesian economics by adding frictions that could explain deviations of actual output from potential output, defined as the pareto-efficient steady-state growth path of the economy.

A few years earlier, Woodford had published *Interest and Prices* (2003), a huge treatise on *Foundations of a Theory of Monetary Policy* which was soon recognized as a landmark contribution to the new synthesis. It revolves around a baseline optimizing model with nominal rigidities that is representative for pre-crisis DSGE and comes in the typical structure of three interconnected building blocks which describe aggregate demand, aggregate supply, and a monetary policy reaction function (Woodford, 2003, pp. 243–247).[1] The functional relations boil down to three eqs. (IS-AS-MP) that determine the dynamics of output, inflation, and the key interest rate in the economy.

The first equation resembles the IS equation of IS-LM analysis, the old neoclassical synthesis model, in that it describes an inverse relationship between income (aggregate demand) and interest. It is obtained by log-linearizing the first-order condition of the representative household's optimal consumption over time:

$$x_t = E_t x_{t+1} - \sigma\big(i_t - E_t p_{t+1} - \rho_t\big) \tag{1}$$

where x denotes the gap between actual output and the "natural rate of output" (actually their levels in logs), $E_t \ldots$ is the operator for rational expectations, σ is the

---

[1]Christiano et al. (2005) and Smets and Wouters (2007) are often cited as standard pre-crisis DSGE models, too, but they are less "foundational." For another ex post-construction of a pre-crisis "New Keynesian benchmark DSGE model," see Vines and Wills (2018).

intertemporal elasticity of substitution, i is the nominal interest rate, p the inflation rate, and ρ is a disturbance term that represents shocks to the "natural rate of interest." The "natural rate of output" is defined as the fictitious output that monopolistic competition would produce under *flexible* prices, in analogy to growth in the perfect-competition setting of RBC theory. The natural rate of interest is "just the real rate of interest required to keep aggregate demand equal at all times to the natural rate of output" (Woodford, 2003, p. 248). Assuming that the representative household holds rational expectations, the gap between actual and potential output is determined by the expected values for future output and inflation, by contemporaneous shocks to the real rate of interest and by shifts of the nominal rate of interest, the policy variable in the system.

The second equation represents aggregate supply (AS) in terms of a New Keynesian Phillips curve in which actual inflation equals expected inflation plus the actual output gap. This is where the adjustment friction comes into the picture:

$$p_t = \beta E_t p_{t+1} + \kappa x_t, \tag{2}$$

where β is a discount factor and κ a rigidity parameter. Current inflation corresponds to rationally expected future inflation plus the current output gap, modified by the rigidity factor. The firms act in monopolistic competition and set their prices in a staggered fashion, "conveniently" modeled as resulting from a Calvo lottery. A significant fraction of the profit-maximizing firms reacts to shocks by varying their output rather than adjusting prices. Price stickiness increases with the degree of strategic complementarity between suppliers, making output effects of shocks large and persistent.

The third equation is a Taylor rule for monetary policy (MP), run by the central bank which is the representative public authority:

$$i_t = i^* + y_p(p_t - p^*) + y_x(x_t - x^*), \tag{3}$$

where i^* is an expression for possible variation in the inflation target p^* as well as in the disturbances that arise from control errors or mismeasurement by the central bank. The two policy coefficients y_p and y_x define the relative weights of inflation and output targets and, hence, the relative intensity of interest rate reactions to the respective gaps. The target value for the output gap is defined as the steady-state value of the output gap that is in accordance with the inflation target, $x^* = (1 - \beta) p^* / \kappa$.[2] The central bank plays the role of the social optimizer with the task of reducing adverse effects of price stickiness, the key friction in the model. The central bank reacts to output gaps and to deviations of inflation from its target value by varying the nominal rate of interest, its control instrument, in the same direction. The Taylor rule closes the model, permitting the simultaneous determination of interest,

---

[2]This makes the Taylor rule internally consistent, as the definition of x^* ensures that i equals i^* whenever the inflation target p^* is achieved.

inflation, and the output gap (i, p, and x in Eqs. 1–3). It implies that targeting (near) zero inflation maximizes social welfare, since price-level stability disarms the critical friction, preventing the nominal rigidities in the AS equation from taking effect. Woodford (2003, p. 238) argues that "[i]n this way it is established that a nonmonetarist analysis of the effects of monetary policy does not involve any theoretical inconsistency of departure from neoclassical orthodoxy."

3   The Trouble with Intertemporal Coordination

Interest and Prices, the title of Woodford's book, echoes the title of Knut Wicksell's *Geldzins und Güterpreise* (1898) which had broken ground in macroeconomics more than a hundred years before. Woodford (2003, p. 49) describes his version of DSGE as "neo-Wicksellian framework," referring explicitly to Wicksell's concepts of a natural rate of interest, a cashless economy, and a policy rule that moves interest rates in reaction to changes in the price level. Even though he does not mention it, the underlying shock-response structure of modeling cyclical fluctuations could be linked to Wicksell's famous explanation of business in terms of the "rocking-horse" metaphor, later popularized by Ragnar Frisch (Boianovsky & Trautwein, 2007).

The IS equation describes intertemporal optimization in terms that can be traced back to Irving Fisher's *Theory of Interest* (Fisher, 1930). The relevant rate of interest is defined by analogy with the Fisher effect, as the difference between nominal interest and expected inflation, albeit in terms of rational expectations, the new classical twist in the quest for (Walrasian) microeconomic foundations of macroeconomics. The new Keynesian Phillips curve in the AS equation helps to meet "the Keynesian goal" of modeling microeconomically rational deviations from optimal output. The MP equation, in turn, does resembles Wicksell's feedback-rule for price-level stabilization. It has, moreover, a monetarist touch insofar as it assigns the task of macroeconomic stabilization to monetary policy alone. The conclusion that aiming for (near) zero inflation minimizes welfare losses is in line with monetarist thought, even if achieved by way of "nonmonetarist analysis."

Frictions or Coordination Failures?

All this seems to support the claims that DSGE modeling provides an all-encompassing framework, a new synthesis that allows to capture in a single consistent structure the most relevant concepts that were developed in macroeconomics over time and often in an antithetical fashion (Woodford, 2009). This could be illustrated by a map of the kind shown in Fig. 1. Yet, that same map could also be used to illustrate the fundamental criticism directed at DSGE both before and after the financial crisis.
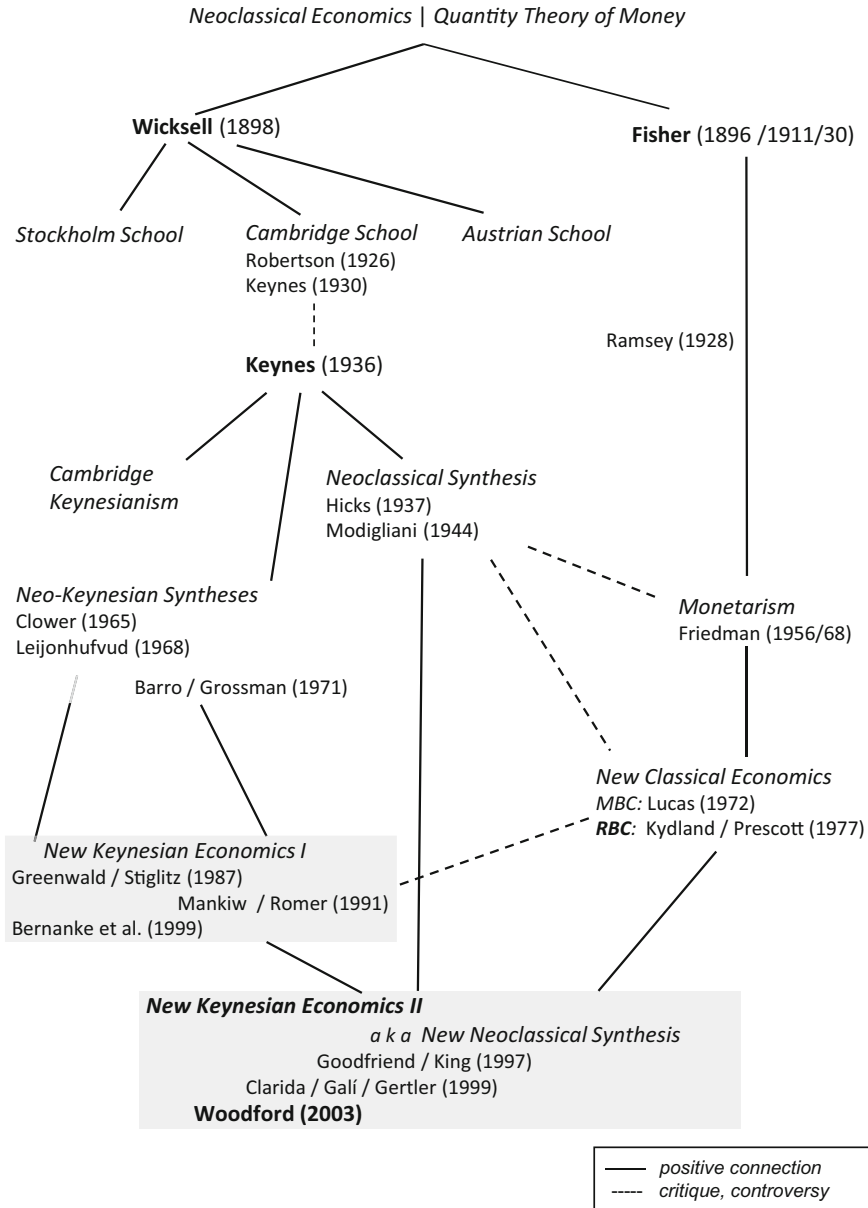
Fig. 1 Routes of thought in macroeconomics (adapted from Sandelin et al., 2014, p. 81)

The upper part of the map is taken from Leijonhufvud (1981) who used it as a "family tree" to distinguish between two traditions in twentieth-century macroeconomics.[3]

One of the traditions is the "Fisher connection," based on Irving Fisher's theory of interest. Along this line of thinking, intertemporal optimization of the representative agent is taken to describe economic activity over time in a general equilibrium framework. Pareto-efficient equilibrium is considered to be globally stable. Inefficiencies are to be explained by frictions in the price mechanism, caused by political interference or other factors exogenous to the model of rational behavior. Money is neutral, at least in the long run, and there is no such thing as financial instability.

The other line of thinking is the "Wicksell connection," in which intertemporal coordination of plans for consumption and production cannot be captured by describing the optimization problem of a representative agent, since it is faced by heterogeneous agents in large and complex systems of interdependent markets. A characteristic of the Wicksellian approach is its focus on recurrent failures of the intertemporal price mechanism to keep the market rate of interest at a level conforming to (pareto-efficient) full employment at a stable value of money. The gaps between the market rate and the equilibrium rate of interest (aka the "natural rate") arise from problems of incomplete information about the latter.

The ensuing imbalances between saving and investment are at the roots of monetary and financial instability which in turn tends to affect real investment and employment. It is a characteristic of most of the theories in the Wicksell connection that they put the global stability of full-employment equilibrium in doubt.[4] However, Leijonhufvud (1981) also argued that Keynes "lost" the Wicksellian theme of saving-investment imbalances on the way from the *Treatise on Money* (Keynes, 1930) to the *General Theory of Employment, Interest and Money* (Keynes, 1936). Keynes detached the coordination problem from the interest-rate mechanism, as his focus on the interaction of investment and aggregate income reduced aggregate saving to a merely passive variable. His liquidity preference theory of interest opened a new line of thinking about financial instability, but Keynes' followers in the neoclassical synthesis by and large discarded it. After Modigliani (1944), the standard "Keynesian" explanation of underemployment equilibrium was essentially reduced to sticky wages, hence to nominal rigidities in the price mechanism. This frictions view brought the neoclassical Keynesians close to the lines of thinking in

---

[3]In Leijonhufvud (1981), the family tree ends with the Cambridge Keynesians, the (old) neoclassical synthesis and Milton Friedman. Frank Ramsey's *Mathematical Theory of Interest* (1928) has been added here, since its intertemporal optimization framework is more explicitly referred to in RBC theory.

[4]Exceptions are the Austrian business cycle theory in the spirit of Hayek's *Prices and Production* (1931; see, e.g., Trautwein, 1996) and the neoclassical synthesis in the spirit of Modigliani (1944). Wicksell kept his theory of monetary instability apart from his (sketchy) view of business cycles (Boianovsky & Trautwein, 2001). He introduced the "rocking-horse" metaphor to describe business cycles, but the "Wicksell connection" in macroeconomics did not proceed from the global-stability presumption.

the Fisher connection, and it became difficult to discuss financial instability within the framework of the neoclassical synthesis. As critics of pre-crisis DSGE have pointed out that same problem reoccurred in the new neoclassical synthesis.
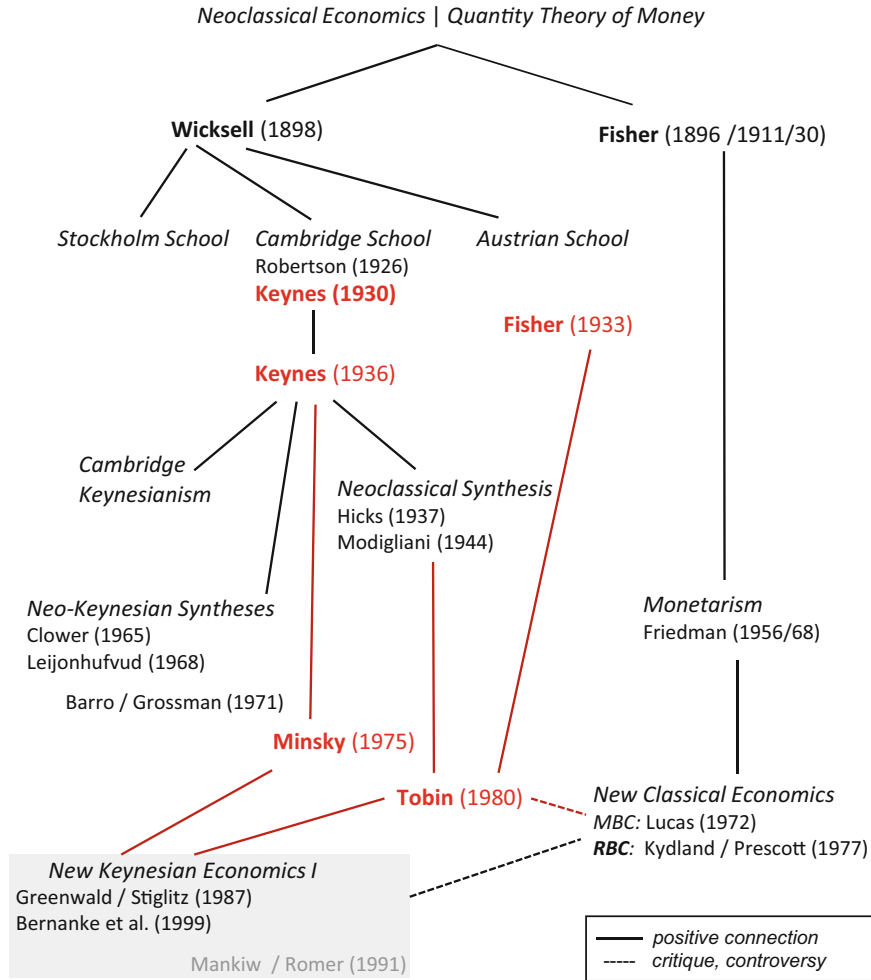
Strange Cousins in the Family Tree

Before taking a closer look at the criticism directed at the IS-AS-MP framework, it should be noted that the post-crisis DSGE literature makes occasional references to earlier theories of financial instability that do not normally figure in Leijonhufvud-style family trees. Suffice it to mention two of those "strange cousins," namely, Irving Fisher's *Debt-Deflation Theory of Great Depressions* (1933) and Hyman Minsky's *Financial Instability Hypothesis* (1975, 1978), and to give them their place in Fig. 2.

Fisher's debt-deflation theory is a strange cousin because it belongs to the Wicksell connection rather than the one under Fisher's own name. It describes a downward cumulative process of deflation in which the real value of nominal debt is rising. Firms that had borrowed to finance their investment are forced into fire sales and illiquidity which sooner or later produces chains of insolvencies of both borrowers and lenders. At the end of the process, the economy may land in a new equilibrium, but this is a state of rest far inferior to the situation before deflation started.

Fisher's debt-deflation theory was revived by James Tobin (1980) to show what was crucially missing in the new classical description of fluctuations of aggregate economic activity as optimal responses to random shocks. Even if Tobin, who labeled himself an "old Keynesian," is normally ranked among the major figures in the old neoclassical synthesis, he did clearly not subscribe to the frictions view nor believe in the global stability of full-employment equilibrium. He belonged to the group of Keynesians who (like Leijonhufvud, 1968) insisted on Keynes' insight (in Chap. 19 of the *General Theory*) that downward flexibility of wages and prices tends to exacerbate rather than counteract the lack of effective demand in recessions.

Minsky's financial instability hypothesis is a strange cousin by affiliation—not so much because it refers to the "animal spirits" motive in Keynes' liquidity preference theory of interest, but because it figures prominently in post-Keynesian economics. The narrative revolves around the endogeneity of credit cycles that run through the stages of hedge, speculative, and Ponzi finance. Despite referring to Keynes (1930, 1936), Minsky's theory is a modern version of older stories about credit cycles that have been recycled ever since the times of Lord Overstone in the mid-nineteenth century. It is an odd bird simply because it has, for a long time, been treated as such in the mainstreams of economic discourse, both in the Fisher and Wicksell connections. It has itself run through cycles, being regularly cited whenever the bursting of financial bubbles reminds observers that another "Minsky moment" has occurred, only to fall into oblivion again shortly thereafter. The reason for such limited attention to Minsky's theory may lie in its affiliation with post-Keynesian

Neoclassical Economics | Quantity Theory of Money

Wicksell (1898)   **Fisher** (1896 /1911/30)

Stockholm School   *Cambridge School*   *Austrian School*
Robertson (1926)
Keynes (1930)

Fisher (1933)

Keynes (1936)

*Cambridge
Keynesianism*

Neoclassical Synthesis
Hicks (1937)
Modigliani (1944)

Neo-Keynesian Syntheses
Clower (1965)
Leijonhufvud (1968)

Monetarism
Friedman (1956/68)

Barro / Grossman (1971)

Minsky (1975)

Tobin (1980) -----   *New Classical Economics*
MBC: Lucas (1972)
RBC: Kydland / Prescott (1977)

New Keynesian Economics I
Greenwald / Stiglitz (1987)
Bernanke et al. (1999)

Mankiw / Romer (1991)

——	*positive connection*
-----	*critique, controversy*

Fig. 2 Some theories of financial instability

Economics, frequently dismissed as a heterodox church of its own. Or it lies in allegations that his financial instability hypothesis is impossible to formalize properly, where "properly" means "compatible with the standard frameworks of macroeconomic theorizing."[5]

Various aspects of Fisher-Tobin and Minsky theories of financial stability have nevertheless found their way into New Keynesian Economics at stage I in the maps

---

[5]It is possible to formalize Minsky's theory in terms of nonlinear dynamics, e.g., in terms of limit cycles *à la* Goodwin, and there is an ample literature about this; for a survey see Ryoo (2010). Modern mainstream macroeconomics has for a long time eschewed nonlinear dynamics because, in addition to being highly demanding, it does not render simple, clear-cut conclusions.

of Figs. 1 and 2. Back then, in the 1980s and early 1990s, the term new Keynesian economics referred to a wide variety of models formed to confront the new classical DSGE paradigm in Robert Lucas' monetary theory of the business cycle and the RBC approach of Finn Kydland and Edward Prescott. While the new classicals insisted on what they considered as standard Walrasian microfoundations, the new Keynesians at stage I made use of more modern microeconomics for rigorous modeling of involuntary unemployment and other Keynesian themes.

Greenwald and Stiglitz (1987) postulated that a general theory must account for the observed fluctuations and persistence of unemployment, carefully distinguish between saving and investment, and show how disturbances in effective demand generate cyclical behavior of macroeconomic variables. They claimed that this could be achieved by models based on incomplete and asymmetric information which feature capital market imperfections, credit rationing, and policies that stabilize not merely the price level but the financial system.

The stage I manifesto of Greenwald and Stiglitz was in line with thinking about intertemporal coordination in the Wicksell connection, but their agenda was soon relegated to the periphery of new Keynesian economics. The reference collection of Mankiw and Romer (1991) neatly divided new Keynesian economics in two lines: "imperfect competition and sticky prices" and "coordination failures and real rigidities." Volume 1 focused "on how friction in price setting at the microeconomic level leads to nominal rigidity at the macroeconomic level." It was this literature that paved the way for new Keynesian economics, stage II. The frictions line gradually converged with new classical RBC theory. The surveys by Goodfriend and King (1997) and Clarida et al. (1999) made it customary to use the label "new Keynesian economics" (NKE) interchangeably with "new neoclassical synthesis" (NNS).

Synthetic Blindness to Financial Instability

The pre-crisis framework of DSGE modeling met with fundamental criticism even before the crisis broke out. It was pointed out that the baseline model had no role for commercial banks or other financial intermediaries, nor for liquidity. Woodford's assumption of complete financial markets implies that money (in terms of the central bank's liabilities) can be perfectly substituted by (other) financial assets. The microeconomic foundations of the power of the central bank to control interest rates in terms of the MP equation were therefore found to be seriously lacking (Boianovsky & Trautwein, 2006).

The primary target of criticism, though, was the IS relation, since it describes the intertemporal coordination of economic activity as the successful outcome of the optimizing choice of a representative household. According to the critics, this precludes any analysis of the coordination problems that cause financial crises, mass unemployment, and other *explananda* of macroeconomic theory (see, e.g., Colander et al., 2008, p. 236; Lux & Westerhoff, 2009; Goodhart & Tsomocos, 2012;

Leijonhufvud, 2014, pp. 770–772; Stiglitz, 1918). Savings-investment imbalances simply cannot be addressed in this framework.[6]

The route to analyzing macroeconomic instabilities is blocked by further requirements that the DSGE technology imposes on the agents in the model, in particular high demands on cognitive competence for intertemporal optimization. Under the assumption of rational expectations, it is postulated that the evolution of the economy is a fully determined, even if stochastic process accurately anticipated by all private-sector agents in the model. That assumption is not even sufficient to make intertemporal optimization fully consistent. While the standard pre-crisis model lets the representative household maximize utility without borrowing constraints, DSGE models with infinitely lived agents routinely require that budget constraints are binding all the time. The transversality condition must hold: no capital, no debt is left over at the end (whatever that means in infinity). Interlocking violations of intertemporal budget constraints, which become manifest in financial crises, are ruled out by assumption (Goodhart & Tsomocos, 2012). Furthermore:

> DSGE models are. . . peculiarly prone to fallacies of composition, most particularly so, of course, in their representative agent versions. The models are blind to the consequences of too many people trying to do the same thing at the same time. The representative lemming is not an intertemporal optimising creature (Leijonhufvud, 2009, p. 753).

In the new Keynesian twist of DSGE, intertemporal optimization does not directly coincide with the social optimum. Output gaps are derived from frictions in the price mechanism that are introduced with the rigidity parameter in the new Keynesian Phillips curve (the AS equation). There are strong doubts about the empirical relevance of the postulated rigidities, as the typical menu costs and strategic complementarities hardly account for the fluctuations of aggregate output observed in the data. The conventional procedure of adding further frictions to make the model fit the data (as popularized by Christiano et al., 2005, and Smets & Wouters, 2007) invites indictments of ad hocery. The links to the "rigorously microfounded" core of the new synthesis are loosened beyond what used to be permissible by earlier modeling conventions.

Even more seriously, the critics point out that introducing nominal rigidities into an RBC-style framework with monopolistic competition has rigorous microfoundations but follows simply from the methodological prescription to minimize deviations from the competitive equilibrium benchmark in order to preserve welfare-theoretical "rigor." In that perspective, unlimited price flexibility, including asset price volatility in financial markets, ensures the proper functioning of the economy, whereas any inflexibility leads to a suboptimal state. The implicit

---

[6]It should be noted that Christiano et al. (2018) disagree, referring to pre-crisis DSGE models that allowed for financial frictions or liquidity-constrained consumers. However, those models received little attention at the time. Christiano et al. (2018, p. 13) inadvertently admit to this when maintaining that, "guided by the post-war data from the U.S. and Western Europe, and experience with existing models of financial frictions, DSGE modelers emphasized other frictions." It is an odd defense to argue that macroeconomists, who routinely assume agents in their models to have forward-looking rational expectations, themselves worked on backward-looking expectations.

assumption of efficient financial markets obfuscates detrimental effects of asset price volatility that typically occur in financial market booms and crises: In addition to redistributing net worth between those who trade in the financial markets, it tightens of aggregate budget constraints in and after financial crises. The economy will not automatically return to the initially pareto-efficient growth path, so the latter cannot be considered as globally stable.

The indictment of pre-crisis DSGE for being guilty of hopeless inadequacy to deal with financial crises and their aftermaths was supported by a sort of "turn Queen's evidence" (remember the royal question!). It was provided by none other than Robert Lucas (2004, p. 23) in a talk on his "Keynesian education" a few years before the outbreak of the global financial crisis:

> The problem is that the new theories, the theories embedded in general equilibrium dynamics of the sort that we know how to use pretty well now—there's a residue of things that they don't let us think about. The don't let us think about the U.S. experience in the 1930s or about financial crises and their real consequences in Asia and Latin America. They don't let us think, I don't think, very well about Japan in the 1990s. We may be disillusioned with the [old] Keynesian apparatus for thinking about these things, but it doesn't mean that this replacement apparatus can do it either. It can't.

4   DSGE with Financial Frictions

It seems that Lucas and the critics have been proved wrong. In the decade following the global financial crisis, the "financial frictions" branch of the DSGE industry has managed to produce stories about financial instability and crises in terms of dynamic stochastic general equilibrium. In the mapping exercise started with Figs. 1 and 2, new Keynesian economics could thus be extended from stage II to III (henceforth NKE III).

Financial Accelerator and Credit Cycles

The basic requirement for modeling financial crises is a distinction between lenders and borrowers plus the introduction of frictions in credit markets. For this, contributions to post-crisis NKE III invoke information asymmetries and moral hazard problems as had been modeled in the market imperfections literature of NKE, stage I. The new literature builds explicitly on the credit cycle approach of Kiyotaki and Moore (1997) and the financial accelerator mechanism in the tradition of Bernanke and Gertler (1989). The former refers to quantity constraints of lending produced by collateral requirements; these are related to borrowers' net worth which in turn fluctuates with asset prices. The latter approach revolves around external finance premia that reflect costs of monitoring loan projects and create a time-varying wedge between the risk-free rate of interest (the policy rate) and the market price of credit.

Net worth changes pro-cyclically and external finance premia move counter-cyclically.

Both approaches can be combined to construct critical feedback loops between fluctuations in investment, output and other activities in the real sphere, and balance-sheet constraints in the financial sphere of the economy. Both approaches provide straightforward routes for integrating banks into DSGE models, in settings that permit to cast light on their ambiguous roles at the center of the system. Banks help lenders to solve problems of costly state verification and repossession of collateral; this provides a microeconomic rationale for building them into the model as a class of intermediate agents. However, they typically run their business of borrowing and lending at levels of leverage higher than those ordinarily sustainable in the nonfinancial sector. Banks can therefore be modeled as amplifying transmitters of real shocks and as originators of financial shocks that affect real economic activity. This is generally done by introducing an agency problem as the key friction: Households who lend to banks are aware of the risk that bankers "cheat" by diverting assets for their own consumption and, hence, have incentives to drive leverage to excessive levels. Quantity constraints in lending can thus go both ways between the financial and nonfinancial sectors. Banks constrain nonfinancial borrowers (households, firms) by collateral requirements that move counter-cyclically and nonfinancial lenders (typically households) force banks into balance-sheet contraction and crisis when responding to (relatively small) shocks by making leverage constraints binding.

DSGE models with financial frictions vary widely in their divergence from the pre-crisis framework. A large assortment of model specifications accounts for balance-sheet constraints that emerge from a range of net worth affecting mechanisms in different sectors.

Table 1 provides a sketchy overview over different approaches derived from NKE stages I and II; the classification of model types is far from complete, and the references are just examples (not necessarily the best) for each class.[7] Proceeding from the top, the table shows a stepwise increasing distance from the pre-crisis DSGE framework, or NKE II, almost following a timeline of chronological progress. Some papers, such as Cúrdia and Woodford (2010), minimize the difference by dividing the household sector into savers and borrowers who use banks as intermediaries. Shocks that generate growth in the volume of lending increase default risks and intermediation costs. This leads to variations in the spread between deposit rates and lending rates, featuring a basic financial accelerator mechanism. In that class of models, the financial friction is merely superimposed on the nominal rigidities of the pre-crisis standard.

By contrast, some of the more advanced and sophisticated DSGE models with financial frictions do without nominal rigidities in the nonfinancial sector. Several of them balance the complexities that arise from modeling a nontrivial financial sector

---

[7]For a variety of surveys that differ in their structuring exercises see, e.g., Brunnermeier et al. (2013), Claessens and Kose (2017), Christiano et al. (2018), and Vines and Wills (2018).

Table 1 Different approaches to financial instability in DSGE

Origins	Frictions	Borrowers	Banks	Exemplary contributions
NKE II + NKE I–*FA*	Sticky prices + time-varying spread-amplifying effects of shocks	Impatient households	Intermediation only	Cúrdia and Woodford (2010)
NKE II + **NKE I–*FA***	Sticky prices + fluctuating risk → ext'l. fin. Premium affecting leverage of borrowers	Firms	Intermediation only	Christiano et al. (2014)
NKE II + **NKE I–*CC***	Sticky prices + agency problem (moral hazard)→ leverage constraint restricting output	Firms	Intermediation only	Gertler and Karadi (2011)
NKE II + NKE I–*FA* + NKE I–*CC*	Sticky prices & wages → ext'l. fin. Premium + leverage constraint leading to large jumps in banks' leverage	Impatient households and firms	Money creation	Boissay et al. (2016); Jakab and Kumhof (2018)
NKE I–*FA* + **NKE I–*CC***	Agency problem (moral hazard) → endogenous cycle with changing leverage constraints and bank runs	Retail banks and wholesale banks	Liquidity expansion by financial innovation (wholesale banks)	Gertler et al. (2016); Fève et al. (2019)

NKE new Keynesian economics; *NKE I–FA* financial accelerator approach at NKE, stage I; *NKE I–CC* "collateral constraint" or "credit cycle" approach at NKE, stage I; *NKE II* "nominal rigidities" approach at NKE, stage II (pre-crisis DSGE); bold and normal letters indicate the relative strength of emphasis on one or other type of frictions

by collapsing households and goods-producing firms into one (nonfinancial) sector. In many papers, the policy analysis papers proceeds beyond variations of the Taylor rule. They include comparisons of quantitative easing, lending of last resort and macroprudential regulation in terms of their effectiveness, or even modeling of their interaction. On the whole, financial frictions, DSGE in NKE III has differentiated itself strongly from the canonical IS-AS-MP anatomy of NKE II.

Post-Crisis DSGE in a Representative Model

Despite the large variety of models in the financial frictions literature, it is possible to refer to a representative framework for testing the validity of the argument that the DSGE approach is inadequate for dealing with financial crises and their aftermaths. In their contribution to the *Handbook of Macroeconomics*, Mark Gertler and Nobuhiro Kiyotaki—two of the pioneers in the financial accelerator and credit cycle literatures of NKE, stage I—claim to "present a simple canonical macroeconomic model of banking crises that [. . .] is representative of the existing literature"

(Gertler, Kiyotaki and Prestipino—henceforth GKP—2016, p. 1348). They point out that their model extends this literature "to feature a role for wholesale banking." This extension serves to mimic the mechanisms underlying the global financial crisis in 2007–2009, when "highly leveraged financial institutions along with highly leveraged households [. . .] were most immediately vulnerable to financial distress" (GKP 2016, p. 1347).

GKP set the focus on the interaction between a consolidated nonfinancial sector and a disaggregated financial sector. The nonfinancial sector consists of households who optimize intertemporal consumption and hold a share of the economy's productive assets directly. The financial sector is disaggregated into retail banks and wholesale banks, with each subsector holding further shares of the productive assets and borrowing from households for this. The wholesale banks correspond roughly to the "shadow banking" system of the pre-crisis boom, with institutions that originated loans to households, securitized them, and funded them by borrowing short-term from more regulated retail banks.[8] In the GKP framework, retail banks take deposits from households and lend them to wholesale banks or to other retail banks in the interbank market. While the wholesale banks have a cost advantage over retail banks in making standardized "nonfinancial loans" for long-term investment in productive assets, the retail banks have a cost advantage over households in lending to wholesale banks and making nonstandardized loans to other retail banks. Efficient specialization and regulatory arbitrage lets the interbank market grow endogenously by way of "financial innovation." Increasing leverage of the wholesale banks improves the liquidity of the system but makes it simultaneously more vulnerable to crises.

In order to analyze vulnerability GKP derive a potentially critical limit to the wholesale banks' ability to raise funds. They interpret the cost advantages in managing nonfinancial loans in terms of asymmetric information that gives rise to an agency problem. Bankers can choose between operating honestly and diverting assets for personal use; retail bankers can do this vis-à-vis households and wholesale bankers vis-à-vis retail bankers. The interplay of efficient increases in leverage and this moral hazard problem make the financial sector susceptible to runs, both unanticipated and anticipated (with a certain probability). Runs start with spontaneous failures of wholesale banks to roll over their short-term loans, with which they finance long-term projects, whenever retail banks decline to lend to them. If retail banks thus choose to "run," the wholesale banks liquidate their capital in fire sales. The proceeds are used by creditors in the retail bank sector to acquire capital or to sell it to households.

In the GKP model, runs on the entire sector occur whenever the liquidation value of the wholesale banks' assets falls below the outstanding liabilities to their interbank creditors. In this case, liquidation tends to turn wholesale banks' net worth negative, making incentive constraints of lenders bind throughout the system. The GKP model demonstrates how such runs can transform small shocks to

---

[8]Fève et al. (2019) construct a very similar model in which they speak explicitly of shadow banks created to circumvent standard regulation of "traditional banks."

productivity, which would produce only minor recessions at normal times, into major disasters in the dimension of the Great Recession. The fire sales in the wholesale bank segment spill over to the retail bank segment, where a fall in the net worth constrains the ability of retail banks to issue safe liabilities. As a consequence, real economic activity is reduced by a rise in the costs of finance.

Gertler et al. (2016) end their exploration of "realistic" extensions to their DSGE model of financial crisis with a discussion of options for stabilization policies. They model and simulate two types of intervention: ex ante in terms of macroprudential leverage restrictions and ex post in terms of lending of last resort. They show how unconventional policies of the latter type (large-scale asset purchases in the interbank market) can prevent runs from happening, if agents anticipate the intervention in response to a rule that refers to a threshold in the spread between the deposit rate and the wholesale banks' return on assets. Compared to the "management of expectations" by a Taylor rule in normal times, as described by pre-crisis DSGE models, this may be considered the "financial frictions" equivalent in times of crisis.

Achievements and Limitations

With the progress that new Keynesians have made in modeling financial crises, they appear to have disproved the verdict of hopeless inadequacy of DSGE. Various contributions to the financial frictions branch of the NKE literature take an even wider historical perspective and reflect on theories of financial instability in the traditions of Fisher (1933), Keynes (1936), Minsky (1975), and others.[9] Yet, it can be argued that some of the critique of DSGE models remains valid even with respect to the present NKE III literature. The following assessment moves from points of criticism that may be considered as outmoded to those that still have some force.

The rigid architecture of pre-crisis DSGE, as represented by Woodford's IS-AS-MP framework, has given way to a variety of structures, in which nominal rigidities are no longer indispensable for explaining output gaps. The single-minded reliance on unexplained price inflexibility has been replaced by emphasis on credit market imperfections that are more appealing with regard to their theoretical foundations and empirical plausibility. In NKE III, the key sources of the short-term adjustment problems in the economy are not rigidities of goods prices or wages, but the volatility of asset prices and credit constraints. Financial frictions DSGE introduces some heterogeneity by splitting the representative household into subclasses of savers and borrowers and by letting additional classes and subclasses of agents (financial intermediaries) solve different optimization problems with behavioral choices ("honest" or "cheating"). In the GKP model, the nonfinancial sector is essentially boiled down to a representative household again, largely in line with

---

[9] See, e.g., Brunnermeier et al. (2013), and Jakab and Kumhof (2015).

the pre-crisis Woodford (2003) model. Yet, even in that simplified fashion, the NKE III literature is obviously able to describe balance-sheet recessions due to the more complex structure of the financial sector. The GKP model, in particular, captures some general characteristics of Minsky's and Fisher's theories of financial instability. It describes a boom generated by a rise in leverage ratios which in turn makes the system increasingly fragile. The ensuing rollover crisis triggers fire sales and downward movements of the system akin to Fisherian debt deflation.

However, while the GKP framework may allow to mimic some of the dynamics of the global financial crisis, it is not easily made compatible with essential features of Fisher- and Minsky-type theories of financial instability. First, modeling the leverage boom as an increase in the efficiency of financial intermediation (reduction of agency cost) appears to contradict Minsky's story about the recurrent moves from hedge finance to speculative and Ponzi finance, which are basically considered as excesses. It may be argued that they, in the GKP model, are exposed as such only with hindsight but that would hardly fit with the underlying assumption of forward-looking rational expectations. Second, GKP-type models of financial frictions do not easily permit to analyze the real effects of Fisherian debt deflation. The occurrence of defaults is restricted to the wholesale bank segment (and the insolvent banks are assumed to be replaced by new entries); there are no nonfinancial firms that can go bankrupt in the course of a revaluation of their debt; nor will households suffer more than a temporary loss in the value of their financial assets.

It may be technically possible to bring the GKP framework closer to the older theories of financial instability. Yet, there are further ambiguities in the progress made in the financial frictions literature. The charge of ad hocery against DSGE frameworks that add frictions to frictions until the model fits the data applies to much work in the "financial frictions" line, too. Many contributors proceed carefully from a baseline model to introduce the frictions step by step, so as to isolate the net effects of each friction; the GKP distinction between unanticipated and anticipated runs is a case in point. Yet, that same example betrays another problem, akin to that indicated above: Unanticipated runs are hardly compatible with the rational expectations assumption that GKP use for intertemporal optimization. In general, the financial frictions literature in NKE III is rather casual about distinctions and interactions between asymmetric and incomplete information, which certainly ought to have consequences for assumptions about the cognitive competence required for intertemporal optimization in these models.

The most serious objection to claims of adequate treatments of financial instability in DSGE is the implicit reliance on the global stability of the ex ante optimal steady-state growth path. Basic assumptions and procedures set bounds that restrict the frameworks from dealing fully with the dynamics of violations of interlocking budget constraints which typically follow from financial crises in the real world. The standard assumption in financial frictions DSGE is an infinite horizon for the intertemporal optimization of the representative household(s). The combination with rational expectation necessitates the setting of a transversality condition that rules out rational bubbles, leftovers of "useless capital," and unsettled debt from defaults. As Goodhart and Tsomocos (2012, p. 51) express it, this requires that:

whatever happens in the future, the debtor will still be able [and willing] to repay. This must logically require complete financial markets, wherein all eventualities, including Donald Rumsfeld's famous "unknown unknowns", can be hedged. How can you price and hedge the unknown? Since the number of potential future outcomes is infinite, any transaction cost, however minute, would make the whole exercise infeasible.

Most DSGE models with financial frictions come with a large formal apparatus in the text parts and appendices, but the transversality conditions have become invisible bolts that keep the structures from falling apart, hidden somewhere deep down in the machine room.[10]

Apart from such heroic assumptions, the standard DSGE technology as such prevents the models from veering far into regions of financial instability. The basic philosophy is to study crises in terms of perturbations around a non-stochastic steady-state growth path of the economy. The models are typically log-linearized around the steady state that represents optimal output. This procedure excludes multiple equilibria and further (in)stability problems to which Axel Leijonhufvud, our representative critic, refers as "violations of interlocking intertemporal budget constraints" and "changing dimensions of the [Edgeworth] box" (Leijonhufvud, 2014, pp. 771–772). Even in the impulse-response graphs of the representative model, for which GKP claim stronger persistence of critical effects than those derived in other studies, the system is back on the original track after 10 years; it reverts to the initial output level after 40 quarters, unless it is impacted by new shocks (GKP 2016, Sects. 4.3–7).

5   Conclusion

In the past decade, DSGE modeling has undoubtedly made progress in addressing issues of financial instability of the type that occurred in the global financial crisis of 2007–2009. To show how actual output may deviate from potential output, new Keynesians do no longer (exclusively) rely on inflexibilities of prices. They are now able to highlight the causes and effects of critical asset price flexibility. They refer to "precursors," such as Fisher, Keynes, and Minsky, and to concepts previously developed but temporarily sidelined, in New Keynesian Economics. Those concepts include agency problems, such as adverse selection, moral hazard or limited commitment of borrowers, following from asymmetric and incomplete information. Such agency problems are invoked, for example, in the representative post-crisis DSGE model crafted by Gertler et al. (2016). What used to be described as "financial market imperfections" at the first stage of new Keynesian economics, back in the 1980s, is named "financial frictions" at the third stage.

---

[10] A few exceptions can be found, for example, in Brunnermeier et al. (2013, p. 17). Models with overlapping generations of finitely lived agents (OLG), which were frequently used in the NKE I literature, can handle asset price bubbles without transversality conditions; but they have other limitations and are outside the scope of this paper.

Does the difference in terminology matter? No and yes. Both terms—frictions and imperfections—are related to an ideal state: perfect competition in the sense of competitive general equilibrium with complete financial markets. The technical conceptualization of agency problems in the NKE III literature does not essentially differ from its origins in the NKE I literature. In that sense, the use of different terms does not seem to matter.

However, there lurks a serious semantic issue here. In the NKE I literature, the notion of "financial market imperfections" was not merely related to asymmetric information that could be handled with incentive-compatible arrangements that set efficient constraints. It referred to more general problems of imperfect information. It had still a connotation of fundamental uncertainty, the "unknown unknowns" that arise from interactive and adaptive behavior in financial markets in which intertemporal plans of savers and investors do not automatically match or converge towards the optimum envisaged ex ante. Some authors in the NKE I literature on the "imperfections" line continued to pursue the theme of intertemporal coordination failures that characterizes the Wicksell–Keynes connection in macroeconomics. Working in that tradition, they sought to uncover the laws of motion of the economy, in which the agents follow specified aims, but do not always succeed in achieving them, while adapting and learning by trial and error. Those older new Keynesians argued that time matters, both for the understanding of adaptive behavior as a process in which the sequencing of decisions is important and for thinking about the evolution of institutions that condition the agents' behavior. In the "imperfections" line of NKE I, full-employment equilibrium (the pareto-efficient steady-state growth path) was not presumed to be globally stable.

The "financial frictions" line of NKE III, on the other hand, remains wedded to the DSGE technology in which the agents' behavior is optimizing ex ante not only by intent but also in terms of outcome. Even in the most sophisticated new Keynesian models of financial crises, the system is bound to return to the predetermined steady-state path. In this respect, DSGE remains inadequate in analyzing the effects of violations of interlocking budget constraints, keeping new Keynesian Economics, stage III stuck in a "stable equilibrium system cum frictions" view. That makes it difficult to gain full analytical scope of financial instability.

References

Barro, R., & Grossman, H. (1971). A general disequilibrium model of income and employment. *American Economic Review, 61*(1), 82–93.

Bernanke, B., & Gertler, M. (1989). Agency costs, net worth and business fluctuations. *American Economic Review, 79*(1), 14–31.

Bernanke, B., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In J. B. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (Vol. 1, pp. 1341–1393). North-Holland.

Boianovsky, M., & Trautwein, H.-M. (2001). Wicksell's lecture notes on economic crises (1902/05). *Structural Change and Economic Dynamics, 12*(3), 343–365.

Boianovsky, M., & Trautwein, H.-M. (2006). Wicksell after Woodford. *Journal of the History of Economic Thought, 28*(2), 171–185.

Boianovsky, M., & Trautwein, H.-M. (2007). Johan Åkerman vs. Ragnar Frisch on quantitative business cycle analysis. *European Journal of the History of Economic Thought, 14*(3), 487–517.

Boissay, F., Collard, F., & Smets, F. (2016). Booms and banking crises. *Journal of Political Economy, 124*(2), 489–538.

Brunnermeier, M., Eisenbach, T., & Sannikov, Y. (2013). Macroeconomics with financial frictions: A survey. In D. Acemoglu, M. Arellano, & E. Dekel (Eds.), *Advances in economics and econometrics II* (pp. 4–94). Cambridge Univ. Press.

Chari, V. V. (2010). Written statement. In *Building a science of economics for the real world* (pp. 34–37) Hearing before the Subcommittee on Investigations and Oversight, Committee on Science and Technology, House of Representatives, 111th Congress, 2nd session, July 20, 2010. U.S. Government Printing Office.

Christiano, L., Eichenbaum, M., & Evans, C. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy, 113*(1), 1–45.

Christiano, L., Motto, R., & Rostagno, M. (2014). Risk shocks. *American Economic Review, 104*(1), 27–65.

Christiano, L., Eichenbaum, M., & Trabandt, M. (2018). On DSGE models. *Journal of Economic Perspectives, 32*(3), 113–140.

Claessens, S., & Kose, A. (2017). Macroeconomic implications of financial market imperfections: A survey. *BIS working papers* no. 677. Bank for International Settlements.

Clarida, R., Galí, J., & Gertler, M. (1999). The science of monetary policy: A new Keynesian perspective. *Journal of Economic Literature, 37*(4), 1661–1707.

Clower, R. (1965). The Keynesian counterrevolution: A theoretical reappraisal. In F. H. Hahn & F. P. R. Brechling (Eds.), *The theory of interest rates* (pp. 103–125). Macmillan.

Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., & Mehrling, P. (2008). Beyond DSGE models: Towards an empirically based macroeconomics. *American Economic Review, 98*(2), 236–240.

Cúrdia, V., & Woodford, M. (2010). Credit spreads and monetary policy. *Journal of Money, Credit and Banking, 42*(s1), 3–35.

Fève, P., Moura, A., & Pierrard, O. (2019). Shadow banking and financial regulation: A small-scale DSGE perspective. *Journal of Economic Dynamics and Control, 101*, 130–144.

Fisher, I. (1896). Appreciation and interest: A study of the influence of monetary appreciation and depreciation on the rate of interest, with applications to the bimetallic controversy and the theory of interest. *Publications of the American Economic Association, 11*(4), 1–98.

Fisher, I. (1911). *The purchasing power of money, its determination and relation to credit, interest and prices*. Macmillan.

Fisher, I. (1930). *The theory of interest, as determined by impatience to spend income and opportunity to invest it*. Macmillan.

Fisher, I. (1933). The debt-deflation theory of great depressions. *Econometrica, 1*(4), 337–357.

Friedman, M. (1956). The quantity theory of money–a restatement. In M. Friedman (Ed.), *Studies in the quantity theory of money* (pp. 3–21). University of Chicago Press.

Friedman, M. (1968). The role of monetary policy. *American Economic Review, 58*(1), 1–19.

Gertler, M., & Karadi, P. (2011). A model of unconventional monetary policy. *Journal of Monetary Economics, 58*(1), 17–34.

Gertler, M., Kiyotaki, N., & Prestipino, A. (2016). Wholesale banking and bank runs in macroeconomic modeling of financial crises. In J. B. Taylor & H. Uhlig (Eds.), *Handbook of macroeconomics* (Vol. 2B, pp. 1345–1425). Elsevier.

Goodfriend, M., & King, R. (1997). The new neoclassical synthesis and the role of monetary policy. In B. S. Bernanke & J. Rotemberg (Eds.), *NBER macroeconomics annual 1997* (Vol. 12, pp. 231–296). Mit Press.

Goodhart, C. A. E., & Tsomocos, D. (2012). The Mayekawa lecture: The role of default in macroeconomics. *Monetary and Economic Studies, 29*, 49–72.

Greenwald, B., & Stiglitz, J. (1987). Keynesian, new classical and new Keynesian economics. *Oxford Economic Papers, 39*(1), 119–133.

Hayek, F. A. (1931). *Prices and production*. Routledge & Sons.

Hicks, J. R. (1937). Mr. Keynes and the classics: A suggested interpretation. *Econometrica, 5*(2), 147–159.

Jakab, Z., & Kumhof, M. (2015). Banks are not intermediaries of loanable funds–why this matters. *Bank of England staff working paper* no. 529. Bank of England.

Jakab, Z., & Kumhof, M. (2018). Banks are not intermediaries of loanable funds–facts, theory and evidence. *Bank of England staff working paper* no. 761. Bank of England.

Keynes, J. M. (1930). *A treatise on money* (Vol. 2 vols). Macmillan.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. Macmillan.

Kiyotaki, N., & Moore, J. (1997). Credit cycles. *Journal of Political Economy, 105*(2), 211–248.

Kydland, F. E., & Prescott, E. C. (1977). Rules rather than discretion. The inconsistency of optimal plans. *Journal of Political Economy, 85*(3), 473–492.

Leijonhufvud, A. (1968). *On Keynesian economics and the economics of Keynes. A study in monetary theory*. Oxford University Press.

Leijonhufvud, A. (1981). The Wicksell connection: Variations on a theme. In A. Leijonhufvud (Ed.), *Information and coordination. Essays in macroeconomic theory* (pp. 131–202). Oxford University Press.

Leijonhufvud, A. (2008). Keynes and the crisis. *CEPR policy insight* no. 23. Centre for Economic Policy Research.

Leijonhufvud, A. (2009). Out of the corridor: Keynes and the crisis. *Cambridge Journal of Economics, 33*(4), 741–757.

Leijonhufvud, A. (2011). *Axel in wonderland: DSGE*. Accessed July 3, 2021, from https://www-ceel.economia.unitn.it/staff/leijonhufvud/files/dsge.pdf

Leijonhufvud, A. (2014). Economics of the crisis and the crisis of economics. *European Journal of the History of Economic Thought, 21*(5), 760–774.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory, 4*(2), 103–124.

Lucas, R. E. (2004). My Keynesian education. In M. De Vroey & K. D. Hoover (Eds.), *The IS-LM model: Its rise, fall, and strange persistence* (pp. 12–24). Duke University Press.

Lux, T., & Westerhoff, F. (2009). Economics crisis. *Nature Physics, 5*(January), 2–3.

Mankiw, G., & Romer, D. (Eds.). (1991). *New Keynesian economics, vol. 1: Imperfect competition and sticky prices; vol. 2: Coordination failures and real rigidities*. MIT Press.

Minsky, H. P. (1975). *John Maynard Keynes*. Macmillan.

Minsky, H. P. (1978). *The financial instability hypothesis: A restatement. Thames papers in political economy*. Autumn. London.

Modigliani, F. (1944). Liquidity preference and the theory of interest and money. *Econometrica, 12*(1), 45–88.

Ramsey, F. (1928). A mathematical theory of saving. *Economic Journal, 38*(152), 543–559.

Robertson, D. (1926). *Banking policy and the price level*. Macmillan.

Ryoo, S. (2010). Long waves and short cycles in a model of endogenous financial fragility. *Journal of Economic Behavior & Organization, 74*(3), 163–186.

Sandelin, B., Trautwein, H.-M., & Wundrak, R. (2014). *A short history of economic thought* (3rd ed.). Routledge.

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review, 97*(3), 586–606.

Stiglitz, J. (1918). Where modern macroeconomics went wrong. *Oxford Review of Economic Policy, 34*(1–2), 70–106.

Tobin, J. (1980). *Asset accumulation and economic activity: Reflections on contemporary economic theory*. University of Chicago Press.

Trautwein, H.-M. (1996). Money, equilibrium and the cycle: Hayek's Wicksellian dichotomy. *History of Political Economy, 28*, 27–55.

Vines, D., & Wills, S. (2018). The financial system and the natural real interest rate: Towards a "new benchmark theory model". *Oxford Review of Economic Policy, 34*(1–2), 252–268.

Wicksell, K. (1898). *Geldzins und Güterpreise. Eine Studie über die den Tauschwert des Geldes bestimmenden Ursachen. Gustav Fischer*; Engl. tr. (1936). *Interest and prices. A study of the causes regulating the value of money*. Macmillan.

Woodford, M. (2003). *Interest and prices. Foundations of a theory of monetary policy*. Princeton University Press.

Woodford, M. (2009). Convergence in macroeconomics: The new synthesis. *American Economic Journal: Macroeconomics, 1*(1), 267–279.

Instability and Structural Dynamics in the Macroeconomy: A Policy Framework

Roberto Scazzieri

1 Introduction

The aim of this chapter is to examine the link between the uneven dynamics of a profit-driven economy and the structural changes inherent to that type of economy. This exploration is complementary to, but distinct from, the strand of research that considers the structural changes that the economic system should follow in order to meet systemic conditions such as full employment and full utilization of productive capacity or maximum feasible expansion.[1] In this chapter, a profit-driven economy subject to dynamic impulses is considered to be open to a *variety of trajectories*, while being constrained by economic actors' visualization of that structure and of its potential for change (Cardinale & Scazzieri, 2019, 2020). The relationship between real and financial investment is a central feature in the long-run dynamics of a profit-driven economy. Moving from the consideration of the profit motive as incentive associated with the discovery and exploitation of price (or cost) differentials, the chapter explores the routes along which those differentials may arise. This point of view suggests looking at the structural dynamic of this type of economy as a process driven by the generation of the price (or cost) differentials that attract investment in

---

[1]Luigi Pasinetti examined the structural changes required on a dynamic path triggered by technical progress and changes in consumer actions under conditions of full employment and full capacity utilization. His analysis is carried out considering the fundamental structural relationships that determine the "natural" dynamic path of the economy independently of its institutional characteristics (Pasinetti, 1981, 1993, 2007; see also Scazzieri, 2012). Alberto Quadrio Curzio investigated the structural transformation path of an economic system identified by access to a given set of production techniques if that system is to grow at its maximum feasible rate under constraints generated by natural or technological scarcities (Quadrio Curzio, 1975, 1986, 1996; see also Scazzieri et al., 2015).

R. Scazzieri (✉)
National Lincei Academy and Department of Economics, University of Bologna, Bologna, Italy
e-mail: roberto.scazzieri@unibo.it

real *or* financial activities. The chapter calls attention to the dual character of that dynamic depending on whether it takes place through the "horizontal" reshuffling of activities in the real economy or the "vertical" shifting of liquidity between real and financial activities. Structural dynamic in the real economy is often associated with technical change and changing consumption patterns. In this case, profit-driven investment may lead to a self-sustaining process of innovation. In other words, the "instability of capitalism" (Schumpeter, 1928), with its waves of creative destruction, may be associated with an "innovation feedback process" (Baumol, 2002) leading to technical improvements in at least some fields of production and some areas of the world economy. On the other hand, structural dynamics as "vertical" liquidity shifts between financial and real activities have a dual character depending on whether liquidity is moved from the financial sphere to the real sphere or vice versa. In the former case, the liquidity injections can support innovation and make industrial revolutions possible (Hicks, 1969, Chap. IX; Neal, 2010). In the latter case, the shift of liquidity away from the real sphere generates an environment in which the profit motive induces short-term capital flows through the purchase and sale of financial assets under conditions of volatile price movements. The search for price differentials in this environment may explain financial innovations triggering increasing uncertainty and allowing increasingly liquid investment (Sen, 2007a, b).

The following section of this chapter highlights the relationship between the profit motive and uneven dynamics and calls attention to the constraints and opportunities for policy-making that arise from that relationship. Section 3 calls attention to the trade-production-finance nexus that can be identified as the mainspring of structural dynamics in a profit-driven economy by considering John Hicks's and Luigi Pasinetti's stylized histories of structural change (Hicks, 1969; Pasinetti, 1965) and combining their analyses with Fernand Braudel's view of structural dynamics as a process taking place through liquidity shifts from one trade or production activity to another *and* through liquidity shifts between the real and the financial spheres of the economy (Braudel, 1977). Section 4 examines policy effectiveness under the distributional changes generated by structural dynamics along a profit-driven path. Section 5 outlines a conceptual framework for policy-making under structural change (*structural policy-making*). This section discusses the conditions for effective policy making in a profit-driven economy by considering in particular the combinations of intended and unintended outcomes characterizing this type of economy under distributional and structural dynamics. Section 6 brings the chapter to close.

2 Uneven Dynamics and Macroeconomic Policy: Prospects and Challenges

The central theme of this volume is the "instability of capitalism," in Joseph Schumpeter's sense of disturbances due to "inherent economic causes" that characterize the historical evolution of capitalist economies (Schumpeter, 1928, p. 361). Schumpeter stressed that the generation of instability in the above sense is what distinguishes capitalism from other types of economic institutions (Schumpeter, 1974, pp. 65–66):

> [i]n so far as the "new combinations" may in time grow out of the old by continuous adjustment in small steps, there is certainly change, possibly growth, but neither a new phenomenon nor development in our sense. In so far as this is not the case, and the new combinations appear discontinuously, then the phenomenon characterising development emerges [...] Development in our sense is then defined by the carrying out of new combinations.

In Schumpeter's view, by "development" one should understand "spontaneous and discontinuous change in the channels of the [circular] flow, disturbance of equilibrium, which forever alters and displaces the equilibrium state previously existing" (Schumpeter, 1974, p. 64). Indeed, development in Schumpeter's sense entails "only such changes in economic life as are not forced upon it from without but arise by its own initiative, *from within*" (ibid., p. 63; emphasis added). In short, it is the *internal generation of changes* "by fits and starts" (ibid., p. 62) what makes "the modern economic system what it is" (ibid.) and distinguishes its dynamics from changes due to "an adaptation to data existing at any time" (ibid.).

As is well known, Schumpeter associated this generating mechanism of changes with innovations (what he called "new combinations") appearing "not [...] evenly distributed through time" but "discontinuously in groups or swarms" (Schumpeter, 1974, p. 223). It is the swarm-like shape of new combinations that allows Schumpeter to interpret the dynamics of a capitalist economy as driven by an internal generating mechanism while avoiding both the view of economic dynamics as *adaptation* to exogenous changes of data and the view of it as a continuous *building up* process of gradual transformation (i.e., as a cumulative process of change). Schumpeter acknowledges the central importance of structural dynamics while at the same time highlighting the need to address it as a process *triggered* by discontinuous, and possibly exogenous, sources of transformation, and *driven* by an endogenous mechanism (the profit motive) characteristic of a capitalist economy. The swarm-like character of innovations highlights the central role of connectivity in this type of economic system. Interdependence between activities is likely to increase the effect of initial perturbations on the whole economic system (see Deane 1980, for historical illustrations of this property in the British first industrialization process). Interdependence leads to different outcomes depending on whether it is primarily associated with connectivity between markets or connectivity between production processes. The diffusion of perturbations through market connectivity is often triggered by speculative actions, whose effects may be transmitted

across markets and may induce upswings and downswings of great amplitude sometimes leading to general crises (Kindleberger, 1978, pp. 47–48). A feature of crises triggered by speculative actions and driven by the connectivity of markets is their one-off character, which makes those perturbations different from crises triggered by the diffusion of innovations (Schumpeter's "new combinations"), and driven by the interdependence of production processes. In the latter case, connectivity generates specific diffusion patterns, which reflect the prevailing types of interdependence between different finished products and/or fabrication stages. This interdependence between production activities may explain the change in the character of economic crises around the third decade of the nineteenth century, when the recurrence of upswings and downswings emerged as a distinctive feature of industrial capitalism (Deane, 1980). Two distinct but ultimately complementary factors were identified, already in the early nineteenth century, as possible explanations for the characteristically uneven dynamics of industrial capitalism. One is the *increasing disconnection* of production from consumption activities, with the associated reliance on the profit motive as the primary signaling device for the allocation of productive capacity between activities. The other is the *mechanization* of production processes, with the increasing utilization of tools and machines made within specialized productive sectors.

Profit motive and mechanization worked hand in hand against the historical background provided by the increasing differentiation of activities and specialization of processes. Two important consequences arise from this combination of factors. On the one hand, the profit opportunities, considered as *a signaling device*, can be inherently biased against an immediate response to mismatches between the demand and supply for goods and services, thereby working as an instability factor. As Jean-Charles-Léonard Sismondi pointed out: "each producer tries to sell at a price lower than that of his colleagues, so that, thanks to the lower price, the buyer would give him preference" (Sismondi 1837–1838, vol. I, p. 74). As a result, each producer "tries to produce more and more, and to produce at a lower price, to produce so much more in order to make up with the quantity [sold] what he loses due to the [lower] price" (ibid.). On the other hand, increasing mechanization leads to further distancing of supply from demand, since the processes making tools and machines, if carried out separately from the processes making final products, may be affected by *lack of synchronization* between their own pattern of supply and the pattern of demand from the potential users of their products.

The combined working of profit motive and increasing mechanization (or, in more recent times, automation) is at the root of the character taken by industrial capitalism around the third decade of the nineteenth century. The distinctive features of its dynamics are the intertwining of medium- and long-term patterns of motion. Medium-term dynamics is characterized by fluctuations of production activity triggered by the recurrent emergence of disproportionalities between circulating and fixed capital (Tugan Baranovsky, 1894; Spiethoff, 1903; Aftalion, 1908–1909, 1913, 1927; Bouniatian, 1908, 1922, 1928; Robertson, 1915; see also Cardinale & Scazzieri, 2017). On the other hand, long-term dynamics shows the recurrence of innovation waves in which, as Schumpeter noted, "new combinations" are

introduced "in groups or swarms" (Schumpeter, 1974, p. 223). The two causal mechanisms may be jointly working, even if their effects have different weights over different time horizons. In any case, the profit motive and increasing specialization are active both in the medium and in the long term. In the medium-term, the failure of the profit motive as stabilizer may be the source of lags leading to increasing disproportionality between productive sectors producing circulating and fixed capital goods respectively.[2] In the long-term, the profit motive may trigger fundamental innovations, while the pattern of interdependence between processes may explain the path and timing of innovation diffusion between different productive sectors. In short, the profit motive and the pattern of interdependence may work hand in hand to explain the uneven dynamics (instability) of industrial capitalism both in the medium term and in the long term. Their joint operation provides a bridge between different time horizons, in the sense that coordination failures in the short or medium term may trigger changes that make themselves felt in the long-term, while a long-term transformation of productive structure may trigger changes in the way disproportionalities drive economic dynamics in the medium term. The intertwining of the profit motive with the dynamics of economic structure both in the medium and in the long term has far-reaching implications for the effectiveness of economic policy. For example, a policy measure targeting macroeconomic aggregates may unintendedly damage certain industrial sectors relative to others, as would be the case with credit conditions making long-term loans more difficult and therefore "patient investment" less likely. On the other hand, a policy measure targeting specific industrial sectors runs the opposite risk of generating unwanted consequences at the systemic level, as would be the case if the speed of technical change in certain sectors were to drastically reduce profit opportunities in less technologically advanced sectors and to eventually generate a contraction at the macroeconomic level. The uneven dynamics of industrial capitalism highlights prospects and challenges for economic policy. In particular, it highlights the need for economic policy to fully grasp the interplay between different time horizons and the intertwining of institutional and technological conditions that determines policy outcomes at any given time and across time.

3   The Mainspring of Structural Dynamics: The Trade-Production-Finance Nexus

Instability and market disruptions are associated with the long-term growth of a profit-driven economy. But is there a systematic linkage between growth, unevenness of dynamic paths, and capitalist institutions? Sir John Hicks addressed this issue

---

[2]This disproportionality is rooted in the hierarchical organization of productive interdependencies in an industrial economy (see, e.g., Masci, 1934, and Lowe, 1976, who discuss in this connection the central role of the machine tool sector; see also Scazzieri, 2018).

in his 1973 Nobel Lecture "The Mainspring of Economic Growth" (Hicks, 1977). There Hicks points out that "[t]he mainspring of economic progress [. . .] is invention; invention that works through the rate of profit. Each invention gives an *Impulse*, as we may call it. But the Impulse of any single invention is not inexhaustible" (ibid., p. 15).[3] In Hicks's view, "an approximation to a steady state is likely to occur, *if there is no further technical change*" (ibid., p. 14; Hicks's emphasis). In this case, wages will be rising, and the rate of profit will decline unless *another invention* lifts up the rate of profit again.

Hicks's argument highlights the relationship between invention and the profit motive, which also provides an analytical link between invention-induced growth and capitalist institutions. This line of reasoning suggests a further look at the profit motive by taking advantage of another line of investigation also pursued by Hicks in his monograph *A Theory of Economic History* (Hicks, 1969). A fundamental thread in Hicks's approach to economic history is the view of the long-run dynamics of market economies in terms of a continuing attempt to obtain economic advantage from the discovery of price differentials (or cost differentials) and their utilization through the trade of goods and services. Hicks describes that situation when considering the intermediation through trade between economic areas that are not otherwise connected with one another: "[T]he merchant is making a profit (in terms of oil) by buying corn at a low price and selling it at a high price; and the trade is unlikely to get started unless, to begin with, it is a handsome profit" (Hicks, 1969, p. 43).

Hicks's view highlights the *interstitial character* of gains from trade as intermediation between "outside areas," and the likelihood of diminishing profits from trade as trade expands without involving new channels of trade (such as new markets or new goods and services). This assessment of gains from trade is akin to Luigi Pasinetti's view of the "phase of trade" in early modern history (Pasinetti, 1965, p. 573):

> The phase of trade is the first to break through [. . .] A few important [. . .] new possibilities of trade open up, with a striking impact on the economic conditions of the whole world. The trading nations are suddenly better off, not because of a rise in world production, but because of a better utilisation of the production which already takes place.

Pasinetti highlights that the switch from a non-trade to a trade situation is a "once-for-all change" (Pasinetti, 1965, p. 574), so that the corresponding advantage, which stems from making "the best use of what one has already" (ibid.), is likely to be "temporary, as it ends when the new equilibrium situation has been reached" (ibid., p. 575). The temporary character of gains from trade explains the "tendency to diminishing returns" that Hicks associates with what we may call *static trade*, that is, trade involving no new actors nor new traded items. However, it is the temporary character of the gains from static trade that explains Hicks's interest in diversification as the fundamental condition for trade expansion in the long term. For diversification

---

[3]Hicks denotes as "double equilibrium path" a path associated with "*both* saving-investment equilibrium and full employment" (Hicks 1977, p. 11, Hicks's emphasis).

of the channels of trade when 'new possibilities of trade open up' (Pasinetti, 1965, p. 573) may allow traders to check diminishing returns from static trade and even to enter a temporary phase of increasing returns due to better organization and lower transaction costs (Hicks, 1969, p. 47).

In short, the switch from static trade to what we may call *dynamic trade* is a necessary condition for gains from trade not to dwindle and for the profit motive to sustain continued trade expansion. This view provides interpretive lenses to explain both the "phase of trade" and the subsequent "phase of industry" in Pasinetti's reconstruction of early modern economic history as stages along a *single sequence* triggered by the same fundamental driving principle. The profit motive may be assigned a central role in explaining the overall dynamic trajectory of a market economy defined as a type of economy whose primary motive force are the actions of those actors (the "merchants") who specialize in *mediating* other actors' needs. In a market economy, the search for profit (as defined above) is relentless, and the search for new price or cost differentials must take place all the time to avoid dwindling profit margins and the onset of diminishing returns from trade. If we follow Hicks's argument, this search explains ancient and modern colonial expansion, the emergence of monetary and credit relationships, the extension of market relationships to include land and labor, and finally the industrial revolution and the emergence of industrial capitalism (Hicks, 1969, Chapters IV, V, VII, VIII, IX). Against this background, the danger of dwindling profit margins and diminishing returns is always looming: "there was a trade network already in existence, in active existence. Like its predecessors, it would be endeavouring to expand; like them, it could only maintain expansion by the continual discovery of new opportunities" (Hicks, 1969, p. 143). In Hicks's narrative, the greater "availability of liquid funds" (ibid., p. 144) in mid-eighteenth-century Britain made possible the "major switch to fixed capital" associated with the transformation from commercial capitalism to industrial capitalism (ibid., p. 148).

Hicks's theory of economic history makes the profit motive (defined as the search for gains arising from price differences or cost differences) into the principal driver of expansion in a market economy, both under commercial capitalism (Pasinetti's "phase of trade") and under industrial capitalism (Pasinetti's "phase of industry"). The profit motive triggers the search for new channels of trade (commercial capitalism) as well as the introduction of less costly production techniques (industrial capitalism). From this point of view, structural dynamics with its inherent instability is a necessary condition for the continuing expansion of economic activities driven by the profit motive. It is the need to avoid diminishing returns from trade that triggers trade diversification, changes in specialization patterns, and technical change. The profit motive explains structural economic dynamics, which in turn allows long-run expansion under market capitalism.

In view of the above interpretive framework, the *industrial* transformation of capitalism (vis à vis the previous economic institutions of commercial capitalism) may have been a solution to the search for satisfactory profit margins. Large-scale capital investment in manufacturing during the First Industrial Revolution took advantage of technical inventions making mechanical production more effective

than previous forms of manufacturing organization. Subsequent industrial revolutions may be interpreted in a similar way as attempts to reorganize manufacturing processes with the aim of thwarting the tendency of profit margins to decline. At the same time, we should not overlook that the tendency to diminishing profit margins is always at work. This means that attempts to avoid declining profits may take place along a *variety* of routes. In this connection, the stylized histories of the unilineal type outlined by Hicks and Pasinetti may be complemented by Fernand Braudel's view of capitalism *as a multilayered structure*, and by Braudel's acknowledgement that, until the nineteenth century, the "great merchant" would never be stuck with a single type of activity and that a distinctive mark of economic success had frequently been the switch from specialization to diversification (Braudel, 1977, Chap. II). Braudel also notes, as an exception, the great merchants' specialization in banking and finance, which is often associated with the shift of business from low to high levels of activity. As a matter of fact, both trajectories (from specialized to flexible or non-specialized, manufacturing, and from unspecialized commerce or manufacturing to specialization in finance or manufacturing) can be seen as instances of the dual route that the avoidance of dwindling profits may take.[4] On the one hand, horizontal specialization, or de-specialization (i.e., the splitting of production activities into specialized tasks, or the integration of activities into more versatile bundles), may trigger market advantages arising from increasing returns *à la* Smith or Babbage.[5] On the other hand, vertical specialization, or de-specialization (i.e., the shift of investment between real and financial activities) may provide a response to declining profits at one or another layer of the market hierarchy. For example, the shift away from manufacturing to finance may be a response to social tension in the production domain, while the opposite shift from finance to manufacturing may be a way to shield investment from the excessive volatility of large-scale loans.

4   Distributional Dynamics and Policy Effectiveness

The above argument suggests that the relationship between profit motive and structural dynamics is a fundamental characteristic of industrial capitalism. The importance of that relationship is enhanced by the fact that historical phases of accelerated growth are generally associated with intense processes of structural transformation, as shown by changes in the relative proportions of productive sectors. A case in point is the dynamics of the share of the British iron industry in

---

[4]This point of view suggests that greater flexibility of manufacturing structures (Bianchi & Labory, 2019) and financial innovation leading to greater liquidity of financial investment may be two sides of the same process driven by the profit motive under conditions of increased volatility.

[5]Charles Babbage added to Smith's advantages (increasing dexterity, saving of time, and greater likelihood of invention) a fourth advantage consisting in the possibility for the "master manufacturer" to "purchase exactly that precise quantity [of the different degrees of skill or of force] which is necessary for each process" (Babbage, 1835, p. 175; see also Scazzieri, 1993).

the British national income, which switches from about 1 or 2 per cent in the 1760s to a value close to 6 per cent in 1805, covering a period of intense expansion in the early decades of the First Industrial Revolution (Deane, 1980, p. 111). In general, variations in the overall rate of growth are closely associated with variations in the growth rate of manufacturing production, and the latter is in turn associated with the growth of certain key sectors, such as the cotton industry in the years following the Napoleonic wars or the railway sectors in the years 1845–1860 (Deane & Cole, 1967,Chap. IX).

The relationship between the profit motive and structural dynamics is akin to a cumulative process: (i) the search for profit margins triggers structural change (through the opening of new channels of trade, technical innovation, institutional and organizational changes reducing transaction costs); (ii) structural change raises profit margins but also starts a diffusion process that may eventually trigger a contraction of profit margins (Hicks's diminishing returns); (iii) to avoid the onset of diminishing returns from investment further structural changes are needed. The cumulative process linking the profit motive with structural dynamics introduces a relationship between structural dynamics and the distribution of income, so that structural dynamics may reflect, in its composition and timing, the institutional and political dynamics of the broader societal setup. In short, we can build on the premises of Hicks's and Pasinetti's stylized histories a conceptual framework for analyzing the instability of a profit-driven economy under conditions of structural change. This framework is built on a multidirectional relationship between: (i) the profit motive as incentive to invest; (ii) the distribution of income and wealth between socioeconomic groups; and (iii) the transformation of profit opportunities as a result of structural dynamics. This triad (profit motive, distribution, structural dynamics) is central to industrial capitalism and is at the root of the opportunities and constraints facing macroeconomic policy in that institutional context. The same triad is at the root of the "narrow path" between wage-led growth and profit-led decline, a path that a capitalist economy needs to follow in order to avoid entering a diminishing returns trajectory.[6]

---

[6]This "narrow path" is emphasized in Shaikh (2019). See also Shaikh (2016) and Shaikh's contribution to this volume. Shaikh develops Richard Goodwin's growth cycle model (Goodwin, 1967), which assumes that only capitalists save and that all savings are invested, so that the rate of capital accumulation equals the rate of profits. This model calls attention to the relationship between the inducement to invest as signaled by the rate of profits and the rate of change of the wage share, defined as "the rate of change of real wages minus the rate of change of productivity" (Shaikh, 2016, p. 642). In this analytical framework, demand stimulus policy may raise real wages above the rate of change of productivity thus raising the wage share to the detriment of the profit share and of the rate of profits with a potentially detrimental effect on the overall level of activity, since "it takes a particular wage share to give the savings rate which will align warranted growth to the natural rate" (ibid., p. 647). The generalization of the model to include workers' savings out of wages (Pasinetti, 1962) would not significantly modify the above result since the rate of profits π compatible with warranted growth at the natural rate n would only reflect the capitalists' propensity to save s_c, that is, $\pi = n / s_c$.

The profit motive-distribution-technical change triad highlights a causal mechanism that may lead to different outcomes depending on which link in that mechanism is strongest. If the link between distribution and the profit motive is most prominent, we are likely to find that an increase in the wage share relative to the profit share may reduce the incentive to invest, thereby bringing into play a factor that may lead to contraction if unchecked. It is for this reason that macroeconomic policy may ultimately be self-defeating if, as a result of expansionary policy, a rising wage share leads to contracting profit margins and eventually to economic stagnation. However, the stagnation trap may be avoided if structural dynamics associated with technical and/or organizational innovation leads to an increase of labor productivity, provided the productivity increase is capable of maintaining a sufficiently high profit motive, and therefore a sufficiently high incentive to invest.

Structural dynamics leading to rising labor productivity is a central feature of capitalist economies, and we may ask under which conditions increases of labor productivity are capable of offsetting the tendency of the wage share to increase, and of the profit margins to shrink, without at the same time triggering economic contraction by other means. We cannot exclude that a stimulus policy leading to an increasing wage share could trigger substitution of machines for labor, which may limit further increases in the wage share and sustain profitability for some time, at the cost of technological unemployment and eventually contraction in levels of activity. In short, the profit motive-distribution-technical change triad suggests that, to maintain a capitalist economy on a growth path associated with a satisfactory level of employment, increases of labor productivity should be achieved without significant displacement of labor by machines, so that profitability is maintained without inducing technological unemployment and the contraction in activity levels that may be associated with it. This argument entails that the relationship between profit motive and structural dynamics may or may not be compatible with economic expansion at full or satisfactory employment depending on the specific character taken by structural dynamics as means to maintaining satisfactory profit margins.

The above argument entails that the causal mechanism embedded in the profit motive-distribution-technical change triad is unlikely to be permanently successful in achieving the dual purpose of maintaining satisfactory profit margins and levels of employment. This makes one wonder whether *other types of policy intervention*, different from standard stimulus measures, could be considered as complements to Keynesian-type aggregate demand management tools. To identify and successfully follow the knife-edge path between wage-led growth and profit-led decline, a differentiated set of mutually compatible policy actions would be required. These actions should aim at the following triplet of objectives: (i) a sufficiently high increase of labor productivity; (ii) a sufficiently high rate of profit leading to a sufficiently high private incentive to invest in real activities; and (iii) a sufficiently high level of effective demand capable of offsetting at systemic level the technological unemployment that may be associated with policy measures leading to a high rate of increase of labor productivity.

To simultaneously achieve the above set of objectives requires moving beyond conventional macroeconomic policy. One important reason is the *internal*

differentiation of productive structure under conditions of "real" (as distinct from "perfect") competition. In this connection, it may be interesting to recall Anwar Shaikh's distinction between "capital-intensive regulating [best practice] capitals" and "non-regulating [non best-practice] capitals," considering that, in Shaikh's view, "capital-intensive [best practice] capitals will tend to have higher wage rates for any given strength of labor organizations" (Shaikh, 2016, p. 751). As a matter of fact, these industries "are more able to tolerate wage increases", since in their case "labor costs are likely to be a smaller portion of their total costs" (ibid.). On the other hand, "Non [best practice] capitals are generally more vulnerable because their prices are determined by the prices of production of the [best practice] capitals, so that increases in their labor cost have more serious effects on their profitability. Therefore, their wage rates will tend to be lower" (ibid.). This internal differentiation of productive structure entails that policy measures adequate to achieve the triplet of objectives mentioned above are likely to be different depending on the types of "capitals" we are considering, which means that different industrial structures are likely to require different combinations of policy tools. It is to be expected that measures inducing a "virtuous" increase of labor productivity and promoting adequate investment would be different from one industry to another. For example, in certain activities, profitability may be guaranteed by a provision of public goods through infrastructural investment in a way that would not be possible in the case of other activities. And the levels and/or compositions of effective demand capable of offsetting technological unemployment may be different depending on whether this objective is achieved via export-led growth or internal market growth.

To conclude, profitability as a limit to active macroeconomic policy is likely to manifest itself in different ways as soon as we move beyond a purely macroeconomic view of the economic system. For example, intermediate levels of aggregation (say, aggregation of productive activities by industries or by vertically integrated sectors) highlight different channels by which any given stimulus policy may exert effects, and therefore different channels along which a "wage-led" growth can induce a "profit-led" decline (see Cardinale, 2018, for a discussion of the different policy routes associated with productive interdependence in terms of interindustry linkages or with productive structure as a collection of vertically integrated sectors). This means that, depending on which transmission channel is most effective in a given context, different policy measures may be required to make the economy to follow the "narrow path" between wage-led growth and profit-led decline. Shaikh's approach revisits the tension in classical economic theory between the profit incentive triggering investment and growth on the one hand, and the employment objective that stimulus policy may try to achieve. This tension, which was already manifest in the Malthus-Ricardo discussions (see Hagemann, 1998), cannot be overcome by conventional stimulus policy. The "narrow path" criterion calls attention to the structural and contextual conditions that stimulus policy must satisfy to be effective.

5 Towards a Structural Policy Framework

The above argument highlights the triad of mutually responsive factors whose interaction shapes the policy domain in a profit-driven economic system. As we have seen, that triad consists of (i) the profit motive; (ii) the distribution of income (and, ultimately, of wealth); and (iii) structural dynamics. We have seen in Sect. 2 that the profit motive may trigger structural dynamics along a plurality of routes depending on historical conditions and institutional constraints.

Contexts in which technological dynamics allow the restructuring of manufacturing forms of production organization may trigger fundamental changes in the "horizontal" pattern of division of labor between production activities (either in the specialization or de-specialization direction). These technological opportunities may go hand in hand with a downward liquidity shift along the vertical route, as when liquid funds originating from the eighteenth-century Financial Revolution in Britain (Neal, 2010) moved into manufacturing activity making large-scale investment in fixed capital possible or when liquid funds generated by economic crises allow liquidity injections into innovative activities (Janeway, 2012). On the other hand, contexts in which technological dynamics is faltering, or in any case insufficient to curb diminishing returns from investment in real activities, may trigger an upward liquidity shift along the vertical direction, as when liquid funds are removed from production and invested in financial assets whose returns are unrelated to production organization in the material sphere.

Upward or downward liquidity shifts along the market hierarchy are associated with changing gains from buying and selling real or financial activities. In this light, the upward shift (in Braudel's sense) inherent to financialization enhances the role of *inequality of positions* as triggers of economic advantage and makes profits virtually undistinguishable from differential rents. This context makes *distribution of ownership* a central factor in explaining income distribution and enhances the intertemporal connection between distribution of wealth and distribution of income (see also Baranzini, 1991, especially Chap. 7). In this case, savings allowing long-run equilibrium growth can be provided by workers and capitalists-rentiers (defined as a socioeconomic group whose income consists of returns on savings, and rents), but the composition of rentiers' income between rent and returns on accumulated savings remains indeterminate (Baranzini & Scazzieri, 1997, pp. 130–131). It is noteworthy that this indeterminacy is consistent with the coexistence of financial growth and real stagnation which characterizes several contemporary economies (Sen, 2007a, 2007b).

The upward shift of liquid funds along the market hierarchy (shift from real to financial activities) is almost unavoidably associated with increasing volatility of the economy. This is due to the fact, which John Maynard Keynes clearly identified in Chap. 12 of the *General Theory*, that:

> as the organisation of investment markets improves, the risk of the predominance of speculation does [. . .] increase. [. . .] Speculators may do no harm as bubbles on a steady stream of enterprise. But the position is serious when enterprise becomes the bubble on a

> whirlpool of speculation [...] These tendencies are a scarcely avoidable outcome of our having successfully organised liquid investment markets (Keynes, 1970, pp. 158–159).

The link between liquid investment and speculation has been investigated by Sunanda Sen who pointed out that "[p]rofits cannot be made on buying-selling unless prices are subject to volatile movements" (Sen, 2007a, p. 69). At the same time, "[w]ith uncertainty extorting a heavy toll on the future, the investor may prefer to move away from long-term productive assets to the short-term liquid types which are primarily speculative" (Sen, 2007b, p. 115). In this scenario, financial innovation may have "a contractionary effect on the real economy" due to the fact that "rising uncertainty in the financial market [...] draws away finance from the real economy" (ibid., p. 116). An important feature of this scenario is that:

> [w]ith free flow of financial assets providing opportunities for speculation, a large part of transactions in the market were no longer subject to national jurisdictions. Classic examples include the expanse of shadow banking which has overtaken, in a large number of countries, the pace of credit flows under the surveillance of the central banks (Sen & Marcuzzo, 2018, p. 3).[7]

This scenario highlights opportunities and constraints for economic policy. The interdependence between profit motive and structural dynamics points to structural changes as principal regulator of a capitalist economy. In this light, the way policy actions affect the direction and speed of structural dynamics is a principal factor determining policy effectiveness in the macroeconomy. However, the interdependence between structural change and the profit motive highlights a causal mechanism that may shift policy outcomes along one or another direction depending on context. For example, industrial policies aimed at avoiding low-growth situations by triggering structural change along a high-technology trajectory may backfire by causing technological unemployment. The narrow path that structural policy should follow to avoid both dwindling profits and declining employment would involve a combination of demand management and targeted support of technical innovation in key sectors. A possible route to achieve this outcome in a profit-driven economy may be to drop a purely macroeconomic approach to stimulus policy and to adopt a policy mix including innovation measures effective at sectoral level (see Amsden, 1997; Sainsbury, 2020). In this case, wages may rise in certain sectors, but the wage share may not necessarily rise in the macroeconomy, and the profit motive may be compatible with a less extensive substitution of capital equipment for direct labor in production processes. For example, policy measures enhancing labor capabilities and the infrastructural endowment of the economy may raise average productivity by other means than substituting machines for labor in specific sectors.

The interaction between profit motive, distribution, and structural dynamics may also highlight other avenues for structural policy-making. For example, credit policy may influence the incentive to invest and the direction of investment by intervening in the vertical division of labor between different layers of the market hierarchy

---

[7] An interesting case is that of the so-called family offices, which are essentially unregulated and managed around $5.9tn in 2019 (Martin et al., 2021, p. 6).

(Braudel, 1977). In this case, a tight credit policy associated with high interest rates may trigger a shift of liquid funds from the bottom to the top layers of that hierarchy and most characteristically a shift of liquidity to "higher level" financial markets (financial markets increasingly removed from the material sphere). Those markets may sustain debt-credit relationships founded on large differences between borrowing costs across different actors and therefore open the way to huge gains (as well as potential losses) for some of those actors. On the other hand, a loose credit policy associated with low interest rates may in principle attract liquid funds from top to bottom layers of the market hierarchy and characteristically trigger liquidity shifts from "higher level" financial markets to financial markets closer to the material sphere. Here, lower borrowing costs could trigger investment in productive processes requiring considerable immobilizations in infrastructure and fixed capital items. In this situation, credit policy may influence the vertical division of labor between market layers and in particular the distribution of liquidity between real and financial assets, without directly affecting the horizontal division of labor in the production sphere. However, the ultimate effect of that shift remains open ended. This is because an expansionary credit policy may shift liquidity away from the top market layers (say, away from higher level financial markets) without necessarily triggering investment and expansion at the intermediate and bottom layers at which production activity takes place. Availability of cheap liquid funds for the material sphere does not necessarily translate into actual investment in that sphere. In fact, greater liquidity may go hand in hand with lower investment, especially if profit margins in productive activities are not significantly affected by the lower cost of loans. Lower interest rates may not necessarily shift liquidity from top to bottom market layers due to sluggish structural dynamics unable to sufficiently lift profit margins in spite of the stimulus coming from the credit sphere. In this case too, policy actions may have unintended consequences due to the complex network of interdependencies within which policy outcomes unfold.

The structural point of view suggests the consideration of policies as actions taking place within a web of connections that may sometimes thwart policies away from intended objectives. Under these conditions, policies may ultimately trigger dynamics very different from their stated goals, if not fully opposed to them. A lesson to be drawn from structural analysis is that policy implementation is fundamentally shaped by the context in which policy actions are taken. Structural dynamics bring to light constraints on the feasibility and likelihood of policy outcomes, as well as policy opportunities arising from changes along one dimension or another of economic structure. This pinpoints the need of coordinating different policy domains in order to increase the likelihood of attaining, in each domain, an outcome as close as possible to the intended one.

6 Conclusion

This chapter has addressed the instability of a profit-driven economy from the point of view of the interaction between the profit motive, changes in trade or production structures, and liquidity shifts between the production and the financial sphere. In particular, the chapter has examined structural dynamics as a process that may take place through changes in trade and/or production arrangements (Hicks, Pasinetti) or through "vertical" shifts of liquidity between different layers of the market hierarchy (Braudel). This dual character of structural dynamics adds flexibility to the search for profit margins, which may at certain times privilege changes in the relative proportions between production or trade activities, and/or in the patterns of specialization between them, and at other times follow the route of liquidity shifts from the real to the financial sphere, or vice versa. This approach highlights the dual role of finance, which may alternatively trigger "patient" long-term investment in the real sphere or generate instability under conditions of increased financialization and market volatility. This open-endedness of structural dynamics has important implication for economic policy. The mutual dependence between profits, distribution, and structural dynamics emphasizes a potential trade-off between incentive to invest and economic expansion that only structural dynamics can overcome. In this light, policy actions have both intended and unintended outcomes. For example, industrial policies targeting the production sphere may sometimes lead to macroeconomic contraction and lower aggregate employment. On the other hand, taxation policies targeting the distribution sphere may lead to declining profit margins in the production sphere and to liquidity shifts from lower to upper levels of the market hierarchy (investment shifts from real to financial assets). The open-endedness of structural dynamics translates into the open-endedness of economic policy under structural change. This makes understanding structural dynamics a fundamental step in alerting policy makers to the manifold routes that policy transmission can take and to the most likely outcomes in different contexts.

References

Aftalion, A. (1908–1909). La réalité des surproduction générales: Essai d'une théorie des crises générales et périodiques. *Revue d'économie politique, 32*(1908), 696–706; *23*(1909), 81–117; 201–229; 241–259.

Aftalion, A. (1913). *Les crises périodiques de surproduction*. Rivière.

Aftalion, A. (1927). The theory of economic cycles based on the capitalistic technique of production. *The Review of Economics and Statistics, 9*(4), 165–170.

Amsden, A. (1997). Bringing production back in: Understanding Government's role in late industrialization. *World Development, 25*(4), 469–480.

Babbage, C. (1835). *On the economy of machinery and manufactures* (2nd ed.). C. Knight.

Baranzini, M. (1991). *A theory of wealth distribution and accumulation*. Clarendon Press.

Baranzini, M., & Scazzieri, R. (1997). Profit and rent in a model of capital accumulation and structural dynamics. In P. Arestis, G. Palma, & M. Sawyer (Eds.), *Capital controversy, post-Keynesian economics and the history of economics. Essays in honour of Geoff Harcourt* (Vol. I, pp. 121–132). Routledge.

Baumol, W. J. (2002). *The free-market innovation machine. Analyzing the growth miracle of capitalism*. Princeton University Press.

Bianchi, P., & Labory, S. (2019). Manufacturing regimes and transitional paths: Lessons for industrial policy. *Structural Change and Economic Dynamics, 48*(March), 24–31.

Bouniatian, M. (1908). *Studien zur Theorie und Geschichte der Wirtschaftskrisen*. Reinhardt.

Bouniatian, M. (1922). *Les crises économiques*. Rivière.

Bouniatian, M. (1928). The theory of economic cycles based on the tendency toward excessive capitalization. *The Review of Economics and Statistics, 10*(2), 67–79.

Braudel, F. (1977). *Afterthoughts on material civilization and capitalism*. Johns Hopkins University Press.

Cardinale, I. (2018). A bridge over troubled water: A structural political economy of vertical integration. *Structural Change and Economic Dynamics, 46*, 172–179.

Cardinale, I., & Scazzieri, R. (2017). Beyond "austerity vs. expansion": Elements for a structural theory of liquidity policy. *Revue d'économie industrielle, 160*(4), 107–136.

Cardinale, I., & Scazzieri, R. (2019). Explaining structural change: Actions and transformations. *Structural Change and Economic Dynamics, 51*, 393–404.

Cardinale, I., & Scazzieri, R. (2020). Interdipendenze produttive, interessi e condizioni sistemiche: elementi per un'economia politica delle strutture industriali. *L'Industria. Review of Industrial Economics and Policy, 41*(1), 21–50.

Deane, P. (1980). *The first industrial revolution* (2nd ed.). Cambridge University Press.

Deane, P., & Cole, W. A. (1967). *British economic growth, 1688–1959. Trends and structure* (2nd ed.). Cambridge University Press.

Goodwin, R. M. (1967). A growth cycle. In C. H. Feinstein (Ed.), *Socialism, capitalism and economic growth* (pp. 54–58). Cambridge University Press.

Hagemann, H. (1998). General glut controversy. In H. Hagemann & N. Salvadori (Eds.), *The Elgar companion to classical economics* (pp. 320–324). Edward Elgar.

Hicks, J. R. (1969). *A theory of economic history*. Clarendon Press.

Hicks, J.R. ([1973] 1977). The mainspring of economic growth. In J. Hicks, *Economic perspectives. Further essays on money and growth* (pp. 1–19). Clarendon Press.

Janeway, W. H. (2012). *Doing capitalism in the innovation economy*. Cambridge University Press.

Keynes, J.M. ([1936] 1970). *The general theory of employment, interest and money*. Macmillan and St. Martin's Press.

Kindleberger, C. P. (1978). *Manias, panics and crashes*. Macmillan.

Lowe, A. (1976). *The path of economic growth*. Cambridge University Press.

Martin, K., Wigglesworth, R., & Fletcher, L. (2021). "They can do what they want": Archegos and the $6tn world of the family office. *Financial Times Weekend, 2*(April), 6.

Masci, G. (1934). Alcuni aspetti odierni dell'organizzazione e delle trasformazioni industriali. In G. Masci (Ed.), *Organizzazione industriale* (pp. 929–958). UTET.

Neal, L. (2010). *The rise of financial capitalism. International capital markets in the age of reason*. Cambridge University Press.

Pasinetti, L. L. (1962). Rate of profit and income distribution in relation to the rate of economic growth. *Review of Economic Studies, 29*(4), 267–279.

Pasinetti, L. L. (1965). A new theoretical approach to the problems of economic growth. In *Econometric approach to development planning* (pp. 571–687). North Holland.

Pasinetti, L. L. (1981). *Structural change and economics growth. A theoretical essay on the dynamics of the wealth of nations*. Cambridge University Press.

Pasinetti, L. L. (1993). *Structural economic dynamics. A theory of the economic consequences of human learning*. Cambridge University Press.

Pasinetti, L. L. (2007). *Keynes and the Cambridge Keynesians: A "revolution in economics" to be accomplished*. Cambridge University Press.

Quadrio Curzio, A. (1975). *Accumulazione del capitale e rendita*. Il Mulino.

Quadrio Curzio, A. (1986). Technological scarcity: An essay on production and structural change. In M. Baranzini & R. Scazzieri (Eds.), *Foundations of economics. Structures of inquiry and economic theory* (pp. 311–338). Basil Blackwell.

Quadrio-Curzio, A. (1996). Production and efficiency with global technologies. In M. Landesmann & R. Scazzieri (Eds.), *Production and economic dynamics* (pp. 105–139). Cambridge University Press.

Robertson, D. H. (1915). *A study of industrial fluctuation: An enquiry into the character and causes of the so-called cyclical movements of trade*. P.S. King.

Sainsbury, D. (2020). *Windows of opportunity. How nations create wealth*. Profile Books.

Scazzieri, R. (1993). *A theory of production. Tasks, processes, and technical practices*. Clarendon Press.

Scazzieri, R. (2012). The concept of "natural economic system": A tool for structural analysis and an instrument for policy design. In R. Arena & P. L. Porta (Eds.), *Structural dynamics and economic growth* (pp. 218–240). Cambridge University Press.

Scazzieri, R. (2018). Innovation and production dynamics: The producibility-scarcity paradox. *Balzan Papers, 1*, 201–207.

Scazzieri, R., Baranzini, M., & Rotondi, C. (2015). Resources, scarcities and rents: Technological interdependence and the dynamics of socio-economic structures. In M. Baranzini, C. Rotondi, & R. Scazzieri (Eds.), *Resources, production, and structural dynamics* (pp. 427–484). Cambridge University Press.

Schumpeter, J. A. (1928). The instability of capitalism. *Economic Journal, 38*(151), 361–386.

Schumpeter, J.A. ([1934] 1974). *The theory of economic development. An inquiry into profits, capital, credit, interest, and the business cycle*. Oxford University Press.

Sen, S. (2007a). *Globalisation and development*. National Book Trust.

Sen, S. (2007b). Global finance. On the move to stock market capitalism. In S. Sen (Ed.), *Dominant finance and stagnant economies* (pp. 113–128). Oxford University Press.

Sen, S., & Marcuzzo, M. C. (2018). Introduction. In S. Sen & M. C. Marcuzzo (Eds.), *The changing face of imperialism. Colonialism to contemporary capitalism* (pp. 1–11). Routledge.

Shaikh, A. (2016). *Capitalism. Competition, conflict, crises*. Oxford University Press.

Shaikh, A. (2019). Profitability, stimulus policy, and finance. *In this volume*.

Sismondi, J.C.L., Simonde de (1837–1838). *Études sur l'économie politique*. Treuttel et Wurtz.

Spiethoff, A. (1903). Die Krisentheorie von M.I. Tugan-Baranowsky und L. Pohle. *Jahrbuch für Gesetzgebung, Verwaltung and Volkswirtschaft, 27*(2), 331–360.

Tugan Baranovsky, M. I. (1894). *Promyshlennye Krizisy v sovremennoj Anglii. St.* I. N. Skorochodova.