Akio Matsumoto · Ferenc Szidarovszky
Toichiro Asada *Editors*

Essays in Economic Dynamics

Theory, Simulation Analysis, and Methodological Study

Institute of Economic Research, Chuo University
IERCU
中央大学経済研究所
1964

Springer

Essays in Economic Dynamics

Akio Matsumoto · Ferenc Szidarovszky
Toichiro Asada
Editors

Essays in Economic Dynamics

Theory, Simulation Analysis, and Methodological Study

Institute of Economic Research, Chuo University
IERCU
中央大学経済研究所
1964

Springer

Editors
Akio Matsumoto
Department of Economics
Chuo University
Hachioji, Tokyo
Japan

Toichiro Asada
Faculty of Economics
Chuo University
Hachioji, Tokyo
Japan

Ferenc Szidarovszky
University of Pecs
Pécs
Hungary

ISBN 978-981-10-1520-5 ISBN 978-981-10-1521-2 (eBook)
DOI 10.1007/978-981-10-1521-2

Library of Congress Control Number: 2016943858

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media Singapore Pte Ltd.

Preface

It is the 9th International Conference on Nonlinear Economic Dynamics (NED2015) that was held in Tokyo, Japan, in June of 2015. The first NED conference started in Odense, Denmark in the year of 2002, which was organized by a small group of scholars having strong desires and interests of applying nonlinear dynamic methods to reveal "new faces" of the traditional dynamic studies of economics and finance. Since then, the NED conference continued in Odense again (Denmark, 2003), Tokyo (Japan, 2004), Urbino (Italy, 2005), Bielefeld (Germany, 2007), Jönköping (Sweden, 2009), Cartagena (Spain, 2011) and Siena (Italy, 2013). Meanwhile at the Jönköping conference, the Nonlinear Economic Dynamic Society was officially founded and has supported the NED conferences since then. It was already announced that the next NED conference will be held at James Madison University in Harrisonburg, USA, 2017.

Following the tradition of the series of conferences, NED2015 aims at bringing together the young and senior researchers who are interested in pursuing research in economic dynamics in a broader sense. The conference was held at Chuo University in Tokyo during June 25–27, 2015 and attracted participants from Australia, Austria, China, Germany, Hungary, Italy, the Netherlands, Singapore, Slovakia, Sweden, Ukraine, the US, and Japan. The talks concerned the recent results of the participants ranging from the pure theory of nonlinear economic dynamics to its applications and practices in various fields of optimization, game theory, finance, regional science, behavioral economics, evolutionary economics, and so forth.

This book is edited by Akio Matsumoto, Ferenc Szidarovszky and Toichoro Asada and consists of three parts including selected 14 papers contributed to NED2015, each of which is refereed and revised. Part I considers the methodological and philosophical implications of the nonlinear dynamics to economics, Part II presents nonlinear models of microeconomic dynamics and discusses the techniques for analyzing the actual economic data, and Part III presents nonlinear models of macroeconomic dynamics. The papers in the book consider economic dynamics from a wide variety of perspective ranging from the monopoly and duopoly in microeconomics to the traditional Keyesian, Kaldorian, and Kaleckian models in macroeconomics. Some papers deepen understandings of the effect caused by delays that inevitably occur in actual economic activities in a real economy while some other papers consider the policy implications in nonlinear dynamic framework.

The conference was financed through various supports: Graduate School of Economics of Chuo University with the MEXT-supported Program for the Strategic Research Foundation at Private University 2013–2017, the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (A) 26242028 and (C) 24530202, 25380238, 26380316) and Chuo University including Institute of Economics Research, International Center and Joint Research Grants. For the preparation of the conference, the Scientific Committee and the Local Organization Committee have been organized. The efforts provided by the members toward the conference are much appreciated. Very special thanks are due to Ayako Kodama, the Secretary of the LOC, without whose efforts the conference would not have been possible. Additional thanks are due to Masato Nakao and Takayuki Mizuno, assistants of the LOC. Finally, needless to say, not least, big thanks are also due to all the participants of the conference. We really wish this book would contribute to the development of nonlinear economic dynamics and its applications.

Tokyo
June 2016

Akio Matsumoto
Ferenc Szidarovszky
Toichiro Asada

Contents

Part III Macroeconomic Models

Part I
Methodological Issue

Simonian Bounded Rationality and Complex Behavioral Economics

J. Barkley Rosser Jr. and Marina V. Rosser

Abstract This chapter will consider the importance of Herbert A. Simon as both the discoverer of the idea of *bounded rationality* and its role in modern behavioral economics and as one of the early developers of complexity theory, especially its hierarchical and computational forms. Bounded rationality was essentially derived from Simon's view of the impossibility of full rationality on the part of economic agents. Modern complexity theory through such approaches as agent-based modeling offers an approach to understanding behavioral economics by allowing for specific behavioral responses to be assigned to agents who interact within this context, even without full rationality. Other parts of modern complexity theory will also be considered in terms of their relationships with behavioral economics. Fundamentally, complexity provides an ultimate foundation for bounded rationality and hence the need to use behavioral economics.

Keywords Bounded rationality · Complexity · Behavioral economics

1 Introduction

It was Herbert A. Simon who developed the idea of *bounded rationality* from his earliest works (Simon 1947, 1955a, 1957), the basis of modern *behavioral economics*. Behavioral economics contrasts with more conventional economics in not assuming full information rationality (or *substantive rationality* in the words of Simon) on the part of economic agents in their behavior, although they may still be capable of a reasonably self-interested *procedural rationality*. In this regard, it draws on insights regarding human behavior from other social science disciplines such as psychology

J.B. Rosser Jr. (✉) · M.V. Rosser
Department of Economics, James Madison University, Harrisonburg, VA 22807, USA
e-mail: rosserjb@jmu.edu

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_1

and sociology, among others, and without question many earlier economists[1] argued that people are motivated by more than mere fully rational selfish maximization. However, it was Simon who crystallized this with his formulation of bounded rationality and behavioral economics.

Aristotle put economic considerations into a context of moral philosophy and proper conduct while also posing a foundational complexity idea of how the whole may be greater than the sum of its parts. The father of political economy, Adam Smith in his *Theory of Moral Sentiments* (1759) saw people deeply influenced by their consciences and obeying social norms while responding to their sympathy for those near to them, even as he also more famously suggested the emergence of market order from interacting individuals pursuing their private economic interest, another fundamental complexity insight. Likewise, institutional economists such as Thorstein Veblen (1898) and Karl Polanyi (1944) saw peoples' economic conduct as embedded within broader social and political contexts. Veblen in particular placed his discussion in the context of establishing an evolutionary economics that would also allow for emergence of new technologies and social orders, also important complexity ideas.

Simon's initiatives led to research over the next few decades, which became influential in business schools and management programs as the rational expectations revolution conquered most of economics during the 1970s and 1980s. Assuming bounded rationality by economic agents led him to the concept of *satisficing*, that while people do not fully maximize their rational self-interest they strive to achieve set goals within constraints. This became accepted in business schools as managers were taught to achieve levels of profit acceptable to owners.

Also arising out of his discovery of bounded rationality was his interest in pursuing more deeply how people think and understand as part of making decisions. This led him to consider how this could be studied through understanding computers, which led him to help found the field of *artificial intelligence* (Simon 1969). While Simon is regarded as one of the more general early leaders of computer science, it was his thinking about the implications of bounded rationality that led him into this nascent field and into artificial intelligence in particular.

Simon also became a leading figure in developing early complexity theory, both its hierarchical complexity version (Simon 1962) with its implications for evolution and emergence, its computational complexity form, and also studying how to estimate power law distributions (Simon 1955b), which he would apply to various phenom-

[1] It must be noted that while Simon received the Swedish Bank Prize in Economic Science in Memory of Alfred Nobel in 1978, usually simply called the "Nobel Prize in Economics," he was not officially an economist in any way. His PhD from the University of Chicago was in political science, and he never was in an economics department during his academic career. At his death in 2001, he was in four different departments at Carnegie Mellon University, where he had been based since 1949 when it was still the Carnegie Institute of Technology: computer science, psychology, cognitive science, and management, and he had earlier been in the philosophy department as well. The first of these authors remembers well from personal communication with him how much Simon disdained conventional economics, and a number of prominent economists expressed public displeasure when he received his prize in 1978.

ena later. Modern complexity theorists see a direct link between complexity of one sort or another and bounded rationality, and thus also with behavioral economics. Complexity can be seen as a foundation for why people have bounded rationality and thus of *complex behavioral economics*.

2 Forms of Complexity

A discussion regarding the relationship between "complexity" and something else clearly requires some discussion of what is meant by this term, or at least what we mean by it. Indeed, this is arguably a weasel term, one that has no clearly agreed-on meaning more generally. The MIT engineer, Seth Lloyd, some time ago famously gathered a list of various different meanings, and this list was at least 45 before he stopped bothering with this effort, or at least making it publicly known Horgan (1997, p. 303). It may be useful therefore to refer to the broadest possible view of complexity that includes all of these and any others as being *meta-complexity*. The definition of this may simply amount to listing all possible meanings that any have ever claimed should be on the list.

If one seeks general definitions or concepts, something often appears in such general definitions is the idea that somehow something that is complex involves a whole that is "greater than the sum of its parts," as the old cliché puts it. Such an idea can be traced as far back as Aristotle, with many since contributing to it. We shall see below that not all the items on Seth Lloyd's list might agree with this, particularly the many that relate to *computational complexity*, arguably the subcategory of complexity with more variations than any other. That those concerned with this subcategory might not have such a view might explain why John von Neumann (1966) did not distinguish complexity from mere *complicatedness*. While some may not wish to make this distinction, many do, with Israel (2005) noting that the two words come from different roots in Latin, *complecti* and *complicare* respectively, the former meaning "to enfold" and the latter "to entangle." Thus, while close and possibly from an identical deeper origin, the former implies some completing in a higher order whereas the latter implies more simply "to confuse" due to the bringing together of many different elements.

In any case, perusing Lloyd's list allows one to lump many of his definitions into higher order subcategories. Arguably the subcategory with the most items on it can be considered forms of *computational complexity*, with at least as many as 15 of them fitting in this category, possibly more.[2] If there is a linking concept through this set of definitions, it involves ideas of size or length, how long a program is or how many distinct units there are within the object such as bits of information. However, the many variations on this do not map onto each other readily. Nevertheless, many of these definitions have the virtue of being clearly measurable, even if there are many

[2]These would include at a minimum those that use the words "algorithm," "information," or "code length."

such definitions. Thus, if one gloms onto one of these, one can argue that it may have a stronger claim to being "scientific" due to this specific clarity than some other fuzzier alternatives. Interestingly, among those fuzzier alternatives listed by Lloyd is the *hierarchical complexity* concept introduced by Herbert Simon (1962), which is relevant to several disciplines.

Within economics and arguably several other disciplines the strongest rival to the varieties of computational complexity can be called *dynamic complexity*, although no item called precisely this appears on Lloyd's list, with perhaps the closest being "self-organization" and "complex adaptive systems." More precisely, Day (1994) defined (dynamic) complexity as arising in nonlinear dynamical systems that due to endogenous causes do not asymptotically approach a point, a non-oscillating growth or decline, or two-period oscillation. Thus such a system will exhibit some form of erratic dynamic behavior arising endogenously from within itself, not due to an erratic exogenous driver. Rosser (1999) adopted this definition for his "broad-tent" complexity that is clearly dynamic.[3]

Within this broad-tent form of dynamic complexity one can observe four well-known subcategories that were identified as being "the four Cs" of *chaoplexity*, according to Horgan (1997, Chap. 11). These were cybernetics, catastrophe theory, chaos, and "small-tent" or agent-based or Santa Fe complexity. Horgan argued that these have all constituted a succession of intellectual fads or bubbles, beginning in the 1950s with Norbert Wiener's cybernetics and moving on successively, with agent-based complexity simply the latest in this succession that was overhyped and then discarded after being shown to be overhyped. However, an alternative view is that these represent an accumulating development of knowledge regarding the nature of nonlinear dynamics, and that students of this development should take Horgan's ridicule and turn it on its head, much as such art movements as Impressionism were originally named critically, only to have them become widely admired. Let the "four Cs" be the focus of a successful ongoing intellectual system.

Norbert Wiener (1948) introduced *cybernetics*, which strongly emphasizes the role of positive and negative feedback mechanisms. Wiener emphasized issues of control, which made cybernetics popular in the Soviet Union and other socialist planned economies long after it had faded from attention in western economies. While Wiener did not emphasize nonlinear dynamics so much, certain close relatives of cybernetics, *general systems theory* (van Bertalanffy 1968) and *systems dynamics* (Forrester 1961) did so more clearly, with Forrester particularly emphasizing how nonlinearities in dynamical systems can lead to surprising and "counterintuitive" results. However, the discrediting of cybernetics and its relatives may have come most strongly from the failure of the limits to growth models based on systems dynamics when they forecast disasters that did not happen (Meadows et al. 1972). Much of the criticism of the cybernetics approaches, which emphasized computer simulations,

[3]Velupillai (2011, p. 553) has referred to this form of complexity as "Day-Rosser complexity," even as he strongly advocates the use of more computationally based forms of complexity as being more useful and scientific. For a fuller presentation of Velupillai's perspective on computational complexity, see Velupillai (2000).

focused on the excessive levels of aggregation in the models, something that more recent agent-based models are not guilty of, with these arguably representing a new improved revival of the older cybernetics tradition.

Catastrophe theory developed out of broader bifurcation theory, and to the extent that formal catastrophe theory may not be applicable in many situations due to the strong assumptions required for it to be applied, broader bifurcation theory can analyze the same fundamental phenomenon, that of smoothly changing underlying control variables having critical values where values of endogenous state variables may change discontinuously. Formal catastrophe theory, based on Thom (1975), provides generic forms for these bifurcation conditions on equilibrium manifolds according to the number of control and state variables, and Zeeman (1974) provided the first application in economics to the analysis of stock market crashes using the cusp catastrophe model that has two control variables and one state variable. Empirical analysis of such models requires the use of multi-modal statistical methods (Cobb et al. 1983; Guastello 2009). A backlash developed as critics argued that the theory was applied to situations that did not fulfill the strict assumptions necessary for the application, but Rosser (2007) has argued that this backlash was overdone, with many avoiding its use who should not do so.[4]

While *chaos theory* can be traced back at least to Poincaré (1890), it became prominent after the identification of sensitive dependence on initial conditions, aka “the butterfly effect,” by the climatologist, Edward Lorenz (1963), probably the most important idea associated with the phenomenon. Applications in economics followed after an important paper by May (1976) that initially suggested some of them. Debates over empirical measurements and problems associated with forecasting have reduced some of the earlier enthusiasm for chaos theory in economics, which probably peaked during the 1980s. However, the fundamental insights derived from it continue to influence economic thinking as well as that in other disciplines.

Figure 1 shows the butterfly effect as found by Lorenz initially, with two very different trajectories arising from the same point, with these distinguished by only small differences in starting conditions from that point. Figure 2 shows a case the combines both catastrophic effects with chaotic dynamics in *chaotic hysteresis*, a figure due originally to Puu (1989), with the axes representing investment and the rate of change of investment for certain parameter values in a Keynesian style macroeconomic model. Rosser et al. (2001) estimated such phenomena for investment in the former Soviet Union over the post-World War II period.[5]

Coming on the heels of the popularity of chaos theory would be agent-based (or “small tent”) dynamic complexity, strongly associated with the Santa Fe Institute. However, its origin is generally traced to the urban segregation model of Schelling (1971), who used a go board rather than a computer to work out the dynamics of a

[4] Vladimir Arnol’d (1992) provides a clear and reasoned overview of the mathematical issues involved while avoiding the controversies.

[5] Ralph Abraham (1985) coined the term *chaostrophe* to describe such combinations, although that has not caught on especially. He also coined the term “chaotic hysteresis” (Abraham and Shaw 1987).

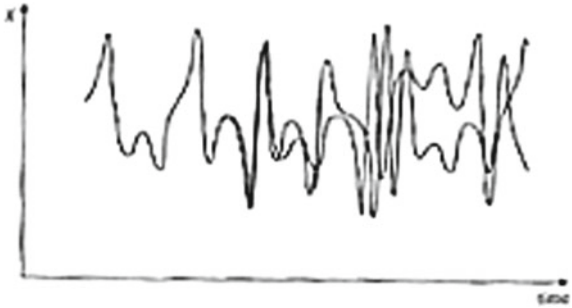


Fig. 1 Sensitive dependence on initial conditions in the Lorenz Climate Dynamics Model

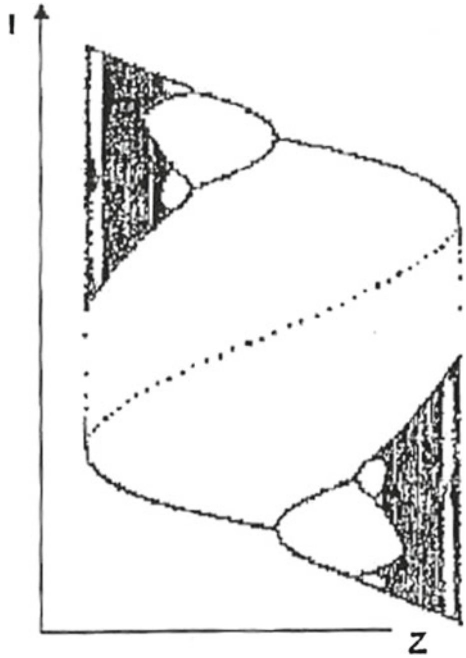


Fig. 2 A contribution of two globally stables states with a bifurcation manifold connecting them and an internal bifurcation mechanism within each of them

city starting out racially integrated and then segregating with only the slightest of incentives through nearest neighbor effects.[6] Such systems are famous for exhibiting self-organization and do not generally converge on any equilibrium, also showing cross-cutting hierarchical interactions and ongoing evolutionary change (Arthur et al. 1997a). Substantial active research in economics using such models is ongoing.

We note that these are only a small subset of the full array of complex dynamics that nonlinear systems can exhibit. Others include *non-chaotic strange attractors* (Lorenz 1992), *fractal basin boundaries* (Abraham et al. 1997), *flare attractors* (Hartmann and Rössler (1998); Rosser et al. 2003), and more.

[6] It is often claimed that Schelling used a chess board, however his board was 19 by 19, which makes it a go board, with go's use of simple black and white stones also fitting the model he developed.

A central point that should be clear is that the presence of such dynamic complexities in economic systems greatly complicates the problem for economic agents of forming rational expectations regarding the future path of such systems. In their presence, it becomes highly unlikely that agents can fulfill the conventional assumption of full information and complete rationality in their decision-making.

3 Herbert Simon and Bounded Rationality

The late Herbert A. Simon is widely considered to be the father of *modern behavioral economics*, at least it was his work to which this phrase was first applied. He was also an early theorist of complexity economics, if not the father per se, and also was one of the founders of the study of artificial intelligence in computer science. Indeed, he was a polymath who published well over 900 academic papers in numerous disciplines, and while he won the Nobel Prize in economics in 1978 for his development of the concept of *bounded rationality*, his Ph.D. was in political science and he was never in a department of economics. We must use the term "modern" before "behavioral economics" because quite a few earlier economists can be seen as focusing on actual human behavior while assuming that people do not behave fully in what we would now call an "economically rational" manner (Smith 1759; Veblen 1898).

We must at this point be clear that by "behavioral economics" we are not assuming a view similar to that of "behavioral psychology" of the sort advocated or practiced by Pavlov or B.F. Skinner (1938). The latter does not view studying what is in peoples' minds or consciousness as of any use or interest. All that matters is how they behave, particularly how they respond to repeated stimuli in their behavior. This is more akin to standard neoclassical economics, which also purports to study how people behave with little interest in what is going on inside their heads. The main difference between these two is that conventional economics makes a strong assumption about what is going on inside peoples' heads: that they are rationally maximizing individual utility functions derived from their preferences using full information. In contrast, behavioral economics does not assume that people are fully rational and particularly does not assume that they are fully informed. What is going on inside their heads is important, and such subjects as *happiness economics* (Easterlin 1974) are legitimate topics for behavioral economics.

In any case, from the beginning of his research with his path-breaking PhD dissertation that came out as a book in 1947, *Administrative Behavior* and on through important articles and books in the 1950s (Simon 1955a, 1957), Simon saw people as being limited in both their knowledge of facts as well as in their ability to compute and solve the difficult problems associated with calculating optimal solutions to problems. They face unavoidable limits to their ability to make fully rational decisions. Thus, people live in a world of *bounded rationality*, and it was this realization that led him into the study of artificial intelligence in computer science as part of his study of how people think in such a world (Simon 1969).

This led Simon to the concept of *satisficing*. People set targets that they seek to achieve and then do not pursue further efforts to improve situations once these targets have been reached, if they are. Thus a firm will not maximize profits, but its managers will seek to achieve an acceptable level of profits that will keep owners sufficiently happy. This idea of satisficing became the central key to the behavioral study of the firm (Cyert and March 1963) and entered into the management literature, where it probably became more influential than it was in economics, for quite a long time.

Some economists, notably Stigler (1961), have taken Simon's position and argued that he is actually a supporter of full economic rationality, but only adding another matter to be optimized, namely minimizing the costs of information. People are still optimizing but take account of the costs of information. However, Stigler's argument faces an unavoidable and ineluctable problem: people do not and cannot know what the full costs of information are. In this regard they face a potential problem of infinite regress (Conlisk 1996). In order to learn the costs of information, they must determine how much time they should spend in this process of learning; they must learn what the costs of learning what the costs of information are. This then leads to the next higher order problem of learning what the costs of learning what the costs of information are, and there is no end to this regress in principle.[7] In the end they must use the sorts of *heuristic* (or "rule of thumb") devices that Simon proposes that people facing bounded rationality must use in order to answer the question. Full rationality is impossible, and the ubiquity of complexity is a central reason why this is the case.

Simon (1976) distinguishes *substantive rationality* from *procedural rationality*. The former is the sort of rationality traditionally assumed by most economists in which people are able to achieve full optimization in their decision-making. The latter involves them selecting procedures or methods by which they can "do their best" in a world in which such full optimization is impossible, the heuristics by which they manage in a world of bounded rationality. In this regard it is not the case that Simon views people as being outright irrational or crazy. They have interests and they generally know what those are and they pursue them. However, they are unavoidably bounded in their ability to do so fully, so they must adopt various essentially ad hoc methods to achieve their satisficing goals.

Among these heuristics that Simon advocated for achieving procedural rationality were trial and error, imitation, following authority, unmotivated search, and following hunches. Pingle and Day (1996) used experiments to study the relative effectiveness of each of these, none of which clearly can achieve fully optimal outcomes. Their conclusion was that each of these can be useful for improving decision-making,

[7]This is a problem that central planners faced: how much time and in what way should planners spend thinking about how they should plan? This problem was discussed in the French and Russian literature on planning, with the French applying the word *planification* to this process of "planning how to plan," although that word was also used for both planning in general as well as for the more specific question of dealing with the problem of aggregating micro level plans into a coherent macro whole (Rosser and Rosser 2004, p. 10).

however, none of them is clearly superior to the others. It is advisable for agents to use several of these and to move from one to another under different circumstances, although as noted above it may be hard to know when to do that and precisely how.

4 Boundedly Rationality and Agent-Based Complexity

Chen and Kao (2015) highlight further the link between Simonian bounded rationality and the agent-based form of complexity that combine to form *complex behavioral economics*. These links draw from the work of Hayek (Vriend 2002) as well as from Schelling (1971), Vinkopić and Kirman (2006), Ostrom et al. From the early days of Simon's formulation of bounded rationality (Simon 2000). These link with Albin's (1998) on how complexity bounds rationality and the idea that agents can be represented by programs, in its hardest form that agents are programs (Newell and Simon 1972; Mirowski 2007; Davis 2013).

The use of agents in genetic programming (Duffy 2006) are marked by their aspirations and the limits of their capabilities (Simon 2000; Chen 2012). All of these represent the real limits of agents that they face in making decisions in real situations in contrast with the overly strong models of standard economic theory.

Drawing on the hierarchy theory of Simon (1962) has been a large literature and set of efforts that emphasize modularity and indecomposability within complex systems using agent-based models (Miller 1956; Simon 1957; Stiglitz and Gallegati 2011; Hommes 2014). Some of this has depended on cellular automata to study the Schelling model and others.

Fundamentally agent-based models based on behavioral assumptions are able to build up aggregate behavior from micro-founded behavior based on bounded rationality (Epstein and Axtell 1996). This potential provides a basis for macroeconomic modeling using an alternative foundation that avoids rational expectations and is consistent with a complex behavioral outcome. This can provide a strong foundation for a genuine behavioral macroeconomics (Akerlof 2002).

5 Imitation and the Instability of Markets

While this list of procedures that can support a boundedly rational pursuit of procedural rationality, a point not clearly made is that excessive focus on one of these rather than others can lead to problems. Clearly following authority can lead to problems when the authority is flawed, as many unfortunate examples in history have shown. Any of these can lead to problems if too intensively followed, but one that has particularly played an unfortunate role in markets is imitation, even though it is a widely used method by many people with a long history of being evolutionarily successful. The problem is particularly acute in asset markets, where imitation can lead to speculative bubbles that destabilize markets and can lead to much broader problems in the economy, as the crisis of 2008 manifestly shows.

A long literature (MacKay 1852; Baumol 1957; Zeeman 1974; Rosser 1997) has recognized that while agents focusing on long-term fundamental values of assets tend to stabilize markets by selling them when their prices exceed these fundamentals and buying when they are below those, agents who chase trends can destabilize markets by buying when prices are rising, thus causing them to rise more, and vice versa. When a rising price trend appears, trend chasers will do better in returns than fundamentalists and imitation of those doing well will lead agents who might have followed stabilizing fundamentalist strategies to follow destabilizing trend chasing strategies, which will tend to push the price further up. And when a bubble finally peaks out and starts to fall, trend chasers can then push the price down more rapidly as they follow each other in a selling panic.

That such a tendency to engage in trend chasing speculation is deeply rooted in the human psyche was initially established by Smith et al. (1988), with many subsequent studies supporting this observation.[8] Even in situations with a finite time horizon and a clearly identified payment that establishes the fundamental value of the asset being traded, in experimental markets it has been repeatedly shown that bubbles will appear even in these simplified and clearcut cases. People have a strong tendency to speculate and to follow each other into such destabilizing speculation through imitation. Procedures that can support procedural rationality in a world of bounded rationality can lead to bad outcomes if pursued too vigorously.

We note that such patterns regularly take three different patterns. One is for price to rise to a peak and then to fall sharply after hitting the peak. Another is for price to rise to a peak and then decline in a more gradual way in a reasonably symmetric manner. Finally, we see bubbles rising to a peak, then declining gradually for awhile, finally collapsing in a panic-driven crash. Kindleberger's classic *Manias, Panics, and Crashes* (2001) shows in its Appendix B that of 47 historical speculative bubbles, each of the first two have five examples, while the remainder, the vast majority, follow the final pattern, which requires heterogeneous agents who are not fully rational for it to occur (Rosser 1997). This shows that complexity is deeply involved in most speculative bubbles.

Figures 3, 4, and 5 show the time path for prices of three bubbles before, during, and immediately after the 2008 crisis. They show the three patterns described above, taken from Rosser et al. 2012. The first is for oil, which peaked at $147 per barrel in July 2008, the highest nominal price ever observed, and then crashed hard to barely over $30 per barrel in the following November. It seems that commodities are more likely to follow this pattern than other assets (Ahmed et al. 2014).

The second pattern was followed by the housing bubble, which peaked in mid-2006 according to this figure, which shows to different indexes, the Case-Shiller

[8]This result is especially significant in that Vernon Smith (1962) has long been an advocate of the idea that free markets work well and lead to rapid convergence in properly structured markets such as double auction arrangements. This insight was a major basis for his receipt of the Nobel Prize in economics, although he clearly understands that markets can behave badly under certain circumstances.

Fig. 3 West Texas intermediate crude oil prices per barrel, 2003–2011

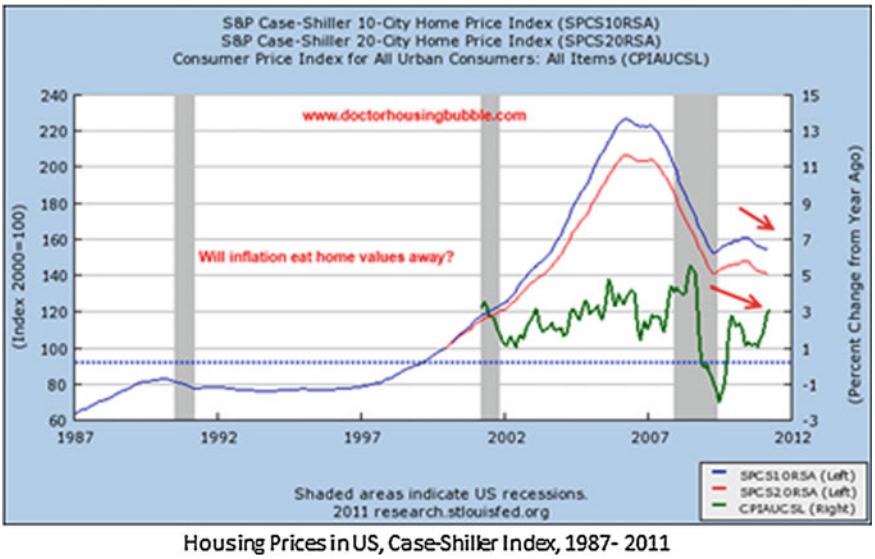


Fig. 4 Housing prices in US, Case-Shiller index, 1987–2011

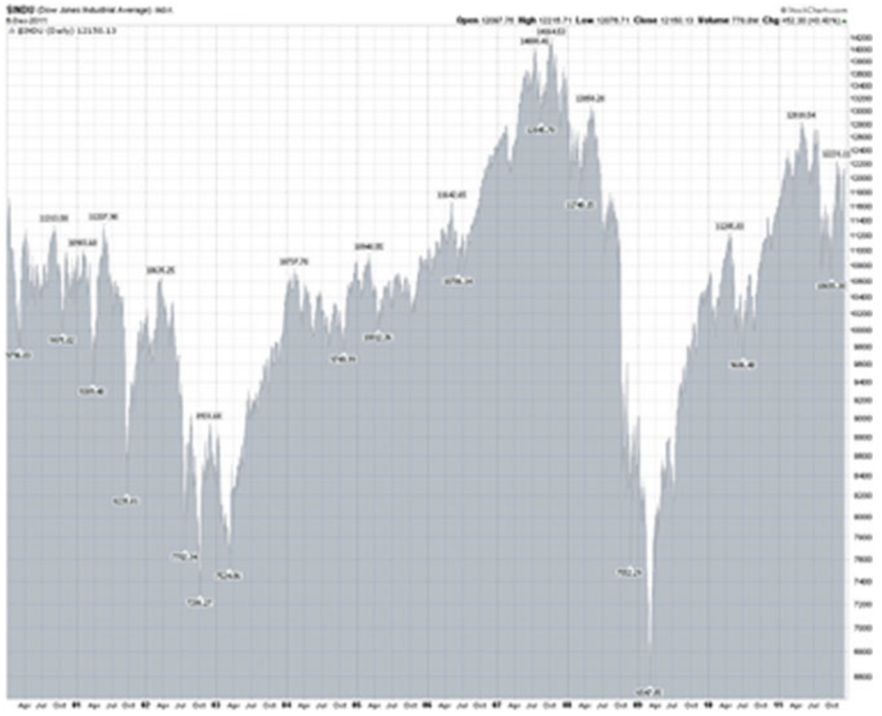


Fig. 5 Dow-Jones industrial average, 2000–2011, daily data

10-city one and their 20-city one as well. Looking closely one can see a bit of roughness around the peak making it look almost like the third pattern, whereas in fact if one looks at housing markets in individual cities, they look as posited by this pattern, with this roughness at the national level reflecting that different cities peaked at different times, with a final round of them doing so as late as January 2007 before they all declined.

This sort of pattern historically is often seen with real estate market bubbles. The more gradual decline than in the other patterns, nearly symmetric with the increase, reflects certain behavioral phenomena. People identify very personally and intensely with their homes and as a result tend not to easily accept that their home has declined in value when they try to sell it during a downturn. As a result they have a tendency to offer prices that are too high and then refuse to lower their prices readily when they fail to sell. The upshot is a more dramatic decline in volume of sales on the downswing compared to the other patterns as people hang on and refuse to lower prices.

The third case shows the US stock market as exhibited by the Dow-Jones average, which peaked in October 2007, only then to crash in September 2008. Such patterns seem to be more common in markets for financial assets. Such patterns show heterogeneity of agents with different patterns of imitation, a smarter (or luckier) group that

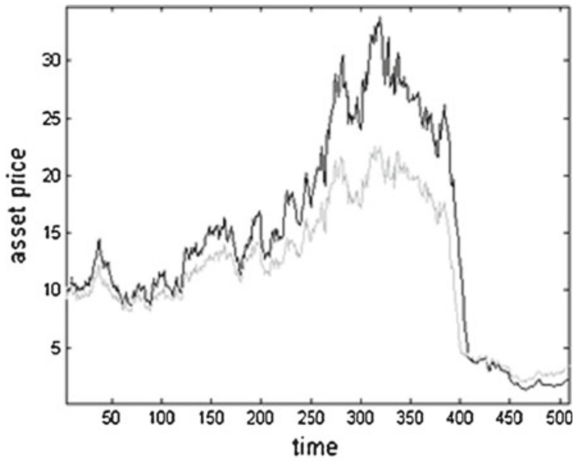


Fig. 6 Increase in J. The two time series share the same random numbers and same parameters but J. In the *grey* time series J = 0.5; in the *black* time series J = 3

gets out earlier at the peak, followed by a less smart (or less lucky) group that hangs on hoping the price will return to rising, only to panic later en masse for whatever reason.

Finally, Fig. 6 shows how this pattern with its *period of financial distress* (Minsky 1972) can be modeled in an agent-based model that has agents shifting from one strategy to another based on their relative successes, although not instantly (Gallegati et al. 2011). This model is based on ideas from Brock and Hommes (1997, 1998) that underlie the so-called Santa Fe stock market model (Arthur et al. 1997b). What triggers the delayed crash is agents running into financial constraints such as happens when individuals must meet margin calls in stock markets. The higher curve shows the pattern when agents imitate each other more strongly, as in a statistical mechanics model when there is a stronger interaction between particles.

6 Hierarchical Complexity and the Question of Emergence

While we can see Herbert Simon's discovery of bounded rationality as an indirect claim to being a "father of complexity," his most direct claim, recognized by Seth Lloyd in his famous list, is his 1962 paper to the American Philosophical Society on "The Architecture of Complexity." In this transdisciplinary essay, he deals with everything from organizational hierarchies through evolutionary ones to those involving "chemico-physical systems." He is much concerned with the problem of the decomposability of higher order systems into lower level ones, noting that production ones, such as for watchmaking, as well as organizational ones, function better when such decomposability is present, which depends on the stability and functionality of the lower level systems.[9]

However, he recognizes that many such systems involve *near decomposability*, perhaps a hierarchical complexity equivalent of bounded rationality. In most of them there are interactions between the subsystems, with the broader evolution of the system depending on aggregated phenomena. Simon provides the example of a building with many rooms. Temperature in one room can change that in another, even though their temperatures may fail to converge. But the overall temperatures that are involved in these interactions are determined by the aggregate temperature of the entire building.

Simon also deals with what many consider to be the most fundamental issue involving complexity, namely that of emergence. His most serious discussion of the emergence of higher levels of hierarchical structure out of lower levels involves biological evolution, where these issues have long been most intensively discussed. He argues that how these higher levels emerged has not reflected teleological processes but strictly random processes. He also argues that even in closed systems, there need be no change in entropy in the aggregate when subsystems emerge within that system. But he also recognizes that organisms are energetically open systems, so that "there is no way to deduce the direction, much less the rate, of evolution from classical thermodynamic considerations" (Simon 1962, p. 8). However, it is the development of stable intermediate forms that is the key for the emergence of yet higher forms.

Simon does not cite this older literature, but this issue was central to the British "emergentist" literature that came out of the nineteenth century to become the dominant discourse in the 1920s regarding the broader story of biological evolution, all embedded within a broader vision fitting this within the emergence of physical and chemical systems from particles through molecules to such higher levels above biological evolution in terms of human consciousness, social systems, and yet higher systems.[10] Simon dealt with this multiplicity of processes without drawing their interconnection as tightly as did these earlier figures. In the 1930s with the *neo-Darwinian synthesis* (Fisher 1930; Wright 1931; Haldane 1932), the emphasis returned to near-continuous Darwinian process of gradual changes appearing from probabilistic changes arising from mutations at the gene level, with the gene the ultimate focus of natural selection (Dawkins 1976).

While Simon avoided dealing with this issue of emergence in biological evolution in 1962, when the reductionist neo-Darwinian synthesis was at the highest level of

---

[9]See Rosser (2010) for a discussion of relations between *multidisciplinary, interdisciplinary, and transdisciplinary* viewpoints. For a discussion of variations on hierarchical relations see Rosser et al. (1994).

[10]This tradition derived from J.S. Mill's (1843) *heteropathic laws* that focused on basic chemical interactions where two molecules come together to form a completely different molecule. Lewes (1875) coined the term *emergence* for such phenomena, with C. Lloyd Morgan (1923) representing its culmination in biological evolutionary theory. In the 1930s this approach would be pushed aside by the neo-Darwinian synthesis (Rosser 2011), which emphasized a reductionist approach to the gene. For further discussion of this point see Rosser and Rosser (2015).

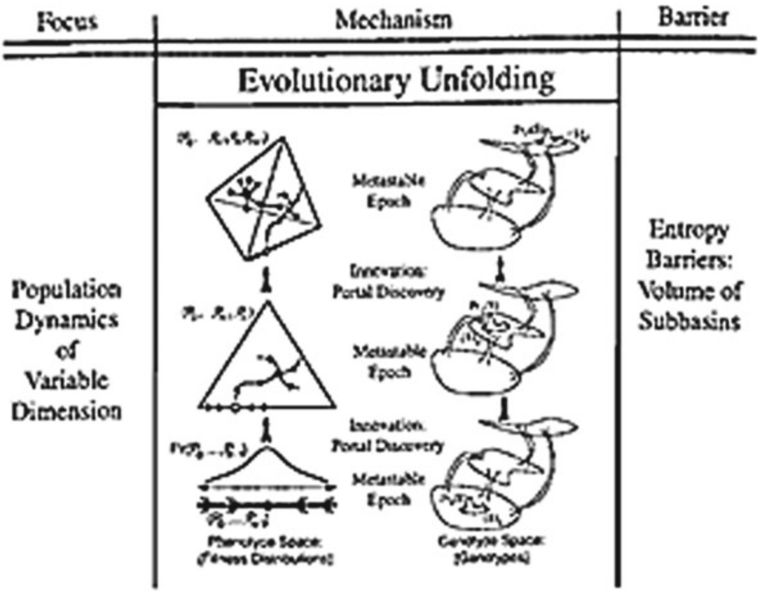


Fig. 7 Hierarchical emergence in entropic evolutionary unfolding

its influence, soon the emergence view would itself re-emerge, based on multi-level evolutionary process (Crow 1955; Hamilton 1964; Price 1970). This would further develop with the study of nonlinear dynamics and complexity in such systems, with such figures as Stuart Kauffman (1993) and James Crutchfield (1994, 2003), who draw on computational models for their depictions of *self-organization* in biological evolutionary systems.

Figure 7 from Crutchfield (2003, p. 116) depicts how an initial genetic level mutation can lead to emergent effects at higher levels. On the right side are genotypes moving upwards from one basin of attraction to another, while on the left side phenotypes are also doing so in a parallel pattern. He introduces the concept of *mesoscales* for such processes, which clearly follow Simon's admonition about the necessity of stable intermediate systems emerging to support the emergence of yet higher order ones.

This view remains questioned by many evolutionists (Gould 2002). While the tradition going through catastrophe theory from D'Arcy Thompson (1917) has long argued for form arising from deep structures in organic evolution, critics have argued that such self-organizing processes are ultimately teleological ones that replicate old pre-evolutionary theological perspectives such as Paley's (1802) in which all things are in their place as they should be due to divine will. Others have criticized that such process lack invariance principles (McCauley 2005). Others coming from a more computational from such processes (Moore 1990). There is no easy resolution of this debate, and even those advocating the importance of emergent self-organization recognize the role of natural selection. Thus, Kauffman (1993, p. 644) has stated, "Evolution is not just 'chance caught on a wing.' It is not just a tinkering of the ad hoc, of bricolage, of contraption. It is emergent order honored and honed by selection."

While the mechanisms are not the same, the problems of emergent self-organization apply as well to socioeconomic systems. Simon's focus tended to be on organizations and their hierarchies. While he may well have sided with the more traditional neo-Darwinian synthesizers when it came to emergence of higher order structures in biological evolution, the role of human consciousness within human socio-economic systems means that the rules are different there, and the formation of higher order structures can become a matter of conscious will and planning, not mere randomness.

7 Conclusions

The late Herbert A. Simon was the "father of behavioral economics," who discovered the ideas of *bounded rationality* and *satisficing*, along with many ideas in many other disciplines such as artificial intelligence, cognitive science, and management. He was also a founder of complexity analysis in its transdisciplinary formulation, in particular hierarchical complexity, although with deep links to both computational complexity through his work on artificial intelligence and also on dynamic complexity. Branching across many disciplines, with its most serious implications relating to evolutionary processes and the problem of the evolutionary emergence of higher order structures in nature. This goes beyond biology to a broader view of the universe, with such emergent evolutionary processes extending to emergence of atoms from subatomic particles to molecules to organic molecules to multi-cellular organisms to human consciousness to societies and to higher order structures beyond those.

A fundamental link between the two concepts is that the existence of complexity provides a foundation for the limits to knowledge and rationality that humans face, and thus why they must operate using bounded rationality. Satisficing is an alternative approach derived from bounded rationality in that one seeks to achieve goals that seem achievable and are socially acceptable rather than striving for all-out optimization. This leads to the use of heuristics such as following authority, imitation, trial and error, and unmotivated search, although overly excessive focus on one of these can lead to problems, as shown by the role of excessive imitation of others in the appearance of damaging speculative bubbles and crashes. While debates continue regarding this and many other matters involving the relation between behavioral economics and complexity, the emergence of a complex behavioral macroeconomics and its apparent influence on policymaking is something to be encouraged.

References

Abraham RH (1985) Chaostrophes, intermittency, and noise. In: Fischer P, Smith WR (eds) Chaos, fractals, and dynamics. Marcel Dekker, New York, pp 3–22

Abraham RH, Gardini L, Mira C (1997) Chaos in discrete dynamical systems: a visual introduction in 2 dimensions. Springer, New York

Abraham RH, Shaw CD (1987) Dynamics: a visual introduction. In: Yates FE (ed) Self-organizing systems: the emergent order. Plenum Press, New York, pp 543–597

Ahmed E, Rosser JB Jr, Uppal J (2014) Are there nonlinear speculative bubbles in commodities prices? J Post Keynes Econo 36:415–438

Akerlof GA (2002) Behavioral macroeconomics and macroeconomic behavior. Am Econ Rev 92:411–433

Albin PS, Foley DK (1998) Barriers and bounds to rationality: essays on complex economics and dynamics in interactive systems. Princeton University Press, Princeton

Arnol'd VI (1992) Catastrophe theory, 3rd edn. Springer, Berlin

Arthur WB, Durlauf SN, Lane DA (1997a) Introduction. In: Arthur WB, Durlauf SN, Lane DA (eds) The economy as an evolving complex system II. Addison-Wesley, Reading, pp 1–14

Arthur WB, Holland JH, LeBaron B, Palmer R, Tayler P (1997b) Asset pricing under endogenous expectations in an artificial stock market. In: Arthur WB, Durlauf SN, Lane DA (eds) The economy as an evolving complex system II. Westview, Boulder, pp 15–44

Baumol WJ (1957) Speculation, profitability, and stability. Rev Econ Stat 39:263–271

Brock WA, Hommes CH (1997) A rational route to randomness. Econometrica 65:1059–1095

Brock WA, Hommes CH (1998) Heterogeneous beliefs and routes to chaos in a simple asset pricing model. J Econ Dyn Control 22:1235–1274

Chen S-H (2012) Varieties of agent-based computational economics: a historical and an interdisciplinary perspective. J Econ Dyn Control 36:1–25

Chen S-H, Kao YF (2015) Herbert Simon and agent-based computational economics. Working Paper, National Chengchi University, Taipei

Cobb L, Koppstein P, Chen NH (1983) Estimation and moment recursion relationships for multimodal distributions of the exponential family. J Am Stat Assoc 78:124–130

Conlisk J (1996) Why bounded rationality? J Econ Lit 34:1–64

Crow JF (1955) General theory of population genetics: synthesis. Cold Spring Harb Symp Quant Biol 20:54–59

Crutchfield JP (1994) The calculi of emergence: Computation, dynamics and induction. Physica D 75:11–54

Crutchfield JP (2003) When evolution is revolution - origins of innovation. In: Crutchfield JP, Schuster P (eds) Evolutionary dynamics: exploring the interplay of selection, accident, neutrality, and function. Oxford University Press, Oxford, pp 101–133

Cyert RM, March JG (1963) A behavioral theory of the firm. Prentice-Hall, Englewood Cliffs

Davis JB (2013) The emergence of agent-based modeling in economics: individuals come down to bits. Filos Econ 1:229–246

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Day RH (1994) Complex economic dynamics, volume I: an introduction to dynamical systems and market mechanisms. MIT Press, Cambridge

Duffy J (2006) Agent-based models and human subject experiments. In: Tesfatsion L, Judd KL (eds) Handbook of computational economics II. Elsevier, Amsterdam, pp 949–1011

Easterlin RA (1974) Does economic growth improve the human lot? Some empirical evidence. In: David PA, Reder MW (eds) Nations and households in economic growth: essays in honor of Moses Abramovitz. Academic Press, New York, pp 89–125

Epstein JB, Axtell R (1996) Growing artificial societies from the bottom up. Brookings Institution Press, Washington

Fisher RA (1930) The general theory of natural selection. Oxford University Press, Oxford

Forrester JW (1961) Industrial dynamics. MIT Press, Cambridge

Gallegati M, Palestrini A, Rosser JB Jr (2011) The period of financial distress in speculative bubbles: Heterogeneous agents and financial constraints. Macroeconomic Dynamics 15:60–79
Gould SJ (2002) The structure of evolutionary theory. Belknap Press of Harvard University Press, Cambridge
Guastello SJ (2009) Discontinuities and catastrophes with polynomial regression. In: Guastello SJ, Gregson RAM (eds) Nonlinear dynamical systems analysis for the behavioral sciences using real data. CRC Press, Boca Raton, pp 251–280
Haldane JBS (1932) The causes of evolution. Longmans and Green, London
Hamilton WD (1964) The genetical evolution of social behavior. J Theor Biol 7:1–52
Hartmann GC, Rössler O (1998) Coupled flare attractors - a discrete prototype for economic modelling. Discret Dyn Nat Soc 2:153–159
Hommes CH (2014) Behaviorally rational expectations and almost self-fulfilling equilibria. Rev Behav Econ 1:76–97
Horgan J (1997) The end of science: facing the limits of knowledge in the twilight of the scientific age. Broadway Books, New York
Israel G (2005) The science of complexity: epistemological implications and perspectives. Sci Context 18:1–31
Kauffman SA (1993) The origins of order: self-organization and selection in evolution. Oxford University Press, Oxford
Kindleberger CP (2001) Manias, panics, and crashes: a history of financial crises, 4th edn. Wiley, New York
Knight FH (1921) Risk, uncertainty, and profit. Houghton Mifflin, Boston
Lewes GH (1875) Problems of life and mind. Kegan Paul Trench Turbner, London
Lorenz EN (1963) Deterministic non-periodic flow. J Atmos Sci 20:130–141
Lorenz H-W (1992) Multiple attractors, complex basin boundaries, and transient motion in deterministic economic systems. In: Feichtinger G (ed) Dynamic economic models and optimal control. North-Holland, Amsterdam, pp 311–430
MacKay C (1852) Memoirs of extraordinary delusions and the madness of crowds. Office of the National Illustrated Library, London
McCauley JL (2005) Making mathematics effective in economics. In: Velupillai KV (ed) Computability, complexity and constructivity in economic analysis. Blackwell, Victoria, pp 51–84
Meadows DH, Meadows DL, Randers J, Behrens WW III (1972) The limits to growth. Universe Books, New York
Mill JS (1843) A system of logic: ratiocinative and inductive. Longmans Green, London
Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63:81–97
Minsky HP (1972) Financial instability revisited: The economics of disaster. Reappraisal of the Fed Reserve Discount Mech 3:97–136
Mirowski P (2007) Markets come to bits: Evolution, computation and markomata in economic science. J Econ Behav Organ 63:209–242
Moore C (1990) Undecidability and unpredictability in dynamical systems. Physical Review Letters 64:2354–2357
Morgan CL (1923) Emergent evolution. Williams and Norgate, London
Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs
Paley W (1802) Natural theology: or evidences of the existence and attributes of the deity, collected from the appearances of nature. R. Faulder, London
Pingle M, Day RH (1996) Modes of economizing behavior: experimental evidence. J Econ Behav Organ 29:191–209
Poincaré H (1890) Sur les équations de la dynamique et le problème de trois corps. Acta Math 13:1–270
Polanyi K (1944) The great transformation. Farrar and Rinehart, New York
Price G (1970) Selection and covariance. Nature 227:520–521
Puu T (1989) Nonlinear economic dynamics. Springer, Berlin

Rosser JB Jr (1997) Speculations on nonlinear speculative bubbles. Nonlinear Dyn Psychol Life Sci 1:275–300
Rosser JB Jr (1999) On the complexities of complex economic dynamics. J Econ Perspect 13:169–192
Rosser JB Jr (2007) The rise and fall of catastrophe theory applications in economics: was the baby thrown out with the bathwater? J Econ Dyn Control 31:3255–3280
Rosser JB Jr (2010) Is a transdisciplinary perspective on economic complexity possible? J Econ Behav Organ 75:3–11
Rosser JB Jr (2011) Complex evolutionary dynamics in urban-regional and ecologic-economic systems: From catastrophe to chaos and beyond. Springer, New York
Rosser JB Jr, Ahmed E, Hartmann GC (2003) Volatility via social flaring. J Econ Behav Organ 50:77–87
Rosser JB Jr, Folke C, Günther F, Isomäki H, Perrings C, Puu T (1994) Discontinuous change in multilevel hierarchical systems. Syst Res 11:77–94
Rosser JB Jr, Rosser MV (2004) Comparative economics in a transforming world economy, 2nd edn. MIT Press, Cambridge
Rosser JB Jr, Rosser MV (2015) Complexity and behavioral economics. Nonlinear Dyn Psychol Life Sci 19:201–226
Rosser JB Jr, Rosser MV, Gallegati M (2012) A Minsky-Kindleberger perspective on the financial crisis. J Econ Issues 45:449–458
Rosser JB Jr, Rosser MV, Guastello SJ, Bond RW Jr (2001) Chaotic hysteresis in systemic economic transformation. Nonlinear Dyn Psychol Life Sci 5:345–368
Schelling TC (1971) Dynamic models of segregation. J Math Sociol 1:143–186
Simon HA (1947) Administrative behavior. Macmillan, New York
Simon HA (1955a) A behavioral model of rational choice. Q J Econ 69:99–118
Simon HA (1955b) On a class of skew distribution functions. Biometrika 42:425–440
Simon HA (1957) Models of man. Wiley, New York
Simon HA (1962) The architecture of complexity. Proc Am Philos Soc 106:467–482
Simon HA (1969) The sciences of the artificial. MIT Press, Cambridge
Simon HA (1976) From substantive to procedural rationality. In: Latsis SJ (ed) Method and appraisal in economics. Cambridge University Press, Cambridge, pp 129–148
Simon HA (2000) Bounded rationality in social sciences: today and tomorrow. Mind Soc 7:25–39
Skinner BF (1938) The behavior of organisms: An experimental analysis. B.F., Skinner Foundation, Cambridge
Smith A (1759) The theory of moral sentiments. Miller, Kincaid, and Bell, London
Smith VL (1962) An experimental study of competitive market behavior. J Polit Econ 70:111–137
Smith VL, Suchanek GL, Williams AW (1988) Bubbles, crashes, and endogenous expectations. Econometrica 56:1119–1151
Stigler GJ (1961) The economics of information. J Polit Econ 69:213–225
Stiglitz J, Gallegati M (2011) Heterogeneous interacting agent models for understanding monetary economies. East Econ J 37:6–12
Thom, R. (1975). Structural stability and morphogenesis: An outline of a theory of models. Reading: Benjamin
Thompson DW (1917) On growth and form. Cambridge University Press, Cambridge
van Bertalanffy L (1968) General systems theory: foundations, development and applications. George Brazilier, New York
Veblen TB (1898) Why is economics not an evolutionary science? Q J Econ 12:373–397
Velupillai KV (2000) Computable economics. Oxford University Press, Oxford
Velupillai KV (2011) Nonlinear dynamics, complexity, and randomness: algorithmic foundations. J Econ Surv 25:547–568
Vinkopić K, Kirman A (2006) Physical analogue of the Schelling model. Proc Natl Acad Sci 103:19261–19265
von Neumann J (1966) Theory of self-reproducing automata. University of Illinois Press, Urbana

Vriend NJ (2002) Was Hayek and ACE? South Econ J 68:811–840
Wiener N (1948) Cybernetics: or control and communication in the animal and machine. MIT Press, Cambridge
Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159
Zeeman EC (1974) On the unstable behavior of the stock exchanges. J Math Econ 1:39–44

An Alternative Proof of the Theorem of Woodford on the Existence of a Sunspot Equilibrium in a Continuous-Time Model

Kazuo Nishimura and Tadashi Shigoka

Abstract Nishimura and Shigoka, Int J Econ Theory 2:199–216, (2006) has proved a continuous-time version of the theorem of Woodford, Stationary sunspot equilibria: the case of small fluctuations around a deterministic steady state, mimeo, (1986) to the effect that there exists a stationary sunspot equilibrium for a continuous-time model with a predetermined variable and with an unstable root, if equilibrium is indeterminate near either a steady state or a closed orbit, and if a stable manifold is well-located in an ambient space. The present study provides this theorem with an alternative proof that is due to Murakami et al. Homoclinic orbit and stationary sunspot equilibrium in a three-dimensional continuous-time model with a predetermined variable forthcoming in: Nishimura K, Venditti A, Yannelis NC (eds) Sunspots and non-linear dynamics. Springer, (2016) and simpler than that of Nishimura and Shigoka, Int J Econ Theory 2:199–216, (2006).

1 Introduction

If for a given deterministic model, there exists a continuum of perfect foresight equilibria, equilibrium is said to be indeterminate. Suppose that fundamental characteristics of an economy are deterministic, but that economic agents believe nevertheless that equilibrium dynamics is affected by random factors apparently irrelevant to the fundamental characteristics (sunspots). This prophecy could be self-fulfilling, and one will get a sunspot equilibrium, if the resulting equilibrium dynamics is subject to a nontrivial stochastic process. A sunspot equilibrium is called a stationary sunspot equilibrium, if the equilibrium stochastic process is stationary. See Shell (1977), Azariadis (1981), and Cass and Shell (1983) for the concept of a sunspot

K. Nishimura (✉)
Research Institute for Economics and Business Administration,
Kobe University, Kobe, Japan
e-mail: nishimura@rieb.kobe-u.ac.jp

T. Shigoka
Institute of Economic Research, Kyoto University, Kyoto, Japan
e-mail: sigoka@kier.kyoto-u.ac.jp

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_2

equilibrium. For a large class of models whose fundamental characteristics are deterministic, if equilibrium is indeterminate, there exists a sunspot equilibrium. See Chiappori and Guesnerie (1991) and Guesnerie and Woodford (1992) for thorough surveys on the sunspot literature. Woodford (1986) has proved that there exists a stationary sunspot equilibrium for a discrete-time model with a predetermined variable and with an unstable root, if equilibrium is indeterminate near a steady state, and if a stable manifold is well-located in an ambient space. Nishimura and Shigoka (2006) treats a three-dimensional continuous-time deterministic model that includes one predetermined variable and two non-predetermined variables and that includes a well-located two-dimensional invariant manifold that might be a stable manifold of either a steady state or a closed orbit, and has constructed a stationary sunspot equilibrium in this model by means of extending the method of Shigoka (1994). This is a continuous-time version of the theorem due to Woodford (1986).

Murakami et al. (2016) treats a three-dimensional continuous-time deterministic model that includes one predetermined variable and two non-predetermined variables and that is amenable to the existence of a homoclinic orbit with multiple steady states, and has constructed a stationary sunspot equilibrium in this model by means of extending the method of Benhabib et al. (2008). The underlying deterministic dynamics in Murakami et al. (2016) is more complex than that in Nishimura and Shigoka (2006). On the other hand, as discussed in Sect. 2.4, the structure of a stochastic differential equation in Benhabib et al. (2008) the extension of which is Murakami et al. (2016) is simpler than that in Shigoka (1994) the extension of which is Nishimura and Shigoka (2006). The present study applies the method of Murakami et al. (2016) to the simpler underlying deterministic dynamics treated by Nishimura and Shigoka (2006), and provides an alternative proof of the existence of a stationary sunspot equilibrium for this model. The alternative proof in this study is simpler than that of Nishimura and Shigoka (2006), because the structure of a stochastic differential equation in the present study is simpler than that in Nishimura and Shigoka (2006).

In Sect. 2.1, we specify an underlying deterministic model. In Sect. 2.2, we specify a stochastic process that generates sunspot variables. In Sect. 2.3, we define a stationary sunspot equilibrium, and state a main theorem that is a continuous-time version of the theorem of Woodford (1986). In Sect. 2.4, we relate our result to those of Shigoka (1994), Nishimura and Shigoka (2006), Benhabib et al. (2008), and Murakami et al. (2016). Section 3 provides the main theorem with a proof the method of which is due to Murakami et al. (2016).

2 Main Result

2.1 Deterministic Equilibrium Dynamics

In the present section, we specify a three-dimensional continuous-time deterministic model that includes one predetermined variable and two non-predetermined variables. We will assume that the model has a well-located two-dimensional invariant manifold that might be a stable manifold of either a steady state or a closed orbit. Since the number of a predetermined variable is one, and since the dimension of the invariant manifold is two, equilibrium is indeterminate near either the steady state or the closed orbit. Let V be a nonempty open subset of $\mathbb{R}^2$ homeomorphic to some convex set. Let I be a nonempty open connected subset of $\mathbb{R}$. Let W be defined as $W := V \times I$. Let $f_i : W \to \mathbb{R}$, $i = 1, 2, 3$, be a continuously differentiable function, i.e., a C^1-function, and let $F : W \to \mathbb{R}^3$ be a C^1-function defined as

$$F(X, u, Q) := \begin{bmatrix} f_1(X, u, Q) \\ f_2(X, u, Q) \\ f_3(X, u, Q) \end{bmatrix},$$

where $(X, u, Q) \in W$. We assume that X is a predetermined variable, whereas u and Q are non-predetermined variables, and that a perfect foresight equilibrium is a solution of an ordinary differential equation $[\dot{X}, \dot{u}, \dot{Q}]^{\mathbf{T}} = F(X, u, Q)$, where $^{\mathbf{T}}$ denotes the transpose of a given vector. We assume:

Assumption 1 There exists a C^1-function $\varphi : V \to I$ such that, for $(X, u) \in V$,

$$f_3(X, u, \varphi(X, u)) = \frac{\partial \varphi}{\partial X}(X, u) f_1(X, u, \varphi(X, u)) + \frac{\partial \varphi}{\partial u}(X, u) f_2(X, u, \varphi(X, u)).$$

Under Assumption 1, $\{(X, u, Q) \in W : (X, u) \in V \wedge Q = \varphi(X, u)\}$ constitutes a two-dimensional manifold invariant under the action of $[\dot{X}, \dot{u}, \dot{Q}]^{\mathbf{T}} = F(X, u, Q)$. Let $G : V \to \mathbb{R}^2$ be a C^1-function defined as

$$G(X, u) := \begin{bmatrix} f_1(X, u, \varphi(X, u)) \\ f_2(X, u, \varphi(X, u)) \end{bmatrix},$$

for $(X, u) \in V$. Under Assumption 1, we further assume that either of the following two assumptions is satisfied.

Assumption 2 There exists a closed subset D of V homeomorphic to the two-dimensional closed unit disk $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ such that the vector field $[\dot{X}, \dot{u}]^{\mathbf{T}} = G(X, u)$ points inward on the boundary ∂D of D, where ∂D is homeomorphic to $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

Assumption 3 There exists a closed subset D of V homeomorphic to the two-dimensional closed doughnut $\{(x, y) \in \mathbb{R}^2 : \frac{1}{2} \leq x^2 + y^2 \leq 1\}$ such that the vector field $[\dot{X}, \dot{u}]^{\mathbf{T}} = G(X, u)$ points inward on the boundary ∂D of D, where ∂D is homeomorphic to $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = \frac{1}{2} \vee x^2 + y^2 = 1\}$.

For some $(\bar{X}, \bar{u}) \in V$, if $(\bar{X}, \bar{u}, \varphi(\bar{X}, \bar{u})) \in W$ is a hyperbolic steady state, and if $\{(X, u, Q) \in W : (X, u) \in V \wedge Q = \varphi(X, u)\}$ constitutes a two-dimensional stable manifold of the steady state, then Assumptions 1 and 2 are satisfied, and equilibrium is indeterminate near the steady state. See Nishimura and Shigoka (2006, pp. 204–205) for the method of assuring that Assumptions 1 and 2 are satisfied, and see

Sect. 3 in Nishimura and Shigoka (2006) for concrete economic models that satisfy Assumptions 1 and 2.

For some closed curve γ in V, if $\hat{\gamma} := \{(X, u, Q) \in W : (X, u) \in \gamma \wedge Q = \varphi(X, u)\}$ is a closed orbit of $[\dot{X}, \dot{u}, \dot{Q}]^{\mathbf{T}} = F(X, u, Q)$, and if $\{(X, u, Q) \in W : (X, u) \in V \wedge Q = \varphi(X, u)\}$ includes a two-dimensional invariant manifold each point of which asymptotically converges to this closed orbit $\hat{\gamma}$, then Assumptions 1 and 3 are satisfied, and equilibrium is indeterminate near the closed orbit. See Sect. 2.4 in Nishimura and Shigoka (2006) for the method of assuring that Assumptions 1 and 3 are satisfied, and see Sect. 3 in Nishimura and Shigoka (2006) for concrete examples that satisfy Assumptions 1 and 3.

2.2 Sunspot Variables

In the present section, we specify a continuous-time stochastic process that generates sunspot variables. We assume that the stochastic process is subject to a *separable two-state Markov process with stationary transition matrices.* A sample function of a random variable subject to this process generates a sequence of discontinuous points such that each discontinuous point is, of itself, a random variable. We will utilize a sequence of discontinuous points in the sample path of a sunspot variable in order to construct a sunspot equilibrium.

Let $\mathbb{T}$ denote the set of all nonnegative real numbers, i.e., $\mathbb{T} := \mathbb{R}_+$. We denote the set of all function from $\mathbb{T}$ to $\{1, 2\}$ by $\{1, 2\}^{\mathbb{T}}$. Let B be some subset of $\{1, 2\}^{\mathbb{T}}$. Let $\varepsilon(t, b)$ denote the tth coordinate of $b \in B$. Let $\mathcal{B}(B)$ be some σ-field on B such that $\varepsilon(t, b)$ is a measurable function of b for each $t \in \mathbb{T}$. And let $\hat{\mathbf{P}}_1 : \mathcal{B}(B) \to [0, 1]$ be some probability measure defined on $\mathcal{B}(B)$. $\varepsilon(t, b)$ that is considered as a function of $t \in \mathbb{T}$ will be called a sample function of $b \in B$. Let $\mathcal{L} = \mathcal{L}(\mathbb{R}^2)$ be the set of all 2×2 real square matrices. Let $\lambda > 0$ be a given positive constant, and let $\mathbf{\Lambda} \in \mathcal{L}(\mathbb{R}^2)$ be given by

$$\mathbf{\Lambda} := \begin{bmatrix} -\lambda & \lambda \\ \lambda & -\lambda \end{bmatrix}.$$

Let $\hat{\mathbf{Q}} : \mathbb{T} \to \mathcal{L}(\mathbb{R}^2)$ be defined as

$$\hat{\mathbf{Q}}(h) := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sum_{k=1}^{\infty} \frac{h^k}{k!} \mathbf{\Lambda}^k,$$

for each $h \in \mathbb{T}$. Let $\hat{B}(\{1, 2\}^{\mathbb{T}})$ be the set of all functions b in $\{1, 2\}^{\mathbb{T}}$ such that b is a piecewise-continuous function of $t \in \mathbb{T}$ and such that b is continuous on the right at each discontinuous point in $\mathbb{T}$. Let $\mathbb{N}$ be the set of all positive integers, i.e., $\mathbb{N} := \{1, 2, \ldots\}$. For $b \in B \cap \hat{B}(\{1, 2\}^{\mathbb{T}})$ and for $m \in \mathbb{N}$, let $\hat{t}(m, b) \in \mathbb{T}$ be the mth discontinuous point of the sample function $\varepsilon(t, b)$ of b if the mth discontinuous point exists. Then, there exists a stochastic process $(B, \mathcal{B}(B), \hat{\mathbf{P}}_1)$ that satisfies the following conditions. See Proposition 1 in Murakami et al. (2016) for the proof and for the reference to the relevant parts of Doob (1953).

Proposition 1 *There exists a continuous-time two-state Markov process* $(B, \mathcal{B}(B), \hat{\mathbf{P}}_1)$ *with stationary transition probabilities that satisfies the following conditions:*

(1) The initial probability is given by

$$\hat{\mathbf{P}}_1\{b \in B : \varepsilon(0, b) = 1\} = \frac{1}{2} \wedge \hat{\mathbf{P}}_1\{b \in B : \varepsilon(0, b) = 2\} = \frac{1}{2}.$$

(2) A family of the stationary transition probabilities is given by a family of the matrices $\hat{\mathbf{Q}}(h)$ *with* $h \in \mathbb{T}$.

(3) $B \subset \hat{B}(\{1, 2\}^{\mathbb{T}})$. *[The separability.]*

(4) For each $b \in B$ *and for each* $m \in \mathbb{N}$*, there exists the* m*th discontinuous point* $\hat{t}(m, b)$ *in the sample function* $\varepsilon(t, b)$ *of* b*, and* $\lim_{m\to\infty} \hat{t}(m, b) = \infty$.

Let $\mathbb{N}_0$ denote the set of all nonnegative integers, i.e., $\mathbb{N}_0 := \{0\} \cup \mathbb{N}$. Let $\tau : \mathbb{N}_0 \times B \to \mathbb{T}$ be a function constructed in the following way. For each $b \in B$, set $\tau(0, b) = 0$. For each $m \in \mathbb{N}$ and for each $b \in B$, set $\tau(m, b) = \hat{t}(m, b)$. Then, by Proposition 1, $\tau(m, b) \in \mathbb{T}$ is well-defined for all $(m, b) \in \mathbb{N}_0 \times B$. By construction, and since $B \subset \hat{B}(\{1, 2\}^{\mathbb{T}})$, $\tau(0, b) = 0$ for each $b \in B$, and $\tau(m + 1, b) - \tau(m, b) > 0$ for each $(m, b) \in \mathbb{N}_0 \times B$. Let $\mathcal{B}_t(B)$ be *the smallest* σ*-field* with respect to which $(\varepsilon(s, b), 0 \leq s \leq t)$ is a family of measurable functions of $b \in B$. For each $m \in \mathbb{N}_0$, if $t \geq \tau(m, b)$, $\tau(m, b)$ is a $\mathcal{B}_t(B)$-measurable function of b, and if $\tau(m, b) > s \geq 0$, $\tau(m, b)$ is not $\mathcal{B}_s(B)$-measurable function of b. Proposition 2 in Murakami et al. (2016) shows that each element of the set of random variables $\{\tau(m + 1, b) - \tau(m, b)\}_{m\in\mathbb{N}_0}$ is independently and identically subject to a exponential distribution with a parameter $\lambda > 0$.

2.3 Stationary Sunspot Equilibrium

In the present section, we assume that the underlying deterministic dynamics $[\dot{X}, \dot{u}, \dot{Q}]^{\mathbf{T}} = F(X, u, Q)$ satisfies Assumption 1 and either of Assumptions 2 and 3. Under these assumptions, we define a stationary sunspot equilibrium *formally*, and state that there exists a stationary sunspot equilibrium thus defined. The proof of the statement will be given in Sect. 3. Let $\hat{D} \subset W$ be defined as $\hat{D} := \{(X, u, Q) \in W : (X, u) \in D \wedge Q = \varphi(X, u)\}$, where φ and D are specified as in Assumption 1 and either of Assumptions 2 and 3, respectively. Let $\mathcal{B}(\hat{D})$ be the Borel σ-field on $\hat{D}$, and let $\hat{\mathbf{P}}_0 : \mathcal{B}(\hat{D}) \to [0, 1]$ be some probability measure on $\mathcal{B}(\hat{D})$. Let $(B, \mathcal{B}(B), \hat{\mathbf{P}}_1)$ be the probability space the existence of which is assured by Proposition 1, and let $\Omega := \hat{D} \times B$ and let $\mathcal{B}_\Omega = \mathcal{B}(\Omega)$ be the product σ-field of $\mathcal{B}(\hat{D})$ and $\mathcal{B}(B)$, i.e., $\mathcal{B}(\Omega) := \mathcal{B}(\hat{D}) \times \mathcal{B}(B)$. Let $\pi_0 : \Omega \to \hat{D}$ be the projection of $\hat{D} \times B$ onto $\hat{D}$. Let $\pi_1 : \Omega \to B$ be the projection of $\hat{D} \times B$ onto B. We have denoted the tth coordinate of $b \in B$ by $\varepsilon(t, b)$. Let $\mathcal{B}_t(\Omega)$ be *the smallest* σ*-field* with respect to which

$(\pi_0(\omega), \varepsilon(s, \pi_1(\omega)), 0 \le s \le t)$ is a family of measurable functions of $\omega \in \Omega$. Let $\mathbf{P} : \mathcal{B}(\Omega) \to [0, 1]$ be defined as the product measure of $\hat{\mathbf{P}}_0$ and $\hat{\mathbf{P}}_1$, and let $\mathbf{E}_t[\cdot]$ be the conditional expectation operator relative to $\mathcal{B}_t(\Omega)$. We denote a set of all functions from $\mathbb{T}$ to $\hat{D}$ by $\hat{D}^{\mathbb{T}}$. Let $\hat{l} : \Omega \to \hat{D}^{\mathbb{T}}$ be a function such that the tth coordinate of $\hat{l}(\omega) \in \hat{D}^{\mathbb{T}}$ is $\mathcal{B}_\Omega$-measurable function of $\omega \in \Omega$ for each $t \in \mathbb{T}$. Let $(X(t, \omega), u(t, \omega), Q(t, \omega))$ denote the tth coordinate of $\hat{l}(\omega) \in \hat{D}^{\mathbb{T}}$. Let $\mathcal{B}(\hat{D}^{\mathbb{T}})$ be some σ-field on $\hat{D}^{\mathbb{T}}$, and let $\hat{\mathbf{P}} : \mathcal{B}(\hat{D}^{\mathbb{T}}) \to [0, 1]$ be some probability measure defined on $\mathcal{B}(\hat{D}^{\mathbb{T}})$. We define a stationary sunspot equilibrium in the following way.

Definition 1 If the probability measure $\mathbf{P} : \mathcal{B}(\Omega) \to [0, 1]$ satisfies the following conditions, then function $\hat{l} : \Omega \to \hat{D}^{\mathbb{T}}$ constitutes a stationary sunspot equilibrium.

(1) For each $t \in \mathbb{T}$, $(X(t, \omega), u(t, \omega), Q(t, \omega)) \in \hat{D}$ is a $\mathcal{B}_t(\Omega)$-measurable function of $\omega \in \Omega$.
(2) The distribution of $(X(0, \omega), u(0, \omega), Q(0, \omega))$ is given by $(\hat{D}, \mathcal{B}(\hat{D}), \hat{\mathbf{P}}_0)$.
(3) There exists a stochastic process $(\hat{D}^{\mathbb{T}}, \mathcal{B}(\hat{D}^{\mathbb{T}}), \hat{\mathbf{P}})$ on $\hat{D}^{\mathbb{T}}$ such that if $\{t_i\}_{i=1}^N$ is a given set of points in $\mathbb{T}$ with $N \ge 1$, and if $\hat{Y}$ is a given Borel subset of $\hat{D}^N$, then

$$\begin{aligned}&\mathbf{P}\{\omega \in \Omega : (\hat{l}(t_1, \omega), \dots, \hat{l}(t_N, \omega)) \in \hat{Y}\}\\ &\quad = \hat{\mathbf{P}}\{\hat{d} \in \hat{D}^{\mathbb{T}} : (\hat{d}(t_1), \dots, \hat{d}(t_N)) \in \hat{Y}\},\end{aligned}$$

where $\hat{l}(t, \omega) := (X(t, \omega), u(t, \omega), Q(t, \omega)) \in \hat{D}$, and $\hat{d}(t)$ denotes the tth coordinate of $\hat{d} \in \hat{D}^{\mathbb{T}}$. [The existence of a stochastic process.]
(4) For each $\omega \in \Omega$, $X(t, \omega)$ is a continuous function of $t \in \mathbb{T}$, and for each $t > 0$,

$$\mathbf{P}\{\omega \in \Omega : \lim_{h \to 0} \frac{X(t+h, \omega) - X(t, \omega)}{h} = f_1(X(t, \omega), u(t, \omega), Q(t, \omega))\} = 1.$$

(5) For each $\omega \in \Omega$, $(u(t, \omega), Q(t, \omega))$ is a piecewise-continuous function of $t \in \mathbb{T}$ and continuous on the right at each discontinuous point in $\mathbb{T}$, and for each $t \in \mathbb{T}$,

$$\begin{bmatrix} \lim_{h \to +0} \frac{X(t+h,\omega) - X(t,\omega)}{h} \\ \mathbf{E}_t[\lim_{h \to +0} \frac{u(t+h,\omega) - u(t,\omega)}{h}] \\ \mathbf{E}_t[\lim_{h \to +0} \frac{Q(t+h,\omega) - Q(t,\omega)}{h}] \end{bmatrix} = \begin{bmatrix} f_1(X(t, \omega), u(t, \omega), Q(t, \omega)) \\ f_2(X(t, \omega), u(t, \omega), Q(t, \omega)) \\ f_3(X(t, \omega), u(t, \omega), Q(t, \omega)) \end{bmatrix}.$$

(6) For each $t > s \ge 0$, $(X(t, \omega), u(t, \omega), Q(t, \omega))$ is not $\mathcal{B}_s$-measurable function of $\omega \in \Omega$.
(7) If $\{t_i\}_{i=1}^N$ is a given set of points in $\mathbb{T}$ with $N \ge 1$, and if $\hat{Y}$ is a given Borel subset of $\hat{D}^N$, for any $h \in \mathbb{R}$ such that $\{t_i + h\}_{i=1}^N \subset \mathbb{T}$,

$$\begin{aligned}&\mathbf{P}\{\omega \in \Omega : (\hat{l}(t_1, \omega), \dots, \hat{l}(t_N, \omega)) \in \hat{Y}\}\\ &\quad = \mathbf{P}\{\omega \in \Omega : (\hat{l}(t_1 + h, \omega), \dots, \hat{l}(t_N + h, \omega)) \in \hat{Y}\},\end{aligned}$$

where $\hat{l}(t, \omega) := (X(t, \omega), u(t, \omega), Q(t, \omega)) \in \hat{D}$. [The stationarity.]

In Sect. 3, we will show the following theorem by means of the method due to Murakami et al. (2016).

Theorem 1 *Suppose that the underlying deterministic dynamics* $[\dot{X}, \dot{u}, \dot{Q}]^T = F(X, u, Q)$ *satisfies Assumption 1 and either of Assumptions 2 and 3. Then, there exists a stationary sunspot equilibrium.*

2.4 On Relations of the Present Result to Other Results

Before leaving Sect. 2, we relate our result to those of Shigoka (1994), Nishimura and Shigoka (2006), Benhabib et al. (2008), and Murakami et al. (2016) here. Let $g_i : V \to \mathbb{R}, i = 1, 2$, be defined as $g_i(X, u) := f_i(X, u, \varphi(X, u)), i = 1, 2$. Suppose that the vector field $[\dot{X}, \dot{u}]^{\mathbf{T}} = G(X, u)$ satisfies either of Assumptions 2 and 3. Let $\{(X(t, \omega), u(t, \omega))\}_{t \in \mathbb{T}}$ be a set of random variables that will have been constructed in Sect. 3. Then, we have the following:

(1) For each $\omega \in \Omega$, $X(t, \omega)$ is a continuous function of $t \in \mathbb{T}$, and for each $t > 0$,

$$\mathbf{P}\{\omega \in \Omega : \lim_{h \to 0} \frac{X(t+h, \omega) - X(t, \omega)}{h} = g_1(X(t, \omega), u(t, \omega))\} = 1.$$

(2) For each $\omega \in \Omega$, $u(t, \omega)$ is a piecewise-continuous function of $t \in \mathbb{T}$ and continuous on the right at each discontinuous point in $\mathbb{T}$, and for each $t \in \mathbb{T}$,

$$\begin{bmatrix} \lim_{h \to +0} \frac{X(t+h,\omega)-X(t,\omega)}{h} \\ \mathbf{E}_t[\lim_{h \to +0} \frac{u(t+h,\omega)-u(t,\omega)}{h}] \end{bmatrix} = \begin{bmatrix} g_1(X(t, \omega), u(t, \omega)) \\ g_2(X(t, \omega), u(t, \omega)) \end{bmatrix}.$$

(3) For each $t > s \geq 0$, $(X(t, \omega), u(t, \omega))$ is not $\mathcal{B}_s$-measurable function of $\omega \in \Omega$.

(4) If $\{t_i\}_{i=1}^N$ is a given set of points in $\mathbb{T}$ with $N \geq 1$, and if Y is a given Borel subset of D^N, for any $h \in \mathbb{R}$ such that $\{t_i + h\}_{i=1}^N \subset \mathbb{T}$,

$$\begin{aligned} &\mathbf{P}\{\omega \in \Omega : (l(t_1, \omega), \ldots, l(t_N, \omega)) \in Y\} \\ &\quad = \mathbf{P}\{\omega \in \Omega : (l(t_1 + h, \omega), \ldots, l(t_N + h, \omega)) \in Y\}, \end{aligned}$$

where $l(t, \omega) := (X(t, \omega), u(t, \omega)) \in D$. [The stationarity.]

The existence of such a set of random variables $\{(X(t, \omega), u(t, \omega))\}_{t \in \mathbb{T}}$ is also the assertion of Theorem 1 in Shigoka (1994).

The present study provides Theorem 1 in Nishimura and Shigoka (2006) with an alternative proof that is due to Murakami et al. (2016). The proof due to Nishimura and Shigoka (2006) is an extension of that due to Shigoka (1994), whereas the proof due to Murakami et al. (2016) is an extension of that due to Benhabib et al. (2008). Shigoka (1994) treats a deterministic model that includes either a steady state or a closed orbit with a two-dimensional stable manifold, whereas Benhabib et al.

(2008) treats a two-dimensional deterministic model that includes one predetermined variable and one non-predetermined variable and that is amenable to the existence of a homoclinic orbit with multiple steady states. The former deterministic model is simpler than the latter deterministic model. On the other hand, the specification of a stochastic differential equation in Shigoka (1994) is more complex than that in Benhabib et al. (2008). According to the specification due to Shigoka (1994),

$$du(t,\omega) = g_2(X(t,\omega), u(t,\omega)) + \bar{s}d\varepsilon(t, \pi_1(\omega)),$$

where $\bar{s}$ is some constant with $\bar{s} \neq 0$, and where $du(t,\omega)$ and $d\varepsilon(t, \pi_1(\omega))$ denote Lebesgue-Stieljes signed measures relative to $t \in \mathbb{T}$, respectively. According to the specification due to Benhabib et al. (2008),

$$\lim_{h \to +0} \frac{u(t+h,\omega) - u(t,\omega)}{h} = g_2(X(t,\omega), u(t,\omega)).$$

Although Benhabib et al. (2008) does not include the proof of the stationarity, Murakami et al. (2016) includes the proof of this. The proof of the present study is simpler than that of Nishimura and Shigoka (2006), because the stochastic differential equation in Benhabib et al. (2008) is simpler than that in Shigoka (1994).

3 Proof of Theorem 1

In the present section, we will prove Theorem 1. We assume that Assumption 1 and either of Assumptions 2 and 3 are satisfied. Let $U \subset \mathbb{R}^2$ be a set of all interior points of D that is the closed subset specified in either of Assumptions 2 and 3. Then there exists an open subset N of $\mathbb{R} \times V$ such that $\mathbb{T} \times D \subset N \subset \mathbb{R} \times V$ and there exists a continuous function $\phi : N \to V$ that satisfies following conditions:

(1) For each $(t, X, u) \in N$, $\phi(t, X, u)$ is C^1-function of t, with $\phi(0, X, u) = (X, u) \in V$.
(2) For each $(X, u) \in D$,

$$\lim_{h \to 0} \frac{\phi(t+h, X, u) - \phi(t, X, u)}{h} = G(\phi(t, X, u)).$$

(3) $\phi(\mathbb{T} \times D) \subset D$, and $\phi(\mathbb{T} \times U) \subset U$.

Let $d : D \to \mathbb{R}_+$ be defined as

$$d(X,u) := \min \sqrt{(X - x_1)^2 + (u - x_2)^2} \text{ subject to } (x_1, x_2) \in \partial D.$$

Since ∂D is a compact set, $d = d(X, u)$ is well-defined, and since $d = d(X, u)$ is a distance between a point in D and the set ∂D, $d = d(X, u)$ is a continuous function of $(X, u) \in D$. Since U is the interior region of D, and since ∂D is the boundary of U, for any $(X, u) \in U$, $d = d(X, u) > 0$, and $(X, u + \frac{1}{2}d(X, u)) \in U$.

Let $\hat{\mathbf{P}}_0 : \mathcal{B}(\hat{D}) \to [0, 1]$ be a probability measure that satisfies

$$\hat{\mathbf{P}}_0\{(X, u, Q) \in \hat{D} : (X, u) \in U\} = 1.$$

We have defined $\mathbf{P} : \mathcal{B}(\Omega) \to [0, 1]$ as the product measure of $\hat{\mathbf{P}}_0$ and $\hat{\mathbf{P}}_1$, where $\hat{\mathbf{P}}_1 : \mathcal{B}(B) \to [0, 1]$ is the probability measure in Proposition 1. Let $\hat{\pi} : \hat{D} \to D$ be the projection of $\hat{D}$ onto D so that $\hat{\pi}(X, u, Q) = (X, u)$ for $(X, u, Q) \in \hat{D}$. Then, we have

$$\mathbf{P}\{\omega \in \Omega : (X, u) \in U\} = 1.$$

Since $\pi_1(\omega) = b$, $\varepsilon(t, \pi_1(\omega))$ and $\tau(m, \pi_1(\omega))$ are measurable functions of $\omega \in \Omega$. We have defined $\mathcal{B}_t(\Omega)$ as the smallest σ-field with respect to which $(\pi_0(\omega)$, $\varepsilon(s, \pi_1(\omega)), 0 \le s \le t)$ is a family of measurable functions of $\omega \in \Omega$. For each $m \in \mathbb{N}_0$, if $t \ge \tau(m, \pi_1(\omega))$, then $\tau(m, \pi_1(\omega))$ is a $\mathcal{B}_t(\Omega)$-measurable function of $\omega \in \Omega$, because $t \ge \tau(m, b)$ so that $\tau(m, b)$ is a $\mathcal{B}_t(B)$-measurable function of $b \in B$.

Since $\phi(\mathbb{T}, U) \subset U$ and since $\tau(0, \pi_1(\omega)) = 0 \wedge \tau(m + 1, \pi_1(\omega)) - \tau(m, \pi_1(\omega)) > 0 \wedge \lim_{m\to\infty} \tau(m, \pi_1(\omega)) = \infty$ for each $(m, \omega) \in \mathbb{N}_0 \times \Omega$, the following constructions are well-defined. Let $f : \mathbb{N}_0 \times \Omega \to U$ and $g : \mathbb{N}_0 \times \Omega \to U$ be defined as follows. For $\omega \in \Omega$, if $\hat{\pi}(\pi_0(\omega)) = (X, u) \in U$, let $f(0, \omega)$ and $g(0, \omega)$ be given by

$$f(0, \omega) := (\hat{\pi}(\pi_0(\omega))),$$
$$g(0, \omega) := f(0, \omega).$$

Choose some specific point $(X^{'}, u^{'})$ from U in advance, and for $\omega \in \Omega$, if $\hat{\pi}(\pi_0(\omega)) = (X, u) \in \partial D$, let $f(0, \omega)$ and $g(0, \omega)$ be given by

$$f(0, \omega) := (X^{'}, u^{'}),$$
$$g(0, \omega) := f(0, \omega).$$

For $(m, \omega) \in \mathbb{N}_0 \times \Omega$, and for given $f(m, \omega)$ and $g(m, \omega)$, let $f(m + 1, \omega)$ and $g(m + 1, \omega)$ be given by

$$f(m + 1, \omega) := \phi(\tau(m + 1, \pi_1(\omega)) - \tau(m, \pi_1(\omega)), g(m, \omega)),$$
$$g(m + 1, \omega) := f(m + 1, \omega) + (0, \frac{1}{2}d(f(m + 1, \omega))).$$

Then, $f(m, \omega) \subset U \wedge g(m, \omega) \subset U$ for each $(m, \omega) \in \mathbb{N}_0 \times \Omega$. For each $m \in \mathbb{N}_0$, if $t \ge \tau(m, \pi_1(\omega))$, $f(m, \omega)$ and $g(m, \omega)$ are $\mathcal{B}_t(\Omega)$-measurable functions of $\omega \in \Omega$.

Note that for each $(t, \omega) \in \mathbb{T} \times \Omega$, there exists a unique element m in $\mathbb{N}_0$ such that $\tau(m, \pi_1(\omega)) \le t < \tau(m + 1, \pi_1(\omega))$. Let $\theta : \mathbb{T} \times \Omega \to U$ be defined as follows. For each $(t, \omega) \in \mathbb{T} \times \Omega$, if $\tau(m, \pi_1(\omega)) \le t < \tau(m + 1, \pi_1(\omega))$, let $\theta(t, \omega)$ be given by

$$\theta(t, \omega) := \phi(t - \tau(m, \pi_1(\omega)), g(m, \omega)).$$

Then, for each $(t, \omega) \in \mathbb{T} \times \Omega$, $\theta(t, \omega) \in U$, and for each $t \in \mathbb{T}$, $\theta(t, \omega)$ is a $\mathcal{B}_t(\Omega)$-measurable function of $\omega \in \Omega$. For each $(t, \omega) \in \mathbb{T} \times \Omega$, let $(X(t, \omega), u(t, \omega), Q(t, \omega))$ be defined as

$$(X(t, \omega), u(t, \omega), Q(t, \omega)) := (\theta(t, \omega), \varphi(\theta(t, \omega)).$$

Then, we can use the same argument as that of Sect. 5.2 in Murakami et al. (2016) to show that a set of random variables $\{(X(t, \omega), u(t, \omega), Q(t, \omega))\}_{t \in \mathbb{T}}$ thus constructed satisfies the conditions (1)–(6) in Definition 1. We can use the same arguments as that of Sect. 5.3 in Murakami et al. (2016) to show that a set of random variables $\{(X(t, \omega), u(t, \omega), Q(t, \omega), \varepsilon(t, \pi_1(\omega)))\}_{t \in \mathbb{T}}$ is subject to a Markov process with stationary transition probabilities and that there is an invariant measure on $\hat{D} \times \{1, 2\}$ such that if we assign this measure as an initial probability measure to $\hat{D} \times \{1, 2\}$, then the resulting stochastic process is stationary, which implies that the condition (7) in Definition 1 is satisfied.

References

Azariadis C (1981) Self-fulfilling prophecies. J Econ Theory 25:380–396

Benhabib J, Nishimura K, Shigoka T (2008) Bifurcation and sunspots in the continuous time equilibrium model with capacity utilization. Int J Econ Theory 4:337–355

Cass D, Shell K (1983) Do sunspots matter? J Polit Econ 91:193–227

Chiappori PA, Guesnerie R (1991) Sunspot equilibria in sequential market models. In: Hildenbrand W, Sonnenschein H (eds) Handbook of mathematical economics, vol 4. North-Holland, Amsterdam, pp 1683–1762

Doob JL (1953) Stochastic processes. Wiley, New York

Guesnerie R, Woodford M (1992) Endogenous fluctuations. In: Laffont JJ (ed) Advances in economic theory, sixth world congress, vol 2. Cambridge University Press, New York, pp 289–412

Murakami H, Nishimura K, Shigoka T (2016) Homoclinic orbit and stationary sunspot equilibrium in a three-dimensional continuous-time model with a predetermined variable. In: Nishimura K, Venditti A, Yannelis NC (eds) Sunspots and non-linear dynamics. Springer (forthcoming)

Nishimura K, Shigoka T (2006) Sunspots and Hopf bifurcations in continuous time endogenous growth models. Int J Econ Theory 2:199–216

Shell K (1977) Monnaie et allocation intertemporelle. Seminarie d'Econometrie Roy-Malinvaud, mimeo. Paris

Shigoka T (1994) A note on Woodford's conjecture: constructing stationary sunspot equilibria in a continuous time model. J Econ Theory 64:531–540

Woodford M (1986) Stationary sunspot equilibria: the case of small fluctuations around a deterministic steady state, mimeo

Part II
Microeconomics Models

Hotelling Duopoly Revisited

Tõnu Puu

Abstract Many articles have been written about Harold Hotelling's model of two competitors on a fixed line segment, competing by choosing mill price and location. Most of them have focused on the paradoxical case of crowding in the middle when demand is totally inelastic. Yet Hotelling conjectured that this would not happen if the consumers not only chose the least expensive supplier, but their demand were dependent on the price charged. However, surprisingly little has been written about the case with elastic demand. Even less has been attempted to put the problem in a dynamic format.

Keywords Hotelling duopoly · Elastic demand · Dynamic adjustment

1 Introduction

The enigmatic Hotelling model of duopoly on a fixed line interval appeared in 1929 Hotelling (1929). In a sense it was the first well structured case of Bertrand oligopoly Bertrand (1883), as heterogeneity of a physically homogenous good was obtained through spatial separation of suppliers and accrued transportation costs.

Unfortunately, Hotelling only analyzed the case where demand was totally inelastic, which resulted in the paradox that both competitors would crowd in the centre. Yet, Hotelling himself in his verbal discussion stated that the paradox would disappear if demand were elastic—there would remain a tendency to gravitate towards the centre, rather than locating at the socially optimal quartiles, though without the extreme crowding.

It is a bit surprising that Hotelling did not himself follow this track. Perhaps he too was under the spell of paradox, like following authors, and so wanted to emphasize this case. There exists an immense literature in his aftermath. Unfortunately, citation

T. Puu (✉)
CERUM, Umeå University, SE-901 87, Umeå, Sweden
e-mail: tonu.pun@umu.se

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_3

databases do not extend as far back as to 1929, so it is difficult to give an exact estimate.[1]

Analytically, the crowding version of the model is problematic, and even inconsistent. With location in the same point, the space Hotelling introduced disappears again, and we are back at the problems that Bertrand pointed out Bertrand (1883). Inconsistence arises from consumers choosing the lowest delivered price, which creates the market areas, but yet each customer buys one unit of commodity, no matter what its price.

Lerner and Singer (1937) provided the first proof that the crowding paradox evaporated if one only assumed the consumers to have a reservation price; if delivered price was higher they would buy nothing, otherwise just one unit as in the original model. The contribution is still most enjoyable through its ingenious use of graphic argument.

Very soon after, Smithies 1941 in two articles, Smithies (1941a, b) put up the problem with a linear downsloping demand function, which seems to correspond to what Hotelling had in mind. However, Smithies claimed that the integrals were too complicated to evaluate.

So it was left to the present author as late as in 2002 Puu (2002) to carry out the formal analysis of Smithies's case. Dependent on the parameter values (maximum demand price, transportation rate and unit production cost) three outcomes were possible:

(1) Disjoint monopolies. (2) Genuine duopoly. (3) Cutting out monopoly.

The first case, occurs when transportation costs are high and the competitors can locate so widely apart that demand goes to zero before they reach eachother's territories. The third case is ill-structured, like all cases of price war, and still needs a truly convincing layout. The second case is genuine duopoly, well structured and worthy of closer study.

In Puu and Gardini (2002), Professor Laura Gardini and the present author studied the dynamics of this genuine duopoly case. It turned out to be very simple; just a contraction to equilibrium, with locations separated in space, not quite at the quartiles, but not crowding in the middle either—quite as conjectured in Hotelling's original article.[2]

---

[1] No doubt the paradox created its popularity. Scientists replaced distance by just "similarity" in some vague sense for competing products, or even for political opinions, all with doubtful measurability. Such vague analogies deprive such a good scientific model of its qualification as science. Further, to escape some consequences of the paradox, the unrealistic and most contrived idea of quadratic transportation costs was launched and surprisingly gained popularity. d'Aspremont et al. (1979).

[2] However, this analysis (without motivation) assumed that the competitors shared the market as a duopoly in a common boundary point, and that *at the other ends each market extended to the boundary points of Hotelling's fixed interval*.

In a communication to the present author Dr. Helge Sanner pointed out that the last may not be true—the competitors might also end the markets where local demand dropped to zero. Unfortunately the present author has not been able to locate any publication by Dr. Sanner to cite on this.

Further thought, however, indicates that this case would never happen. If one endpoint only extends to the point where demand vanishes, the competitor in question would always profit from

It remains to formulate the entire model under elastic demand and to consider its global dynamics. To provide a basis for such is the very purpose of this contribution.

The first task is to formulate the map, describe its equilibria and study its dynamic.

2 The Model

2.1 Assumptions and Notation

Consider a line interval $[-1, 1]$ in which two competitors take locations x_1, x_2. They charge mill prices p_1, p_2. Given a transport cost rate k, local price at x is $p_i + k\,|x - x_i|$ depending on from which supplier $i = 1, 2$ the commodity is bought.[3] The boundary point is

$$x = \frac{x_1 + x_2}{2} \mp \frac{p_1 - p_2}{2k} \tag{1}$$

(sign depending on whether $x_1 < x_2$ or $x_1 > x_2$).

The boundary point would be halfway between the locations of the competing firms if their mill prices were equal. If not, the point is dragged in the direction of the firm charging the higher mill price, thereby decreasing its market share.

However, it is possible that demand goes to zero before this boundary point is reached. When this happens the firms can have disjoint non-competing monopoly areas.

Assume the linear demand function $q_i = \max\left(a - (p_i + k\,|x - x_i|)\,, 0\right)$.[4] Accordingly, demand for each competitor goes to zero when

$$x = x_i \pm \frac{a - p_i}{k} \tag{2}$$

We could call the expression $\frac{a-p_i}{k}$ market radius. Note that this applies both when we have two disjoint monopolies, and when one firm cuts the other out to establish one single monopoly. Market radius decreases with higher mill price and with higher

(Footnote 2 continued)
moving its location until an endpoint of the fixed interval is reached. This is because greater spatial symmetry of the market resulting from this always increases profit.

[3] Mill pricing, where consumers pay for full transportation costs is our case. However, it is by no means the only possibility. As the competitors are monopolists in their market areas, they can also themselves provide for delivery and apply price discrimination, provided they do not charge more for transport than its actual cost. It is well known that with linear demand perfect discrimination implies charging for exactly half the transportation cost.

[4] A more general formula for demand is $q_i = \max\left(a - b\,(p_i + k\,|x - x_i|)\,, 0\right)$, but experience from work with linear models shows that parameter b has no independent influence, so we simplify by putting $b = 1$.

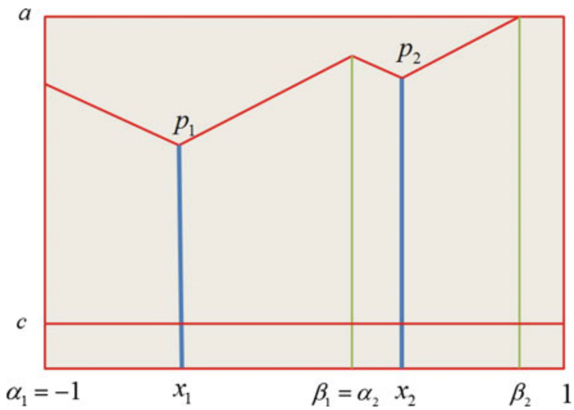


Fig. 1 Picture of the duopoly market case. For variety we displayed the market area of the *left firm* as extending to the interval boundary, whereas that of the *right firm* falls short of it, going only to the point where local demand vanishes due to a too high delivered price

transportation cost. High transportation cost would therefore favour establishing disjoint monopolies.

From (1) we get two candidates for market area endpoints, and (2) adds further four. Taking in account the fixed endpoints of the entire interval, chosen as ∓ 1, we are dealing with eight potential endpoints for market areas.

The geometry of the format is shown in Fig. 1.

2.2 *Total Demand and Profits*

Let us so introduce symbols for these market area endpoints: α_i, β_i. Given these, we can calculate total demand

$$Q_i = \int_{\alpha_i}^{\beta_i} q_i dx = \int_{\alpha_i}^{\beta_i} a - (p_i + k\,|x - x_i|)\, dx$$

Despite Smithies's doubts, integration is simple:

$$Q_i = (a - p_i)(\beta_i - \alpha_i) - \frac{k}{2}\left((x_i - \alpha_i)^2 + (\beta_i - x_i)^2\right) \tag{3}$$

Profits then are

$$G_i = (p_i - c)\, Q_i \tag{4}$$

where c denotes constant unit production cost—for simplicity taken constant and equal for both competitors.[5]

[5]It should be noted that Hotelling assumed *production cost to be zero*, like many other oligopoly theorists, such as Cournot Cournot (1838), and von Stackelberg von Stackelberg (1934). Equal, but nonzero costs do not complicate things notably, and provide a more reasonable first approximation.

Note in particular that the price minus cost factor in (4) does not depend on location. Hence, whenever we look for an optimal location, we can just maximize (3).

2.3 Market Area Endpoints

From our previous digression, it is obvious that the left endpoint of a market area α_i can take on the following values:

$$\alpha_i = \begin{cases} -1 \\ x_i - \frac{a-p_i}{k} \\ \frac{x_i+x_j}{2} + \frac{p_i-p_j}{2k} \end{cases} \tag{5}$$

and the right endpoint β_i

$$\beta_i = \begin{cases} \frac{x_i+x_j}{2} - \frac{p_i-p_j}{2k} \\ x_i + \frac{a-p_i}{k} \\ 1 \end{cases} \tag{6}$$

Obviously we can combine each left market endpoint α_i with each right endpoint β_i, except the case $\alpha_i = \frac{x_i+x_j}{2} + \frac{p_i-p_j}{2k}$, $\beta_i = \frac{x_i+x_j}{2} - \frac{p_i-p_j}{2k}$ as in a shared market the left endpoint of the firm to the right cannot be combined with the right endpoint of the firm to the left.

This leaves eight cases[6]:

$$[\alpha_i, \beta_i] = \begin{cases} \left[-1, \frac{x_i+x_j}{2} - \frac{p_i-p_j}{2k}\right], & \mathbf{1} \\ \left[-1, x_i + \frac{a-p_i}{k}\right], & \mathbf{2} \\ [-1, 1], & \mathbf{3} \\ \left[x_i - \frac{a-p_i}{k}, \frac{x_i+x_j}{2} - \frac{p_i-p_j}{2k}\right], & \mathbf{4} \\ \left[x_i - \frac{a-p_i}{k}, x_i + \frac{a-p_i}{k}\right], & \mathbf{5} \\ \left[x_i - \frac{a-p_i}{k}, 1\right], & \mathbf{6} \\ \left[\frac{x_i+x_j}{2} + \frac{p_i-p_j}{2k}, x_i + \frac{a-p_i}{k}\right], & \mathbf{7} \\ \left[\frac{x_i+x_j}{2} + \frac{p_i-p_j}{2k}, 1\right], & \mathbf{8} \end{cases} \tag{7}$$

[6] Actually they represent even more cases as a generic firm i could locate to left or right. Further a monopoly case can represent disjoint monopolies, or cutting out the competitor.

This list exhausts all possibilities for total market demand, as only endpoints matter in (3). In most cases, the evaluation is quite simple. A graphic summary of the regions is displayed in Fig. 5.

Also recall that only total demand matters for the choice of optimal location.

2.4 Interpretation of the Cases

2.4.1 Shared Duopoly Market

The duopoly cases are listed as **1**, **4**, **7** and **8**. In all they are four, because given the location of one competitor, the other may choose to locate to the left or to right of it. Further, its market at the outer end can extend to the respective interval endpoint, or stop at the point where local demand vanishes. These are the most interesting cases of the Hotelling model, as they deal with genuine duopoly. As mentioned, cases **1** and **8** were investigated in Puu and Gardini (2002), and it was about the cases **4** and **7** that the author was alerted by Dr. Sanner (Figs. 2, 3 and 4).

2.4.2 Monopolies

Cases **2**, **5** and **6** represent monopoly. We can either have two coexistent disjoint monopolies, not sharing any common boundary point, or one competitor cutting out the other. Case **3** too is a cutting out monopoly that covers the entire fixed interval. This case is what has been focused in the literature.

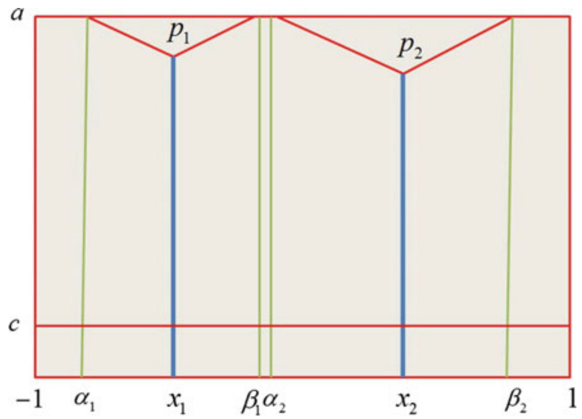


Fig. 2 A case of disjoint monopolies on the fixed interval that do not compete. Such a case is most likely to occur when transportation costs are relatively high, so that the monopoly areas within which local demand remains positive can be accommodated within the fixed interval

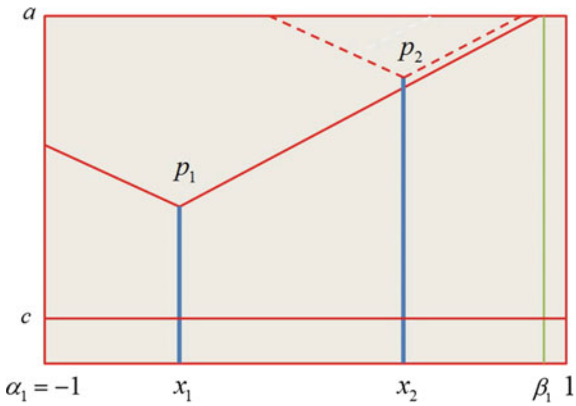


Fig. 3 Cutout monopoly. Through lowering mill price the *left firm* makes delivered price at the location of the *right firm* lower than the mill price of the latter, so taking the entire market. This will, of course, be contested by the cut out firm in further moves

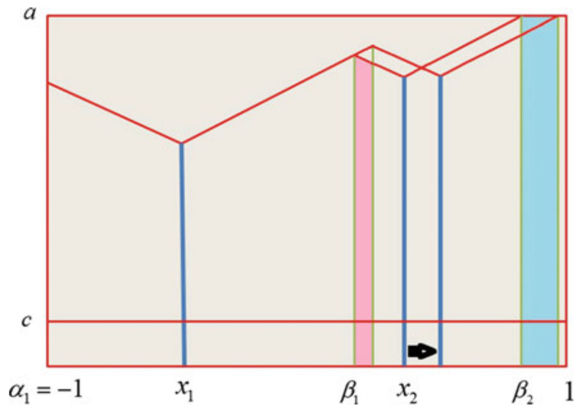


Fig. 4 Gain (*right*) and loss (*left*) strips of total sales when the firm to the right takes a step towards the right boundary of the fixed interval. Note that it does not quite stretch out the market for the firm to the right to this right boundary—so, there remains a little further possible gain in sales

2.5 Demand and Optimal Location

We can now use the endpoints listed in (7) to calculate total demand for the firms according to (3). Recall that the choice of location in all cases can be done through maximizing demand, as the multiplicative factor $(p_i - c)$ in the expression for profits (4) does not include location.

After finding optimal locations that maximize demand, we can substitute it out, and obtain expressions for profits that only depend on mill price. So, let us take the list (7) case by case.

2.5.1 Cases 1 and 8 (Duopoly)

These true oligopoly situations are the only tough cases. It was these that, however, were investigated in Puu and Gardini (2002) and shown to be simple contractions in the dynamic adjustment process.

In cases **1** and **8** the market of firm 1 extends from the left endpoint $\alpha_1 = -1$ of the interval to a common market boundary point $\beta_1 = \frac{x_1+x_2}{2} - \frac{p_2-p_1}{2k}$. Substituting in (3), we get

$$Q_1 = (a - p_1)(1 + \beta_1) - \frac{k}{2}\left((1 + x_1)^2 + (\beta_1 - x_1)^2\right)$$

Maximizing through putting the derivative $\frac{\partial Q_1}{\partial x_1} = 0$, we can solve for a new optimum location

$$x_1' = \frac{x_2 - 4}{5} + \frac{2a - 3p_1 + p_2}{5k} \tag{8}$$

Note that, as a consequence of choosing x_1' according to (8), the right market boundary point is changed to

$$\beta_1' = \frac{3x_2 - 2}{5} + \frac{a - 4p_1 + 3p_2}{5k} \tag{9}$$

This consequence of location choice is perceived by the firm, so, along with x_1' from (8), β_1' from (9), must be substituted back in its estimate of demand.

Defining

$$\lambda_1 = p_2 + k(1 + x_2) \tag{10}$$

the resulting total demand becomes

$$Q_1' = \frac{1}{10k}\left(6(a - p_1)^2 - 4(a - p_1)(a - \lambda_1) - (a - \lambda_1)^2\right) \tag{11}$$

Likewise for firm 2, whose market extends from $\alpha_2 = \frac{x_1+x_2}{2} - \frac{p_2-p_1}{2k}$ to $\beta_2 = 1$, substitution in (3) yields

$$Q_2 = (a - p_1)(1 - \alpha_2) - \frac{k}{2}\left((\alpha_2 - x_2)^2 + (1 - x_2)^2\right)$$

Putting $\frac{\partial Q_2}{\partial x_2} = 0$ and solving, we obtain

$$x_2' = \frac{x_1 + 4}{5} - \frac{2a - 3p_2 + p_1}{5k} \tag{12}$$

Again this changes the left market boundary point to

$$\alpha_2' = \frac{3x_1 + 2}{5} - \frac{a - 4p_2 + 3p_2}{5k} \tag{13}$$

which, again, along with substitution of the new location, must be substituted back to get a total demand estimate. Defining

$$\lambda_2 = p_2 + k\,(1 - x_2) \tag{14}$$

we get

$$Q_2' = \frac{1}{10k}\left(6\,(a - p_2)^2 - 4\,(a - p_2)\,(a - \lambda_2) - (a - \lambda_2)^2\right) \tag{15}$$

Note two things: (i) the slight change of sign in (10) as compared to (14), and that, given this, (11) and (15) have exactly the same form.

Also, note that λ_1, λ_2 contain all information about the competitor's location and mill price. About these any firm can only have expectations, and, as usual, we will take the simplest possibility—that each firm believes the other to retain their previous moves, even if they do not in a dynamic process.

This makes total demand, given optimal location choice, quadratic in its own mill price for each firm.

Then multiplication by $(p_i - c)$ makes profit a cubic in mill price. So, once we proceed to optimize with respect to mill price, first-order conditions result in a quadratic to solve—though it is obvious which of its two solution roots provides profit maximum. This analysis is carried out below.[7]

2.5.2 Cases 4 and 7 (Duopoly)

We can now pass to the cases where the firms still share a common market boundary point, but the other market endpoint falls short of the interval limit because local demand goes to zero before it is reached. I will call it the Sanner Case.

Again there are two cases, depending on which side of the prelocated competitor the firm chooses.

Surprisingly these cases are considerably simpler than the previous. Taking case **4**, the endpoints of the market are $\alpha_1 = x_1 - \frac{a - p_1}{k}$, the point where local demand vanishes and $\beta_1 = \frac{x_1 + x_2}{2} - \frac{p_1 - p_2}{2k}$, where the delivered prices of the firms break even.

Substituting in (3) and putting $\frac{\partial Q_1}{\partial x_1} = 0$, we get the optimal location

$$x_1' = x_2 - \frac{2a - p_1 - p_2}{k} \tag{16}$$

which substituted back in (3) yields

$$Q_1' = \frac{1}{k}\,(a - p_1)^2 \tag{17}$$

This is a formula we will encounter repeatedly in the sequel.

---

[7] We chose to denote the firm to the left 1, and the one to the right 2. This works in each single step, but it is fully possible, and shows up in numerical experiment, that each firm can move to the other side of its competitor, and so we must keep track of the numbering.

In case **7**, location is to the right of the competitor, between points $\alpha_2 = \frac{x_1+x_2}{2} + \frac{p_1-p_2}{2k}$ and, $\beta_2 = x_2 + \frac{a-p_2}{k}$, again substituting in (3), and putting $\frac{\partial Q_2}{\partial x_2} = 0$ we get optimal location

$$x_2' = x_1 + \frac{2a - p_i - p_j}{\kappa} \tag{18}$$

Substituting back in (3) we again find the same formula as (17)

$$Q_2' = \frac{1}{k}(a - p_2)^2 \tag{19}$$

Now there is a simple argument why the Sanner Case will never be chosen in duopoly: Given any mill price, total demand will increase with the degree of symmetry of the market, which is obtained through approaching the interval limit ∓ 1 on the side of the firm. Then, either such move results in monopoly if there is space enough, or in duopoly with a larger market demand. Further, the competitor has no reason to retaliate at any stage because by such move the firm decreases competition in the boundary point.

Digression on the Sanner Case

We can make this clear by a simple graphic argument. Take the case illustrated in Fig. 4, where the market for the firm to the right does not stretch out to the right interval endpoint. Then assume the firm changes location as indicated by the arrow.

The vertical shaded vertical strip to the right indicates the gain of total sales, the one to the left, the loss to the competitor. As we see, the loss area is much smaller than the gain area.

When the firm changes location, the gain area on the right has its boundary pushed exactly the same distance as the change of location. However, the left market area is moved much less, due to the sloping delivered price line of the competitor. This is, of course, due to transportation cost, but the exact level for it does not matter—the left loss strip remains narrower than the right gain strip. Further, in terms of area (total sales), it is bounded above by the delivered price line, whereas the gain strip to the right goes all way up to maximum chargeable price.[8]

2.5.3 Cases 2, 6 and 5 (Disjoint Monopolies)

The remaining cases are monopolies, coexistent and disjoint, or single cutting out monopolies. The endpoint alternatives are the same, whether there is coexistence or cutting out. Intuitively it seems that, given transportation costs are high, and the firms can accommodate their relatively small market areas in the preassigned segment, then there is no point in going to extremes such as cutting out.

[8]Of course, a change of location will be followed by a change of mill price. But this is a further issue, location can be chosen so as to maximize sales alone.

Let us start with the monopoly cases where one market endpoint extends to the fixed interval limit. If to the left, case **2**, we have market endpoints $\alpha_1 = -1$ and $\beta_1 = x_1 + \frac{a-p_1}{k}$. Then, from (3)

$$Q_1 = (a - p_1)\left(1 + x_1 + \frac{a - p_1}{k}\right) - \frac{k}{2}\left((1 + x_1)^2 + \left(\frac{a - p_1}{k}\right)^2\right)$$

Putting $\frac{\partial Q_i}{\partial x_1} = 0$,

$$x_1' = \frac{a - p_1}{k} - 1 \tag{20}$$

which substituted back yields

$$Q_1 = \frac{1}{k}(a - p_1)^2 \tag{21}$$

i.e. again the same formula as (17).

For case **6**, the firm locates on the right, with market limits $\alpha_2 = x_2 - \frac{a-p_2}{k}$ and $\beta_2 = 1$,

$$Q_2 = (a - p_2)\left(1 - x_2 + \frac{a - p_2}{k}\right) - \frac{k}{2}\left((1 - x_2)^2 + \left(\frac{a - p_2}{k}\right)^2\right)$$

The optimal location is

$$x_2' = 1 - \frac{a - p_2}{k} \tag{22}$$

Upon substitution back it again yields

$$Q_1 = \frac{1}{k}(a - p_1)^2 \tag{23}$$

quite as (21).

Finally, let us so consider case **5**, a monopoly where the market does not extend to any of the interval endpoints. It extends from $\alpha_i = x_i - \frac{a-p_i}{k}$ to $\beta_i = x_i + \frac{a-p_i}{k}$, i.e. as far as where demand vanishes at either end. Now there is no need to distinguish between locations to left or right.

Substituting for the endpoints in (3), we obtain

$$Q_i = \frac{1}{k}(a - p_i)^2 \tag{24}$$

i.e. again the same formula as in (21), (23), (17) and (19).[9]

---

[9]Notably, this time we did not to optimize with respect to location to get the formula. Demand, and therefore profit as well, is *independent* of location, and location itself *indeterminate*. Therefore, we

2.5.4 Case 3 (Cutting Out Monopoly)

Remains just one case **3**, a monopoly extending over the whole interval $[-1, 1]$. This can only be a cutting out monopoly. From (3), we immediately obtain

$$Q_i = 2\,(a - p_i) - \frac{k}{2}\left((1 + x_i)^2 + (1 - x_i)^2\right)$$

Optimizing with respect to location, we find[10]

$$x_i' = 0 \tag{25}$$

and substituting back the simple expression

$$Q_i' = 2\,(a - p_i) - k \tag{26}$$

2.6 *Profit Maximization and Mill Price*

We now have all the total demand expressions for all possible market area endpoint $[\alpha_i, \beta_i]$ combinations. Likewise, we derived optimal location choices, unless they were indeterminate. We can hence pass to considering profits as functions of mill prices, and optimize with respect to these.

2.6.1 Cases 1, 8

Start with the toughest cases, where the interval is split in duopoly, and the market areas stretch out to the interval endpoints at their far ends. Demand for cases **1** and **8** was given in (11) and (15), which had the same form given the different definitions (10) and (14)

$$Q_i' = \frac{1}{10k}\left(6\,(a - p_i)^2 - 4\,(a - p_i)\,(a - \lambda_i) - (a - \lambda_i)^2\right)$$

(Footnote 9 continued)
have the disadvantage of not getting a definite location choice for the map we want to formulate. To solve this problem, consider that we deal with coexistent disjoint monopolies, whose maximum profits do not depend on location. We can choose either -1 or 1. This actually means merging the case with cases **2** or **6**.

[10]Note that this case can only occur when transportation cost is very low. As the entire interval $[-1, 1]$ must be covered, market radius $\frac{a-p}{k}$ must exceed unity, i.e. half the interval.

Substituting in the profit expression $G_i' = (p_i - c)Q_i'$, it becomes a cubic in p_i,

$$G_i' = (p_i - c)Q_i'$$

The optimality condition $\frac{\partial G_i}{\partial p_i} = 0$ then yields a quadratic in p_i:

$$18p_i^2 - 4(4a + 3c + 2\lambda_i)p_i + a^2 + 2a(4c + 3\lambda_i) + \lambda_i\,(4c - \lambda_i) = 0 \tag{27}$$

The quadratic, as usual, has two roots, of which the smaller provides profit maximum:

$$p_i' = \frac{4a + 3c + 2\lambda_i}{9} - \frac{\sqrt{36(a-c)^2 - 24\,(a-c)\left(a - \lambda_j\right) + 34\left(a - \lambda_j\right)^2}}{18} \tag{28}$$

Locations could now be obtained from (8) and (12) through substitution of p_i, but if we write them down, they just look messy. In numerical work, it is easy to let the computer do the job.

2.6.2 Cases 2, 4, 5, 6, 7

As we saw, demand in all these cases is given by one single formula, $Q_i = \frac{1}{k}\,(a - p_i)^2$. These are two cases of duopoly with one end of the market limited by vanishing local demand (Sanner cases, **4** and **7**), and three cases of monopoly (**2, 5,** and **6**), either coexistent or cutting out.

In all these cases profits are

$$G_i = \frac{1}{k}\,(p_i - c)\,(a - p_i)^2 \tag{29}$$

Thus the optimum condition

$$\frac{\partial G_i}{\partial p_i} = \frac{1}{k}\,(a - p_i)^2\,(a + 2c - 3p_i) = 0$$

has two zeros $p_i = a$, and

$$p_i = \frac{1}{3}a + \frac{2}{3}c \tag{30}$$

The first is a minimum resulting in zero profit as mill price is equal to maximum chargeable price, which results in vanishing demand everywhere, even at the location of the mill. Expression (30) is a well-known formula for such linear models—to charge a weighted average of maximum price and unit production cost, with weight one third for the first and two thirds for the latter. See Beckmann (1968).

Also maximum profit using (30) is now easy to calculate

$$G_i = \frac{4}{27k}(a-c)^3 \tag{31}$$

Locations Once we know optimal price for these five cases, we can calculate optimal locations from (16), (18), (20), (22) and (25).[11]

Thus

$$\mathbf{2,6}\ x_i = \pm\frac{2}{3k}(a-c) \mp 1$$
$$\mathbf{4,7}\ x_1 = \pm\frac{1}{k}(p_2 + kx_2) \mp \frac{1}{3k}(5a-2c)$$
$$\mathbf{5}\ x_i = \pm\left(\frac{2}{3k}(a-c) - 1\right)$$

2.6.3 Case 3

From (26) we get profit

$$G_i = (p_i - c)(2(a-p_i) - k)$$

which, optimized with respect to price, yields

$$p_i = \frac{1}{2}a + \frac{1}{2}c - \frac{1}{4}k$$

Substituted back, we obtain

$$G_i = \frac{1}{2}\left(a - c - \frac{k}{2}\right)^2$$

This is a cutting out case with location chosen in the midpoint of the interval as we see from (25). The case corresponds to the Hotelling paradox, and will no doubt be contested by the competitor. Recall what was said about the conditionality of this case on a low transportation cost, and also that there are other instances of cutout behaviour.

[11] Note that though price according to (30) is in the right interval between c and a, it is by no means certain that the location according to these formulas is reasonable—depending on parameters an "optimal" location may even fall outside the admissible interval $[-1, 1]$. We will therefore have to run a check of relevance with respect to region for each alternative once we proceed to formulating the map. Note further that in the monopoly cases (except for cutting out) the location like optimal price is independent of the competitor's moves, whereas in the shared market cases **4** and **6** the best reply depends on the expected move by the competitor.

3 Summary for the Cases

To interpret Fig. 5, note that, taking x_2, p_2 as given, we can use the same type of diagram as in all the pictures above as phase space for x_1', p_1', the dash, as usual, denoting the next move considered. Several features should be noted.

The empty areas in the rectangle, below c, and in the little triangle on top of the competitor's mill price, are such which firm 1 never considers—a mill price lower than unit cost yields no profit, and neither does a combination of location and mill price where the firm cuts itself out.

The structure seems quite complicated, but we see that the regions are separated by parallel straight lines, three downsloping and three upsloping, whose formulas are easy to state.

Define $r_i = p_i + kx_i, \quad s_i = p_i - kx_i$. Then the downsloping lines have formulas

$$r_1 = 2a - r_2$$
$$r_1 = r_2$$
$$r_1 = a - k$$

and the upsloping

$$s_1 = 2a - s_2$$
$$s_1 = s_2$$
$$s_1 = a - k$$

Further study, using our formulas for optimal location and mill pricing choice, show that the *optimal* x_1', p_1' points in the plane can only be located on stretches of these sloping lines, which, of course, simplifies analysis a lot. We, however, do not present the argument here as it leads into further detail.

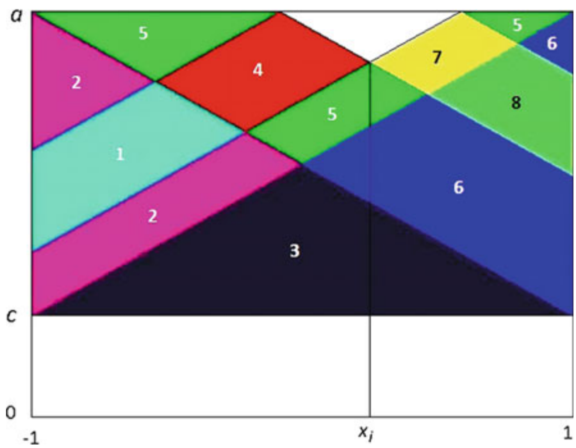


Fig. 5 Possible moves x_i', p_i' with regions relevant for the eight *endpoint* combinations **(1)**,…**(8)**

Notably, there are three areas labelled with **5**, those on top representing local monopolies, left, respectively, right of the competitor. The area mid in the picture is a single cutting out monopoly. Also regions **2** and **6** are in duplicate, local monopoly extending to the interval boundary on top, and a corresponding cutting out case lower down. In all we see 12 coloured regions, but three of them are further split by the vertical line at the competitor's location, so if we count locations to the left and right of the competitor, we arrive at 15 different cases.

After computing profits, we choose the reaction which returns the highest profit.

Note that for dynamics the results are quite sufficient for writing a simple computer programme, though the map is too complicated with all its pieces to formulate any elegant closed form iteration formula.

4 Equilibria

All numerical experiments show that there are no periodic or more complicated orbits for the map, provided we ignore the cutting out cases, which remain too ill-structured to be included. However, there exist two different kinds of equilibrium states; genuine duopoly with a common boundary, and disjoint non-competing monopoles.

4.1 Duopoly

The most interesting, though messy, case is oligopoly when the firms share the total market between them with some competition at the common boundary, cases **1** and **8**.

4.1.1 Locations

As the firms are identical (facing the same demand and having the same production and transport costs), it is a reasonable conjecture that in equilibrium they also charge the same mill prices, i.e. $p = p_1 = p_2$. Substituting in (8) and (12), putting $x_1' = x_1, \quad x_2' = x_2$, now taking the equations as a simultaneous system, and solving, we get

$$x_1 = \frac{1}{3}\frac{a-p}{k} - \frac{2}{3} \tag{32}$$

$$x_2 = \frac{2}{3} - \frac{1}{3}\frac{a-p}{k} \tag{33}$$

Adding (32) and (33)

$$x_1 + x_2 = 0$$

Obviously the firms locate symmetrically around the centre, which seems plausible by intuition.

Let us pay attention to two particular cases; (i) if $\frac{a-p}{k} = \frac{1}{2}$ then $x_1 = -\frac{1}{2}$, $x_2 = \frac{1}{2}$ and (ii) if $\frac{a-p}{k} = 2$ then $x_1 = x_2 = 0$. In the first case the firms locate socially optimally at the quartiles, each at the centre of its half-interval, in the second they crowd both in the centre, Hotelling's main case.

Hence, we see that according to (32)–(33) both these outcomes are possible under duopoly action, as are any cases between these. The locations depend on one single compound expression, $\frac{a-p}{k}$, the difference of maximum chargeable price and actual mill price, divided by the transportation rate.

4.1.2 Demand

Note that with equal mill prices, and symmetric locations from (9) and (13) we have $\beta_1 = \alpha_2 = 0$, so the market areas are $\alpha_1 = -1$, $\beta_1 = 0$ and $\alpha_2 = 0$, $\beta_2 = 1$.

Further, from (10) and (14)

$$\lambda = \lambda_1 = \lambda_2 = \frac{4p - a + 5k}{3} \tag{34}$$

Substituting this in (11) and (15), along with $p_1 = p_2 = p$, we obtain

$$Q_1 = Q_2 = \frac{6\,(a-p)^2 - 4\,(a-p)\,(a-\lambda) - (a-\lambda)^2}{10k} \tag{35}$$

4.1.3 Price and Profit

From (4), then

$$G_i = (p - c)\,Q_i \tag{36}$$

We already obtained the optimal duopoly prices in (28) above. As we deal with equilibrium, we can remove the dash, delete the index on costs and mill prices and substitute for λ_1, λ_2 from (34).

By this (28) becomes quite messy. Furthermore, substitution for λ_1, λ_2 introduces mill prices under the root sign. Therefore, we have to square out the root again and solve for mill price anew. As the resulting equation is quadratic in mill price, we again have two roots and have to choose the relevant one, obtaining:

$$p = \frac{2a + 3c + 8k}{5} - \frac{3\sqrt{4(a-c)^2 - 8\,(a-c)\,k + 34k^2}}{10} \tag{37}$$

We can restate this as

$$\frac{a-p}{k} = \frac{3}{5}\kappa - \frac{8}{5} + \frac{3}{10}\sqrt{4\kappa^2 - 8\kappa + 34}$$

where

$$\kappa = \frac{a-c}{k} \tag{38}$$

From equations (32)–(33) let us determine x_1, x_2 once we know $\frac{a-p}{k}$. This ratio was already designated as an important compound, but it depends on mill price, which is a decision variable. The ratio $\kappa = \frac{a-c}{k}$ has the same form, but is more fundamental as it only contains parameters.

We can now use (37) and (34) in (35). Using the convenient (38), total demands become

$$Q_1 = Q_2 = \frac{k}{50}\left((22 - 2\kappa)\sqrt{4\kappa^2 - 8\kappa + 34} + 2\kappa^2 - 24\kappa + 67\right)$$

Further, using (36),

$$G_1 = G_2 = \frac{k^2}{250}((2\kappa^2 - 44\kappa + 377)\sqrt{4\kappa^2 - 8\kappa + 34} + 4\kappa^3 - 92\kappa^2 + 482\kappa - 2194)$$

- not so elegant, but closed form expressions for total demands and profits anyhow.

As mentioned, this case was analyzed in Puu and Gardini (2002). Any dynamic process defined by Eqs. (8), (12), (10), (14) and (28) is just a contraction. In Fig. 6, we illustrate how fast this global approach to equilibrium goes.

In only five steps, the system has converged so close to equilibrium that further iterations can no longer be distinguished. Not seen in the picture, but showing up

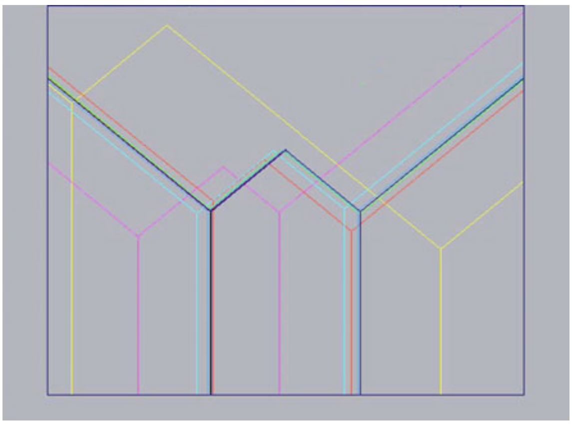


Fig. 6 Illustration of the contraction to equilibrium duopoly

when we print the locations, is that in the first move the firms swap left/right positions. Firm 2 starts so far to the right that it finds it better to move to the left of firm 1.

4.2 Monopolies

As for the monopoly cases, we already derived optimal prices, profits and locations, unless they were indeterminate. Given parameter κ is high, for instance resulting from a high transportation rate, only narrow market areas delimited by positive local demand can be established. These can be easily accommodated in the fixed interval, and there is no point in crowding to a duopoly. In these cases location is indeterminate, because its exact choice does not influence demand or profit, provided the fixed interval endpoints and the other competitor's monopoly area are not touched.

Should the competitor be too centrally located, the firm will temporarily have to choose a duopoly, but the competitor will then find it more profitable to establish a monopoly area more to its side. This is confirmed by numerics. See Fig. 7.

We can even use the iterative system designed for duopoly, as it also works for disjoint monopolies. Several interesting things can be noted in this case. The convergence is again rather fast. Now, we take a starting position where the firms are initially located very close to the centre. Further, the initial prices are so chosen that firm 1 is actually cut out by firm 2. Yet *both* firms in the next move jump wide apart—even so wide that they hit the interval boundaries. This, we already noted, is inoptimal, so after that the firms start converging to their final positions, though definite positions as defined by the duopoly process. As a matter of fact, once they are disjoint monopolies, their exact final positions are indeterminate.

Fig. 7 Iteration to disjoint monopolies, using the duopoly map

5 Further Issues

In a first analysis one can assume, as Hotelling originally, that both competitors are free to choose location *and* mill price in each move.

5.1 Different Time Scales

However, we could also consider a combination of short run dynamics where only prices change but location is fixed, and a long run dynamics where both location and price are variable. In such a case, it is natural to connect location change with capital formation. A firm is unlikely to change location as long as it has capital invested in some plant. But capital wears out with time, and when it is time for renewal, then also location can be changed.

The present author has in mind capital with constant rate of decay and "sudden death", combined with a Leontief type of production function which produces capacity limits.

5.2 Different Pricing Policies

Above we mentioned that price discrimination was not considered. Any competitors with local monopoly market areas could consider price discrimination, arranging for transportation and charging less than actual transportation cost. It is even possible to combine discrimination in an inner area with mill pricing at the fringe towards the competitor. See Beckmann (1968). Also note that uniform pricing is a case of discrimination, though utterly unsuited for the Hotelling case as one does not get any definite market boundary points.

5.3 Geographical Space

An intriguing problem is, of course, extending the Hotelling model to the two dimensions of geographical space. Hotelling, like almost all economists dealing with spatial issues stopped short of this extension, because the complexity increases, and economists as a rule cannot use the mathematical tools needed for the study of phenomena in two dimensions.

The first issue is, of course, to make clear what one means by the Hotelling problem in two dimensions. In a sense a generalization from two firms on an interval could be three firms in a triangle.

However, as soon as one moves from one to two dimensions, one has to consider the distance metric on which transportation costs depend. Now, all integrations using the Euclidean distance become immensely messy. Much easier to deal with is, for instance a Manhattan metric, as an idealization of an infinitely dense net of west–east and south–north roads. Such calculations become much easier. But to obtain an analog to the Hotelling problem one would need a region which conforms to the equidistance loci of the metric, which would be a tilted square. But then the choice would be four competing firms rather than three. Is this really the next generalization step?

How about the triangle and three competitors? Does there exist any metric that produces triangular distance loci? Yes, there is one, with roads in three different directions. However, with lanes in both directions, the result is hexagonal distance loci. However, if we let all roads be one-way, then we actually get distance loci that are triangles.

Anyhow, the two dimensions of geographical space provide a real challenge, and it is by no means obvious what the generalization of Hotelling's model would look like. The present author reviewed quite a number of extensions of the model to 2D, unfortunately, none convincing.

References

Beckmann MJ (1968) Location theory. Random House Inc, New York
Bertrand J (1883) Theorie mathématique de la richesse sociale. Journale des Savants 48:499–508
Cournot A (1838) Récherces sur les principes mathématiques de la théorie des ricesses. Paris
d'Aspremont C, Gabszewitsc JJ, Thisse JF (1979) On Hotelling's "stability in competition". Econometrica 47:1145–1150
Hotelling H (1929) Stability in competition. Econ J 39:41–57
Lerner AP, Singer HW (1937) Some notes on duopoly and spatial competition. J Polit Econ 45:145–186
Puu T (2002) Hotelling's 'ice cream dealers' with elastic demand. Ann Reg Sci 36:1–17
Puu T, Gardini L (2002) Hotelling type duopoly, in Puu and Sushko, Oligopoly Dynamics. Springer, Heidelberg, pp 265–310
Smithies A (1941a) Optimum location in spatial competition. J Polit Econ 49:423–439
Smithies A (1941b) Monopolistic price policy in a spatial market. Econometrica 9:63–73
von Stackelberg H (1934) Marktform und Gleichgewicht. Julius Springer, Heidelberg

Learning in Monopolies with Delayed Price Information

Akio Matsumoto and Ferenc Szidarovszky

Abstract The intersection of the price function with the vertical axis is called the *maximum price* and the slope of the price function is called the *marginal price*. It is assumed that a monopoly has full information about the marginal price and its own cost function but is uncertain about the maximum price. Based on repeated price observations an adaptive learning process can be developed for the maximum price. If the price observations have fixed delays, then the learning process can be described by a delayed differential equation. In the cases of one or two delays, the asymptotic behavior of the associated dynamic process is examined. Stability conditions are derived and the occurrence of Hopf bifurcation is shown at the critical values. The nonlinear learning process can generate complex dynamics in the case of local instability when the delay is sufficiently long.

Keywords Bounded rationality · Monopoly dynamics · Fixed time delay · Adaptive learning · Hopf bifurcation

JEL Classification C62 · C63 · D21 · D42

The authors highly appreciate the financial supports from the MEXT-Supported Program for the Strategic Research Foundation at Private Universities 2013–2017, the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 24530202, 25380238 and 26380316) and Chuo University (Joint Research Grant). The usual disclaimers apply.

A. Matsumoto (✉)
Department of Economics, International Center for Further Development of Dynamic Economic Research, Chuo University, 742-1, Higashi-Nakano, Hachioji, Tokyo 192-0393, Japan
e-mail: akiom@tamacc.chuo-u.ac.jp

F. Szidarovszky
Department of Applied Mathematics, University of Pécs, Ifjúság u 6, Pécs 7624, Hungary
e-mail: szidarka@gmail.com

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_4

1 Introduction

This paper is based on the familiar monopoly model in which there is only one firm having linear price and cost functions. Its main purpose is to show how cyclic and erratic dynamics can emerge from quite simple economic structures when uncertainty, information delays, and behavioral nonlinearities are present. Implicit in the textbook approach is an assumption of complete and instantaneous information availability on price and cost functions. In consequence, the textbook-monopoly can choose its optimal choices of price and quantity to maximize profit with one shot. Thus the monopoly model is static in nature. The assumption of such a rational monopoly is, however, questionable and unrealistic in real economies, since there are always uncertainty and a time delay in collecting information and determining optimal responses, and in addition, function relations such as the market price function cannot be determined exactly based on theoretical consideration and observed data. Getting closer to the real world and improving the monopoly theory, we replace this extreme but convenient assumption with the more plausible one. Indeed, the monopoly firm is assumed, first, to have only limited knowledge on the price function and, second, to obtain it with time delay. As a natural consequence of these alternations, the firm gropes for its optimal choice by using data obtained through market experiences. The modified monopoly model becomes dynamic in nature.

In the recent literature, it has been demonstrated that a boundedly rational monopoly may exhibit simple as well as complex dynamic behavior. Nyarko (1991) solves the problems of a profit maximizing monopoly without knowing the slope and intercept of a linear demand and shows that using Bayesian updating leads to cyclic actions and beliefs if the market demand is mis-specified. Furthermore, in the framework with discrete-time scale, Puu (1995) shows that the boundedly rational monopoly behaves in an erratic way under cubic demand with a reflection point. In the similar setting, Naimzada and Ricchiuti (2008) represent that complex dynamics can arise even if cubic demand does not have a reflection point. Naimzada (2012) exhibits that delay monopolistic dynamics can be described by the well-known logistic equation when the firm takes a special learning scheme. More recently Matsumoto and Szidarovszky (2014a, b) demonstrate that the monopoly equilibrium undergoes to complex dynamics through either a period-doubling or a Neimark-Sacker bifurcation.

This paper considers monopoly dynamics in continuous-time scale and presents a new characterization of a monopoly's learning process under a limited knowledge of the market demand. It is, in particular, a continuation of Matsumoto and Szidarovszky (2012) where the monopoly does not know the price function and fixed time delays are introduced into the output adjustment process based on the gradient of the marginal expected profit. It also aims to complement Matsumoto and Szidarovszky (2014a, b, 2015) where uncertain delays are modeled by continuously distributed time delays when the firm wants to react to average past information instead of sudden market changes.[1] Gradient dynamics is replaced with an adaptive learning scheme based on profit maximizing behavior. Although there is price uncertainty and the price information is delayed, the monopoly is still able to update its estimate on the price function via the usage of price observations and its optimal price beliefs. We will consider the cases of a single delay and two delays, respectively, and then demonstrate a variety of dynamics ranging from simple cyclic oscillations to complex behavior involving chaos.

This paper develops as follows. The mathematical model is formulated in Sect. 2. The case of a single delay is examined in Sect. 3, when the firm uses the most current delayed price information to form its expectation about the maximum price. In Sect. 4, it is assumed that the firm formulates its price expectation based on two delayed observations by using a linear prediction scheme. Complete stability analysis is given, the stability regions are determined and illustrated. The occurrence of Hopf bifurcation is shown at the critical values of the bifurcation parameter, which is the length of the single delay or one of the two delays. The last section offers conclusions and further research directions.

2 The Mathematical Models

Consider a single product monopoly that sells its product to a homogeneous market. Let q denote the output of the firm, $p(q) = a - bq$ the price function and $C(q) = cq$ the cost function.[2] Since $p(0) = a$ and $|\partial p(q)/\partial q| = b$, we call a the *maximum price* and b the *marginal price*. There are many ways to introduce uncertainty into this framework. In this study, it is assumed that the firm knows the marginal price but does not know the maximum price. In consequence it has only an estimate of it at each time period, which is denoted by $a^e(t)$. So the firm believes that its profit is

$$\pi^e = (a^e - bq)q - cq$$

[1]There are two different ways to model time delays in continuous-time scale: fixed time delay and continuously distributed time delay (fixed delay and continuous delay henceforth). The choice of the type of delay results in the use of different analytical tools. In the cases of fixed delay, dynamics is described by a delay differential equation whose characteristic equation is a mixed polynomial-exponential equation with infinitely many eigenvalues. Bellman and Cooke (1956) offer methodology of complete stability analysis in such models. On the other hand, in the cases of continuous delay, Volterra-type integro-differential equations are used to model the dynamics. The theory of continuous delays with applications in population dynamics is offered by Cushing (1977). Since Invernizzi and Medio (1991) have introduced continuous delays into mathematical economics, its methodology is used in analyzing many economic dynamic models.

[2]Linear functions are assumed only for the sake of simplicity. We can obtain a similar learning process to be defined even if both functions are nonlinear. It is also assumed for the sake of simplicity that the firm has perfect knowledge of production technology (i.e., cost function).

and its best response is

$$q^e = \frac{a^e - c}{2b}.$$

Further, the firm expects the market price to be

$$p^e = a^e - bq^e = \frac{a^e + c}{2}. \tag{1}$$

However, the actual market price is determined by the real price function

$$p^a = a - bq^e = \frac{2a - a^e + c}{2}. \tag{2}$$

Using these price data, the firm updates its estimate. The simplest way for adjusting the estimate is the following. If the actual price is higher than the expected price, then the firm shifts its believed price function by increasing the value of a^e, and if the actual price is the smaller, then the firm decreases the value of a^e. If the two prices are the same, then the firm wants to keep its estimate of the maximum price. This adjustment or learning process can be modeled by the differential equation

$$\dot{a}^e(t) = k\left[p^a(t) - p^e(t)\right],$$

where $k > 0$ is the speed of adjustment. Substituting relations (1) and (2) reduces the adjustment equation to a linear differential equation with respect to a^e as

$$\dot{a}^e(t) = k\left[a - a^e(t)\right]. \tag{3}$$

In another possible learning process, the firm revises the estimate in such a way that the growth rate of the estimate is proportional to the difference between the expected and actual prices. Replacing $\dot{a}^e(t)$ in Eq. (3) with $\dot{a}^e(t)/a^e(t)$ yields a different form of the adjustment process

$$\frac{\dot{a}^e(t)}{a^e(t)} = k\left[a - a^e(t)\right]$$

or multiplying both sides by $a^e(t)$ generates the logistic model

$$\dot{a}^e(t) = ka^e(t)\left[a - a^e(t)\right] \tag{4}$$

which is a nonlinear differential equation.

If there is a time delay τ in the estimated price, then we can rewrite the estimated price and market price at time t based on information at time $t - \tau$ as

$$p^e(t; t-\tau) = a^e(t-\tau) - bq^e(t; t-\tau)$$

and

$$p^a(t; t-\tau) = a - bq^e(t; t-\tau)$$

where $q^e(t; t-\tau)$ is the delay best reply,

$$q^e(t; t-\tau) = \frac{a^e(t-\tau) - c}{2b}.$$

Then Eqs. (3) and (4) have to be modified, respectively, as

$$\dot{a}^e(t) = k\left[a - a^e(t-\tau)\right] \tag{5}$$

and

$$\dot{a}^e(t) = ka^e(t)\left[a - a^e(t-\tau)\right]. \tag{6}$$

If the firm uses two past price information, then the delay dynamic equations turn to be

$$\dot{a}^e(t) = k\left[a - \omega a^e(t-\tau_1) - (1-\omega)a^e(t-\tau_2)\right] \tag{7}$$

and

$$\dot{a}^e(t) = ka^e(t)\left[a - \omega a^e(t-\tau_1) - (1-\omega)a^e(t-\tau_2)\right], \tag{8}$$

where τ_1 and τ_2 denote the delays in the price information. If the firm uses interpolation between the observations, then $0 < \omega < 1$, and if it uses extrapolation to predict the current price, then the value of ω can be negative or even greater than unity. Notice that for $\omega = 0$ and $\omega = 1$, Eqs. (7) and (8) reduce to Eqs. (5) and (6). If $0 < \omega < 1$, then the cases of $\omega \leq 1/2$ are the same as $\omega \geq 1/2$ because of the symmetry of the model between τ_1 and τ_2. Similarly, if $\omega < 0$, then $1 - \omega > 1$, so the cases of $\omega < 0$ and $\omega > 1$ are also equivalent. Therefore in models (7) and (8), we will consider only the case of $1/2, = \omega < 1$.

By introducing the new variable $z(t) = a^e(t) - a$, Eq. (5) and the linearized version of Eq. (6) are written as

$$\dot{z}(t) + \alpha z(t-\tau) = 0 \tag{9}$$

where $\alpha = k$ or $\alpha = ak$. By the same way, Eq. (7) and the linearized version of Eq. (8) are modified as

$$\dot{z}(t) + \alpha\omega z(t-\tau_1) + \alpha(1-\omega)z(t-\tau_2) = 0. \tag{10}$$

In the following sections, we will examine the asymptotic behavior of the trajectories of Eqs. (9) and (10).

3 Single Fixed Delay

If there is no delay, then $\tau = 0$ and Eq. (9) becomes an ordinary differential equation with characteristic polynomial $\lambda + \alpha$. So, the only eigenvalue is negative implying the global asymptotic stability of the steady state $\bar{z} = 0$ if the original equation is linear and the local asymptotic stability if nonlinear. The steady state corresponds to the true value of the maximum price. If $\tau > 0$, then the exponential form $z(t) = e^{\lambda t} u$ of the solution gives the characteristic equation,

$$\lambda + \alpha e^{-\lambda\tau} = 0. \tag{11}$$

Since the only eigenvalue is negative at $\tau = 0$, we expect asymptotical stability for sufficiently small values of τ and loss of stability for sufficiently large values of τ. If the steady state becomes unstable, then stability switch must occur when $\lambda = i\nu$. If λ is an eigenvalue, then its complex conjugate is also an eigenvalue. In consequence we can assume, without any loss of generality, that $\nu > 0$. So Eq. (11) can be written as

$$i\nu + \alpha e^{-i\nu\tau} = 0.$$

By separating the real and imaginary parts, we have

$$\alpha \cos \nu\tau = 0$$

and

$$\nu - \alpha \sin \nu\tau = 0.$$

Therefore

$$\cos \nu\tau = 0 \text{ and } \sin \nu\tau = \frac{\nu}{\alpha}$$

with $\nu = \alpha$ leading to infinitely many solutions,

$$\tau = \frac{1}{\alpha}\left(\frac{\pi}{2} + 2n\pi\right) \text{ for } n = 0, 1, 2, \ldots \tag{12}$$

The solution τ with $n = 0$ forms a downward-sloping curve with respect to α,

$$\tau^* = \frac{\pi}{2\alpha} \text{ with } \alpha = k \text{ or } \alpha = ak.$$

Applying the main theorem in Hayes (1950) or the same result obtained differently in Matsumoto and Szidarovszky (2013), we can find that this curve divides the nonnegative (α, τ) plane into two subregions; the real parts of the roots of the characteristic equation are all negative in the region below the curve and some roots are positive in the region above. This curve is often called a *partition curve* separating the stability region from the instability region. Notice that the critical value of

τ decreases with α, so a larger value of α caused by the high speed of adjustment and/or the larger maximum price makes the steady state less stable.

We can easily prove that all pure complex roots of Eq. (11) are simple. If λ is a multiple eigenvalue, then

$$\lambda + \alpha e^{-\lambda\tau} = 0$$

and

$$1 + \alpha e^{-\lambda\tau}(-\tau) = 0$$

implying that

$$1 + \lambda\tau = 0$$

or

$$\lambda = -\frac{1}{\tau}$$

which is a real and negative value.

In order to detect stability switches and the emergence of a Hopf bifurcation, we select τ as the bifurcation parameter and consider λ as function of τ, $\lambda = \lambda(\tau)$. By implicitly differentiating Eq. (11) with respect to τ, we have

$$\frac{d\lambda}{d\tau} + \alpha e^{-\lambda\tau}\left(-\frac{d\lambda}{d\tau}\tau - \lambda\right) = 0$$

implying that

$$\frac{d\lambda}{d\tau} = -\frac{\lambda^2}{1 + \tau\lambda}$$

and as $\lambda = i\nu$, its real part becomes

$$\text{Re}\left(\frac{d\lambda}{d\tau}\right) = \text{Re}\left(\frac{\nu^2}{1 + i\tau\nu}\right)$$

$$= \frac{\nu^2}{1 + (\tau\nu)^2} > 0.$$

At the critical value of τ, the sign of the real part of an eigenvalue changes from negative to positive and it is a Hopf bifurcation point of the nonlinear learning process (6) with one delay that has a family of periodic solutions. Thus we have the following results:

Theorem 1 *(1) For the linear adjustment process (5), stability of the steady state is lost at the critical value of* τ,

$$\tau^* = \frac{\pi}{2k}$$

and cannot be regained for larger values of $\tau > \tau^*$. *(2) For the logistic adjustment process (6), stability of the steady state is lost at*

$$\tau^{**} = \frac{\pi}{2ak}$$

and limit cycles appear through Hopf bifurcation for $\tau > \tau^{**}$.

An intuitive reason why stability switch occurs only at the critical value of τ with $n = 0$ is the following. Notice first that the delay differential equation has infinitely many eigenvalues and second that their real parts are all negative for $\tau < \tau^*$. When increasing τ arrives at the partition curve, then the real part of one eigenvalue becomes zero and its derivative with respect to τ is positive implying that the real part changes its sign to positive from negative. Hence the steady state loses stability at this critical value. Further increasing τ crosses the (α, τ) curve defined by Eq. (12) with $n = 1$. There the real part of another eigenvalue changes its sign to positive from negative. Repeating the same arguments, we see that stability cannot be regained and therefore no stability switch occurs for any $n \geq 1$. Hence stability is changed only when τ crosses the partition curve.

Theorem 1 is numerically confirmed. In Fig. 1a, three cyclic trajectories generated by the linear delay equation (5) are depicted under $a = 1$ and $k = 1$. The initial functions are the same, $\phi(t) = 0.5$ for $t \leq 0$ but lengths of delay are different. The blue trajectory with $\tau = \tau^* - 0.1$ shows damped oscillation approaching the steady state, the black trajectory with $\tau = \tau^*$ converges to a limit cycle (i.e., a degenerated Hopf cycle) and the red trajectory with $\tau = \tau^* + 0.1$ cyclically diverges away from the steady state (The color versions of the figures of this chapters can be found in the e-version of this book). On the other hand, in Fig. 1b, one limit cycle generated by the logistic delay equation (6) is illustrated under the same parametric conditions, the same initial function and $\tau = \tau^{**} + 0.05$. By comparing these numerical results, it

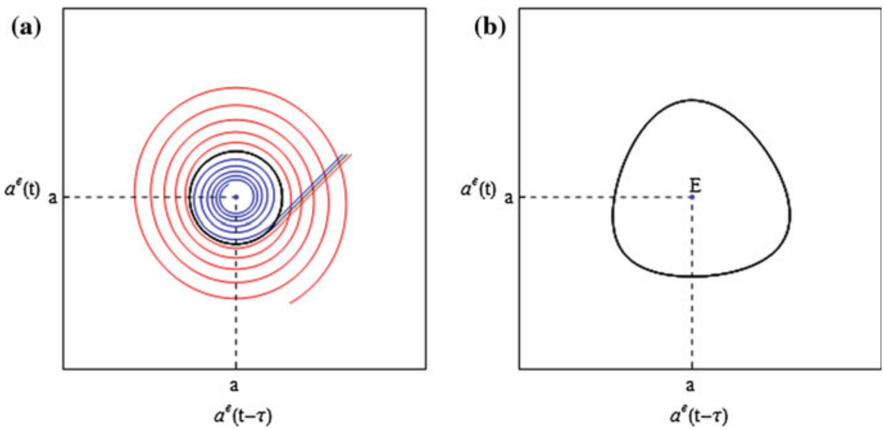


Fig. 1 Cyclic oscillations. **a** Linear learning. **b** Nonlinear learning

is quite evident that nonlinearity of the logistic equation can be a source of persistent fluctuations when the steady state loses its stability.

4 Two Fixed Delays

In this section, we draw attention to asymptotic behavior of differential equation (8) with two delays. Its characteristic equation is obtained by substituting the exponential form $q(t) = e^{\lambda t}u$ into Eq. (10) and arranging the terms:

$$\lambda + \alpha\omega e^{-\lambda\tau_1} + \alpha(1-\omega)e^{-\lambda\tau_2} = 0$$

or

$$\bar{\lambda} + \omega e^{-\bar{\lambda}\gamma_1} + (1-\omega)e^{-\bar{\lambda}\gamma_2} = 0, \tag{13}$$

where $\bar{\lambda} = \lambda/\alpha,\ \gamma_1 = \alpha\tau_1$ and $\gamma_2 = \alpha\tau_2$. If $\gamma_1 = \gamma_2 = 0$ (or $\tau_1 = \tau_2 = 0$), then the steady state is asymptotically stable, since the only eigenvalue is negative. In order to find stability switches, we assume again that $\bar{\lambda} = i\nu$ with $\nu > 0$. Then Eq. (13) becomes

$$i\nu + \omega e^{-i\nu\gamma_1} + (1-\omega)e^{-i\nu\gamma_2} = 0.$$

Separating the real and imaginary parts yields

$$\omega\cos\nu\gamma_1 + (1-\omega)\cos\nu\gamma_2 = 0 \tag{14}$$

and

$$\nu - \omega\sin\nu\gamma_1 - (1-\omega)\sin\nu\gamma_2 = 0 \tag{15}$$

when $\omega \geq 1/2$ and $\omega \neq 1$ are assumed. We first examine two boundary cases, one with $\gamma_1 = 0$ and the other with $\gamma_2 = 0$, to obtain the following two results.

Theorem 2 *If $\gamma_1 = 0$, then the steady state is locally asymptotically stable for all $\gamma_2 > 0$.*

Proof In the case of $\gamma_1 = 0$, Eq. (14) is reduced as

$$\cos\nu\gamma_2 = -\frac{\omega}{1-\omega}.$$

If $\omega = 1/2$, then $\cos\nu\gamma_2 = -1$ so $\sin\nu\gamma_2 = 0$ implying that $\nu = 0$, which is a contradiction. If $1/2 < \omega < 1$, then $-\omega/(1-\omega) < -1$, so no solution exists. Therefore Eq. (13) has no pure imaginary roots that cross the imaginary axis when γ_2 increases from zero. ■

Putting this result differently, we can say that for $\gamma_1 = 0$, delay γ_2 is *harmless* implying that the steady state is locally asymptotically stable regardless of the values of γ_2. We now proceed to the other case.

Theorem 3 *If* $\gamma_2 = 0$ *and* $\omega > 1/2$*, then the steady state is locally asymptotically stable for* $0 < \gamma_1 < \gamma_1^*$ *and locally unstable if* $\gamma_1 > \gamma_1^*$ *where the critical value of* γ_1 *is defined as*

$$\gamma_1^* = \frac{\cos^{-1}\left(-\frac{1-\omega}{\omega}\right)}{\sqrt{2\omega - 1}}.$$

Proof In the case of $\gamma_2 = 0$, (14) and (15) are reduced to

$$(1-\omega) + \omega \cos \nu\gamma_1 = 0$$

and

$$\nu - \omega \sin \nu\gamma_1 = 0.$$

Moving $1 - \omega$ and ν to the right-hand sides, squaring both sides and adding the resulted equations yield

$$\nu^2 = 2\omega - 1.$$

The positive solution of ν is

$$\nu = \sqrt{2\omega - 1}$$

where $\omega > 1/2$. Substituting this value into the first equation above and solving the resultant equation, we have γ_1^*, the critical value of γ_1. In the same way as in proving Theorem 1, we arrive at the stability result by applying Hays' theorem or our result in Matsumoto and Szidarovszky (2012). ■

We now examine the general case of $\gamma_1 > 0$ and $\gamma_2 > 0$. By introducing the new variables

$$x = \sin \nu\gamma_1 \text{ and } y = \sin \nu\gamma_2,$$

Equation (14) implies that

$$\omega^2(1 - x^2) = (1-\omega)^2(1 - y^2)$$

or

$$-\omega^2 x^2 + (1-\omega)^2 y^2 = 1 - 2\omega \tag{16}$$

and from (15),

$$y = \frac{\nu - \omega x}{1 - \omega}. \tag{17}$$

By combining (16) and (17), we get an equation for x,

$$-\omega^2 x^2 + (1-\omega)^2 \left(\frac{\nu - \omega x}{1-\omega}\right)^2 = 1 - 2\omega$$

implying that

$$x = \frac{\nu^2 + 2\omega - 1}{2\nu\omega}$$

and from (17),

$$y = \frac{\nu^2 - 2\omega + 1}{2\nu(1-\omega)}.$$

These two equations provide a parametric representation in the (γ_1, γ_2) plane:

$$\sin \nu\gamma_1 = \frac{\nu^2 + 2\omega - 1}{2\nu\omega} \text{ and } \sin \nu\gamma_2 = \frac{\nu^2 - 2\omega + 1}{2\nu(1-\omega)}. \tag{18}$$

The feasibility of solutions requires that

$$-1 \le \frac{\nu^2 + 2\omega - 1}{2\nu\omega} \le 1 \tag{19}$$

and

$$-1 \le \frac{\nu^2 - 2\omega + 1}{2\nu(1-\omega)} \le 1. \tag{20}$$

Consider first condition (19), which is a pair of quadratic inequalities in ν,

$$\nu^2 + 2\nu\omega + 2\omega - 1 \ge 0 \tag{21}$$

and

$$\nu^2 - 2\nu\omega + 2\omega - 1 \le 0 \tag{22}$$

The roots of (21) are -1 and $1 - 2\omega$, and the roots of (22) are $2\omega - 1$ and 1. Condition (20) is also a pair of quadratic inequalities,

$$\nu^2 + 2\nu(1-\omega) + (1-2\omega) \ge 0 \tag{23}$$

and

$$\nu^2 - 2\nu(1-\omega) + (1-2\omega) \le 0 \tag{24}$$

with -1 and $2\omega - 1$ being the roots of (23) and $1 - 2\omega$ and 1 being the roots of (24). So if $\omega < 1$, then the value of ν has to satisfy the following relation:

$$2\omega - 1 \le \nu \le 1.$$

As it was explained earlier, we have assumed that $\omega \geq 1/2$. The remaining part of this section is divided into two subsections. First, the nonsymmetric case (i.e., $1 > \omega > 1/2$) is examined in Sect. 4.1, then the symmetric case (i.e., $\omega = 1/2$) in Sect. 4.2.

4.1 The Case of $\frac{1}{2} < \omega < 1$

In this subsection, we assume that $1/2 < \omega < 1$. Since from (14), we see that the signs of cos $\nu\gamma_1$ and cos$\nu\gamma_2$ are different, Eq. (18) imply that

$$L_1(k,n) : \begin{cases} \gamma_1 = \dfrac{1}{\nu}\left(\sin^{-1}\left(\dfrac{\nu^2+2\omega-1}{2\nu\omega}\right)+2k\pi\right) (k \geq 0) \\ \gamma_2 = \dfrac{1}{\nu}\left(\pi - \sin^{-1}\left(\dfrac{\nu^2-2\omega+1}{2\nu(1-\omega)}\right)+2n\pi\right) (n \geq 0) \end{cases} \tag{25}$$

or

$$L_2(k,n) : \begin{cases} \gamma_1 = \dfrac{1}{\nu}\left(\pi - \sin^{-1}\left(\dfrac{\nu^2+2\omega-1}{2\nu\omega}\right)+2k\pi\right) (k \geq 0) \\ \gamma_2 = \dfrac{1}{\nu}\left(\sin^{-1}\left(\dfrac{\nu^2-2\omega+1}{2\nu(1-\omega)}\right)+2n\pi\right) (n \geq 0) \end{cases} \tag{26}$$

which gives two sets of parametric curves in the (γ_1, γ_2) plane. The domain of ω is the interval $[2\omega - 1, 1]$. At the initial point $\nu = 2\omega - 1$, we have

$$\frac{\nu^2+2\omega-1}{2\nu\omega} = 1 \text{ and } \frac{\nu^2-2\omega+1}{2\nu(1-\omega)} = -1$$

and at the end point $\nu = 1$, we have

$$\frac{\nu^2+2\omega-1}{2\nu\omega} = 1 \text{ and } \frac{\nu^2-2\omega+1}{2\nu(1-\omega)} = 1.$$

Therefore the initial and end points of $L_1(k, n)$ are

$$I_1(k,n) = \left(\frac{1}{2\omega-1}\left(\frac{\pi}{2}+2k\pi\right),\ \frac{1}{2\omega-1}\left(\frac{3\pi}{2}+2n\pi\right)\right)$$

and

$$E_1(k,n) = \left(\frac{\pi}{2}+2k\pi,\ \frac{\pi}{2}+2n\pi\right).$$

and similarly, the initial and end points of $L_2(k, n)$ are

$$I_2(k, n) = \left(\frac{1}{2\omega - 1}\left(\frac{\pi}{2} + 2k\pi\right),\ \frac{1}{2\omega - 1}\left(-\frac{\pi}{2} + 2n\pi\right)\right)$$

and

$$E_2(k, n) = \left(\frac{\pi}{2} + 2k\pi,\ \frac{\pi}{2} + 2n\pi\right).$$

Notice that $E_1(k, n) = E_2(k, n)$ and $I_1(k, n) = I_2(k, n + 1)$, that is, $L_1(k, n)$ and $L_2(k, n)$ have the same end points and $L_1(k, n)$ and $L_2(k, n + 1)$ have the same initial points. Hence the segments

$$(L_2(k, 0),\ L_1(k, 0), L_2(k, 1),\ L_1(k, 1), L_2(k, 2),\ L_1(k, 2), \ldots)$$

starting at $I_2(k, 0)$ and passing through points

$$E_2(k, 0) = E_1(k, 0),\ I_1(k, 0) = I_2(k, 1),\ E_2(k, 1) = E_1(k, 1), \ldots$$

form a continuous curve.

Figure 2 illustrates the loci $L_1(k, n)$ and $L_2(k, n)$ with the value of ν varying from $2\omega - 1$ to unity for $n = 0, 1, 2$ and $k = 0$. The parameter value $\omega = 0.8$ is selected. The red curves are $L_1(k, n)$ and the blue curves are $L_2(k, n)$. The red and blue curves shift upward when n increases and rightward when k increases. There the initial point $I_2(0, 0)$ is infeasible, and $L_2(0, 0)$ is feasible only for $\nu \geq \sqrt{2\omega - 1}$. Notice that $I_1(0, n) = I_2(0, n + 1)$ at point I_n for $n = 0, 1$ and $E_1(0, n) = E_2(0, n)$ at point E_n for $n = 0, 1, 2$. $\gamma_1^m \simeq 1.493$ is the minimum γ_1-value of the segment $L_1(0, n)$ while $\gamma_1^M \simeq 2.733$ is the maximum γ_1-value of the segment $L_2(0, n)$. It can be checked that $\gamma_1^0 \simeq 2.354$ is the γ_1-value of the intersection of the segment $L_2(0, 0)$ with the horizontal axis and is identical with γ_1^* given in Theorem 3. The

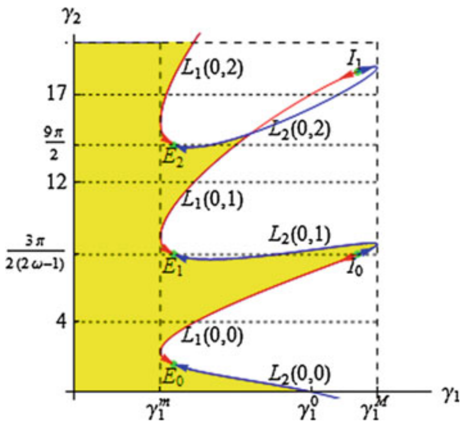


Fig. 2 Partition curves with $k = 0$ and $n = 0, 1, 2$

partition curve stays within the interval $[\gamma_1^m, \gamma_1^M]$ when $k = 0$ but its shape could be different for a different value of ω. It will be shown that the steady state is locally asymptotically stable in the yellow region of Fig. 2.

Apparently $\gamma_1^m < \gamma_1^0$ is implying that the steady state is locally asymptotically stable for $0 < \gamma_1 < \gamma_1^m$ and $\gamma_2 = 0$ due to Theorem 3. Since the segment $L_i(0, n)$ is in the interval $[\gamma_1^m, \gamma_1^M]$, there are no eigenvalues changing the sign of the real part when γ_2 increases. In other words, the real parts of the eigenvalues are negative for $\gamma_1 < \gamma_1^m$ and any $\gamma_2 > 0$. Hence the steady state is locally asymptotically stable. Such delays of (γ_1, γ_2) do not affect asymptotic behavior of the steady state and thus are harmless. We summarize this result in the following theorem:

Theorem 4 *If $k = 0$ and $0 < \gamma_1 < \gamma_1^m$, then delay γ_2 is harmless, so the steady state is locally asymptotically stable.*

We now move to the case of $\gamma_1 > \gamma_1^m$ and examine the directions of the stability switches by selecting γ_1 as the bifurcation parameter with fixed value of γ_2. So we consider the eigenvalue as a function of the bifurcation parameter, $\bar{\lambda} = \bar{\lambda}(\gamma_1)$. By implicitly differentiating Eq. (13) with respect to γ_1, we have

$$\frac{d\bar{\lambda}}{d\gamma_1} + \omega e^{-\bar{\lambda}\gamma_1}\left(-\frac{d\bar{\lambda}}{d\gamma_1}\gamma_1 - \bar{\lambda}\right) + (1-\omega)e^{-\bar{\lambda}\gamma_2}\left(-\frac{d\bar{\lambda}}{d\gamma_1}\gamma_2\right) = 0$$

implying that

$$\frac{d\bar{\lambda}}{d\gamma_1} = \frac{\omega\bar{\lambda}e^{-\bar{\lambda}\gamma_1}}{1 - \omega\gamma_1 e^{-\bar{\lambda}\gamma_1} - (1-\omega)\gamma_2 e^{-\bar{\lambda}\gamma_2}}. \tag{27}$$

Since from (13)

$$(1-\omega)e^{-\bar{\lambda}\gamma_2} = -\bar{\lambda} - \omega e^{-\bar{\lambda}\gamma_1},$$

the right-hand side of Eq. (27) can be rewritten as

$$\frac{d\bar{\lambda}}{d\gamma_1} = \frac{\omega\bar{\lambda}e^{-\bar{\lambda}\gamma_1}}{1 + \bar{\lambda}\gamma_2 + \omega(\gamma_2 - \gamma_1)e^{-\bar{\lambda}\gamma_1}}.$$

If $\bar{\lambda} = i\nu$, then

$$\frac{d\bar{\lambda}}{d\gamma_1} = \frac{i\nu\omega\,(\cos\nu\gamma_1 - i\sin\nu\gamma_1)}{1 + i\nu\gamma_2 + \omega(\gamma_2 - \gamma_1)\,(\cos\nu\gamma_1 - i\sin\nu\gamma_1)}. \tag{28}$$

Its real part is

$$\operatorname{Re}\left[\frac{d\bar{\lambda}}{d\gamma_1}\right] = \frac{\nu\omega\,[\sin\nu\gamma_1 + \nu\gamma_2\cos\nu\gamma_1]}{(1 + \omega(\gamma_2 - \gamma_1)\cos\nu\gamma_1)^2 + (\nu\gamma_2 - \omega(\gamma_2 - \gamma_1)\sin\nu\gamma_1)^2}, \tag{29}$$

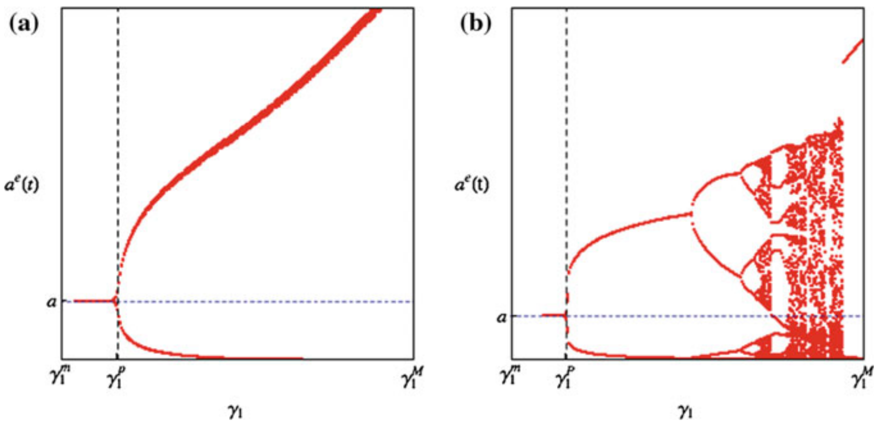


Fig. 3 Bifurcation diagrams I with different values of γ_2. **a** $\gamma_2 = 4$. **b** $\gamma_2 = 12$

As is already shown in Theorem 3, the steady state is asymptotically stable for $\gamma_1 = 0$ and any $\gamma_2 > 0$. Gradually increasing the value of γ_1 with fixed value of γ_2, the horizontal line crosses either $L_1(0, n)$ or $L_2(0, n)$. Consider first the intercept with the segment $L_1(0, n)$. Since both $\sin \nu\gamma_1$ and $\cos \nu\gamma_1$ are positive for $\nu\gamma_1 \in (0, \pi/2)$,

$$\mathrm{Re}\left[\frac{d\bar{\lambda}}{d\gamma_1}\right] > 0.$$

This inequality implies that as γ_1 increases, stability is lost when γ_1 crosses the segment $L_1(0, n)$. In the case of the linear learning process, local instability leads to global instability. However this is not necessary true if the learning process is nonlinear. To confirm global behavior, two bifurcation diagrams generated by the delay logistic equation (8) are illustrated.[3] In particular, we vary γ_1 from γ_1^m to γ_1^M in 0.01 increments, calculate 1000 values for each value of γ_1 and use the last 300 values to get rid of the transients. The local maximum and minimum values of the trajectory are plotted against γ_1 to construct a bifurcation diagram of a^e with respect to γ_1. In Fig. 3a, γ_2 is fixed at 4 and γ_1 increases along the lowest horizontal dotted line of Fig. 2. The stable steady state loses stability at $\gamma_1^P \simeq 1.691$, the intersection with the segment $L_1(0, 0)$.[4] The bifurcation diagram in Fig. 3a takes a distorted C-shape and its upper part is thick, indicating that trajectories are quasi-periodic with

---

[3]The linear equation (7) with two delays generates the same simple dynamics as the linear equation (3) with one delay. So no further considerations are given to it.

[4]Using the second equation of (25), we solve equation

$$4 = \frac{1}{\nu}\left(\pi - \sin^{-1}\left(\frac{\nu^2 - 2\omega + 1}{2\nu(1-\omega)}\right)\right)$$

for ν and then substitute the solution into the first equation of (25) to obtain the value of γ_1^P.

minor fluctuations in their local maximum values for $\gamma_1 > \gamma_1^P$. It is further seen that the periodic cycle expands as γ_1 becomes larger. In Fig. 3b, γ_2 is changed to 12 and γ_1 increases along the third highest horizontal dotted line of Fig. 2. The stability is lost at $\gamma_1^P \simeq 1.80$ where the dotted line crosses the segment $L_1(0, 1)$ from the left.[5] As γ_1 gets larger than γ_1^P, erratic (chaotic) behavior emerges via a period-doubling bifurcation. Further increasing γ_1 suddenly reduces complex dynamics to simple periodic oscillations for γ_1 being close to γ_1^M. Only the values of γ_2 are different between these two diagrams. So a larger γ_2 can be a source of erratic oscillations of $a^e(t)$. These results are numerically confirmed and thus summarized as follows:

Proposition 1 *The steady state loses local stability when increasing γ_1 crosses the segment $L_1(0, n)$ from left for the first time and global behavior of the unstable steady state exhibits simple oscillations if γ_2 is relatively small and complex oscillations if γ_2 is relatively large.*

Assume next that γ_1 crosses a segment $L_2(0, n)$ where $\cos \nu\gamma_1 < 0$ as $\nu\gamma_1 \in (\pi/2, \pi)$. It is clear from (29) that

$$\text{Re}\left[\frac{d\bar{\lambda}}{d\gamma_1}\right] = -\frac{\nu^3\omega \cos \nu\gamma_1}{(1 + \omega(\gamma_2 - \gamma_1)\cos \nu\gamma_1)^2 + (\nu\gamma_2 - \omega(\gamma_2 - \gamma_1)\sin \nu\gamma_1)^2}\frac{\partial\gamma_2}{\partial\nu} \tag{30}$$

where, by implicitly differentiating the second equation in (18) and using (14), we have

$$\frac{\partial\gamma_2}{\partial\nu} = -\frac{1}{\nu^2}\frac{\sin \nu\gamma_1 + \nu\gamma_2 \cos \nu\gamma_1}{\cos \nu\gamma_1}.$$

Since the first factor with the minus sign of (30) is positive, stability is switched to instability when $\partial\gamma_2/\partial\nu > 0$ and instabiliy might be switched to stability when $\partial\gamma_2/\partial\nu < 0$. Although it is possible to confirm analytically the sign of the derivative $\partial\gamma_2/\partial\nu$ on the segment $L_2(0, n)$, we numerically check it.[6] We also examine responses of the nonlinear learning as a function of γ_1 for two different fixed values of γ_2,

$$\gamma_2 = \frac{3\pi}{2(2\omega - 1)} \text{ and } \gamma_2 = \frac{9\pi}{2}.$$

We start with $\gamma_2 = 3\pi/(2(2\omega - 1)) \simeq 7.85$. Although it is not clear in Fig. 2, the segment $L_2(0, 1)$ takes a convex–concave shape. So the dotted horizontal line at $3\pi/(2(2\omega - 1))$ could have multiple intersects with $L_2(0, 1)$. The second equation of (26) determines a value of γ_2. So solving the following equation:

$$\frac{3\pi}{2(2\omega - 1)} = \frac{1}{\nu}\left(\sin^{-1}\left(\frac{\nu^2 - 2\omega + 1}{2\nu(1 - \omega)}\right) + 2\pi\right)$$

[5] See footnote 4 for detailed arguments to determine γ_1^P.

[6] See Matsumoto and Szidarovszky (2012) for analytical examinations of the sign of the derivative.

with $\omega = 0.8$ for ν yields three solutions,

$$\nu_a = 1,\ \nu_b \simeq 0.84, \text{ and } \nu_c = 0.6,$$

each of which is then substituted into the first equation of (26) to obtain three corresponding values of γ_1,

$$\gamma_1^a = \frac{\pi}{2} \simeq 1.57,\ \gamma_1^b \simeq 2.15 \text{ , and } \gamma_1^c = \frac{\pi}{2(2\omega - 1)} \simeq 2.62.$$

Notice that γ_1^a and γ_1^c are the γ_1 -values of the points E_1 and I_0 in Fig. 2. Fixing the parameter γ_2 at $3\pi/(2(2\omega - 1))$, we perform simulations of Eq. (8) for different γ_1 values to confirm two dynamic results; one is the appearance and disappearance of a limit cycle for $\gamma_1^a < \gamma_1 < \gamma_1^b$ and the other is initial point dependency of dynamics for $\gamma_1^b < \gamma_1 < \gamma_1^c$. We will discuss these results in detail.

Characterization of bifurcation occurring along the dotted line is given. Start with Fig. 4a. The steady state is locally stable for $\gamma_1^m < \gamma_1 < \gamma_1^a$ and loses stability for $\gamma_1 = \gamma_1^a$ at which $\partial\gamma_2/\partial\nu > 0$ implying that the real part of an eigenvalue is positive for $\gamma_1 > \gamma_1^a$. With further increasing γ_1, it bifurcates to a limit cycle which first expands, then shrinks and finally merges to the steady state to regain stability for $\gamma_1 = \gamma_1^b$. There $\partial\gamma_2/\partial\nu < 0$ implies that the real part of the same eigenvalue becomes negative again for $\gamma_1 > \gamma_1^b$. For $\gamma_1^b < \gamma_1 < \gamma_1^c$, the steady state is locally asymptotically stable as the dotted line is in the yellow region of Fig. 2. In order to examine global behavior, we simulate the nonlinear learning process with a constant initial function defined for $t \le 0$ having slightly different constant values, $\varphi(t) = 0.2$, $\varphi(t) = 0.4$, $\varphi(t) = 0.6$ and $\varphi(t) = 0.8$. Two results are numerically confirmed: the first is that the learning process generates the same dynamics regardless of the different initial functions if $\gamma_1 \le \gamma_1^b$ or $\gamma_1 > \gamma_1^c$; the second is that for $\gamma_1^b < \gamma_1 < \gamma_1^c$, a trajectory converges to the

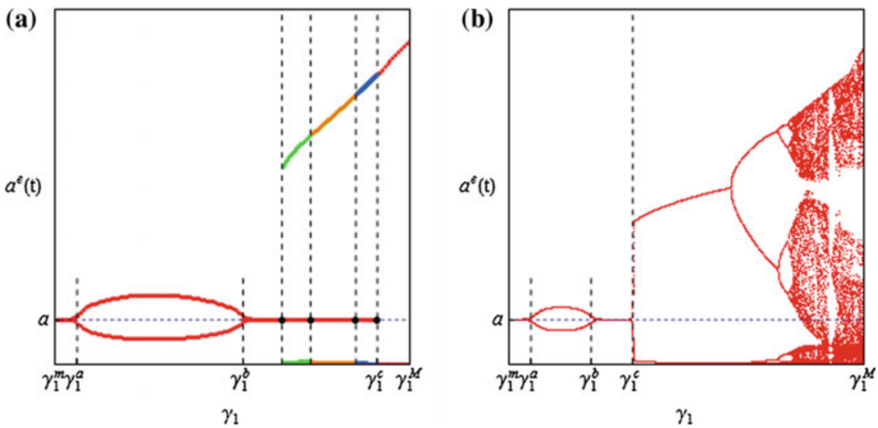


Fig. 4 Bifurcation diagram II with different values of γ_2. **a** $\gamma_2 = \frac{3\pi}{2(2\omega-1)}$. **b** $\gamma_2 = \frac{9\pi}{2}$

steady state when $\varphi(t) = 0.8$[7] while it bifurcates to a periodic cycle by discontinuous jump when any other initial function is selected and further, the jumping value of γ_1 depends on the selection of the initial function. In particular, the green trajectory with $\varphi(t) = 0.2$ jumps to the periodic cycle at the first dotted point in Fig. 4a, the orange trajectory with $\varphi(t) = 0.4$ at the second dotted point, the blue trajectory with $\varphi(t) = 0.6$ at the third dotted point and finally the red trajectory with $\varphi(t) = 0.8$ at the fourth dotted point. Depending on a choice of the initial function, the same delay equation generates different global dynamics.

In Fig. 4b, γ_2 is increased to $9\pi/2 \simeq 14.14$ and the horizontal dotted line at $9\pi/2$ crosses the segment $L_2(0, 2)$ twice at the points

$$\gamma_1^a = \frac{\pi}{2} \simeq 1.57 \text{ and } \gamma_1^b \simeq 1.78.$$

Stability is lost at $\gamma_1 = \gamma_1^a$ for which the dotted line crosses $L_2(0, 2)$ with $\partial\gamma_2/\partial\nu > 0$ and regained at $\gamma_1 = \gamma_1^b$ for which the dotted line crosses $L_2(0, 2)$ with $\partial\gamma_2/\partial\nu < 0$. The dotted line also crosses the segment $L_1(0, 1)$ with $\partial\gamma_2/\partial\nu > 0$ at

$$\gamma_1^c \simeq 1.93$$

for which stability is lost again. Taking $\varphi(t) = 0.9$, we simulate the model and obtain the following results. The appearance and disappearance of a limit cycle for $\gamma_1^a < \gamma_1 < \gamma_1^b$ is observed again. Although the initial point dependency is observed in the interval (γ_1^b, γ_1^c) in this case as well, we omit it from Fig. 4b to avoid the messy bifurcation diagram. Instead, it should be noticed that complex behavior emerges via a period-doubling bifurcation for $\gamma_1^c < \gamma_1 < \gamma_1^M$. Comparing these bifurcation diagrams where only the values of γ_2 are different leads to the same conclusion that a larger γ_2 can be a source of erratic oscillations of $a^e(t)$.

Proposition 2 *When the horizontal line of γ_1 crosses the segment of $L_2(0, n)$ for the first time from left, then three different dynamics emerge; (1) the steady state bifurcates to a limit cycle that expands, shrinks and then merges to the steady state for $\gamma_1^a < \gamma_1 < \gamma_1^b$; (2) depending on a choice of the initial functions, it becomes locally asymptotically stable or bifurcates to a periodic oscillation for $\gamma_1^b < \gamma_1 < \gamma_1^c$; (3) it proceeds to a periodic oscillation or chaos via a period-doubling cascade for $\gamma_1^c < \gamma_1 < \gamma_1^M$ according to whether γ_2 is small or large.*

In the next simulation, γ_2 is increased to 17 and then kept fixed. γ_1 is increased along the highest horizontal dotted line of Fig. 2. As can be seen, the horizontal line has three intercepts. At the first one with $L_1(0, 2)$ the steady state becomes unstable. At the second one with $L_1(0, 1)$ the real part of one more eigenvalue becomes positive. At the third intercept with $L_2(0, 2)$ the real part of only one of the two eigenvalues changes back to negative. Therefore stability cannot be regained

[7] It is numerically verified that trajectories converge to the steady state for any initial values close to a (i.e., unity).

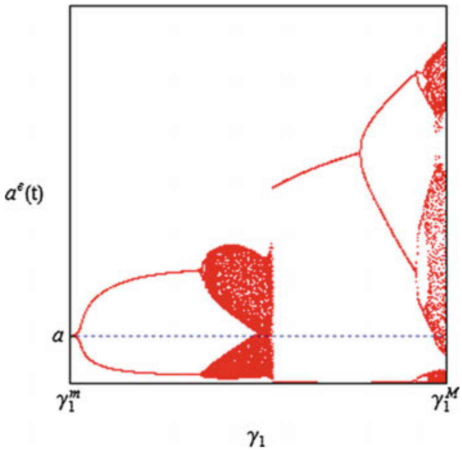


Fig. 5 Bifurcation diagram with $\gamma_2 = 17$

at this point, so no stability switch occurs. According to Fig. 5, the steady state is replaced with a periodic oscillation just after it becomes unstable and there is a very short period-doubling cascade to chaos. As we can see further, interesting dynamics begins; complex dynamics suddenly disappears and a periodic oscillation appears and undergoes a period-doubling bifurcation cascade to chaos again.

Let us summarize the main point that has been made so far.

Proposition 3 *(1) Given* $k = 0$, *the boundary of the stable region consists of the envelop of the segments* $L_1(0, n)$ *and* $L_2(0, n)$ *for* $n \geq 0$; *(2) depending values of* (γ_1, γ_2), *the nonlinear learning process can generate a wide spectrum of dynamics ranging from simple periodic oscillations to complex aperiodic oscillations when the steady state loses local stability.*

We can also show that at stability switches only one eigenvalue changes the sign of its real part, that is, the pure complex eigenvalues are single. Assume not, then $\bar{\lambda} = i\nu$ solves both equations

$$\bar{\lambda} + \omega e^{-\bar{\lambda}\gamma_1} + (1 - \omega)e^{-\bar{\lambda}\gamma_2} = 0 \tag{31}$$

and

$$1 - \omega\gamma_1 e^{-\bar{\lambda}\gamma_1} - (1 - \omega)\gamma_2 e^{-\bar{\lambda}\gamma_2} = 0 \tag{32}$$

from which we have

$$e^{-\bar{\lambda}\gamma_1} = \frac{1 + \bar{\lambda}\gamma_2}{(\gamma_1 - \gamma_2)\omega} \text{ and } e^{-\bar{\lambda}\gamma_2} = \frac{-1 - \bar{\lambda}\gamma_1}{(\gamma_1 - \gamma_2)(1 - \omega)}. \tag{33}$$

If $\bar{\lambda} = i\nu$, then by comparing the real and imaginary parts, we conclude that

$$\sin \nu\gamma_1 + \nu\gamma_2 \cos \nu\gamma_1 = \sin \nu\gamma_2 + \nu\gamma_1 \cos \nu\gamma_2 = 0$$

or

$$\tan \nu\gamma_1 + \nu\gamma_2 = \tan \nu\gamma_2 + \nu\gamma_1 = 0. \tag{34}$$

Let $L_i(k, n)$ denote the segment containing (γ_1, γ_2), then this point is a common extremum of $L_i(k, n)$ with respect to γ_1 and γ_2, which is impossible, since $L_i(k, n)$ is a differentiable curve.

Until this point, we examined the curves $L_1(k, n)$ and $L_2(k, n)$ for $k = 0$. If $k \geq 1$, then these curves are shifted to the right and slightly modified resulting in similar patterns. If we fix the value of γ_2 and gradually increase the value of γ_1 from zero, then it is unknown theoretically how the stability region, if any, looks like behind the $L_1(0, n)$ and $L_2(0, n)$ curves. By performing repeated simulations no stability region was found here.

Consider now a point (γ_1^*, γ_2^*) with positive coordinates, and consider the horizontal line $\gamma_2 = \gamma_2^*$ and its segment for $0 \leq \gamma_1 \leq \gamma_1^*$. There are finitely many intercepts of this horizontal segment with the set

$$L = \cup_{n=0}^{\infty} \cup_{k=0}^{\infty} \{L_1(k, n) \cup L_2(k, n)\}.$$

Let $s(\gamma_1^*, \gamma_2^*)$ denote the number of intercepts where stability is lost and $\boldsymbol{g}(\gamma_1^*, \gamma_1^*)$ the number of intercepts where stability can be regained. With (γ_1^*, γ_2^*) the system is asymptotically stable if $\boldsymbol{g}(\gamma_1^*, \gamma_2^*) \geq s(\gamma_1^*, \gamma_2^*)$ and unstable otherwise. The stability region is illustrated as the yellow domain in Fig. 2.

4.2 The Symmetric Case of $\omega = \frac{1}{2}$

If $\omega = 1/2$, then Eqs. (14) and (15) become

$$\cos(\nu\gamma_1) + \cos(\nu\gamma_2) = 0$$

and

$$\nu - \frac{1}{2}\left(\sin(\nu\gamma_1) + \sin(\nu\gamma_2)\right) = 0$$

and the segments $L_1(k, n)$ and $L_2(k, n)$ are simplified as follows:

$$L_1(k, n) : \begin{cases} \gamma_1 = \dfrac{1}{\nu}\left(\sin^{-1}(\nu) + 2k\pi\right) & (k \geq 0) \\[2ex] \gamma_2 = \dfrac{1}{\nu}\left(\pi - \sin^{-1}(\nu) + 2n\pi\right) & (n \geq 0) \end{cases} \tag{35}$$

and

$$L_1(k,n):\begin{cases}\gamma_1=\dfrac{1}{\nu}\left(\pi-\sin^{-1}(\nu)+2k\pi\right) & (k\geq 0)\\ \\ \gamma_2=\dfrac{1}{\nu}\left(\sin^{-1}(\nu)+2n\pi\right) & (n\geq 0).\end{cases} \tag{36}$$

Clearly ν has to be in the unit interval in order to have feasible solutions. The segments $L_1(k,n)$ and $L_2(k,n)$ for small values of k and n are illustrated as the red and the blue curves in Fig. 6. When $k=n=0$, these segments construct a hyperbolic curve passing through the point $(\pi/2,\pi/2)$ which is the common point of $L_1(0,0)$ and $L_2(0,0)$. It divides the first quadrant of the (γ_1,γ_2) plane into two subregions: in the yellow region under the curve, the steady state is locally asymptotically stable and in the white region above, it is locally unstable. Note that the curve is symmetric with respect to the diagonal and asymptotic to the line $\gamma_i=1$ for $i=1,2$. This implies that any delay $\gamma_i>0$ is harmless if $\gamma_j\leq 1$ for $i,j=1,2$ and $i\neq j$.

Two numerical simulations are done with two different values of γ_1. In the first simulation depicted in Fig. 7a, we increase the value of γ_2 from zero to 20 along the vertical dotted line at $\gamma_1=\pi/2$. The steady state loses stability at $\gamma_2=\pi/2$ and bifurcates to cyclic oscillations with finite number of periodicity as γ_2 becomes larger. As far as the simulations are concerned, only periodic cycles can be born. In the second simulation shown in Fig. 7b, we change the value of γ_1 to $(\pi/2)+1$. A limit cycle emerges after stability is lost at $\gamma_2=\gamma_2^a$ when the increasing value of γ_2 crosses the hyperbolic curve, then exhibits aperiodic oscillations for γ_2 being about 14 and returns to periodic oscillations afterwards.

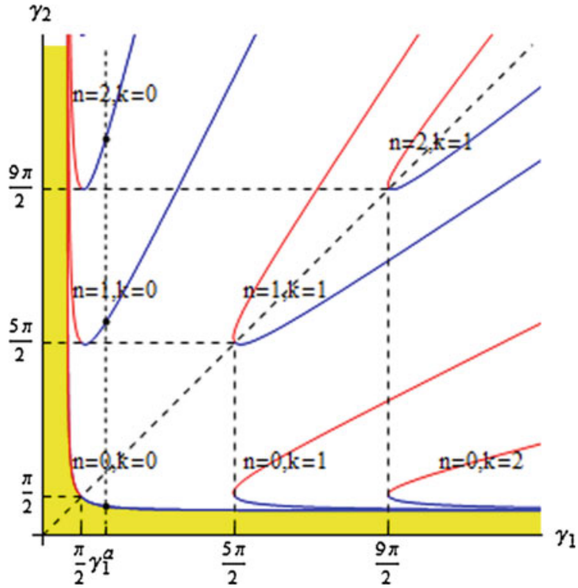


Fig. 6 Partition curvs with $\omega=0.5$

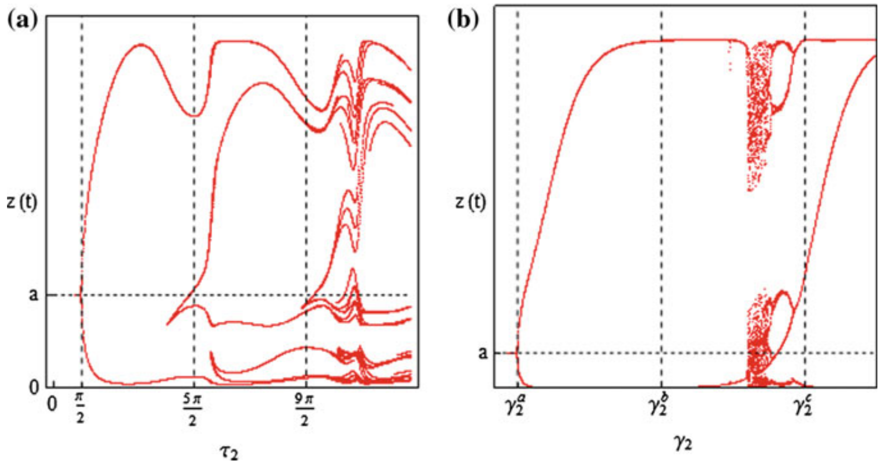


Fig. 7 Bifurcation diagrams with different values of γ_1. **a** $\gamma_1 = \frac{\pi}{2}$. **b** $\gamma_1 = \frac{\pi}{2} + 1$

5 Conclusion

An adaptive learning process is introduced when the monopoly knows its cost function, the marginal price, and uncertain about the maximum price. It is able to update repeatedly its belief of the maximum price by comparing the actual and predicted market prices. It is assumed that the firm's prediction is either the most current delayed price information or it is obtained by interpolation or by extrapolation based on two delayed data. The asymptotical stability of the resulted dynamic learning process is examined. If it is asymptotically stable, then the beliefs of the firm about the maximum price converge to the true value, so successful learning is possible. Stability conditions are derived, the stability regions are determined and illustrated. The global behavior of the trajectory is examined by using simulation.

The dynamic models (5) and (7) are linear, when local asymptotical stability implies global asymptotical stability. However (6) and (8) are nonlinear, where only local asymptotic stability can be guaranteed under the derived conditions. The learning processes (3) and (4) can be generalized as

$$\dot{a}^e(t) = \boldsymbol{g}(a - a^e(t))$$

where function $\boldsymbol{g}$ is sign preserving, that is, for all $\delta \neq 0$,

$$\delta \boldsymbol{g}(\delta) > 0.$$

In our future research, different types of such nonlinear learning schemes will be introduced in our model and we will investigate the asymptotical behavior of the resulted dynamics. Uncertainty and learning of other model parameters will be additional subjects of our study.

References

Bellman R, Cooke K (1956) Differential-difference equations. Academic Press, New York

Cushing J (1977) Integro-difference equations and delay models in population dynamics. Springer, Berlin

Hayes N (1950) Roots of the transcendental equations associated with a certain difference-differential equation. J Lond Math Soc 25:226–232

Invernizzi S, Medio A (1991) On lags and chaos in economic dynamic models. J Math Econ 20:521–550

Matsumoto A, Szidarovszky F (2012) Nonlinear delay monopoly with bounded rationality. Chaos Solitons Fractals 45:507–519

Matsumoto A, Szidarovszky F (2013) An elementary study of a class of dynamic systems with single time delay. CUBO: Math J 15:1–7

Matsumoto A, Szidarovszky F (2014a) Discrete-time delay dynamics of boundedly rational monopoly. Decis Econ Financ 37:53–79

Matsumoto A, Szidarovszky F (2014b) Bounded rational monopoly with single continuously distributed time delay. In: Dieci R et al (eds) Nonlinear dynamics and financial modeling. Springer, Heidelberg, pp 83–107

Matsumoto A, Szidarovszky F (2015) Dynamic monopoly with multiple continuously distributed time delays. Math Comput Simul 108:99–118

Naimzada A (2012) A delay economic model with flexible time lag, mimeo

Naimzada A, Ricchiuti G (2008) Complex dynamics in a monopoly with a rule of thumb. Appl Math Comput 203:921–925

Nyarko Y (1991) Learning in mis-specified models and the possibility of cycles. J Econ Theory 55:416–427

Puu T (1995) The chaotic monopolist. Chaos Solitons Fractals 5:35–44

Different Modelling Approaches for Time Lags in a Monopoly

Luca Gori, Luca Guerrini and Mauro Sodini

Abstract This chapter considers different modelling approaches to study nonlinear monopolies with a downward and concave demand function based on the model by Naimzada and Ricchiuti (Appl Math Comput 203:921–925, 2008). In particular, the article characterises the dynamics of continuous time models with delays related to several assumptions regarding the bounded rationality of the monopolist. Some results about global dynamics are also obtained through simulations.

Keywords Chaos · Monopoly · Time delays

JEL Classification C62 · D43 · L13

1 Introduction

Monopoly perhaps represents the simplest market structure amongst the several major market forms analysed by economists. In fact, in first-year microeconomics courses it is often presented as the first and most intuitive kind of behaviour of producers and consumers in the market (Schotter 2008). In a monopoly, there is only one producer of a good or service and the inverse market demand (i.e. the marginal willingness to pay of consumers) is generally downward sloping. However, despite its relative simplicity, it has been shown that in the absence of perfect knowledge on the market

L. Gori (✉)
Department of Political Science, University of Genoa, Piazzale E. Brignole, 3a, 16125 Genoa (GE), Italy
e-mail: luca.gori@unige.it; dr.luca.gori@gmail.com

L. Guerrini
Department of Management, Polytechnic University of Marche, Piazza Martelli 8, 60121 Ancona (AN), Italy
e-mail: luca.guerrini@univpm.it

M. Sodini
Department of Economics and Management, University of Pisa, Via Cosimo Ridolfi, 10, 56124 Pisa (PI), Italy
e-mail: mauro.sodini@unipi.it

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_5

demand by the producer, the introduction of naïve adjustment mechanisms on the quantities produced by the monopolist can actually generate dynamics that do not converge towards the optimal solution of the problem. In this regard, in fact, Puu (1995) builds on a model with a downward sloping market demand having two inflection points and analyses in a discrete-time dynamic set up monopoly dynamics driven by a Newton-like adjustment mechanism, i.e. a mechanism based on the profit differential realised in two subsequent periods.

It is quite peculiar to note that since Puu (1995), perhaps motivated by an interest related to the study of game theory and strategic interaction, a vast and burgeoning literature is emerged to deepen our knowledge of duopoly dynamics, but with sporadic contributions on the dynamics of a monopoly with bounded rationality. Only recently, there has been a renewed interest in this issue in both a discrete-time framework (difference equation) and continuous-time framework with delays (delay differential equations). With specific regard to the former set up, we mention the work of Naimzada and Ricchiuti (2008) that introduces a rule based on the gradient dynamics in a monopoly with a nonlinear market demand and the work Naimzada and Ricchiuti (2011), which takes into account a monopoly with a nonlinear demand and a linear approximation of it on the side of the monopolist. More recently, Cavalli and Naimzada (2015) considers a behavioural rule equal to the one adopted by Naimzada and Ricchiuti (2008) and studies the dynamics of a monopoly with isoelastic demand, whereas some analytical developments of Puu (1995) were proposed by Al-Hdaibat et al. (2015). In the latter set up, instead, i.e. continuous-time models with delays, it is important to mention the works of Matsumoto and Szidarovszky (2012, 2014a, b, 2015a, b) and Matsumoto et al. (2013). Specifically, Matsumoto and Szidarovszky (2012) present a first attempt to model out the existence of discrete-time delays in a dynamic monopoly under the assumption of information lags on profits, linear market demand and non-constant speed of adjustment of the quantity produced. Also, Matsumoto and Szidarovszky (2014a) study a model with a nonlinear demand and non-constant marginal costs of production, implying strongly decreasing returns to scale. A special feature of this work is that of considering an asymmetric information structure between marginal revenues (that enter with delays) and marginal costs (incurred immediately by the firm). Then, the article of Matsumoto and Szidarovszky (2014b) builds on a continuous-time model with delays by taking into account the discrete-time set up as a starting point and using the method proposed by Berezowski (2001). Differently, the works of Matsumoto et al. (2013) and Matsumoto and Szidarovszky (2015a) propose two nonlinear monopoly models with linear demand where the monopolist adjusts production according with the gradient mechanism and delays are continuously distributed. Finally, in the work of Matsumoto and Szidarovszky (2015b) a feedback from the market on the expectation formation on the price of the monopolist is introduced.

The aim of the present article is to point out how different assumptions on the bounded rationality of the monopolist lead to models with strongly different dynamic properties. To this purpose, we take the discrete-time model of Naimzada and Ricchiuti (2008) as a starting point, that is a monopoly with a decreasing and concave demand function and constant returns to scale technology. The article shows a range of results depending on the way a discrete-time model is transformed into a continuous-time model with delays. In particular, we get two polar scenarios (within which the other models presented are included) in two distinct contexts. In the former case, by considering that the monopolist knows the market price as well as the linear approximation of the market demand in a neighbourhood of the current value of the price, the equilibrium of the system is always stable. In the latter case, by following the method proposed by Berezowski (2001) and considering that at every date the monopolist is not able to perfectly realise the production plan arranged in the previous period (because of frictions due to the long time required for production), the long-term outcomes of the economy may be characterised by chaotic behaviour. Our results, therefore, stresses the importance of the theoretical modelling framework used as a device that may dramatically change the long-term findings of an economy.

The rest of the article proceeds as follows. Section 2 presents several models to allow a discrete-time nonlinear monopoly model to be transformed into a continuous-time model with delays and gives some economic interpretations in each case analysed. Section 3 studies the dynamic properties of the different ways of transforming a discrete-time dynamic set up into a continuous-time model with delays. Section 4 outlines the conclusions.

2 The Model

We consider a market served by a monopolist who produces a homogeneous product. The market demand is described by the following downward sloping and concave function:

$$p = P(q) := a - bq^3, \tag{1}$$

where $q \geq 0$ is the quantity demanded by consumers, $p \geq 0$ is their corresponding marginal willingness to pay (price), $a > 0$ is the market size and $b > 0$. Output q is produced by a constant-return-to-scale technology, so that the cost function of the monopolist is $C(q) = cq$, where $c \geq 0$ is the average and marginal cost.[1] Profits are therefore given by

$$\Pi(q) = \left(a - bq^3\right) q - cq. \tag{2}$$

By standard calculations, the market equilibrium in a static context by assuming that the monopolist has perfect knowledge about the market demand is the value of q such that $\Pi(q)$ is maximized, that is

$$q_* = \sqrt[3]{(a - c)/4b}. \tag{3}$$

[1] In order to guarantee the nonnegativity of prices, we assume that $q < \sqrt[3]{a/b}$ holds from now on.

Note that we assume $a > c$ holds in order to guarantee the existence of a positive stationary equilibrium value of the quantity produced by the monopolist. For what follows, it is important to recall that

$$\frac{\partial \Pi(p)}{\partial q} = P'(q)q + P(q) - c. \tag{4}$$

Equation (4) implies that marginal profits depend on components related to the market demand ($P'(q)$ and $P(q)$) as well as on a component related to the quantity produced (q). We note that the component associated with the cost function enters in a relatively simple way in the marginal profit equation. In fact, given our assumption of constant returns to scale technology, the marginal cost is a constant (see Matsumoto and Szidarovszky 2014a, for a study of a monopoly dynamic model where the marginal cost is not constant).

Dynamics. Consider a dynamic setting in which we will take into account several assumptions about limited knowledge of the monopolist. However, all the models that we will build on later in this article share the assumption for which the monopolist chooses future production by considering the rule of thumb formerly introduced by Baumol and Quandt (1964) and recently used by Naimzada and Ricchiuti (2008) in a discrete-time monopoly to show that it may give birth to nonlinear dynamics. We now briefly recall that the dynamic model of Naimzada and Ricchiuti (2008) is described by the following equation:

$$q(t+1) = \phi(q(t)) := q(t) + k\frac{\partial \Pi(t)}{\partial q(t)}, \tag{5}$$

where $k > 0$ is the constant speed of adjustment of output and $\partial \Pi(t)/\partial q(t)$ is the marginal profit. This implies that at time t the firm knows its own marginal profitability and uses this information to produce at time $t + 1$. In other words, at time t the monopolist chooses to start with a production that will be effective in the time interval $[t, t + 1)$. This production process will actually bring to the market the quantity $q(t + 1)$ at time $t + 1$. We note that with this kind of modelling approach markets are open at discrete-time intervals and no trading takes place in the interval of time $(t, t + 1)$. One of the advantages of dealing with a continuous-time set up is precisely the one of allowing markets to be always open, so that through time delays it is possible to consider the existence of some lags related to either the information set of the firm or technology of production (gestation lags).

Consider now some alternative modelling approaches to transform a discrete-time model into a continuous-time model with delays. In this section, we will focus on the description of each specific scheme adopted as well as on its economic interpretation. Later, we will analyse the dynamic properties of the models just introduced.

Case 1. After a simple algebraic manipulation, we note that (5) can be rewritten as follows:

$$[q(t+1) - q(t)] = k\frac{\partial \Pi(t)}{\partial q(t)}. \tag{6}$$

Then, by interpreting the term within brackets on the left-hand side of (6) as an approximation of $\partial q(t)/\partial t$ and assuming the existence of a delay about the knowledge of the expression at the right-hand side of such an equation, we derive the following time-delayed model:

$$\dot{q} = k(a - c - 4bq_d^3), \tag{7}$$

where $q_d := q(t - \tau)$ is the variable q by considering one delay $\tau \geq 0$ (the time index has been omitted for simplicity). From an economic point of view, (7) tells us that the instantaneous changes in production ($\dot{q}$) depend on the marginal profitability at time $t - \tau$. In other words, we are assuming that production is immediately available but there exists a time lag between the time at which the firm computes its own marginal profit ($t - \tau$) and the time at which such a marginal profit is used to produce final output (t). In fact, it is reasonable to assume that the monopolist repeatedly uses a market research in order to adjust production but also that such an analysis takes a non-negligible amount of time, so that production choices are affected by a time lag. We note that the explicit knowledge of production at time t is not required with this kind of adjustment mechanism. Another possible interpretation is that the technology requires a time τ for bringing production to completion and that marginal profit is computed at time t by considering the quantity produced at time $t - \tau$. In other words, the decision to produce q_d is revealed on the market at time t. Then, with this kind of interpretation we do not have any information lag.

Case 2. Let us assume that the monopolist produces without delays and chooses to use a mechanism similar to the one detailed in *Case 1* above based on marginal profits as a behavioural rule, that is if the marginal profit is positive (resp. negative), the firm increases (resp. decreases) production. Assume also that the firm knows (1) the price components that enter the marginal profit with a time delay τ, and (2) the quantity currently produced by starting from (6) at time t. Then, we get

$$\dot{q} = k(a - c - 3bqq_d^2 - bq_d^3). \tag{8}$$

Case 3. Let us now assume as in *Case 2* that the monopolist produces without delays and chooses to use a mechanism analogous to the one described in *Case 1* and *Case 2* as a behavioural rule based on marginal profits. However, different from the previous case we assume that the firm does not know current production but knows the stock produced in a previous period. In addition, the firm also knows the market price and the linear approximation of the market demand in a neighbourhood of the current value of the price. Then, by starting from (6) we get

$$\dot{q} = k(a - c - 3bq_dq^2 - bq^3). \tag{9}$$

Case 4. An alternative way to switch from a discrete-time model to a continuous-time model is given by the following procedure. To this purpose, we note that (5) is also equivalent to the equation

$$[q(t+1) - q(t)] = \phi(q(t)) - q(t). \tag{10}$$

As in (6), by assuming that the term in brackets represents an approximation of $\partial q(t)/\partial t$ and also that there exists a time delay about the knowledge of the expression on the right-hand side of (10), we get

$$\dot{q} = q_d + k(a - c - 4bq_d^3) - q. \tag{11}$$

From an economic point of view, (11) tells us that the instantaneous variation of production is based on the differential existing between the target (based on past information with a delay τ) and current production. If such a differential is positive (resp. negative), production will tend to increase (resp. reduce). Different from the mechanism detailed previously, the explicit knowledge of production at time t is now necessary to adjust production.

Case 5. Amongst the several ways to transform a discrete-time model into a continuous-time model, an interesting technique is the one proposed by Berezowski (2001). Assume that at time $t - \tau$ the monopolist plans production for time t by using the rule of thumb specified in (5). In particular, we consider that at every time t the monopolist is not able to perfectly realise the production plan arranged at time $t - \tau$ (this is because of the existence of frictions due to the long time required for production). The technique proposed by Berezowski (2001) implies that the dynamics of the model are described by the following equation:

$$\sigma\dot{q} + q = q_d + k\left(a - c - 4bq_d^3\right), \tag{12}$$

where $\sigma \geq 0$ is a parameter that measures the degree of friction in production. Equation (12) tells us that in a phase of output growth, i.e. $\dot{q} > 0$, the monopolist is not able to realise a sufficiently large amount of products, meaning that realised production is smaller than the one planned in the previous period. The opposite holds in a phase of recession.

Remark 1 If $\sigma = 0$, Eq. (12) boils down to the discrete-time model described by (5).

Remark 2 If $\sigma = 1$, Eq. (12) boils down to the model described by Eq. (11). Then, from an mathematical point of view (but not from an economic point of view) the model described by (11) can be viewed as a sub-case of the model described by (12).

Case 6. In this last case we begin with the behavioural rule adopted by Puu (1995) in a discrete-time dynamic monopoly. In particular, he assume that the firm knows the profit obtained at two subsequent dates (i.e. the profit obtained in the current period and the one obtained in the previous period) and uses this information to choose next period output. Formally, this rule implies

$$q_{t+1} = q_t + k\frac{\Pi(q_t) - \Pi(q_{t-1})}{q_t - q_{t-1}}. \tag{13}$$

By generalising (13) to an arbitrarily time interval l, we get

$$\left[q_{t+l} - q_t\right] = k\frac{\Pi(q_t) - \Pi(q_{t-l})}{q_t - q_{t-l}}. \tag{14}$$

By assuming that the left-hand side of (14) represents an approximation of the first derivative and that there exists a time delays d_2 in the knowledge of the expression q_t at the right-hand side of (14), then by using the profit equation (2) after some algebraic manipulations one gets the following equation:

$$\dot{q} = k\left[a - c - b\left(q_{d_1}^3 + q_{d_2}^3 + q_{d_1}^2 q_{d_2} + q_{d_1} q_{d_2}^2\right)\right], \tag{15}$$

where $d_1 = d_2 + l$. The possible economic interpretations of (15) are similar to those detailed in *Case 1*, where we have substituted out the marginal profit with the ratio between the profit differential and the quantity differential at two subsequent dates. In the light of Remark 2, in the following section we will ignore the analysis of *Case 4*.

3 Dynamic Properties

The aim of this section is to analyse the local properties of the models introduced previously and to present some simulations also to infer about some global properties of the different systems. We note that in all cases considered in Sect. 2 there is only one delay and there exists a unique positive equilibrium $q_* = \sqrt[3]{(a-c)/4b}$, obtained by solving $\dot{q} = 0$ with $q_d = q = q_*$. The same result is obtained when there are two delays by considering that $\dot{q} = 0$ with $q_{d_1} = q_{d_2} = q = q_*$. In order to avoid to make the exposition tedious, we will present all analytical details only for Cases 1, 5 and 6. In fact, as all models from 1 to 5 are related to the equation $\dot{q}(t) = z(q(t), q(t-\tau))$, for every model presented the corresponding characteristic equation will be of the form $\lambda + A + Be^{-\lambda\tau} = 0$, where A and B are appropriate parameters. However, we will see that by passing from one model to another the results just obtained may be quite different, both from an economic and mathematical point of views and by considering local and global aspects.

Case 1. The local stability of the unique positive equilibrium point q_* of Eq. (7) is governed by the roots of the corresponding characteristic equation for Eq. (7).

By linearising (7) at q_*, we obtain

$$\dot{q} = -12bkq_*^2(q_d - q_*). \tag{16}$$

The characteristic equation associated to the linearised equation is

$$\lambda + 12bkq_*^2 e^{-\lambda\tau} = 0. \tag{17}$$

When $\tau = 0$, one gets the real eigenvalue $\lambda = -12bkq_*^2 < 0$. Thus, q_* is locally asymptotically stable. Let $\tau > 0$ and assume $\lambda = i\omega$ $(\omega > 0)$ to be a root of (17). Then we have

$$i\omega + 12bkq_*^2 \left[\cos(\omega\tau) - i\sin(\omega\tau)\right] = 0. \tag{18}$$

Separating the real and imaginary parts of Eq. (18) yields

$$\omega = 12bkq_*^2 \sin(\omega\tau) \text{ and } \cos(\omega\tau) = 0.$$

Therefore, the characteristic equation (17) has purely imaginary roots $\lambda = \pm i\omega_0$ when $\tau = \tau_0$, where

$$\omega_0 = 12bkq_*^2 \text{ and } \tau_0 = \pi/(2\omega_0). \tag{19}$$

These roots are simple. In order to consider the way of the complex roots of Eq. (17) crossing through the imaginary axis when $\tau = \tau_0$, let $\lambda(\tau) = \nu(\tau) + i\omega(\tau)$ be the root of Eq. (17) near $\tau = \tau_0$ with $\nu(\tau_0) = 0$ and $\omega(\tau_0) = \omega_0$. Then, one can derive that

$$sign\left\{\left.\frac{d(Re\lambda)}{d\tau}\right|_{\tau=\tau_0}\right\} = sign\left\{Re\left.\left(\frac{d\lambda}{d\tau}\right)^{-1}\right|_{\tau=\tau_0}\right\} = sign\left\{\frac{1}{\omega_0^2}\right\}.$$

Theorem 3 *Let τ_0 be defined by (19). Then the positive equilibrium q_* of Eq. (7) is locally asymptotically stable when $\tau \in [0, \tau_0)$ and unstable for $\tau > \tau_0$. Furthermore, Eq. (7) undergoes a Hopf bifurcation at $q = q_*$ when $\tau = \tau_0$.*

Remark 4 As is known, the result of Theorem 3 implies that for values of τ a few larger than τ_0 there exists an invariant curve of class C_1 that surrounds the equilibrium. It is possible to show (but we will show only some numerical evidence for this result) that the bifurcation is super-critical, that is the invariant curve attracts all trajectories starting from close enough to it. In addition, numerical evidence allows us also to show that even moving away from τ_0, results remain unchanged from a qualitative point of view and there still exist nonexplosive trajectories.

We note that it is possible to obtain a result about the global stability of the equilibrium by taking sufficiently small values of the delay. This result is summarised in the following theorem (see also Fig. 1d).

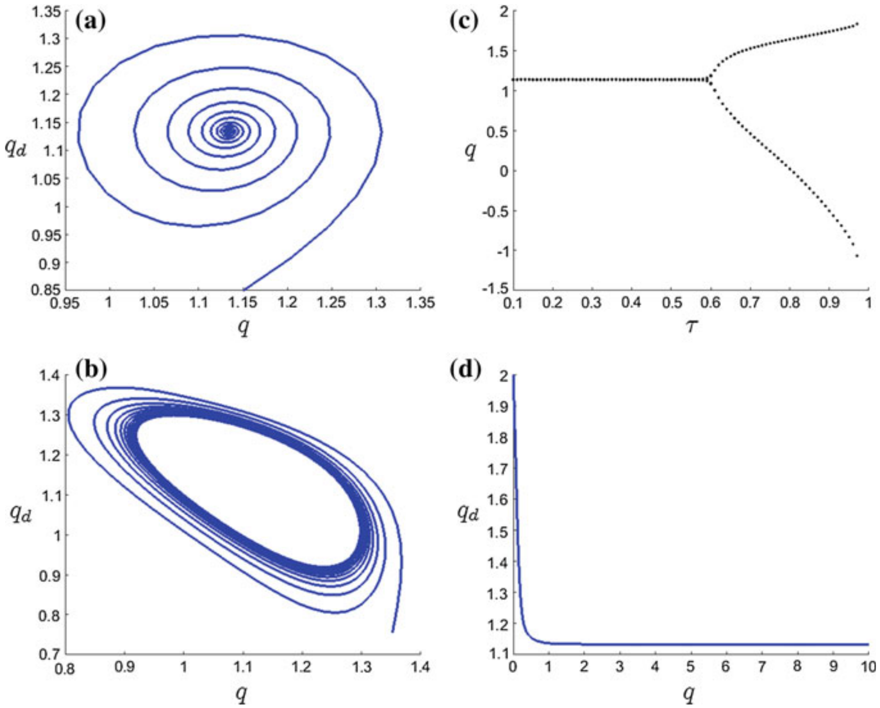


Fig. 1 Parameter set: $a = 4$, $b = 0.6$, $c = 0.5$ and $k = 0.28$. **a** A trajectory convergent towards the equilibrium for $\tau = 0.53$. **b** A trajectory convergent towards the invariant curve generated by the Hopf bifurcation for $\tau = 2$. **c** Bifurcation diagram for τ ($\tau_0 \cong 0.6058867855$). We note that this bifurcation diagram is economically meaningful only for values of τ smaller than almost 0.73. In fact, for larger values of τ the quantity would be negative and then economically meaningless. This result is due to the particular form of the demand curve that becomes negative for sufficiently large values of the quantity demanded by consumers. This problem can actually exist also in discrete-time models and can be overcome by accounting for (adequate) constraints on prices or quantities. However, the study of this issue is rather complicated and we avoid to deepen this point in the present article, as it may be subject to possible future research. **d** A trajectory that starts far away from the stationary equilibrium converges towards it ($\tau < 1/12bkq_*^2e^{-\lambda\tau} \cong 0.1418982769$)

Theorem 5 *Assume that* $\tau < 1/\left(12bkq_*^2e\right)$. *Then the positive steady-state solution of Eq. (11) is globally asymptotically stable.*

Proof Let $x = q - q_*$. Then Eq. (7) becomes $\dot{x} = -12bkq_*^2x_d - 4bkx_d^2(3q_* + x_d)$ with $x_* = 0$. The statement follows from Theorem 4.2 of Györi (1990) with $d = 12bkq_*^2$, $\sigma = \tau$, $f(x) = -4bkx^2(3q_* + x)$. Note that condition (a_2) $xf(x) \geq 0$ is verified since it yields $-3q_* \leq x \leq 0$, i.e. $0 < q \leq q_*$, whether the condition $|f(x)| < d\,|x|$ holds true since it gives $x^2 + 3q_*x - 3q_*^2 < 0$. Thus, $-(3 + \sqrt{21})/2 < x < 0$, i.e. $0 < q < q_*$. ■

Case 2. By considering the linearised system at q_* and by studying the solutions of the associated characteristic equation, we obtain the following result:

Theorem 6 *Let* $\tau_0 = \frac{1}{\omega_0}\left[\arctan\left(-\frac{\omega_0}{3bkq_*^2}\right) + \pi\right]$, *where* $\omega_0 = 6\sqrt{2}bkq_*^2$. *Then, the positive equilibrium* q_* *of Eq. (8) is locally asymptotically stable for* $\tau \in [0, \tau_0)$ *and unstable for* $\tau > \tau_0$. *Equation (7) undergoes a Hopf bifurcation*[2] *at the equilibrium* $q = q_*$ *when* $\tau = \tau_0$.

Also in this case, it is possible to show that for a small enough value of τ the stationary equilibrium q_* is globally asymptotically stable and that the Hopf bifurcation shown in Theorem 6 is super-critical. Now, by a direct comparison, it is possible to verify that in this case the bifurcation value of τ is larger than the one of Case 1. This implies (ceteris paribus) that the equilibrium of this model loses stability (and the dynamics are captured by a limit cycle) for a value of the delay larger than in the previous model and then the equilibrium is much more stable in this case. In this regard, for the same parameter values used in Fig. 1a–c, we have that the bifurcation value of τ is 1.042, whereas in the previous examples it was 0.6.

Case 3. The characteristic equation of the linearisation of Eq. (9) at the unique positive equilibrium q_* of (9) writes as

$$\lambda + 9bkq_*^2 + 3bkq_*^2 e^{-\lambda\tau} = 0. \tag{20}$$

In absence of delay, q_* is locally asymptotically stable since the root $\lambda = -12bkq_*^2 < 0$ is the unique solution of (20). If $\tau > 0$ and $\lambda = i\omega$ $(\omega > 0)$ is a root of (20), we can derive that

$$\omega = 3bkq_*^2 \sin(\omega\tau) \ \text{ and } 9bkq_*^2 = -3bkq_*^2 \cos(\omega\tau).$$

By squaring these equations and then adding them together, we find $\omega^2 = -72b^2k^2q_*^4$. Hence, Eq. (20) has no purely imaginary root. In conclusion, the positive equilibrium q_* of Eq. (9) is locally asymptotically stable for all $\tau \geq 0$.

Similar to Case 2 it is possible to show for Case 3 that q_* is globally asymptotically stable for sufficiently low values of τ.

Remark 7 The knowledge of the current market price and the use of a linear approximation of the market demand are actually stabilising devices that make the equilibrium locally asymptotically stable irrespective of the information on the delays about the production process (gestation lags).

We do not explicitly deal with Case 4 as it is a sub-case of Case 5, which is discussed below.

Case 5. In this case we will present all the mathematical steps because, as we shall see from the next Theorem 8, the results are more articulated. First of all, let us rewrite Eq. (12) in the following way:

$$\dot{q} = -\frac{q}{\sigma} + \frac{q_d}{\sigma} + \frac{k(a - c - 4bq_d^3)}{\sigma}, \tag{21}$$

[2] The transversality condition can easily be verified.

where $q_d := q(t-\tau)$, $\sigma > 0$ and $0 < q < \sqrt[3]{a/b}$. In order to analyse the stability of the equilibrium q_*, we linearise Eq. (21) at q_* and get

$$\dot{q} = -\frac{1}{\sigma}(q - q_*) + \frac{(1 - 12bkq_*^2)}{\sigma}(q_d - q_*). \tag{22}$$

The characteristic equation of the linearised equation at q_* takes the following form:

$$\lambda + \frac{1}{\sigma} - \frac{(1 - 12bkq_*^2)}{\sigma} e^{-\lambda\tau} = 0. \tag{23}$$

For $\tau = 0$, (23) becomes $\lambda = -12bkq_*^2/\sigma < 0$. Thus, we can conclude that, in the absence of delay, the positive equilibrium q_* is locally asymptotically stable. For $\tau > 0$, it is well known that q_* is locally asymptotically stable if all roots of Eq. (22) have negative real parts. Let $\lambda = i\omega$ ($\omega > 0$) be a root of (23). Then

$$i\omega + \frac{1}{\sigma} - \frac{(1 - 12bkq_*^2)}{\sigma}\left[\cos(\omega\tau) - i\sin(\omega\tau)\right] = 0.$$

Separating the real and imaginary parts, we obtain

$$\omega = -\frac{(1 - 12bkq_*^2)}{\sigma}\sin(\omega\tau) \;\text{ and }\; \frac{1}{\sigma} = \frac{(1 - 12bkq_*^2)}{\sigma}\cos(\omega\tau),$$

which imply

$$\omega^2 = \frac{(1 - 12bkq_*^2)^2 - 1}{\sigma}.$$

There exists $\omega > 0$ if $\left|1 - 12bkq_*^2\right| > 1$, that is if $6bkq_*^2 > 1$. For $k > 1/(6bq_*^2)$, set

$$\omega_0 = \frac{\sqrt{(1 - 12bkq_*^2)^2 - 1}}{\sigma} \;\text{ and }\; \tau_0 = \frac{1}{\omega_0}\left[\arctan(-\omega_0) + \pi\right]. \tag{24}$$

A direct calculation shows $\lambda = i\omega_0$ to be a simple root of (23). Let $\lambda(\tau) = \nu(\tau) + i\omega(\tau)$ denote the root of Eq. (23) near $\tau = \tau_0$ satisfying $\nu(\tau_0) = 0$ and $\omega(\tau_0) = \omega_0$, with ω_0 and τ_0 defined in (24). By substituting $\lambda(\tau)$ into Eq. (23), and differentiating both sides of (23) with respect to τ, it follows that

$$\left(\frac{d\lambda}{d\tau}\right)^{-1} = -\frac{\sigma}{\lambda(\sigma\lambda + 1)} - \frac{\tau}{\lambda}.$$

This implies

$$sign\left\{\left.\frac{d\,(Re\lambda)}{d\tau}\right|_{\tau=\tau_0}\right\} = sign\left\{Re\left.\left(\frac{d\lambda}{d\tau}\right)^{-1}\right|_{\tau=\tau_0}\right\} = sign\left\{\frac{\sigma^2}{\sigma^2\omega_0^2+1}\right\}.$$

Thus, we have the transversal condition $[d\,(Re\lambda)\,/d\tau]_{\tau=\tau_0} > 0$. Accordingly, a Hopf bifurcation at $q = q_*$ occurs when $\tau = \tau_0$.

Summarising the discussion above we have the following.

Theorem 8 *Let τ_0 be defined as in (24). For Eq. (11) the following hold.*

(1) If $k \le 1/(6bq_*^2)$, the positive equilibrium q_* is locally asymptotically stable for $\tau \ge 0$.
(2) If $k > 1/(6bq_*^2)$, the positive equilibrium q_* is locally asymptotically stable for $\tau \in [0, \tau_0)$ and unstable for $\tau > \tau_0$. Equation (11) undergoes a Hopf bifurcation at q_* when $\tau = \tau_0$.

In this case, we present a result about global stability based on the work of Cooke et al. (1999). It is summarised in the following theorem, which is a consequence of Theorem 3.1 of Cooke et al. (1999).

Theorem 9 *Let τ_0 be defined as in (24).*

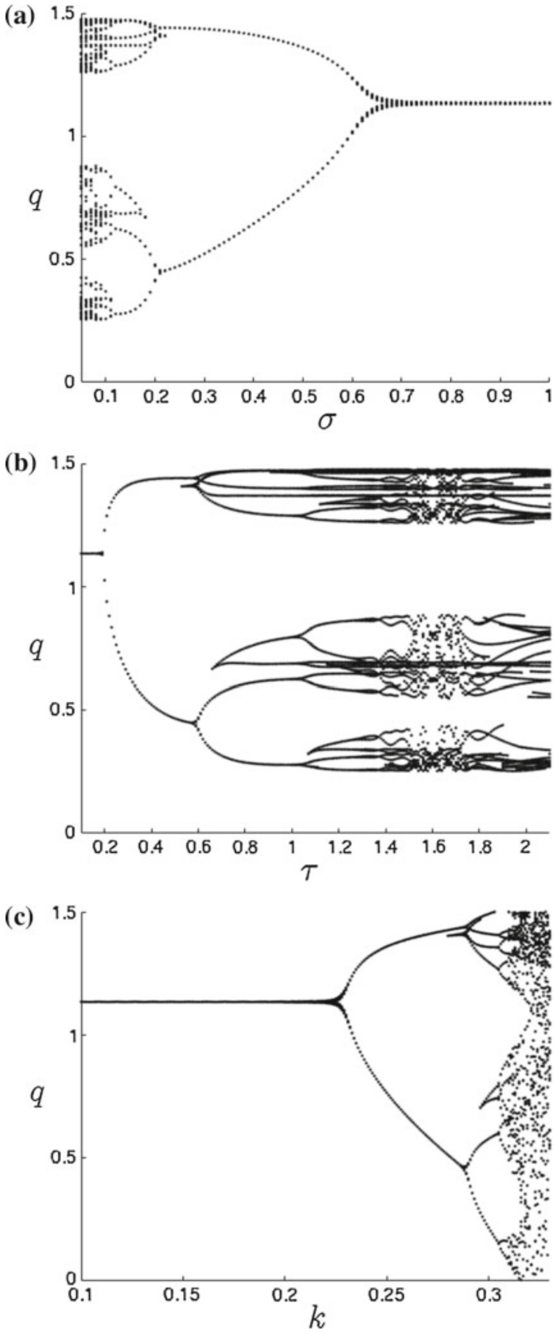
(1) If $k \le 1/\left\{12b\,(a/b)^{2/3}\right\}$ or if $q < 1/\left(2\sqrt{3bk}\right)$ and $1/\left\{12b\,(a/b)^{2/3}\right\} < k \le 1/(6bq_*^2)$, the positive equilibrium q_* is globally asymptotically stable for all $\tau \ge 0$.
(2) If $q < 1/\left(2\sqrt{3bk}\right)$ and $k > 1/(6bq_*^2)$, the positive equilibrium q_* of Eq. (11) is globally asymptotically stable for all $\tau \in [0, \tau_0)$.

Proof Write Eq. (11) as

$$\dot{q} = -\frac{q}{\sigma} + q_d\left[\frac{1}{\sigma} + \frac{k\,(a-c)}{\sigma q_d} - \frac{4bkq_d^2}{\sigma}\right].$$

Applying Theorem 3.1 in Cooke et al. with $d = 1$, $B(q) = 1/\sigma + k\,(a-c)\,/(\sigma q) - 4bkq^2/\sigma$ and $d_1 = 0$, we get the following conditions for the global stability: $1 - 4bkq^2 + k\,(a-c)\,/q > 0$ and $1 - 12bkq^2 > 0$. Since $1 - 4bkq^2 > 1 - 12bkq^2$, these conditions reduce to $1 - 12bkq^2 > 0$, i.e. $q < 1/\left(2\sqrt{3bk}\right)$. Recalling that $q < \sqrt[3]{a/b}$, the statement follows considering the two cases $\sqrt[3]{a/b} \le 1/\left(2\sqrt{3bk}\right)$ and $\sqrt[3]{a/b} > 1/\left(2\sqrt{3bk}\right)$. Note that $\sqrt[3]{a/b} \le 1/\left(2\sqrt{3bk}\right)$ yields $k \le 1/\{12b\,(a/b)^{2/3}\}$ and one has $1/\left\{12b\,(a/b)^{2/3}\right\} < \sqrt[3]{2}/\left\{3b\,[(a-c)/b]^{2/3}\right\} = 1/(6bq_*^2)$. ■

Fig. 2 Parameter set: $a = 4$, $b = 0.6$ and $c = 0.5$. **a** Bifurcation diagram for σ ($k = 0.29$ and $\tau = 1$). **b** Bifurcation diagram for τ ($k = 0.29$ and $\sigma = 0.2$). **c** Bifurcation diagram for k ($\sigma = 0.2$ and $\tau = 1$)



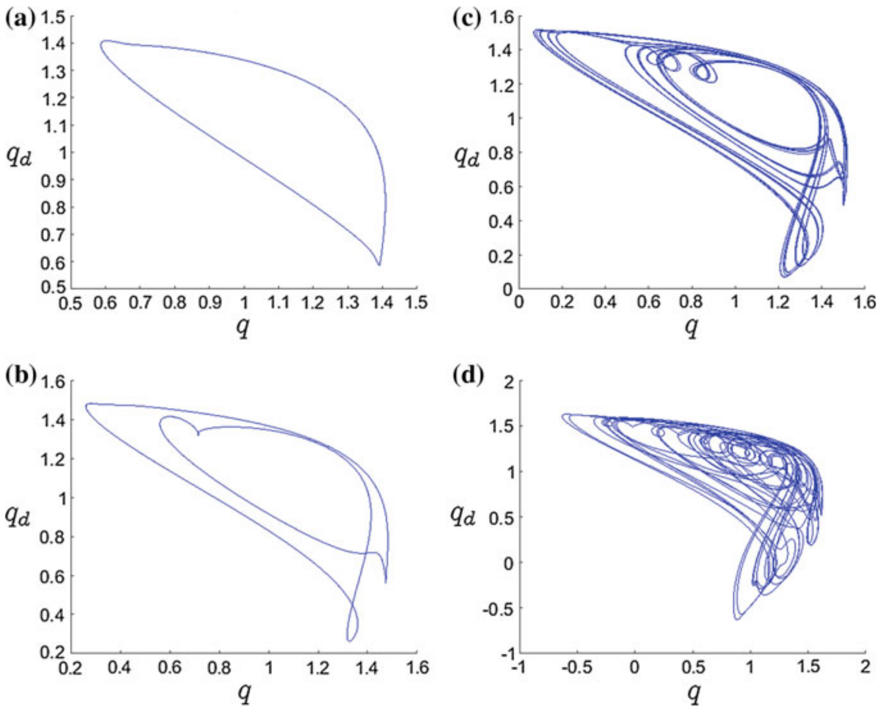


Fig. 3 Parameter set: $a = 4$, $b = 0.6$, $c = 0.5$, $\sigma = 0.2$ and $\tau = 1$. Evolution of the attractor when k varies. **a** $k = 0.27$. **b** $k = 0.29$. **c** $k = 0.31$. **d** $k = 0.35$

From a mathematical point of view, the technique proposed by Berezowski (2001), introduces a perturbation of the discrete dynamic system. By taking this fact into account, we know that from a qualitative point of view the dynamic behaviours of trajectories of the continuous-time system with delays for values of σ close enough to 0 will be similar to trajectories of the discrete-time system (see Fig. 2a). In particular, Fig. 2b, c show the stabilising role of τ and k, respectively. We note that it is possible to have dynamics with oscillations characterised by several maximum and minimum values or the presence of a chaotic attractor. To this purpose, Fig. 3a–d show the evolution of the attractor of the system when k varies. For $k = 0.27$, the projection of the attractor in (q, q_d) plane is a closed curve of class C_1 without auto intersections. In this case, the long-term trajectory of q is characterised by a unique maximum value and a unique minimum value. For $k = 0.29$, we note the presence of an auto intersection of the curve and this is a sign of the birth of further relative maximum and minimum values in the long-term trajectory of q. When k increases further ($k = 0.31$) we observe an increase in the number of auto intersections of the curve to which correspond more complicated trajectories, until we get chaotic trajectories when $k = 0.35$.

Case 6. The characteristic equation of the linearisation of Eq. (15) at q_* is

$$\lambda = -6bkq_*^2 e^{-\lambda\tau_1} - 6bkq_*^2 e^{-\lambda\tau_2}. \tag{25}$$

For $\tau_2 = 0$, (25) has the form

$$\lambda = -6bkq_*^2 e^{-\lambda\tau_1} - 6bkq_*^2. \tag{26}$$

If $\lambda = i\omega$, with $\omega > 0$, is a solution of (26), then separating the real and imaginary parts gives

$$\omega = 6bkq_*^2 \sin\omega\tau_1, \quad 6bkq_*^2 = -6bkq_*^2 \cos\omega\tau_1.$$

Hence, we obtain $\omega^2 = 0$. In case $\tau_2 = 0$, the stationary solution q_* of Eq. (15) is locally asymptotically stable for arbitrary $\tau_1 > 0$. By continuity, for sufficiently small $\tau_2 > 0$, q_* remains locally asymptotically stable. Let $\tau_2 > 0$. Then, Eq. (25) has a purely imaginary root $\lambda = i\omega$, $\omega > 0$, if the following equations are satisfied:

$$\omega - 6bkq_*^2 \sin(\omega\tau_1) = 6bkq_*^2 \sin(\omega\tau_2), \quad 6bkq_*^2 \cos(\omega\tau_1) = -6bkq_*^2 \cos(\omega\tau_2).$$

Squaring and adding up both equations, we get

$$\sin(\omega\tau_1) = \frac{\omega}{12bkq_*^2}. \tag{27}$$

For any $\tau_1 > 0$, Eq. (27) has finite number of positive solutions ω_j, $j = 1, 2, \ldots, m$. For every arbitrary chosen $\tau_1 > 0$ and for each ω_j there exist an infinite number of τ_2 such that $6bkq_*^2 \cos(\omega_j\tau_1) = -6bkq_*^2 \cos(\omega_j\tau_2)$. For all $j = 1, 2, \ldots, m$, we define

$$\tau_2^j = \min\left\{\tau_2 > 0 : 6bkq_*^2 \cos(\omega_j\tau_1) = -6bkq_*^2 \cos(\omega_j\tau_2)\right\}.$$

Set

$$\bar{\tau}_1^0 = \frac{1}{12bkq_*^2} \text{ and } \bar{\tau}_2^0 = \frac{\tau_2^0}{6bkq_*^2} \tag{28}$$

where $\tau_2^0 = \min\left\{\tau_2^j : j = 1, 2, \ldots, m\right\}$.

Theorem 10 *Let $\bar{\tau}_1^0$ and $\bar{\tau}_2^0$ be defined by (28).*

(1) If $\tau_1 \in [0, \bar{\tau}_1^0]$ and $\tau_2 > 0$ or if $\tau_1 > \bar{\tau}_1^0$ and $\tau_2 \in [0, \bar{\tau}_2^0)$, then the positive equilibrium q_* of Eq. (15) is locally asymptotically stable.

(2) If $\tau_1 > \bar{\tau}_1^0$ and $6bkq_*^2\omega_0\bar{\tau}_2^0 \in \cup_{l\in\mathbb{N}}[2l\pi, 2l\pi + \pi/2]$, then the Hopf bifurcation occurs at $\tau_2 = \bar{\tau}_2^0$.

Proof The proof can be found in Theorem 2 of Piotrowska (2007). ■

Corollary 11 *If* $\tau_1 \in (\pi/(12bkq_*^2), \pi/(6bkq_*^2)]$*, then there exists* $\bar{\tau}_2^0$ *such that for* $\tau_2 \in [0, \bar{\tau}_2^0)$ *the positive equilibrium* q_* *of Eq. (15) is locally asymptotically stable and for* $\tau_2 = \bar{\tau}_2^0$ *the Hopf bifurcation occurs.*

Proof The proof can be found in Corollary 3 of Piotrowska (2007). ■

We note that due to the symmetry of (26), it is possible to obtain a result about the stability of the stationary equilibrium q_* in a nicely way. In particular, we may use Theorem 4.2 of Kuang (1993, p. 87), to characterise the stability region. Then, we have the following theorem:

Theorem 12 *The positive equilibrium* q_* *is locally asymptotically stable if and only if*

$$12bkq_*^2(\tau_1 + \tau_2)\cos\left(\frac{\pi}{2}\frac{\tau_1 - \tau_2}{\tau_1 + \tau_2}\right) - \pi < 0 \tag{29}$$

Figure 4 illustrates the result stated in Theorem 12.

In the light of Corollary 11 and Theorem 12, in Fig. 5a, b we present some numerical simulations to clarify the dynamic behaviour of this model.

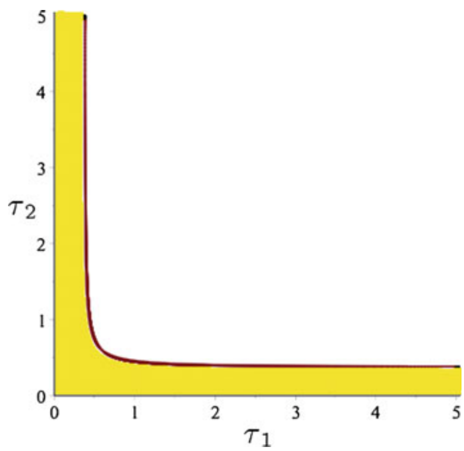


Fig. 4 Parameter set: $a = 4$, $b = 0.6$, $c = 0.5$ and $k = 0.285$. Switching curves in (τ_1, τ_2) plane. In the *yellow region* the stationary equilibrium is stable (color figure online)

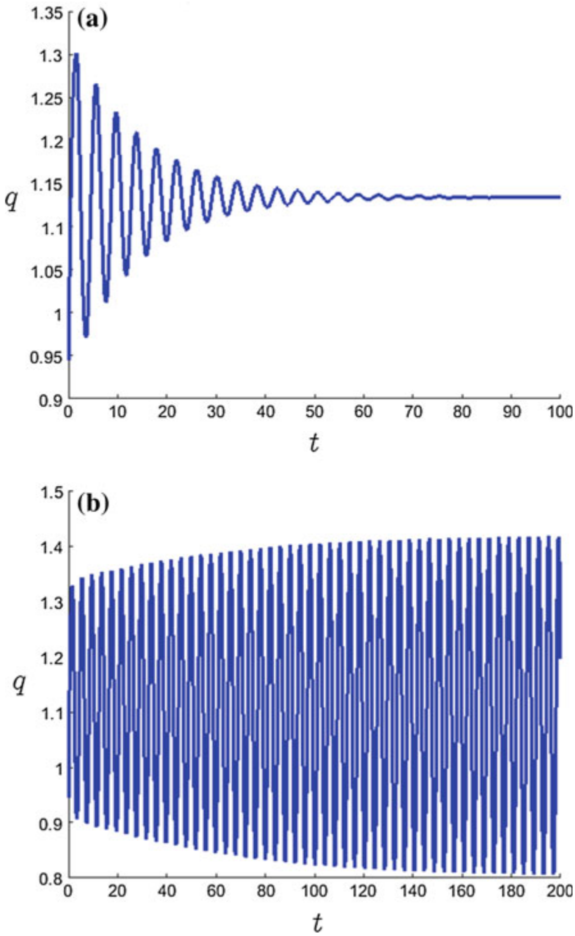


Fig. 5 Parameter set: $a = 4$, $b = 0.6$, $c = 0.5$ and $\tau_1 = 1.6 \in (\pi/(12bkq_*^2), \pi/(6bkq_*^2)]$. **a** Time series of q associated with a trajectory that converges towards the long-term equilibrium for $\tau_2 = 0.3$ ($\bar{\tau}_2^0 \cong 0.38$). **b** Time series of q associated with a trajectory that converges towards the limit cycle for $\tau_2 = 0.42 > \bar{\tau}_2^0$

4 Conclusions

By taking the discrete-time nonlinear monopoly model of Naimzada and Ricchiuti (2008) as a starting point, in this article we have proposed a continuous-time version of it with discrete delays. We have shown that depending on the specific way of transforming a discrete-time model in a continuous-time model with delays, it is possible to obtain several results about local stability of the equilibrium. In particular, there are two polar cases: a model which is always stable (when the monopolist knows the market price and the linear approximation of the market demand in a neighbourhood of the current value of the price) and a model that can actually display chaotic dynamics (this is the case that follows Berezowski 2001, according to which at every time t the monopolist is not able to perfectly realise the production plan arranged at time $t - \tau$ because of frictions due to the long time required for production). Our findings suggest that from a theoretical point of view it is important to have a good description of the information set and the assumptions made as these hypotheses have important consequences on the main dynamic results. In order to infer about the long-term dynamics of the model, it would be of importance also to have available empirical evidence of the economic agents in a monopoly to understand which of the theoretical models presented is the one that better captures realistic economic behaviours.

Acknowledgments The authors are grateful to an anonymous reviewer for valuable comments on an earlier draft. The usual disclaimer applies.

References

Al-Hdaibat B, Govaerts W, Neirynck N (2015) On periodic and chaotic behavior in a two-dimensional monopoly model. Chaos Solitons Fractals 70:27–37
Baumol WJ, Quandt RE (1964) Rules of thumb and optimally imperfect decisions. Am Econ Rev 54:23–46
Berezowski M (2001) Effect of delay time on the generation of chaos in continuous systems. One-dimensional model. Two-dimensional model - tubular chemical reactor with recycle. Chaos Solitons Fractals 12:83–89
Cavalli F, Naimzada AK (2015) Effect of price elasticity of demand in monopolies with gradient adjustment. Chaos Solitons Fractals 76:47–55
Cooke K, van den Driessche P, Zou X (1999) Interaction of maturation delay and nonlinear birth in population and epidemic models. J Math Biol 39:332–352
Györi I (1990) Interaction between oscillations and global asymptotic stability in delay differential equations. Differ Integral Equ 3:181–200
Kuang Y (1993) Delay differential equations with applications in population dynamics. Academic Press, New York
Matsumoto A, Szidarovszky F (2012) Nonlinear delay monopoly with bounded rationality. Chaos Solitons Fractals 45:507–519
Matsumoto A, Szidarovszky F (2014a) Complex dynamics of monopolies with gradient adjustment. Econ Model 42:220–229
Matsumoto A, Szidarovszky F (2014b) Discrete and continuous dynamics in nonlinear monopolies. Appl Math Comput 232:632–642
Matsumoto A, Szidarovszky F (2015a) Dynamic monopoly with multiple continuously distributed time delays. Math Comput Simul 108:99–118
Matsumoto A, Szidarovszky F (2015b) Learning monopolies with delayed feedback on price expectations. Commun Nonlinear Sci Numer Simul 28:151–165
Matsumoto A, Chiarella C, Szidarovszky F (2013) Dynamic monopoly with bounded continuously distributed delay. Chaos Solitons Fractals 47:66–72
Naimzada AK, Ricchiuti G (2008) Complex dynamics in a monopoly with a rule of thumb. Appl Math Comput 203:921–925
Naimzada AK, Ricchiuti G (2011) Monopoly with local knowledge of demand function. Econ Model 28:299–307
Piotrowska MJ (2007) A remark on the ODE with two discrete delays. J Math Anal Appl 329:664–676
Puu T (1995) The chaotic monopolist. Chaos Solitons Fractals 5:35–44
Schotter A (2008) Microeconomics: a modern approach. South-Western College, California

Dynamic Oligopoly Models with Production Adjustment and Investment Costs

Akio Matsumoto, Ugo Merlone and Ferenc Szidarovszky

Abstract A modified version of single-product discrete Cournot oligopolies is introduced, where the additional costs of decreasing or increasing the output level from the previous time period as well as increasing the output level from the already built up capacity limit are included. In this way the costs of laying off or hiring new workers and making investments for increasing capacity are essential parts of the model. The best response functions of the firms are analytically determined, which are non-differentiable even if the price and the production cost functions are linear. The set of all equilibria are characterized by a system of linear inequalities, which is illustrated in the case of duopoly. The asymptotic properties of the equilibria are examined by using computer simulation. Finally, it is shown how some specific models can be derived from the general approach.

Keywords Oligopolies · Repeated games · Complex dynamics · Workforce flexibility · Investment costs

1 Introduction

Since the pioneering work of Cournot (1960) the theory of oligopoly became one of the most frequently discussed subjects in mathematical economy. Oligopoly was considered as an n-person non-cooperative game, the existence and uniqueness of the

A. Matsumoto
Department of Economics, Chuo University, 742-1, Higashi–Nakano,
Tokyo, Hachioji 192-0393, Japan
e-mail: akiom@tamacc.chuo-u.ac.jp

U. Merlone (✉)
Department of Psychology, Center for Cognitive Science, University
of Torino, via Verdi 10, I 10124 Torino, Italy
e-mail: ugo.merlone@unito.it

F. Szidarovszky
Department of Applied Mathematics, University of Pécs, Ifjúság u. 6,
Pécs H-7624, Hungary
e-mail: szidarka@gmail.com

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_6

Nash-equilibrium was the central issue at the beginning. Different variants of the classical Cournot model were introduced and analyzed including oligopolies with product differentiation, multi-product models, labor-managed oligopolies among others. The investigation of the dynamic extensions was the next stage of research. First linear models were considered, the asymptotic behavior of which are relatively simple, since local asymptotic stability implies global stability. The most important results up to the mid 70s were presented in Okuguchi (1976), and their multiproduct generalizations were introduced and discussed in Okuguchi and Szidarovszky (1999). More recently nonlinear models became the major focus of research. A comprehensive summary of the methodology and its applications in oligopoly theory were presented in Bischi et al. (2010).

The conditions assumed in the classical models were already criticized in the literature and modified models were formulated. Production adjustment costs were introduced in Szidarovszky and Yen (1995), Reynolds (1987, 1991), Howroyd and Rickard (1981), Macleod (1985). The main issue was the effect of adjustment costs on the stability of the equilibrium. In Driskill and McCafferty (1989) the authors developed a differential game model including production adjustment costs. A complete equilibria analysis was performed in Zhao and Szidarovszky (2008), where the best response functions were non-differentiable. In Szidarovszky and Matsumoto (2016) an oligopoly model with discontinuous cost functions was introduced when the setup cost was included for cleaning and depositing waste. The best response functions were also discontinuous, and those as functions of the total industry output could be even multiple valued. In Burr et al. (2015) adjustment constraints were introduced resulting in discontinuous best responses. In addition to complete equilibrium analysis the dynamic extension of the model was also examined.

Another way of developing more realistic models was the consideration of cartelizing groups and antitrust thresholds Matsumoto et al. (2008, 2010a, b), contingent workforce and investment costs Merlone and Szidarovszky (2015), as well as unemployment insurance systems Matsumoto et al. (2015b). In these works both static and dynamic models were examined.

In this paper a generalized model is introduced considering the layoff and hiring of new workers, and additional investments. After the model is formulated the best response functions are determined in Sect. 2 and the set of the equilibria is described in Sect. 3. It will be also demonstrated how the specific models can be reduced from the general approach. The asymptotic behavior of the equilibria is examined by using computer simulation in Sect. 4. Section 5 concludes the paper with future research directions.

2 The General Model and Best Responses

Consider an n-firm single-product oligopoly without product differentiation. If x_k is the output of firm k, then $s = \sum_{k=1}^{n} x_k$ is the industry output. The price function is assumed to be linear: $p(s) = A - Bs$, and the production cost of firm k, $C_k(x_k) =$

$c_k + d_k x_k$ is also linear in x_k. In addition to the production costs we consider the following additional cost types.

Laying off or hiring new workers:

$$\bar{C}_k\left(x_k, x_k\left(t-1\right)\right) = \begin{cases} \delta_k\left(x_k\left(t-1\right)-x_k\right) \text{ if } x_k < x_k\left(t-1\right) \\ \gamma_k\left(x_k - x_k\left(t-1\right)\right) \text{ if } x_k \geq x_k\left(t-1\right) \end{cases}$$

which includes the unemployment insurance, severance pays, hiring and training new workers. The cost of increasing production level beyond the already built up capacity limit is

$$\overline{\overline{C}}_k\left(x_k, X_k\left(t-1\right)\right) = \begin{cases} 0 & \text{if } x_k \leq X_k\left(t-1\right) \\ \alpha_k\left(x_k - X_k\left(t-1\right)\right) & \text{if } x_k > X_k\left(t-1\right) \end{cases}$$

where

$$X_k\left(t-1\right) = \max_{0\leq\tau\leq t-1}\left\{x_k\left(\tau\right)\right\} \quad \left(\geq x_k\left(t-1\right)\right)$$

is the already built up capacity limit. So the profit function of firm k has the form:

$$\Pi_k = \begin{cases} x_k\left(A - Bx_k - Bs_k\right) - \left(c_k + d_k x_k\right) - \delta_k\left(x_k\left(t-1\right) - x_k\right) & \text{if } 0 \leq x_k < x_k\left(t-1\right) \\ x_k\left(A - Bx_k - Bs_k\right) - \left(c_k + d_k x_k\right) - \gamma_k\left(x_k - x_k\left(t-1\right)\right) & \text{if } x_k\left(t-1\right) < x_k \leq X_k\left(t-1\right) \\ x_k\left(A - Bx_k - Bs_k\right) - \left(c_k + d_k x_k\right) - \gamma_k\left(x_k - x_k\left(t-1\right)\right) \\ \quad -\alpha_k\left(x_k - X_k\left(t-1\right)\right) & \text{if } X_k\left(t-1\right) < x_k \leq L_k \end{cases} \tag{1}$$

where L_k is the maximum possible capacity limit of the firm which cannot be increased.

Let φ_1, φ_2 and φ_3 denote these functions. Notice first that

$$\begin{aligned} \varphi_1'\left(x_k\right) &= A - 2Bx_k - Bs_k - d_k + \delta_k \\ \varphi_2'\left(x_k\right) &= A - 2Bx_k - Bs_k - d_k - \gamma_k \\ \varphi_3'\left(x_k\right) &= A - 2Bx_k - Bs_k - d_k - \gamma_k - \alpha_k \end{aligned}$$

so $\varphi_1'\left(x_k\right) > \varphi_2'\left(x_k\right) > \varphi_3'\left(x_k\right)$ for each feasible x_k.

Furthermore

$$\begin{aligned} &\varphi_1'\left(0\right) = A - Bs_k - d_k + \delta_k \\ &\varphi_1'\left(x_k\left(t-1\right)\right) = A - 2Bx_k\left(t-1\right) - Bs_k - d_k + \delta_k \\ &\varphi_2'\left(x_k\left(t-1\right)\right) = A - 2Bx_k\left(t-1\right) - Bs_k - d_k - \gamma_k \\ &\varphi_2'\left(X_k\left(t-1\right)\right) = A - 2BX_k\left(t-1\right) - Bs_k - d_k - \gamma_k \\ &\varphi_3'\left(X_k\left(t-1\right)\right) = A - 2BX_k\left(t-1\right) - Bs_k - d_k - \gamma_k - \alpha_k \\ &\varphi_3'\left(L_k\right) = A - 2BL_k - Bs_k - d_k - \gamma_k - \alpha_k \end{aligned}$$

So in determining the best response R_k of player k we have to consider the following cases:

(i) $\varphi_1'(0) \leq 0$ occurs when $s_k \geq \frac{A-d_k+\delta_k}{B}$ and now $R_k = 0$

(ii) $\varphi_1'(0) > 0$ and $\varphi_1'(x_k(t-1)) \leq 0$ occur when

$$\frac{A - 2Bx_k(t-1) - d_k + \delta_k}{B} \leq s_k < \frac{A - d_k + \delta_k}{B}$$

and R_k is the stationary point between 0 and $x_k(t-1)$:

$$R_k = \frac{A - Bs_k - d_k + \delta_k}{2B}$$

(iii) $\varphi_1'(x_k(t-1)) > 0$ and $\varphi_2'(x_k(t-1)) \leq 0$ is the case when

$$\frac{A - 2Bx_k(t-1) - d_k - \gamma_k}{B} \leq s_k < \frac{A - 2Bx_k(t-1) - d_k + \delta_k}{B}$$

with $R_k = x_k(t-1)$

(iv) $\varphi_2'(x_k(t-1)) > 0$ and $\varphi_2'(X_k(t-1)) \leq 0$ occur when

$$\frac{A - 2BX_k(t-1) - d_k - \gamma_k}{B} \leq s_k < \frac{A - 2Bx_k(t-1) - d_k - \gamma_k}{B}$$

and R_k is the stationary point between $x_k(t-1)$ and $X_k(t-1)$:

$$R_k = \frac{A - Bs_k - d_k - \gamma_k}{2B}$$

(v) $\varphi_2'(X_k(t-1)) > 0$ and $\varphi_3'(X_k(t-1)) \leq 0$ is the case when

$$\frac{A - 2BX_k(t-1) - d_k - \gamma_k - \alpha_k}{B} \leq s_k < \frac{A - 2BX_k(t-1) - d_k - \gamma_k}{B}$$

and now $R_k = X_k(t-1)$

(vi) $\varphi_3'(X_k(t-1)) > 0$ and $\varphi_3'(L_k) \leq 0$ occur when

$$\frac{A - 2BL_k - d_k - \gamma_k - \alpha_k}{B} \leq s_k < \frac{A - 2BX_k(t-1) - d_k - \gamma_k - \alpha_k}{B}$$

and R_k is the stationary point between $X_k(t-1)$ and L_k:

$$R_k = \frac{A - Bs_k - d_k - \gamma_k - \alpha_k}{2B}$$

(vii) $\varphi_3'(L_k) > 0$ occurs when

$$s_k < \frac{A - 2BL_k - d_k - \gamma_k - \alpha_k}{B}$$

in which case $R_k = L_k$.

These cases with the possible shapes of the profit function of firm k are illustrated in Fig. 1.

The best response of firm k is therefore given as follows:

$$R_k = \begin{cases} 0 & \text{if } s_k \geq \frac{A-d_k+\delta_k}{B} \\ \frac{A-Bs_k-d_k+\delta_k}{2B} & \text{if } \frac{A-2Bx_k(t-1)-d_k+\delta_k}{B} \leq s_k < \frac{A-d_k+\delta_k}{B} \\ x_k(t-1) & \text{if } \frac{A-2Bx_k(t-1)-d_k-\gamma_k}{B} \leq s_k < \frac{A-2Bx_k(t-1)-d_k+\delta_k}{B} \\ \frac{A-Bs_k-d_k-\gamma_k}{2B} & \text{if } \frac{A-2BX_k(t-1)-d_k-\gamma_k}{B} \leq s_k < \frac{A-2Bx_k(t-1)-d_k-\gamma_k}{B} \\ X_k(t-1) & \text{if } \frac{A-2BX_k(t-1)-d_k-\gamma_k-\alpha_k}{B} \leq s_k < \frac{A-2BX_k(t-1)-d_k-\gamma_k}{B} \\ \frac{A-Bs_k-d_k-\gamma_k-\alpha_k}{2B} & \text{if } \frac{A-2BL_k-d_k-\gamma_k-\alpha_k}{B} \leq s_k < \frac{A-2BX_k(t-1)-d_k-\gamma_k-\alpha_k}{B} \\ L_k & \text{if } s_k < \frac{A-2BL_k-d_k-\gamma_k-\alpha_k}{B}. \end{cases} \quad (2)$$

In the case of interior $x_k(t-1)$ the shape of the best response function is given in Fig. 2. If $x_k(t-1)$ is on the boundary, then one or more segments of R_k are omitted.

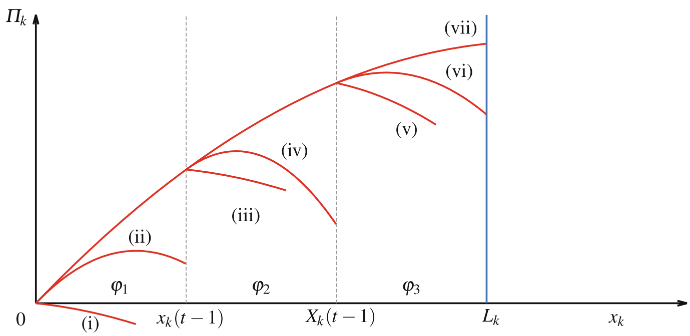


Fig. 1 The possible shapes of the profit function of firm k

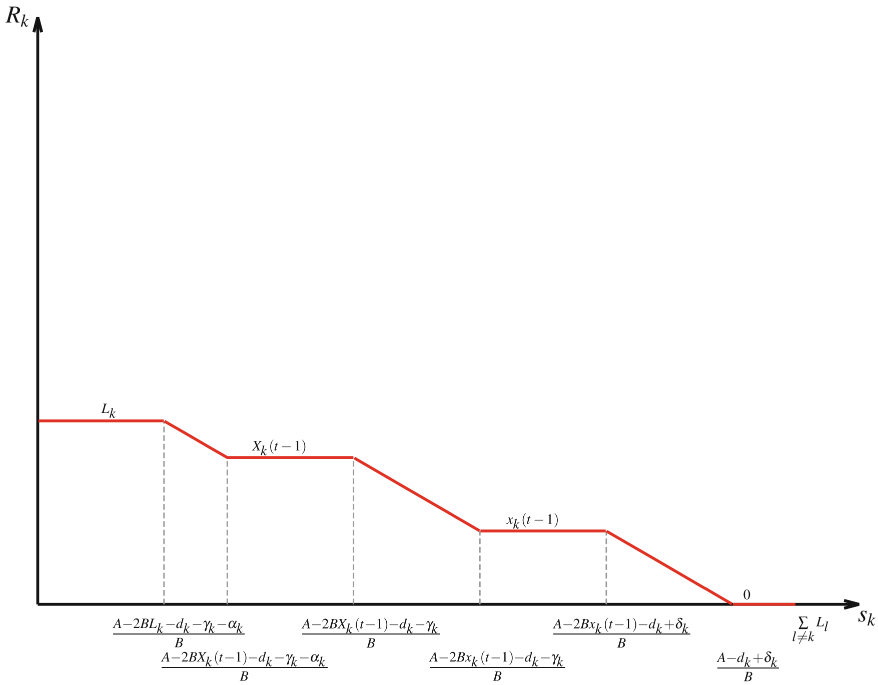


Fig. 2 Best response of firm k when $x_k\ (t-1)$ is interior

3 Equilibrium Analysis

Let $\bar{x}_k$ denote an equilibrium output level of firm k, and $\bar{s}_k = \sum_{l \neq k} \bar{x}_l$ the equilibrium output level of the rest of the industry.

Case (i) implies that

$$\bar{x}_k = 0 \quad \text{if} \quad \bar{s}_k \geq \frac{A - d_k + \delta_k}{B},$$

case (vii) implies that

$$\bar{x}_k = L_k \quad \text{if} \quad \bar{s}_k \leq \frac{A - 2BL_k - d_k - \gamma_k}{B}$$

and case (iii) implies that $\bar{x}_k$ is interior if

$$\frac{A - 2B\bar{x}_k - d_k - \gamma_k}{B} \leq \bar{s}_k \leq \frac{A - 2B\bar{x}_k - d_k + \delta_k}{B}$$

In the second case we used the fact that if $\bar{x}_k = L_k$, then the third segment of Π_k disappears. We can illustrate the equilibrium set in the case of a duopoly, when $\bar{s}_1 = \bar{x}_2$ and $\bar{s}_2 = \bar{x}_1$. So

$$\bar{x}_1 = 0 \quad \text{if} \quad \bar{x}_2 \geq \frac{A - d_1 + \delta_1}{B},$$

$$\bar{x}_2 = 0 \quad \text{if} \quad \bar{x}_1 \geq \frac{A - d_2 + \delta_2}{B},$$

$$\bar{x}_1 = L_1 \quad \text{if} \quad \bar{x}_2 \leq \frac{A - 2BL_1 - d_1 - \gamma_1}{B},$$

$$\bar{x}_2 = L_2 \quad \text{if} \quad \bar{x}_1 \leq \frac{A - 2BL_2 - d_2 - \gamma_2}{B},$$

$\bar{x}_1$ is interior if

$$\frac{A - 2B\bar{x}_1 - d_1 - \gamma_1}{B} \leq \bar{x}_2 \leq \frac{A - 2B\bar{x}_1 - d_1 + \delta_1}{B}$$

and finally, $\bar{x}_2$ is interior if

$$\frac{A - 2B\bar{x}_2 - d_2 - \gamma_2}{B} \leq \bar{x}_1 \leq \frac{A - 2B\bar{x}_2 - d_2 + \delta_2}{B}.$$

These cases are shown in Fig. 3, where the shaded region shows the set of the interior equilibria $(\bar{x}_1, \bar{x}_2)$. In the cases of certain parameter values (such as shown in the figure) we might have two boundary equilibria $(0, L_2)$ and $(L_1, 0)$ in addition.

In the profit function (1) there are three marginal cost factors: δ_k, γ_k and α_k, the earlier discussed models can be obtained by the special choices of the values of these parameters.

By selecting $\alpha_k = 0$, output adjustment cost occurs if x_k differs from the previous output level $x_k\ (t - 1)$. If the output level decreases, then the layoff of workers might result in additional cost, and if it increases, then the hiring of new workers adds additional cost by training them and maybe paying higher wages. This model was investigated in Matsumoto et al. (2015a).

By selecting $\delta_k = \alpha_k = 0$ the model Matsumoto et al. (2015b) with contingent workforce and unemployment insurance system is obtained. If $\delta_k = 0$, then the model of Merlone and Szidarovszky (2015) with contingent workforce and investment cost is derived.

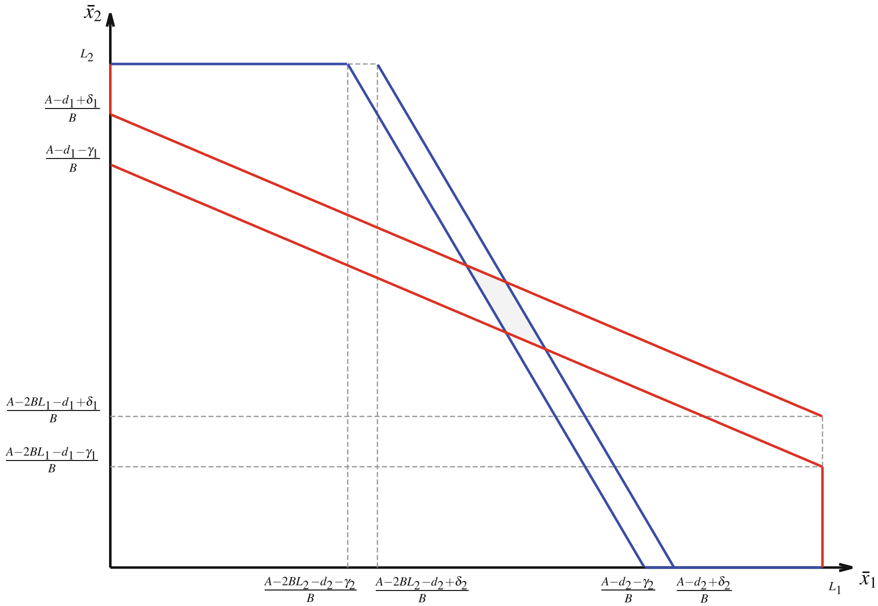


Fig. 3 Equilibrium set in the duopoly case

4 Simulation Study

In the simulation study we assume discrete time scales and that the firms adjust their output quantities partially towards best responses. Let K_k denote the speed of adjustment of firm k, then this adjustment process can be described by the first order difference equations:

$$x_k(t) = x_k(t-1) + K_k\left(R_k\left(X_k(t-1), x_k(t-1)\right) - x_k(t-1)\right) \tag{3}$$

for $k = 1, 2, \ldots, N$, where we assume that $0 < K_k \leq 1$. As it is well known, the selection of $K_k = 0$ would lead to constant trajectories, and the case of $K_k = 1$ corresponds to best response dynamics. For the sake of simplicity we assume the speed of adjustment is the same for all the firms, i.e., $K_1 = K_2 = \cdots = K_N = K$.

Further, we select the semisymmetric case of N firms ($N > 1$) in which $N-1$ firms are identical and the Nth firm is different. We assume that the price function is $p(X) = 20 - 2X$, the common cost function of firms k ($k = 1, 2, \ldots, N-1$) is $C_k(x_k) = x_k$ and that of the Nth firm is $C_N(x_N) = 2x_N$. For both types of firms $L_k = 10$. It is also assumed that the coefficients α_k, γ_k and δ_k are the same for all the firms. If the initial outputs of the first $N-1$ firms are identical then they have the same trajectories for all future times. Therefore, this model is two-dimensional where x_1 can denote the common output of the first $N-1$ firm, and x_N the output of the last firm.

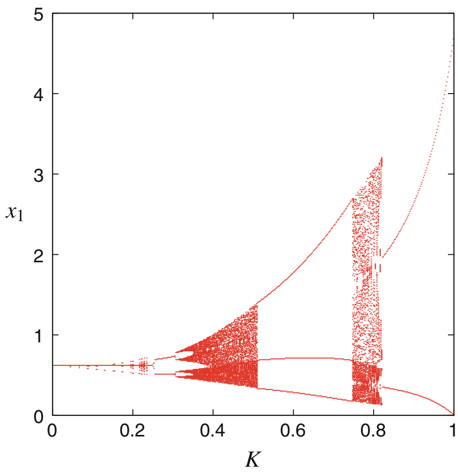
The dynamic properties of this system can be conveniently illustrated by the bifurcation diagrams with zero initial output levels. In our earlier papers Matsumoto et al. (2015b); Merlone and Szidarovszky (2015) we examined the dynamic behavior of the special models with different values of N, so in this paper we select $N = 15$ since this case generates interesting dynamics in the special models.

The bifurcation diagrams with respect to the adjustment cost coefficients (α_k, γ_k and δ_k) are not interesting since the qualitative properties of the dynamics do not change with these parameters. More interesting conclusions can be easily reached form the bifurcation diagrams with respect to the common speed of adjustment K.

Figure 4 shows the bifurcation diagram with respect to K in the case of $\alpha_k = \gamma_k = \delta_k = 0$. For small values of K ($K < 0.22$) the system is asymptotically stable, then period doubling like bifurcation occurs which leads to chaotic behavior which suddenly (at $K \simeq 0.5$) turns into three-cycle and then for a short interval around 0.8 becomes chaotic and finally turns into a two-cycle with increasing amplitude.

Figure 5 shows four bifurcation diagrams: in case (a) only $\alpha = 1$, the other two coefficients are zeros; in case (b) only $\gamma = 1$ with the other two coefficients are zeros; in case (c) only $\delta = 1$ where the other two coefficients are zeros and, finally, in case (d) all coefficient are equal to one. Notice that the dynamic behavior of the system is qualitatively identical in the second and third cases however the first and the fourth cases are quite different. The first case is very similar to the one without adjustment costs, in the second and third cases the lengths of the chaotic intervals become much smaller and between the chaotic intervals the system is asymptotically stable. In the fourth case the first chaotic interval disappears and only the very short second chaotic interval shows up. In all cases after the second chaotic interval the system presents a two-cycle and clearly this is the case for best response dynamics when $K = 1$ as well.

Fig. 4 The case of $N = 15$ firms: bifurcation diagram with $\alpha = \gamma = \delta = 0$



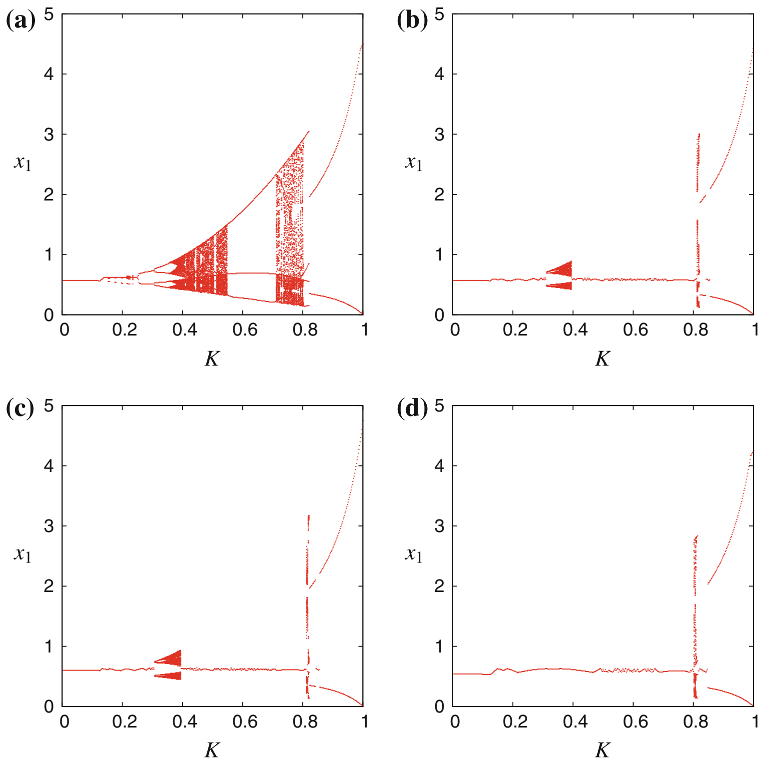


Fig. 5 The case of $N = 15$ firms: bifurcation diagrams with different costs configurations. **a** $\alpha = 1,\ \gamma = 0$ and $\delta = 0$, **b** $\alpha = 0,\ \gamma = 1$ and $\delta = 0$, **c** $\alpha = 0,\ \gamma = 0$ and $\delta = 1$, **d** $\alpha = 1,\ \gamma = 1$ and $\delta = 1$

5 Conclusions

A general linear oligopoly model with additional production adjustment and investment costs was introduced, which contains some of the earlier studied models as special cases. The profit functions of the firms are continuous and non-differentiable. The best responses of the firms were analytically determined as non-increasing, continuous and non-differentiable functions of the rest of the industry. The set of all equilibria was characterized by a system of linear inequalities which was illustrated in the case of a duopoly. The dynamic extension of the model was investigated by using computer simulation. A semisymmetric case was investigated with 15 firms, when the first 14 firms were assumed to be identical so the system became two-dimensional. The bifurcation diagrams with respect to speed of adjustment K were determined. For small values of K the system is asymptotically stable in all cases, for large values of K the system shows two-cycle dynamics, and in between we might have intervals in which chaotic like behavior, three-cycle and stability might occur.

It is interesting to extend this study to the cases of nonlinear price and cost functions as well as to nonlinear production adjustment and investment costs. This will be the subject of our next research project.

References

Bischi GI, Chiarella C, Kopel M, Szidarovszky F (2010) Nonlinear oligopolies: stability and bifurcations. Springer, Berlin

Burr C, Gardini L, Szidarovszky F (2015) Discrete time dynamic oligopolies with adjustment constraints. J Dyn Game 2(1):65–87

Cournot A (1960) Recherches sur les Principes Mathematiques de la Theorie des Richesses. Hachette, Paris, (1838). Kelly, New York

Driskill R, McCafferty S (1989) Dynamic oligopoly with adjustment costs: a differential game approach. J Econ Theor 49:324–338

Howroyd TD, Rickard J (1981) Cournot oligopoly and adjustment costs. Econ Lett 7:113–117

Macleod WB (1985) On adjustment costs and the stability of equilibria. Rev Econ Stud 52:575–591

Matsumoto A, Merlone U, Szidarovszky F (2008) Cartelizing groups in dynamic linear oligopoly with antitrust threshold. Int Game Theor Rev 10(4):399–419

Matsumoto A, Merlone U, Szidarovszky F (2010a) Cartelising groups in dynamic hyperbolic oligopoly with antitrust threshold. Aust Econ Pap 49(4):289–300

Matsumoto A, Merlone U, Szidarovszky F (2010b) Dynamic oligopoly with partial cooperation and antitrust threshold. J Econ Behav Organ 73:259–272

Matsumoto A, Merlone U, Szidarovszky F (2015a) Extended oligopoly with contingent workforce. Submitted to a Special Issue of the Journal of Evolutionary Economics

Matsumoto A, Merlone U, Szidarovszky F (2015b) Oligopolies with contingent workforce and unemployment insurance systems. Commun Nonlinear Sci Numer Simul 27:52–65

Merlone U, Szidarovszky F (2015) Dynamic oligopolies with contingent workforce and investment costs. Math Comput Simul 108:144–154

Okuguchi K (1976) Expectations and stability in oligopoly models. Springer, Berlin

Okuguchi K, Szidarovszky F (1999) The theory of oligopoly with multi-product firms. Springer, Berlin

Reynolds SS (1987) Capacity investment, preemption and commitment in an infinite horizon model. Int Econ Rev 28:69–88

Reynolds SS (1991) Dynamic oligopoly with capacity adjustment costs. J Econ Dyn Control 15:491–514

Szidarovszky F, Matsumoto A (2016) On a discontinuous Cournot oligopoly. In: von Mouche P, Quartieri F (eds) Equilibrium theory for Cournot oligopolies and related games: essays in honour of Koji Okuguchi. Springer, Berlin, pp 97–112

Szidarovszky F, Yen J (1995) Dynamic Cournot oligopoly with production adjustment costs. J Math Econ 24:95–101

Zhao J, Szidarovszky F (2008) N-firm oligopolies with production adjustment costs: best responses and equilibrium. J Econ Behav Organ 68:87–99

A Stylized Model for Long-Run Index Return Dynamics

Natascia Angelini, Giacomo Bormetti, Stefano Marmi and Franco Nardini

Abstract We introduce a discrete-time model of stock index return dynamics grounded on the ability of Shiller's Cyclically Adjusted Price-to-Earning ratio to predict long-horizon market performances. Specifically, we discuss a model in which returns are driven by a fundamental term and an autoregressive component perturbed by external random disturances. The autoregressive component arises from the agents' belief that expected returns are higher in bullish markets than in bearish markets. The fundamental term, driven by the value towards which fundamentalists expect the current price should revert, varies in time and depends on the initial averaged price-to-earnings ratio. The actual stock price may deviate from the perceived reference level as a combined effect of an idyosyncratic noise component and local trends due to trading strategies. We demonstrate both analytically and by means of numerical experiments that the long-run behavior of our stylized dynamics agrees with empirical evidences reported in literature.

Keywords Fundamental value · Shiller's cape

N. Angelini
School of Economics, Management and Statistics, University of Bologna,
Via Angherà 22, 40127 Rimini, Italy
e-mail: natascia.angelini@unibo.it

G. Bormetti · F. Nardini (✉)
Department of Mathematics, University of Bologna, Viale Filopanti 5,
40126 Bologna, Italy
e-mail: franco.nardini@unibo.it

G. Bormetti
e-mail: giacomo.bormetti@unibo.it

S. Marmi
Scuola Normale Superiore and C.N.R.S. UMI 3483, Laboratorio Fibonacci,
Piazza Dei Cavalieri 7, 56126 Pisa, Italy
e-mail: stefano.marmi@sns.it

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_7

1 Introduction

"I never have the faintest idea what the stock market is going to do in the next six months, or the next year, or the next two. But I think it is very easy to see what is likely to happen over the long term", Buffet (2001).

The first part of Buffett's statement clearly explains why nowadays it is widely accepted that stock prices and stock market indexes behave like stochastic processes. Such a long lived popularity is further supported by two different arguments. The first one is the argument put forth by Fama (1965) that financial markets are "informationally efficient". One can not achieve returns in excess of average market returns on a risk-adjusted basis, given the information available at the time the investment is made since the instantaneous adjustment property of an efficient market implies that successive price changes in individual securities may be assumed independent for any practical purpose, see Fama (1965) and Samuelson (1965). The second argument is the possibility—within the formal framework of stochastic processes—to develop pricing models.

Despite the second part of the statement may be tracked back to Buffett's mentor Benjamin Graham, it has been obscured by the efficient market hypothesis for decades. Campbell and Shiller are two of the few scholars long skeptical about the latter; as early as 1988 they found statistical evidence that "the present value of future dividends is, for each year, roughly a weighted average of moving-average earnings and current real price" which has implication for the present-value model of stock prices and for recent results that long-horizon stock returns are highly predictable. At the very beginning of 2000 Robert Shiller wrote "we do not know whether the market level makes any sense, or whether they are indeed the result of some human tendency that might be called irrational exuberance", Shiller (2000). He reached his conclusion through an innovative test of the appropriateness of prices in the stock market: the Cyclically Adjusted Price-to-Earning ratio (CAPE), which he proved to be a powerful predictor of future long run performances of the market. The performance of the test is quite satisfactory in the case of the US market from the end of 19th century up to today. For a detailed discussion refer to Campbell and Yogo (2006) and references therein.

It is clear that modeling a Shiller-type price dynamics requires a time scale completely different from those considered when pricing options. The latter scales range from several days to several months (see Cont and Tankov (2004) p. 3), time scales for which "the full effects of new information on intrinsic values to be reflected "instantaneously" in actual prices" (see Fama 1965 p. 56). On the other hand, introducing in the model some mean reverting mechanism would not be enough to generate stock prices which "have a life of their own; they are not simply responding to earnings or dividends. Nor does it appear that they are determined only by information about future" earnings or dividends, see Shiller (2000) p. 183 and Zhong et al. (2003).

To cope with this evidence a model of stock price dynamics should be able

i. to generate a significant transitory component around the rationally expected equilibrium value of the asset. This component requires the action of at least two different contrasting forces: One pushing the price towards its equilibrium and the other pointing at the opposite direction;

ii. to determine whether the trajectory is wandering far from the fundamentals. To do so the model should explicitly take into account macroeconomic variables such as the CAPE.

Surprisingly enough to our knowledge very few such models have been so far put forth. Boswijk et al. (2007) propose a model in which agents have different beliefs about the persistence of deviations of stock prices from the publicly known fundamental value. Quite recently, following Koijen et al. (2009), He et al. (2014) propose an asset pricing model which incorporates a mean reversion process and a moving average momentum component into the drift of a standard geometric Brownian motion. They prove that the profitability of different investment strategies depends on different time horizons and on the market state. In all these models the fundamental value is constant at its (very) long-run historical mean.

Obviously how the fundamental price is determined is a very delicate issue: The initial assumption of a known constant fundamental price may be regarded as a preliminary simplifying hypothesis. A more realistic assumption is that the fundamental value follows itself a random walk (see Lux and Marchesi 1999; Chiarella et al. 2008) and agents know it only approximately due to their bounded rationality. In Westerhoff (2004) agents make estimates by starting from an initial value that is adjusted as time goes on. Thanks to this assumption the model can exhibit prolonged phases of under and over valuation.

Here we choose to follow the approach suggested in a similar context by Biagini et al. (2013), who describe the effects at an aggregate level of the interaction at a micro-level of different types of agents. In particular they assume that "the perceived fundamental value" shifts in time because of the varying share of optimists in the market.[1] Differently from all the above cited papers, we do not try to a priori guess how the mood of the market dictates "the perceived fundamental value". Instead, we allow the fundamental value, towards which fundamentalists expect that the current price should revert, to vary in time and to depend on the initial averaged price-to-earnings ratio as on an initial anchor (see Tversky and Kahneman 1974).

In our model the price growth depends on three components

1. an autoregressive component, naturally justified in terms of agents' expectation that expected returns are higher in bullish markets than in bearish ones;
2. a fundamental component, proportional to the level of the logarithmic averaged Earnings-to-Price ratio (for brevity log EP ratio) and the perceived fundamental value;
3. a stochastic component ensuring the diffusive behavior of stock prices.

[1] A similar assumption of possible shifts of the perceived fundamental value is proposed in Lengnick and Wohltmann (2010) where financial and real markets are taken into account. In De Grauwe and Kaltwasser (2012) traders switch between optimistic and pessimistic views about the fundamental value.

We show that with a suitable choice of the parameters the assumptions of Lengnick and Wohltmann (2010) are in some sense corroborated by our model. Initially the fundamentalists' perception of the fundamental value is biased in the direction of the most recent performance of the market, i.e., if prices are high (low) the fundamental stock price is perceived to lie above (below) its true counterpart. However optimism (pessimism) does not last for ever, as in Biagini et al. (2013) (see p. 10), and within approximately 11 or 12 years it reverts to a value independent of the initial one and compatible with the long-run mean observed by Shiller.

Moreover, we are able to prove that, if we consider a sufficiently large number of periods, the expected rate of return and the expected gross return are linear in the initial time value of log EP, and their variance converges to zero with rate of convergence consistent with a diffusive behavior. This means that, in our model, the stock prices dynamics may exhibit significant and persistent upwards and downwards deviations form the long run mean value of the averaged earning-to-price ratio, nevertheless the averaged earning-to-price ratio is a good predictor of future long-run returns, as claimed by Campbell and Shiller (1988a), Shiller (2000). The result holds for both returns and gross returns; in the latter case we assume that the log dividend-to-price ratio follows a stationary stochastic process as in Campbell and Shiller (1988a, b). Our results are also in keeping with Hodrick (1992), who "demonstrates that a relatively large amount of long-run predictability is consistent with only a small amount of short-run predictability".

2 The Model

We refer to the inflation adjusted price of the stock index measured at the beginning of time period t with P_t, while D_t denotes the real dividend paid between t and $t+1$. Accordingly, we write the real log gross return on the index held from time t until time $t+1$ as

$$H_t = \log (P_{t+1} + D_t) - \log P_t .$$

The description we provide of the return dynamics is on a monthly basis. Thus, the notation $t+1$ refers to time t increased by one month and the real gross yield over a period of length h months corresponds to

$$y_{t,h} = \frac{1}{h} \sum_{i=0}^{h-1} H_{t+i} . \tag{1}$$

We also introduce the index log price $p_t = \log P_t$, in terms of which the gross yield can be rewritten as

$$y_{t,h} = \frac{1}{h} \sum_{i=0}^{h-1} (p_{t+i+1} - p_{t+i}) + \frac{1}{h} \sum_{i=0}^{h-1} \log \left(1 + \frac{D_{t+i}}{P_{t+1+i}} \right), \tag{2}$$

where the telescopic sum is equivalent to $(p_{t+h} - p_t)/h$. The latter term on the right hand side represents a non linear function of the logarithmic dividend-to-price ratio. Campbell and Shiller argue that the log dividend-to-price ratio $d_t - p_{t+1} \doteq \log D_t - \log P_{t+1}$ follows a stationary stochastic process (see page 666 of Campbell and Shiller 1988b). In light of this evidence the dynamics of the log dividend-to-price is given by

$$\Delta(d_{t-1} - p_t) = -\theta(d_{t-1} - p_t - \log \mathcal{G}) + \sigma_{\mathrm{d}} W_t^{\mathrm{d}}, \tag{3}$$

with initial time condition equal to $d_{-1} = \log D_{-1}$. The AR(1) coefficient is given by $1 - \theta$, σ_{d} is a positive volatility constant, $\{W_t^{\mathrm{d}}\}$ are independent identically distributed (i.i.d.) Gaussian increments with zero mean and unit variance, and $\log \mathcal{G}$ is the fixed mean. By means of a first-order Taylor expansion centred in $\log \mathcal{G}$, the quantity $\log(1 + D_t/P_{t+1})$ appearing in Eq. 2 can be replaced by a linear function of the log dividend-to-price ratio.

The dependent variable dealt with throughout the paper is the gross return of the stock index, while as a predictive quantity we consider the log price-to-earnings ratio $cape_t \doteq p_t - \log \langle e \rangle_t^{10}$. The symbol $\langle e \rangle_t^{10}$ refers to the moving average of real earnings over a time window of ten years. The use of an average of earnings in computing the price ratios has been strongly pushed by the literature in recognition of the cyclical variability of earnings.

In Campbell and Shiller (1988a, b) the regression of real and excess stock returns on explanatory variables which are known at the start of the year t shows that the log dividend-to-price ratio and the log earnings-to-price ratio have good predictive capabilities. The ratio variables are used as indicators of fundamental value relative to price. The basic idea is that if stocks are under-priced relative to fundamental value, returns tend to be high subsequently, while the converse holds if stocks are overpriced. Consistently, we describe the dynamics of the log price assuming the existence of an exogenous fundamental component given by a mean-reverting term whose long-run target level depends linearly on the current value of the earnings-to-price ratio.

We model the dynamics of the log price by means of the linear system of stochastic difference equations

$$\begin{cases} p_{t+1} = p_t + \mu_t + \xi_t, \\ \mu_{t+1} = \gamma \mu_t + \kappa\,(\mathcal{H} + \mathcal{F}t - cape_t) + \sigma_\mu W_t^\mu, \\ \xi_{t+1} = \xi_t + \sigma_\xi W_t^\xi, \end{cases} \tag{4}$$

with initial time conditions equal to $p_0 = \log P_0$, and μ_0. The quantities $\{W_t^\mu\}$, and $\{W_t^\xi\}$ for $t = 0, \ldots, h$ are i.i.d. Gaussian increments with zero mean and unit variance, and σ_μ, and σ_ξ are positive volatility constants. The system of equations (4) determines the evolution of log prices as a superposition of a local drift μ_t and a noise component ξ_t. The latter is a zero mean process originating from ξ_0 which ensures the diffusive behavior of stock prices. The most relevant component corresponds to the equation driving the local drift

$$\mu_{t+1} = \gamma \mu_t + \kappa \left(\mathcal{H} + \mathcal{F}t - cape_t \right) + \sigma_\mu W_t^\mu . \tag{5}$$

The dependence of the future level of μ_{t+1} on the value μ_t prevailing at the previous time step is expressed in terms of an autoregressive component whose intensity is determined by the agents' sensitivity to the market trend, γ. This effect can be justified in terms of the expectation that returns are higher in bullish markets than in bearish markets. Competing with the latter effect we add a second mechanism which affects the drift from a fundamental perspective. The second term in the right hand side of Eq. (5) represents the exogenous "fundamental" component given in terms of a mean reverting term. The actual stock price may deviate from the long-run behavior as a combined effect of both random external disturbances and short-term speculative component. Eventually this disequilibrium becomes apparent causing stock prices to move in the direction that reduces the deviation. In modeling the fundamental effect we bear in mind that "in reality it is very difficult (if not impossible) to identify the true fundamental value of any stock" (see Lengnick and Wohltmann 2010). Consistently we allow the mean reversion target to vary in time. Finally, we assume that the evolution of the averaged earnings is exogenous and follows an exponential law, i.e. $\langle e \rangle_t^{10} = \langle e \rangle_0^{10} \exp(gt)$.

The main theoretical result of this paper characterises the asymptotic behavior of the first and second moment of the log-price gross returns.

Proposition 1 *The expected gross yield over h months is asymptotically linear in $\mathcal{F}$ and $\mathcal{G}$*

$$\mathbb{E}_0 \left[y_{0,h} \right] = g + \mathcal{F} + \mathcal{G} + O(\frac{1}{h}), \tag{6}$$

while the variance converges to zero as predicted by a diffusive model

$$Var_0 \left[y_{0,h} \right] = \frac{\sigma_p^2}{h} + \frac{1}{h} \frac{\mathcal{G}^2 \sigma_{\mathrm{d}}^2}{\theta(2-\theta)} + o(\frac{1}{h}), \tag{7}$$

with $\sigma_p = \sigma_\xi (1-\gamma)/\kappa$.

Equation (6) provides an insightful decomposition of the return growth in three components: the growth of earnings, g, a fundamental term, $\mathcal{F}$, ascribable to the price-over-earnings ratio, and the long-run level of the dividend-to-price ratio, $\mathcal{G}$. In this respect Proposition 1 sheds light on the economic constituents of the expected gross yield and matches John Bogle's suggestion for forecasting the long-term performance of stock markets. At the beginning of the 1990s in an article entitled "Investing in the 1990s" he propose to forecast long-run behavior on the basis of three variables: The initial dividend yield, the expected growth of earnings, and the expected change in the price-to-earnings ratio, Bogle (1991). More recently, Estrada (2007) extends Bogle's proposal including a fourth variable, the expected growth of dividends, providing a simple framework for the decomposition of returns similar in spirit to our findings. Proposition 1 also clarifies the long-run behavior of the gross yield's variance. Equation 7 states that it asymptotically reduces to zero with a rate of convergence which is coherent with the diffusive behavior of stock returns.

3 Numerical Computations

In light of the result Eq. (6) and supported by the evidence provided in Campbell and Shiller (1988a, b) that the long-run expected gross yield is a linear function of the initial CAPE, we assume the following

$$\mathcal{F} = \alpha_{\mathcal{F}} - \beta_{\mathcal{F}}\, cape_0,$$
$$\mathcal{G} = \alpha_{\mathcal{G}} - \beta_{\mathcal{G}}\, cape_0.$$

Coherently, we also assume that $\mathcal{H}$ depends linearly on the initial ratio

$$\mathcal{H} = \alpha_{\mathcal{H}} - \beta_{\mathcal{H}}\, cape_0.$$

It is interesting to comment how the perceived fundamental value $\mathcal{H} + \mathcal{F}t$ evolves in time. If $\beta_{\mathcal{F}} > 0$ and $\beta_{\mathcal{H}} < 0$, its initial value is smaller (larger) the lower (higher) $cape_0$, but it gradually reverts toward larger and larger (smaller and smaller) values as time elapses and within $-\beta_{\mathcal{H}}/\beta_{\mathcal{F}}$ months reaches a value independent of the initial level of the value ratio. This behavior is confirmed by Fig. 1 where we plot the evolution of the price-to-earnings ratio over 12 years computed by means of fifty Monte Carlo simulations with different initial values. Figure 1a corresponds to an initial price-to-earnings equal to 11, Fig. 1b–16.6 and 1c–22. The red line is the target of the mean reversion. All paths are sampled with $\alpha_{\mathcal{F}} = 0.033$, $\alpha_{\mathcal{H}} = 0.84$, $\beta_{\mathcal{F}} = 0.006$, $\beta_{\mathcal{H}} = -0.84$, $g = 0.0012$, and $\kappa = 0.037$. In all figures there are paths which exhibit long transients wandering away from the long run value of the price-to-earnings, but finally most of the paths end in the same interval around the long run value of 16.6, irrespective of the initial ratio. These values are close to those considered by Campbell and Shiller (1988a, b) to prove the forecasting ability of long-term stock returns. Coherently with their findings our model captures the mechanism for which an initially under-priced market is driven to the higher long-run level by means of the fundamental anchor. Conversely, keeping fixed all the parameter values, an initially over-priced market is deflated to the long-run price-to-earnings ratio of 16.6 within a transient period of nearly 12 years ($-\beta_{\mathcal{H}}/\beta_{\mathcal{F}} \simeq 141$ months).

Figure 2 is obtained using a data sample consisting of prices, earnings, and dividends for the Standard and Poor Composite Stock Price Index (S&P) on a monthly basis. The data are discussed in Campbell and Shiller (1987, 1988a, b), and are freely available from Robert J. Shiller's webpage http://www.econ.yale.edu/. These time series cover the entire period from January 1871 until December 2012. Figure 2a shows the empirical yields for a time horizon of two years. The dashed line corresponds to a linear regression on the logarithmic CAPE. Figure 2b–h report the

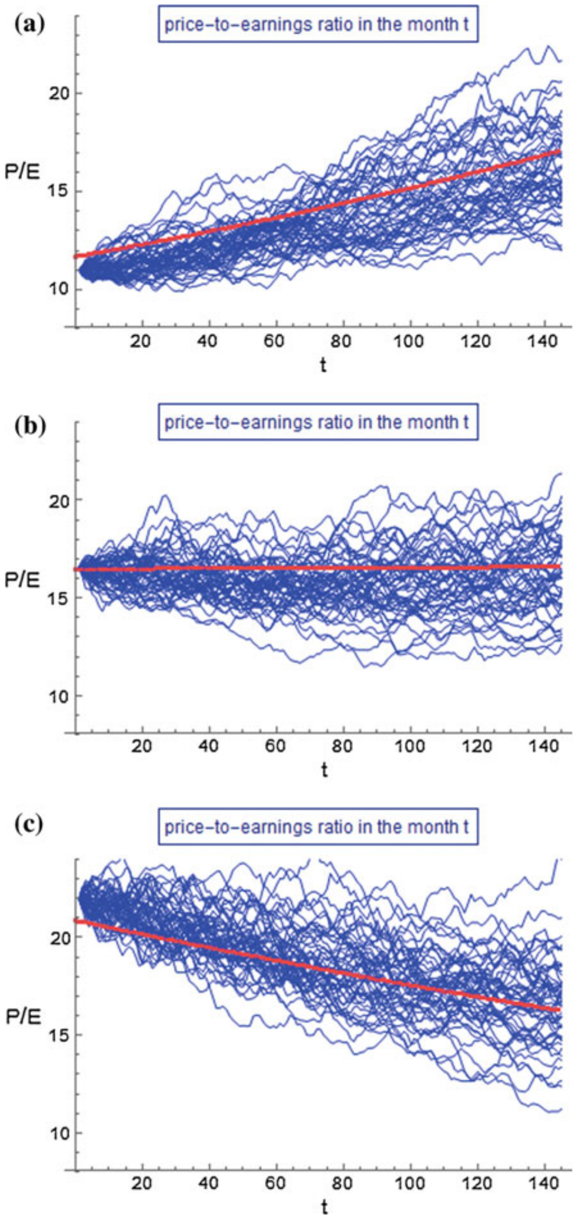


Fig. 1 Fifty simulated paths of price-to-earnings ratio for initial CAPE equal to 11 (Fig. 1a), 16.6 (Fig. 1b), and 22 (Fig. 1c). The *solid lines* correspond to the target of mean reversion

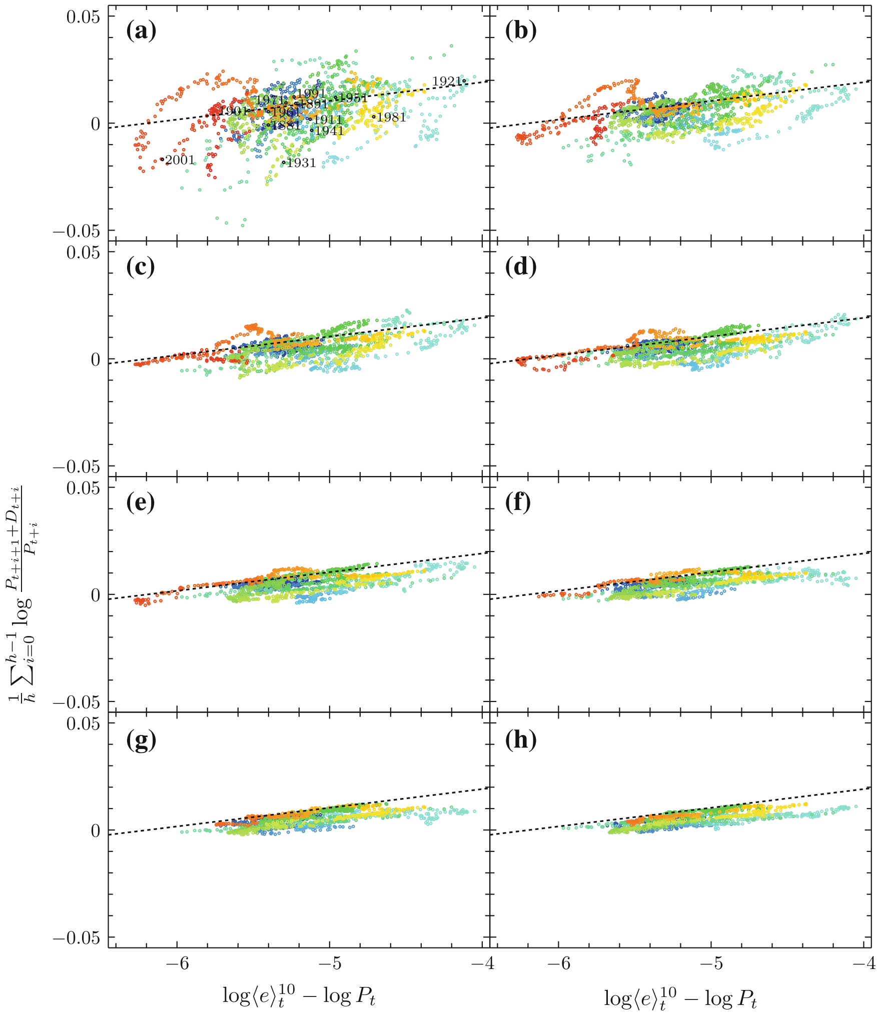


Fig. 2 Empirical yields for the Standard & Poor's 500 from January 1871 until December 2012 for a 2, 4, 6, 8, 10, 12, 14, and 16 year time horizon (figures a, b, c, d, e, f, g, and h, respectively) (color figure online)

same as Fig. 2a with time horizons increasing from 4 to 16 years. Points are given in chronological order according to the color scale ranging from dark blue to red passing through light blue, green, yellow, and orange; labels in the top left figure refer to points which correspond to the first month of the specified year. In Fig. 3 we present a Monte Carlo simulation of the model given by Eqs. (3) and (4). The dashed line corresponds to the long-run behavior predicted by the Eq. (6) and the dotted lines to the boundaries of the 95 % confidence level region. The parameter values chosen for the simulation are given in Table 1. Values reported in the second column

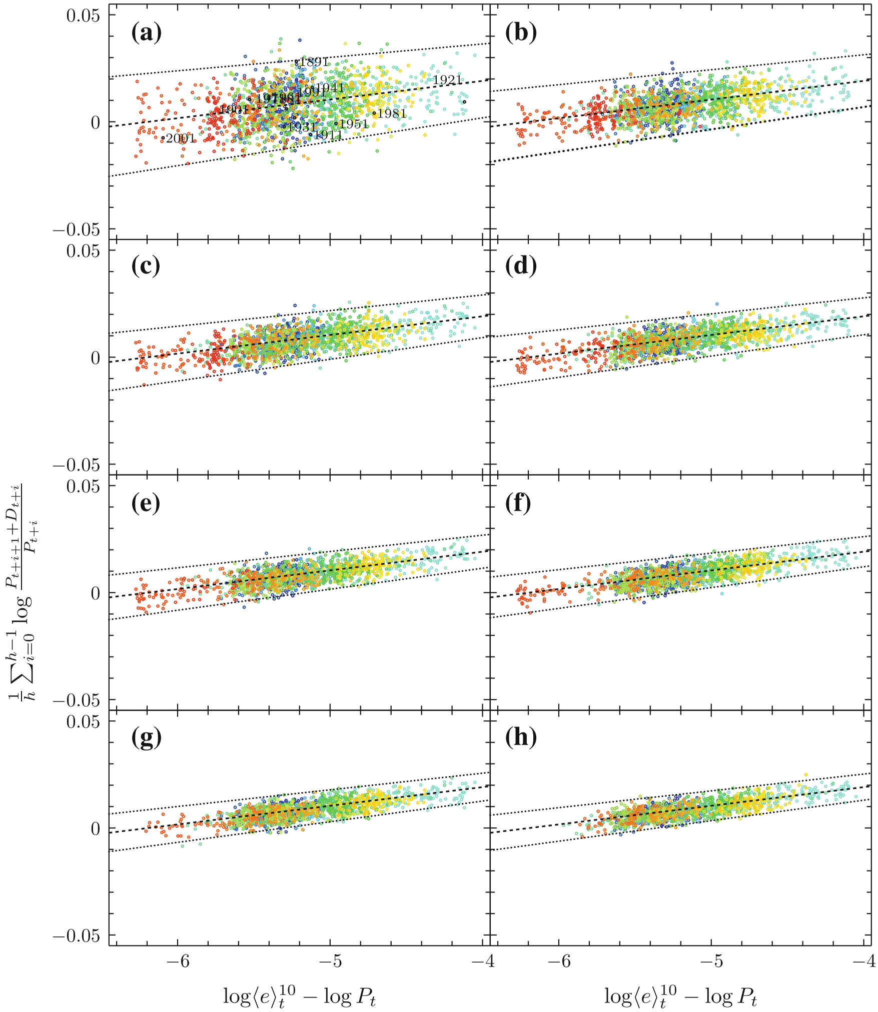


Fig. 3 Monte Carlo scenarios generated with parameter values given in Table 1 and initial time conditions as in Fig. 2

are obtained from the linear regressions displayed in Fig. 2. The monthly earning growth, g, is consistent with the historical long-run growth, while κ provides the typical scale of mean reversion of the fundamental component consistent with the results discussed in Campbell and Shiller (1988a, b). The autoregressive coefficient γ reflects the positive empirical autocorrelation measured from equity index monthly returns. In line with the strong evidence that the log dividend-to-price ratio follows a near unit root process, we set θ equal to 0.025. Finally, all values of the variance coefficients are set equal to 18 bps yielding 15 % yearly volatility for the market index. The color scale determines the initial time condition prevailing historically at

Table 1 Common set of parameter values used in the numerical analysis

$\alpha_{\mathcal{F}}$ ($month^{-1}$)	0.033	g ($month^{-1}$)	0.0012
$\beta_{\mathcal{F}}$ ($month^{-1}$)	0.006	θ ($month^{-1}$)	0.025
$\alpha_{\mathcal{G}}$ ($month^{-1}$)	0.021	γ	0.25
$\beta_{\mathcal{G}}$ ($month^{-1}$)	0.003	κ ($month^{-1}$)	0.037
$\alpha_{\mathcal{H}}$	0.84	σ_{d}^2 ($month$)	18 (bps)
$\beta_{\mathcal{H}}$	-0.84	σ_{μ}^2 ($month$)	18 (bps)
		σ_p^2 ($month$)	18 (bps)

the beginning of each month. Since Monte Carlo scenarios are generated under the same initial time conditions, the remarkable agreement of the color distributions in Figs. 3 and 2 confirms the ability of the model to capture the long-run behavior of the market index.

4 Conclusions and Perspectives

This paper proposes a simple dynamic model for the long-run behavior of stock index returns for the U.S. market. The log price dynamics depend on two market forces: A positive autoregressive component typical for stock index returns and a mean-reverting term whose long-run level is fixed exogenously on the basis of the predictive ability of Shiller's CAPE. Accordingly, we show that the long-run expected growth of the market index can be decomposed in three components: The earning growth, the log dividend-to-price ratio long-run level, and a fundamental term ascribable to the price-over-earnings ratio.

Substantial evidence of the importance of fundamentals in the valuation of international stock markets has been accumulated by the proponents of fundamental indexation e.g. Arnott et al. (2005). Practitioners and academicians alike have been using several valuation measures for estimating the intrinsic value of a stock index. For example, in Table 2 of Poterba and Samwick (1995) the ratio of market value of corporate stock to GDP, the year-end price-to-earnings ratio, the year-end price-to-dividend ratio and Tobin's q are reported from 1947 to 1995 in an effort of alerting the reader on the possible overvaluation of the index. In particular Tobin's q has been proposed as another efficient method of measuring the value of the stock market, with an efficiency comparable to the CAPE (see Smithers 2009). The q ratio is the ratio of price to net worth at replacement cost rather than the historic or book cost of companies. It therefore allows for the impact of inflation, much alike the CAPE which averages real earnings over a ten year span. It would be interesting to carry out an empirical analysis of the relationship between Tobin's q and future stock index returns as far as to extend the present approach to countries other than the U.S. Both perspectives are worth to be followed but require high quality long-term time series. As a possible future extension to model the emergence of explosive bubbles, we plan to relax the assumption of stationarity of the log dividend price ratio process following the approach recently investigated by Engsted et al. (2012).

References

Arnott RD, Hsu J, Moore P (2005) Fundamental indexation. Financ Anal J 61(2):83–99
Biagini F, Föllmer H, Nedelcu S (2013) Shifting martingale measures and the birth of a bubble as a submartingale. To appear in Finance and Stochastics
Bogle J (1991) Investing in the, (1990s). J Portf Manag Spring 5–14
Boswijk HP, Hommes CH, Manzan S (2007) Behavioral heterogeneity in stock prices. J Econ Dyn Control 31(6):1938–1970
Buffet W (2001) Fortune magazine
Campbell JY, Shiller RJ (1987) Cointegration and tests of present value models. J Political Econ 95(5):1062–1088
Campbell JY, Shiller RJ (1988a) The dividend-price ratio and expectations of future dividends and discount factors. Rev Financ Stud 1(3):195–228
Campbell JY, Shiller RJ (1988b) Stock prices, earnings and expected dividends. J Financ XLIII(3):661–676
Campbell JY, Yogo M (2006) Efficient tests of stock return predictability. J Financ Econ 81(1):27–60
Chiarella C, He X-Z, Wang D, Zheng M (2008) The stochastic bifurcation behaviour of speculative financial markets. Phys A: Stat Mech Appl 387(15):3837–3846
Cont R, Tankov P (2004) Financial modelling with jump processes. Chapman & Hall/CRC, London
De Grauwe P, Kaltwasser PR (2012) Animal spirits in the foreign exchange market. J Econ Dyn Control 36(8):1176–1192
Engsted T, Pedersen TQ, Tanggaard C (2012) The log-linear return approximation, bubbles, and predictability. J Financ Quant Anal 47:643–665
Estrada J (2007) Investing in the twenty-first century: with Occam's razor and Bogle's wit. Corp Financ Rev
Fama EF (1965) The behavior of stock-market prices. J Bus 38:34–105
He X, Li K, Li Y (2014) Optimality of momentum and reversal. SSRN 2467392
Hodrick RJ (1992) Dividend yields and expected stock returns: alternative procedures for inference and measurement. Rev Financ Stud 5(3):357–386
Koijen RS, Rodriguez JC, Sbuelz A (2009) Momentum and mean reversion in strategic asset allocation. Manag Sci 55(7):1199–1213
Lengnick M, Wohltmann HW (2010) Agent-based financial markets and new keynesian macroeconomics: a synthesis. J Econ Interact Coord pages 1–32
Lux T, Marchesi M (1999) Scaling and criticality in a stochastic multi-agent model of a financial market. Nature 397(6719):498–500
Poterba JM, Samwick AA (1995) Stock ownership patterns, stock market fluctuations, and consumption. Brook Pap Econ Act 1995(2):295–372
Samuelson P (1965) Proof that properly anticipated prices fluctuate randomly. Ind Manag Rev 6(2):41–49
Shiller RJ (2000) Irrational Exuberance. Princeton University Press, New Jersey
Smithers A (2009) Wall street revalued imperfect markets and inept central bankers. Wiley, Hoboken
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185(4157):1124–1131
Westerhoff FH (2004) Multiasset market dynamics. Macroecon Dyn 8(05):596–616
Zhong M, Darrat AF, Anderson DC (2003) Do US stock prices deviate from their fundamental values? some new evidence. J Bank Financ 27(4):673–697

A Non-Walrasian Microeconomic Foundation of the "Profit Principle" of Investment

Hiroki Murakami

Abstract In this chapter, microeconomic foundation of the "profit principle" of investment is discussed from a non-Walrasian/Keynesian perspective. A non-Walrasian "quantity constraint" is introduced in the intertemporal profit maximization problem to consider non-Walrasian/Keynesian excess supply situations. Consequently, we find that it is possible to provide microeconomic foundation for the profit principle in the case of static expectations but it may not in the case of more general types of expectations. We also clarify that Tobin's q can also be defined in non-Walrasian/Keynesian excess supply situations.

Keywords Non-Walrasian analysis · Profit principle · Quantity constraint · Tobin's q

JEL Classification D50 · E12 · E22

1 Introduction

The "profit principle" of investment states that the current investment demand is determined by the current level of income (or the current level of profit) and the current volume of capital stock. Since Kalecki (1935, 1937) and Kaldor (1940) utilized it to formalize the investment function in their business cycle models, it

H. Murakami (✉)
Graduate School of Economics, University of Tokyo, 7-3-1, Hongo,
Bunkyo-ku, Tokyo 113-0033, Japan
e-mail: leafedclover@gmail.com

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_8

has widely been employed as a principle governing corporate fixed investment.[1] For instance, the investment function that obeys the profit principle has been formulated in mathematical refinements and extensions of Kaldor's (1940) business cycle theory (e.g., Chang and Smyth 1971; Varian 1979; Semmler 1986; Asada 1987, 1995; Skott 1989; Murakami 2014, 2015) and in the post-Keynesian theory of economic growth (e.g., Robinson 1962; Malinvaud 1980; Rowthorn 1981; Dutt 1984; Skott 1989; Marglin and Bhaduri 1990; Lavoie 1992). Moreover, it was confirmed by, for instance, Blanchard et al. (1993) and Cummins et al. (2006) that the profit principle fits well with empirical facts. Thus, the profit principle has played an important role in both theoretical and empirical studies.

Unfortunately, however, there have only a few attempts to provide the microeconomic foundation of the profit principle. Grossman (1972) and Skott (1989, Chap. 6) derived the investment function that obeys the profit principle as a solution of the firm's intertemporal profit maximization problem, but they had theoretical flaws in that adjustment costs of installing new capital (e.g., Eisner and Strotz 1963; Lucas 1967; Gould 1968; Treadway 1969; Uzawa 1969) were ignored and that the optimal level of capital and that of investment were not distinguished.[2] Recently, Murakami (2015) was successful in providing a microeconomic foundation of the profit principle in the presence of adjustment costs of investment, but the analysis was confined to the situation where static expectations prevail.

The purpose of this paper is to examine the possibility of microeconomic foundation of the profit principle of investment in general situations. The starting point of our analysis is Murakami's (2015) microeconomic foundation of the profit principle, and we extend the analysis to more general situations. Following Murakami (2015) we introduce not only adjustment costs of investment but also a non-Walrasian "quantity constraint" (e.g., Barro and Grossman 1971; Drèze 1975; Benassy 1975; Grandmont and Laroque 1976; Malinvaud 1977; Hahn 1978; Negishi 1979), which describes the situation where the firm cannot sell all it can produce due to the deficiency of demand for its product, i.e., where Keynes' (1936) principle of effective demand holds true. By so doing, we intend to demonstrate that the profit principle of investment is closely related to non-Walrasian/Keynesian excess supply situations.

This paper is organized as follows. In Sect. 2, we set up a model of optimal decisions on investment and derive the investment function that obeys the profit principle, by following Murakami (2015). In Sect. 3, we attempt to generalize the

---

[1]The profit principle of investment is often confused with the "acceleration principle" of investment, which was used by Harrod (1936), Samuelson (1939), Hicks (1950) and Goodwin (1951) in their business cycle models, but as Kaldor (1940, p. 79, f. n. 3) pointed out, they are different from each other because the latter asserts that investment demand is determined by the *rate of changes* in income, not by the level of income. In reviewing Hicks (1950), Kaldor (1951, p. 837) also argued that the profit principle is more akin to Keynes' (1936) marginal efficiency theory of investment than the acceleration principle is. Moreover, the acceleration principle is not a theoretical consequence but an empirical law. For these reasons, in this paper, we focus on the microeconomic foundation of the profit principle.

[2]The difference between the concepts of capital and of investment was pointed out by, for example, Lerner (1944) and Haavelmo (1960).

analysis in Sect. 2 by relaxing the assumption of static expectations and verify that the profit principle may not be microeconomically founded without the assumption of static expectations. In Sect. 4, we compare our non-Walrasian microeconomic foundation with several theories of investment: Keynes' (1936) marginal efficiency theory, Tobin's (1969) *q* theory, the neoclassical optimal capital theory (e.g., Jorgenson 1963, 1965) and the neoclassical adjustment cost theory (e.g., Eisner and Strotz 1963; Lucas 1967; Gould 1968; Treadway 1969; Uzawa 1969). In Sect. 5, we conclude this paper by mentioning the possibility of microeconomic foundation of the "utilization principle" of investment.[3]

2 The Model Under Static Expectations

In this section, we present a model of decisions on investment of a price-taking firm. As specified below, the firm is assumed to incur adjustment costs associated with installing new capital and face with a non-Walrasian "quantity constraint." Since we assume that the firm is a price-taker, we can normalize the price of the firm's product as unity.[4]

[3]We mean by the utilization principle of investment that investment demand is determined by the rate of utilization. Along with the profit principle, use has intensively been made of it in the post-Keynesian analysis (e.g., Steindl 1952, 1979; Rowthorn 1981; Dutt 1984, 2006; Amadeo 1986; Skott 1989; Marglin and Bhaduri 1990; Lavoie 1992; Sasaki 2010; Murakami 2016).

[4]The existence of "quantity constraint" may seem incompatible with the assumption of a price-taking firm. Certainly, as Arrow (1959, pp. 45–47) clarified, if a supplier of a commodity cannot sell all he can produce, i.e., if he faces a quantity constraint, he may act as if he were a monopolist, who takes account of the (perceived) inverse demand function of his product in his decision-making. As Negishi (1979) maintained, however, the assumption of a price-taker can be defended even in non-Walrasian excess supply situations, by introducing the assumption of a kinked demand curve (*á la* Sweezy). Indeed, Negishi (1979) stated as follows:

> More important for oligopolistic price rigidity is, therefore, the fact that, as Sweezy stated, any shift in demand will clearly first make itself felt in a change in the quantity sold at the current price. In other words, a shift in demand changes the position of the starting point *P* at which the kink occurs to the right or left without affecting the price. If the marginal cost is not increasing rapidly, the equilibrium price remains unchanged while shifts in demand are absorbed by changes in the level of output. (pp. 80–81)
>
> Although Arrow did not mention it explicitly, such an imperfectly demand curve must be considered to have a kink at the currently realized point or the starting point in the sense of Sweezy. Firstly, perceived demand curves generally have kinks in a non-Walrasian monetary economy where information is not perfect. (p. 87)
>
> When demand falls short of supply, the model of competitive suppliers, is therefore, very much like the Sweezy model of oligopoly, at least in some formal aspects. (p. 88)

If the firm has a perceived demand with kinks due to, for instance, lack of information, as Negishi explained, it is rational for the firm to respond to changes in the demand conditions by adjusting the quantity of its output (which corresponds to the output-capital ratio in our case) rather than by varying the price. In this respect, the assumption of a price-taker is compatible with the existence

First, the production technology of the firm is specified. We assume that the production function of the firm is represented as

$$Y(t) = F(K(t), N(t)). \tag{1}$$

In (1), Y, K, and N stand for the firm's quantity of output, stock of capital and labor employment, respectively.

For the analysis, we make the following standard assumption:

Assumption 1 The real valued function $F : \mathbb{R}^2_{++} \to \mathbb{R}_{++}$ is homogeneous of degree one and twice continuously differentiable with

$$F_N(1, n) > 0,\ F_{NN}(1, n) < 0, \quad \text{for every } n \in \mathbb{R}_{++}, \tag{2}$$

$$\lim_{n\to 0+} F(1, n) = 0,\ \lim_{n\to\infty} F(1, n) = \infty,\ \lim_{n\to\infty} F_N(1, n) = 0,\ \lim_{n\to 0+} F_N(1, n) = \infty. \tag{3}$$

Condition (2) means that the marginal productivity of labor is positive but strictly decreasing, while condition (3) is the so-called Inada condition.

Let $n = N/K$ and define $f(n) = F(1, n)$. Then, Assumption 1 implies that $f : \mathbb{R}_{++} \to \mathbb{R}_{++}$ satisfies

$$f'(n) > 0,\ f''(n) < 0, \quad \text{for every } n \in \mathbb{R}_{++}, \tag{4}$$

$$\lim_{n\to 0+} f(n) = 0,\ \lim_{n\to\infty} f(n) = \infty,\ \lim_{n\to\infty} f'(n) = 0,\ \lim_{n\to 0+} f'(n) = \infty. \tag{5}$$

Second, we introduce the concept of adjustment costs associated with increasing stock of capital (e.g., Eisner and Strotz 1963; Lucas 1967; Gould 1968; Treadway 1969; Uzawa 1969). Following Uzawa (1969) in particular, we assume that the effective cost, including the adjustment cost, of capital accumulation Φ is represented as follows:

$$\Phi(t) = \varphi(z(t))K(t). \tag{6}$$

In (6), z stands for the ratio of gross capital accumulation (including replacement investment) to capital stock. In other words, z satisfies the following equation:

$$\dot{K}(t) = [z(t) - \delta]K(t), \tag{7}$$

where δ is a positive constant which represents the rate of depreciation of capital.

As regards the effective cost function of capital accumulation φ, the following standard assumption is made.[5]

(Footnote 4 continued)
of quantity constraint. Thus, in what follows, it is implicitly assumed that the firm faces a kinked demand curve in the Sweezy–Negishi sense.

[5]Grossman (1972) derived the optimum level of *capital stock* from the profit maximization problem and then formalized investment as a discrepancy between the optimum and current levels of capital

Assumption 2 The real valued function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is twice continuously differentiable with

$$\varphi(0) = 0, \varphi'(z) > 0, \varphi''(z) > 0, \quad \text{for every } z \in \mathbb{R}_+. \tag{8}$$

Condition (8) states that the effective cost function is strictly convex and means that the effective cost of investment is greater as the rate of capital accumulation increases.

Third, we assume that the firm perceives the upper limit of demand for its product due to the existing deficiency of demand, i.e., that the firm faces some kind of "quantity constraint" (*á la* the non-Walrasian theory). In the analysis below, the "quantity constraint" is specified as follows:

$$f(n(t)) \leq x, \tag{9}$$

where x is a positive constant. In (9), x stands for the ratio of the firm's perceived upper limit of demand (X) to its capital stock K. From (9), x can be viewed as the *perceived maximum average productivity of capital* by the firm. The index of x reflects the firm's expectation on the future demand condition.

The quantity constraint can be written in the form of

$$F(K(t), N(t)) \leq X, \tag{10}$$

where X is a positive constant. In (10), X stands for the perceived upper limit of *level* of demand.[6] Constraint (9) is similar to (10) but differs from it. The former means that the firm anticipates that its productivity of capital cannot exceed the upper limit x, while the latter that the firm expects that its level of production cannot be larger than the upper limit X. In discussing growing (resp. shrinking) economies, constraint (10) is unnatural one because, in growing (resp. shrinking) economies, it is natural that the firm expects that the demand for its product increases (resp. decreases) in accordance with a rise (resp. decline) in the scale of the economy, which is measured by, for instance, the number of population or aggregate capital stock. However, this problem can be avoided if constraint (9) is adopted. Moreover, it seems a natural assumption that in making decisions on investment, the firm cares more about the (maximum) productivity of capital, which measures the profitability of capital[7] than about the level of demand. For these reasons, we adopt (9) instead of (10) as a "quantity constraint." As we will see below, this constraint plays a significant role in the firm's decision-making on investment.

(Footnote 5 continued)
stock, while we directly derive the optimum investment from the profit maximization problem by introducing the concept of adjustment costs. In his approach, the optimum level of capital stock can be rationalized but investment itself cannot.

[6]This constraint was adopted by Grossman (1972).

[7]As we will see in (14), the ratio x is related to the (expected) rate of profit ρ.

In the rest of this section, we consider the optimal capital accumulation in the case where the firm's expectations on the perceived maximum average productivity of capital x, the real wage w and the rate of interest r are all static, i.e., where their values are constant over time.

Since the price of the firm's product is normalized as unity, the firm's optimal investment plan (made at time 0) can be represented as a solution $\{z(t)\}_{t=0}^{\infty}$ of the following problem (Problem (M)):

$$\max_{\{z(t),n(t)\}_{t=0}^{\infty}} \int_0^{\infty} [f(n(t)) - wn(t) - \varphi(z(t))]K(t)e^{-rt}dt \qquad \text{(M)}$$
$$\text{s.t. (7) and (9),}$$

where $K(0)$, x, w and r are given positives.

As Murakami (2015) proved, the case in which the quantity constraint (9) binds can be characterized by the situation in which the marginal productivity of labor corresponding to the perceived maximum average productivity of capital x is larger than the real wage w.

Proposition 1 *Let Assumptions 1 and 2 hold. Assume that the following condition is satisfied*[8]*:*

$$f'(f^{-1}(x)) > w, \qquad (11)$$

Then, for every solution to Problem (M), $n(t)$ is equal to the positive constant n^* for all $t \geq 0$ such that

$$f(n^*) = x. \qquad (12)$$

Proof See Murakami (2015, p. 29, Proposition 2.1). Since this proposition is a corollary to Proposition 3, see also the proof of Proposition 3. □

Proposition 1 states that if the marginal productivity of labor corresponding to the perceived maximum productivity of capital, which is determined by the firm's expectation on future demand conditions, exceeds the (expected) real wage, the firm's optimum production per unit of capital is reduced to the perceived upper limit of demand per unit of capital. In this respect, condition (11) is the one that characterizes Keynes' (1936) principle of effective demand in the long run.[9]

The expected rate of gross profit $\rho = f(n^*) - wn^*$ can be defined as a function of x and w[10]:

[8]It can be verified that, under (4) and (5) deduced from Assumption 1, the inverse functions $f^{-1}, f'^{-1} : \mathbb{R}_{++} \to \mathbb{R}_{++}$, exist.

[9]The same condition as (11) can be found in Barro and Grossman (1971, p. 85), which characterizes non-Walrasian excess supply situations. However, their analysis was static in nature.

[10]The expected rate of gross profit ρ is, in principle, identical to Keynes' (1936, chap. 11) marginal efficiency of capital. However, as long as condition (11) holds, it is generally not equal to the marginal productivity of capital.

$$\rho(x,w)=\begin{cases} x-wf^{-1}(x) & \text{if condition (11) holds} \\ f(f'^{-1}(w))-wf'^{-1}(w) & \text{otherwise} \end{cases}. \tag{13}$$

The partial derivatives of ρ are given by:

$$\rho_x=\begin{cases} 1-w/f'(f^{-1}(x))>0 \\ 0 \end{cases},\ \rho_w=\begin{cases} -f^{-1}(x)<0 & \text{if condition (11) holds} \\ -f'^{-1}(w)<0 & \text{otherwise} \end{cases}. \tag{14}$$

It follows from (14) that the real valued function $\rho:\mathbb{R}^2_{++}\to\mathbb{R}_{++}$ is continuously differentiable. Unlike in the neoclassical theory of investment (e.g., Jorgenson 1963, 1965; Lucas 1967; Gould 1968; Treadway 1969; Uzawa 1969), the (expected) rate of gross profit ρ is affected by the perceived maximum average productivity of capital x, i.e., by the firm's expectation on the future demand condition.

Thanks to Proposition 1, the firm's optimal expected profit maximization problem, Problem (M), can be reduced to

$$\max_{\{z(t)\}_{t=0}^{\infty}} \int_0^{\infty}[\rho(x,w)-\varphi(z(t))]K(t)e^{-rt}dt \quad \text{s.t. (7)} \tag{M}$$

In the following, Problem (M), redefined above, is examined.

To solve Problem (M), we impose the following constraint:

$$\int_0^{\infty}[r+\delta-z(t)]dt=\infty. \tag{15}$$

Condition (15) is satisfied if the rate of net capital accumulation $z(t)-\delta$ is less than the rate of interest r. To see what condition (15) implies, suppose, for the time being, that $z(t)$ is a constant z over time. Under this assumption, condition (15) reduces to $z-\delta<r$ and the marginal effect of an increase in z on the discounted present value of the firm's profit at $t=\infty$ is zero because we have

$$\lim_{t\to\infty}\varphi'(z)K(t)e^{-rt}=\lim_{t\to\infty}\varphi'(z)K(0)e^{-(r+\delta-z)t}=0.$$

In the case where $z(t)$ is constant over time, condition (15) has the same meaning as that of the usual transversality condition.

As Murakami (2015) verified, the solution of Problem (M), $\{z(t)\}_{t=0}^{\infty}$, is constant over time under condition (15).

Proposition 2 *Let Assumptions 1 and 2 hold. Assume that condition (15) holds. Then, if there exists a solution of Problem (M), $z(t)$ is equal to the constant z^* for $t\geq 0$ such that*

$$\frac{\rho(x,w)-\varphi(z^*)}{r+\delta-z^*}=\varphi'(z^*). \tag{16}$$

Proof See Murakami (2015, p. 30, Proposition 2.2). □

Proposition 2 states that, if the firm's expectations on the perceived maximum of productivity of capital (the index of the firm's expectations on demand) x, on the rate of interest r and on the real wage w are static, the optimum rate of capital accumulation z^* is unique and constant. Figure 1 illustrates geometrically how the optimal rate of capital accumulation z^* is determined.

Proposition 2 implies that the optimum rate of capital accumulation z^* can be represented as a function of x, w and r in the following form:

$$z^* = g(x, w, r). \tag{17}$$

Furthermore, it follows from (8), (14)–(16) that the partial derivatives of g are given by

$$\begin{aligned} g_x &= \frac{\rho_x}{\varphi'(z^*)(r+\delta-z^*)} \begin{cases} > 0 & \text{if condition (11) holds} \\ = 0 & \text{otherwise} \end{cases}, \\ g_w &= \frac{\rho_w}{\varphi'(z^*)(r+\delta-z^*)} < 0,\ g_r = -\frac{\varphi''(z^*)}{\varphi'(z^*)(r+\delta-z^*)} < 0. \end{aligned} \tag{18}$$

Condition (18) says that the optimal rate of capital accumulation g is strictly increasing in the index of the expectation on future demand x if the marginal productivity of labor corresponding to x is greater than the given real wage w, while g is inelastic to x otherwise. This implies that investment demand is affected by the firm's expectation on demand provided that the marginal productivity of labor exceeds the real wage. In the sense that the firm's expectation on future demand has an influence on investment demand, the formula given in (16) may be regarded as a natural expression of Keynes' (1936, Chap. 11) theory of investment.

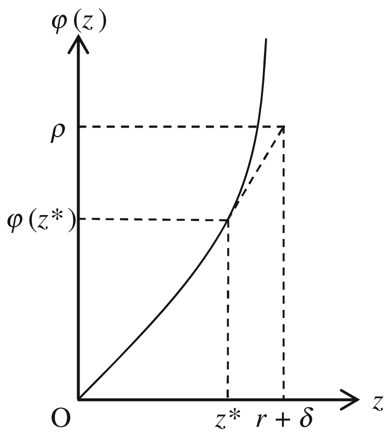


Fig. 1 Optimal rate of capital accumulation

Tobin's (1969) q of investment can also be defined in our context. Since the optimal rate of capital accumulation $z(t)$ is shown to be equal to a unique constant z^* over time, the discounted present value of the firm's expected profit associated with the optimal plan of capital accumulation, or V^*, can be calculated as follows:

$$V^* = K(0)\frac{\rho(x, w) - \varphi(z^*)}{r + \delta - z^*}.$$

Since the replacement cost of capital is $K(0)$ at time 0 because the price is normalized as unity, we know from (16) that Tobin's (1969) q can be defined as follows:

$$q = \frac{V^*}{K(0)} = \frac{\rho(x, w) - \varphi(z^*)}{r + \delta - z^*} = \varphi'(z^*). \tag{19}$$

As the q ratio defined in (19) allows for the case in which the non-Walrasian quantity constraint (9) binds, it may be interpreted as Tobin's q in *non-Walrasian/Keynesian excess supply situations*. This definition of Tobin's q ratio is different from those of Yoshikawa (1980) and Hayashi (1982) in that a non-Walrasian quantity constraint was not taken into account in their definitions and that their definitions dealt only with the case of full employment. Unlike in the interpretations of the q theory by Yoshikawa (1980) and Hayashi (1982), the role of expectation on the future demand condition (represented by x) is taken into account in our interpretation. Since Tobin (1969) defined the q ratio to discuss corporate investment in non-Walrasian/Keynesian excess supply situations, our interpretation of the q theory conforms more to Tobin's (1969) original definition than those in the preceding works.

According to (19), the optimal rate of capital accumulation z^* may also be represented as an increasing function of q. Moreover, if $\varphi'(0) = 1$, which is often assumed as a property of the effective cost function φ, condition (16) implies that when $\rho(x, w) - \delta \leq r$, we have $z^* = 0$. This means that when the (expected) rate of profit (net of the rate of depreciation) is less than or equal to the rate of interest, no new investment is carried out. This result is consistent with Tobin's (1969) q theory of investment and Keynes' (1936) marginal efficiency theory of investment.

In what follows, we proceed to derive the investment function that obeys the profit principle of investment.

Since Proposition 1 implies that the firm's current output is represented as $Y(0) = xK(0)$ under condition (11), non-Walrasian/Keynesian excess supply situations can be characterized by

$$f'\left(f^{-1}\left(\frac{Y(0)}{K(0)}\right)\right) > w. \tag{20}$$

On the other hand, when the quantity constraint (9) does not bind, the optimal level of production is so determined that the marginal productivity of labor would be equal to the real wage. So if condition (20) is not met, the following condition is fulfilled:

$$f'\Big(f^{-1}\Big(\frac{Y(0)}{K(0)}\Big)\Big) = w. \tag{21}$$

The situations in which the *notional* labor demand is met are characterized by (21).

Condition (18) can thus be replaced by

$$g_{Y/K} \begin{cases} > 0 & \text{when condition (20) holds} \\ = 0 & \text{when condition (21) holds} \end{cases}, \ g_r < 0, g_w < 0. \tag{22}$$

In particular, condition (22) says that investment demand g is positively influenced by the current (average) productivity of capital $Y(0)/K(0)$ in non-Walrasian excess supply situations (in the case of (20)) but not affected by it in the situations where the notional labor demand is fulfilled (in the case of (21)). The negative effects of the rate of interest and of the real wage on investment demand are also confirmed by (22).

Therefore, the gross capital accumulation function and the investment expenditure function, Z and I, can be defined, respectively, as follows:

$$Z(Y(0), K(0), w, r) = g\Big(\frac{Y(0)}{K(0)}, w, r\Big)K(0), \tag{23}$$

$$I(Y(0), K(0), w, r) = \varphi\Big(g\Big(\frac{Y(0)}{K(0)}, w, r\Big)\Big)K(0). \tag{24}$$

It follows from (22) that the partial derivatives of Z and I are given by

$$\begin{aligned} &Z_Y = g_{Y/K} \begin{cases} > 0 \\ = 0 \end{cases}, \ Z_K = g - g_{Y/K}\frac{Y}{K} \begin{cases} \lesseqgtr 0 & \text{when condition (20) holds} \\ \geq 0 & \text{when condition (21) holds} \end{cases}, \\ &Z_w = g_w K, Z_r = g_r K < 0, \end{aligned} \tag{25}$$

$$\begin{aligned} &I_Y = \varphi'(g)g_{Y/K} \begin{cases} > 0 \\ = 0 \end{cases}, \ I_K = \varphi(g) - \varphi'(g)g_{Y/K}\frac{Y}{K} \begin{cases} \lesseqgtr 0 & \text{when condition (20) holds} \\ \geq 0 & \text{when condition (21) holds} \end{cases}, \\ &I_w = \varphi'(g)g_w K < 0, I_r = \varphi'(g)g_r K < 0. \end{aligned} \tag{26}$$

The investment functions Z and I ((23) and (24)) can be considered to obey the profit principle of investment, because the current investment demand is a function of the current income $Y(0)$ and the current capital stock $K(0)$. Furthermore, since the (current) level of income $Y(0)$ has a positive impact on the current investment demand only in the case of (20), i.e., only in the case where the equality of the quantity constraint (9) holds, the profit principle of investment is closely related to non-Walrasian/Keynesian excess supply situations.

In this section, we have verified that the profit principle of investment can be rationalized by following Murakami (2015), but we have assumed that the firm's expectations (on x, w and r) are static. In the next section, we shall drop the assumption of static expectations and investigate whether the profit principle can also be rationalized under more general types of expectations.

3 The Model Under General Expectations

In this section, we explore the possibility that the profit principle of investment is microeconomically founded in general expectations.

To allow for the case in which the firm's expectations on the perceived maximum of productivity of capital x, on the real wage w and on the rate of interest r vary with time t, they are represented as functions of t : $x(t)$, $w(t)$, $r(t)$.

Under the general expectations, the quantity constraint (9) is replaced with

$$f(n(t)) \leq x(t). \tag{27}$$

Note that constraint (27) includes (10) as a special case.

For given expectations $\{x(t), w(t), r(t)\}_{t=0}^{\infty}$, the firm's expected profit maximization problem (Problem (G)) can be formalized as follows:

$$\max_{\{z(t),n(t)\}_{t=0}^{\infty}} \int_0^{\infty} [f(n(t)) - w(t)n(t) - \varphi(z(t))]K(t) \exp\Big(-\int_0^t r(s)ds\Big)dt \qquad \text{(G)}$$
$$\text{s.t. (7) and (27),}$$

where $K(0) > 0$; $x(t) > 0$, $w(t) > 0$ and $r(t) > 0$ for all $t \geq 0$.

Let $\lambda(t) \geq 0$ be the Lagrange multiplier concerning (27) and set the Hamiltonian as follows:

$$\begin{aligned} H(K(t), n(t), z(t); \lambda(t), \mu(t)) = {} & [f(n(t)) - w(t)n(t) - \varphi(z(t))]K(t) \\ & + \lambda(t)[x(t) - f(n(t))] + \mu(t)[z(t) - \delta]K(t). \end{aligned}$$

Then, we know from the Kuhn–Tucker condition and the maximum principle that, if a solution of Problem (G) $\{z(t), n(t)\}_{t=0}^{\infty}$ exists, it satisfies (7), (27) and

$$\begin{aligned} &\frac{\partial H(t)}{\partial n(t)} = 0, \\ &\frac{\partial H(t)}{\partial z(t)} = 0, \\ &\lambda(t)[x(t) - f(n(t))] = 0, \\ &\dot{\mu}(t) = r(t)\mu(t) - \frac{\partial H(t)}{\partial K(t)}, \end{aligned}$$

or

$$\left[1-\frac{\lambda(t)}{K(t)}\right]f'(n(t)) = w(t), \tag{28}$$

$$\dot{z}(t) = \frac{1}{\varphi''(z(t))}\{[r(t)+\delta-z(t)]\varphi'(z(t)) - [f(n(t)) - w(t)n(t) - \varphi(z(t))]\}, \tag{29}$$

$$\lambda(t)[x(t) - f(n(t))] = 0. \tag{30}$$

One can find from (30) that if $\lambda > 0$, the equality of (9) is fulfilled. In other words, if $\lambda > 0$, the firm's output per unit of capital $f(n)$ is restricted to the perceived upper limit of demand per unit of capital x. As in Sect. 2, we can characterize the situation in which the quantity constraint (27) binds.

Proposition 3 *Let Assumptions 1 and 2 hold. Assume that the following condition is satisfied at t:*

$$f'(f^{-1}(x(t))) > w(t). \tag{31}$$

Then, for every solution of Problem (G), $n(t)$ satisfies the following equation at t:

$$f(n(t)) = x(t). \tag{32}$$

Proof Suppose, for the sake of contradiction, that $f(n(t)) < x(t)$ (note that $n(t)$ must satisfy (27)).

By the implicit function theorem, conditions (4) and (5) imply that the inverse function $f^{-1} : \mathbb{R}_{++} \to \mathbb{R}_{++}$ of f exists and satisfies $df^{-1}/dx > 0$. Then, the above assumption implies

$$n(t) < f^{-1}(x(t)).$$

Because of (4), we have

$$f'(n(t)) > f'(f^{-1}(x(t))). \tag{33}$$

Since we have $\lambda(t) = 0$ by condition (30) and the above assumption, conditions (28) and (33) imply

$$w(t) = f'(n(t)) > f'(f^{-1}(x(t))).$$

But this contradicts (31).

Therefore, condition (32) is fulfilled. □

Proposition 3 is a generalized version of Proposition 1 and condition (31) is a generalized version of (11) and characterizes the situation where Keynes' (1936) principle of effective demand holds.

With the help of Proposition 3, Problem (G) can thus be reformulated as follows:

$$\max_{\{z(t)\}_{t=0}^{\infty}} \int_0^{\infty} [\rho(x(t), w(t)) - \varphi(z(t))]K(t)\exp\Big(-\int_0^t r(s)ds\Big)dt \quad \text{s.t. } (7), \tag{G}$$

where ρ is defined by (13).

Then, the first order conditions for optimality in Problem (G), (28)–(30), can be reduced to

$$\dot{z}(t) = \frac{1}{\varphi''(z(t))}\{[r(t) + \delta - z(t)]\varphi'(z(t)) - [\rho(x(t), w(t)) - \varphi(z(t))]\}. \tag{34}$$

This fact is summarized in the following proposition:

Proposition 4 *Let Assumptions 1 and 2 hold. Then, if there exists a solution of Problem (G),* $\{z(t)\}_{t=0}^{\infty}$, *it satisfies (34).*

As in Sect. 2, the first order condition for optimality, (34), can also be interpreted *à la* Tobin's (1969) q theory. To verify this fact, assume that the following transversality condition holds:

$$\lim_{t\to\infty} \varphi'(z(t))K(t)\exp\Big(-\int_0^t r(s)ds\Big) = 0. \tag{35}$$

By (7) and (35), we find that a solution of Problem (G), $\{z(t)\}_{t=0}^{\infty}$, satisfies

$$\begin{aligned}&\int_0^{\infty} \varphi''(z(t))\dot{z}(t)K(t)\exp\Big(-\int_0^t r(s)ds\Big)dt \\ &\quad = -\varphi(z(0))K(0) + \int_0^{\infty} \varphi'(z(t))[r(t) + \delta - z(t)]\Big(-\int_0^t r(s)ds\Big)dt.\end{aligned} \tag{36}$$

Moreover, it follows from (34) that along a solution of Problem (G), $\{z(t)\}_{t=0}^{\infty}$, we have

$$\begin{aligned}&\int_0^{\infty} \varphi''(z(t))\dot{z}(t)K(t)\exp\Big(-\int_0^t r(s)ds\Big)dt \\ &\quad = \int_0^{\infty} \{[r(t) + \delta - z(t)]\varphi'(z(t)) - [\rho(x(t), w(t)) \\ &\quad - \varphi(z(t))]\}K(t)\exp\Big(-\int_0^t r(s)ds\Big)dt.\end{aligned} \tag{37}$$

Comparing (36) and (37) with each other, we find that a solution of Problem (G), $\{z(t)\}_{t=0}^{\infty}$, satisfies

$$\int_0^{\infty} [\rho(x(t), w(t)) - \varphi(z(t))]K(t) \exp\Big(-\int_0^t r(s)ds\Big)dt = \varphi'(z(0))K(0). \quad (38)$$

Letting V be the left-hand side of (38), we know that a solution of Problem (G), $\{z(t)\}_{t=0}^{\infty}$, fulfills

$$q = \frac{V}{K(0)} = \varphi'(z(0)). \quad (39)$$

The left hand side of (39) is Tobin's (1969) q because V is the discounted present value of the firm's expected profit and $K(0)$ is the replacement cost of capital at time 0. Condition (38) says that the discounted present value of the firm's expected profit is equal to the marginal cost of installing new capital. Since the firm's expectations are general, the q given by (39) can be interpreted as a generalized version of Tobin's q in *non-Walrasian/Keynesian excess supply situations*.

We are now ready to check if the profit principle of investment can be rationalized even under general expectations. In the below, it is demonstrated that, without the assumption of static expectations, the investment function that obeys the profit principle may not be obtained. Since the profit principle states that the current investment demand or the current rate of capital accumulation $z(0)$ is a function of the current income $Y(0) = x(0)K(0)$ and the current stock of capital $K(0) > 0$, it suffices for our purpose to show that, even if the current capital stock $K(0)$ and the current perceived upper limit of productivity of capital $x(0)$ are specified, the current optimal rate of capital accumulation $z(0)$ is not uniquely determined without the assumption of static expectations. To do so, we assume that the real wage $w(t)$ and the rate of interest $r(t)$ are positive constants w and r, respectively, for all $t \geq 0$ and consider the optimal rate of capital accumulation $z(0)$ for the following two time paths of $\{x(t)\}_{t=0}^{\infty}$; the one is $x_0(t) = x$ for all $t \geq 0$; the other is $x_1(t) = x$ for $t \in [0, t_0]$ or $t \geq t_1$ and $x_1(t) > x$ for $t \in (t_0, t_1)$, where x is a positive constant that satisfies (11) and t_0 and t_1 are positive with $t_0 < t_1$. Let $\{z_i(t), K_i(t)\}_{t=0}^{\infty}$ and V_i be the optimal plans of capital accumulation and capital stock corresponding to $\{x_i(t)\}_{t=0}^{\infty}$ and the discounted present value of the firm's expected profit obtained along $\{z_i(t)\}_{t=0}^{\infty}$, respectively, for $i = 0, 1$. Then, we have

$$\begin{aligned} V_0 &= \int_0^{\infty} [\rho(x_0(t), w) - \varphi(z_0(t))]K_0(t)e^{-rt}dt \\ &< \int_0^{\infty} [\rho(x_1(t), w) - \varphi(z_0(t))]K_0(t)e^{-rt}dt \\ &\leq \int_0^{\infty} [\rho(x_1(t), w) - \varphi(z_1(t))]K_1(t)e^{-rt}dt = V_1 \end{aligned}$$

Because the optimal plan of capital accumulation $\{z_i(t)\}_{t=0}^{\infty}$ satisfies (38), we obtain

$$\varphi'(z_0(0))K_0(0) < \varphi'(z_1(0))K_1(0).$$

Since we have $K_0(0) = K_1(0) = K(0) > 0$, we find from (8) that

$$z_0(0) < z_1(0). \tag{40}$$

Noting that $K_0(0) = K_1(0) = K(0)$ and $x_0(0) = x_1(0) = x$, inequality (40) implies that, even when $K(0)$ and $x(0)$ are specified, $z(0)$ may not uniquely be determined. This consequence suggests that the profit principle may not be rationalized without the assumption of static expectations. Therefore, the profit principle of investment may not necessarily be obtained as the optimal plan of capital accumulation under general expectations.

In this section, we have extended the argument in Sect. 2 to the case where more general expectations prevail and checked if the profit principle of investment can be microeconomically founded. We have revealed that the optimal plan of capital accumulation and Tobin's (1969) q can be derived in the existence of non-Walrasian quantity constraint even under general types of expectations but that the profit principle of investment may not be rationalized without static expectations. This clarifies that the assumption of static expectations plays a critical role in the profit principle of investment.

4 Discussion on Non-Walrasian Microeconomic Foundation of Investment

We have so far explored non-Walrasian microeconomic foundation of the investment function. In particular, we have inquired into the possibility of microeconomic foundation of the profit principle of investment. In this section, we turn to the advantage of non-Walrasian microeconomic foundation of investment. For this purpose, we compare our non-Walrasian microeconomic foundation with the other representative theories on investment: the Keynesian theories (Keynes' marginal efficiency theory and Tobin's q theory) and the neoclassical theories (the neoclassical optimal capital theory and the neoclassical adjustment cost theory).

First, we compare our results with Keynes' (1936) marginal efficiency theory of investment. Keynes' (1936, chap. 11) maintained that the optimal investment is subjected largely to changes in the marginal efficiency of investment.[11] Our results are in favor of Keynes' (1936) argument because the (expected) rate of profit $\rho(x, w)$,

[11] Lerner (1944) argued that the term "marginal efficiency of capital" used in Keynes' (1936) *General Theory* should be renamed "marginal efficiency of investment" because the concepts of optimal capital stock and optimal investment are different from each other.

which can be regarded as identical with the marginal efficiency, has a positive effect on investment demand. In this respect, the results of our analysis may be viewed as an appropriate expression of Keynes' theory of investment.

Second, the relationship between Tobin's (1969) q and our results is examined. As we have seen in Sects. 2 and 3 (especially (39)), our results can provide a support for Tobin's q theory. Since Tobin (1969) originally coined the concept of q on the basis of Keynes' (1936) theory, it is a natural consequence that our results also constitute a foundation of Tobin's q theory. Furthermore, the microeconomic foundation of Tobin's q theory derived from our analysis is more comprehensive than those by Yoshikawa (1980) and Hayashi (1982) in that excess supply (underemployment) situations can also be discussed in ours unlike in Yoshikawa (1980) and Hayashi (1982).

Third, our results are contrasted with the neoclassical optimal capital theory of investment (e.g., Jorgenson 1963, 1965). In this theory, the optimum investment is discussed in the framework of dynamic optimization, but this theory has been criticized for its failure to describe investment as an optimum activity because, in this theory, investment is explained as an activity to fill the gap between the optimum and current levels of capital stock.[12] Since we have introduced the concept of adjustment costs, we can escape from this kind of criticism. What is more, the neoclassical optimal capital theory is, in general, an investment theory in full employment situations (neoclassical situations) and does not discuss investment plans in non-Walrasian excess supply situations. In this respect, our analysis has an advantage over this theory.

Fourth, the differences between the neoclassical adjustment cost theory of investment (e.g., Eisner and Strotz 1963; Lucas 1967; Gould 1968; Treadway 1969; Uzawa 1969) are mentioned. As we have seen in Sects. 2 and 3, full employment situations alone are considered in the neoclassical adjustment cost theory, but our results accommodate both full employment and underemployment (non-Walrasian excess supply) situations and so include the results obtained in the neoclassical adjustment cost theory as a special case because we have allowed for the case where the quantity constraint (9) or (27) binds (if this constraint does not bind, the results obtained in our analysis reduce to those obtained in the neoclassical adjustment cost theory).

Thus, our analysis has advantages as a general investment theory in that it incorporates both Keynesian and neoclassical aspects by imposing the quantity constraint (9) or (27) and takes into consideration the difference between capital and investment by introducing the adjustment cost function. In this sense, our analysis can be said to retain generality.

[12] The difference between the concepts of capital and of investment was pointed out by, for example, Lerner (1944) and Haavelmo (1960).

5 Concluding Remarks

We summarize the analysis in this paper.

In Sect. 2, we have formalized a model of optimal investment that emphasizes adjustment costs of investment and non-Walrasian "quantity constraint," by following Murakami (2015). We have found through the analysis in this section that Tobin's (1969) q can be extended to non-Walrasian/Keynesian excess supply situations and that the profit principle of investment can be derived as the intertemporally optimal plan of capital accumulation. However, the argument in this section has been confined to the case of static expectations.

In Sect. 3, we have generalized the argument in Sect. 2 by allowing for more general types of expectations. We have made clear that the formula of optimal capital accumulation and Tobin's q can also be derived even under the assumption of general expectations but that the profit principle of investment may not necessarily be microeconomically founded without the assumption of static expectations. By so doing, we have verified that the assumption of static expectations is vital for the profit principle.

In Sect. 4, our non-Walrasian microeconomic foundation of investment has been compared with other investment theories: Keynes' marginal efficiency theory, Tobin's q theory, the neoclassical optimal capital theory and the neoclassical adjustment cost theory. It has been confirmed that our microeconomic foundation provides natural and appropriate expressions of Keynes' marginal efficiency theory and of Tobin's q theory, includes the neoclassical adjustment cost theory as a special case and is superior to the neoclassical optimal capital theory.

Before concluding this paper, we shall mention the possibility of microeconomic foundation of the utilization principle of investment. In this paper, for the sake of simplicity, we have not explicitly taken account of (variations of) the rate of utilization. So our analysis does not directly contribute to microeconomic foundation of the utilization principle. However, since the rate of utilization u is usually measured by the output-capital ratio Y/K and the results in Sect. 2 indicate that the optimal rate of capital accumulation z^* is influenced positively by the ratio Y/K, our analysis may partially make a contribution to microeconomic foundation of the utilization principle.

Acknowledgments The author would like to thank a referee for his/her comments on the earlier version of this paper. Needless to say, the author is solely responsible for possible remaining errors in this paper. This work was financially supported by the Japan Society for the Promotion of Science (Grant-in-Aid for JSPS Fellows, Grant Number 14J03350).

References

Amadeo EJ (1986) The role of capacity utilization in long-period analysis. Polit Econ 2(2):147–160
Arrow KJ (1959) Toward a theory of price adjustment. In: Abramovitz M (ed) The allocation of economic resources. Stanford University Press, Stanford, pp 41–51

Asada T (1987) Government finance and wealth effect in a Kaldorian cycle model. J Econ 47(2):143–166
Asada T (1995) Kaldorian dynamics in an open economy. J Econ 62(3):239–269
Barro RJ, Grossman HI (1971) A general disequilibrium model of income and employment. Am Econ Rev 61(1):82–93
Benassy JP (1975) Neo-Keynesian disequilibrium theory in a monetary economy. Rev Econ Stud 42(4):503–523
Blanchard OJ, Rhee C, Summers LH (1993) The stock market, profit, and investment. Q J Econ 108(1):115–136
Chang WW, Smyth DJ (1971) The existence and persistence of cycles in a non-linear model: Kaldor's 1940 model re-examined. Rev Econ Stud 38(1):37–44
Cummins JG, Hassett KA, Oliner SD (2006) Investment behavior, observables expectations, and internal funds. Am Econ Rev 96(3):796–810
Drèze JH (1975) Existence of an exchange equilibrium under price rigidities. Int Econ Rev 16(2):301–320
Dutt AK (1984) Stagnation, income distribution and monopoly power. Camb J Econ 8(1):25–40
Dutt AK (2006) Aggregate demand, aggregate supply and economic growth. Int Rev Appl Econ 20(3):319–336
Eisner R, Strotz RH (1963) The determinants of business investment. Impact on monetary policy: commission on money and credit. Prentice-Hall Inc., Englewood Cliffs, pp 59–337
Goodwin RM (1951) The non-linear accelerator and the persistence of business cycles. Econometrica 19(1):1–17
Gould JP (1968) Adjustment costs in the theory of investment of the firm. Rev Econ Stud 35(1):47–55
Grandmont JM, Laroque G (1976) On temporary Keynesian equilibria. Rev Econ Stud 43(1):53–67
Grossman HI (1972) A choice-theoretic model of an income-investment accelerator. Am Econ Rev 62(4):630–641
Haavelmo T (1960) A study in the theory of investment. University of Chicago Press, Chicago
Hahn FH (1978) On non-Walrasian equilibria. Rev Econ Stud 45(1):1–17
Harrod RF (1936) The trade cycle: an essay. Clarendon Press, Oxford
Hayashi F (1982) Tobin's marginal q and average q: a neoclassical interpretation. Econometrica 50(1):213–224
Hicks JR (1950) A contribution to the theory of the trade cycle. Clarendon Press, Oxford
Jorgenson D (1963) Capital theory and investment behavior. Am Econ Rev 53(2):129–155
Jorgenson D (1965) Anticipation and investment behavior. In: Duesenberry JS, Fromm G, Klein LR, Kuh E (eds) The Brookings quarterly econometric model of the United States. Rand MacNally, Chicago
Kaldor N (1940) A model of the trade cycle. Econ J 50(197):78–92
Kaldor N (1951) Mr. Hicks on the trade cycle. Econ J 61(244):833–847
Kalecki M (1935) A macrodynamic theory of business cycle. Econometrica 3(3):327–344
Kalecki M (1937) A theory of the business cycle. Rev Econ Stud 4(2):77–97
Keynes JM (1936) The general theory of employment, interest and money. MacMillan, London
Lavoie M (1992) Foundations of post-Keynesian economic analysis. Edward Elgar, Aldershot
Lerner AP (1944) The economics of control: principles of welfare economics. MacMillan, New York
Lucas RE (1967) Adjustment costs and the theory of supply. J Polit Econ 75(4):321–334
Malinvaud E (1977) The theory of unemployment reconsidered. Basil Blackwell, London
Malinvaud E (1980) Profitability and unemployment. Cambridge University Press, Cambridge
Marglin SA, Bhaduri A (1990) Profit squeeze and Keynesian theory. In: Marglin A, Schor JB (eds) The golden age of capitalism: reinterpreting the postwar experience. Clarendon Press, Oxford, pp 153–186
Murakami H (2014) Keynesian systems with rigidity and flexibility of prices and inflation-deflation expectations. Struct Change Econ Dyn 30:68–85

Murakami H (2015) Wage flexibility and economic stability in a non-Walrasian model of economic growth. Struct Change Econ Dyn 32:25–41
Murakami H (2016) Inflation-deflation expectations and economic stability in the Kaleckian system, Mimeo
Negishi T (1979) Microeconomic foundations of Keynesian macroeconomics. North-Holland, Amsterdam
Robinson J (1962) Essays in the theory of economic growth. Macmillan, London
Rowthorn RE (1981) Demand, real wages and economic growth. Thames Pap Polit Econ 22:1–39
Samuelson PA (1939) Interactions between the multiplier analysis and the principle of acceleration. Rev Econ Stat 21(2):75–78
Sasaki H (2010) Endogenous technical change, income distribution, and unemployment with interclass conflict. Struct Change Econ Dyn 21:123–134
Semmler W (1986) On nonlinear theories of economic cycles and the persistence of business cycles. Math Soc Sci 12:47–76
Skott P (1989) Effective demand, class struggle in economic growth. Cambridge University Press, Cambridge
Steindl J (1952) Maturity and stagnation in American capitalism. Basil Blackwell, Oxford
Steindl J (1979) Stagnation theory and stagnation policy. Camb J Econ 3(1):1–14
Tobin J (1969) A general equilibrium approach to monetary theory. J Money Credit Bank 1(1):15–29
Treadway AB (1969) On rational entrepreneurial behavior and the demand for investment. Rev Econ Stud 36(2):227–239
Uzawa H (1969) Time preference and the Penrose effect in a two-class model of economic growth. J Polit Econ 77(4):628–652
Varian HR (1979) Catastrophe theory and the business cycle. Econ Inq 17(1):14–28
Yoshikawa H (1980) On the "q" theory of investment. Am Econ Rev 70(4):739–743

Part III
Macroeconomic Models

The Stability of Normal Equilibrium Point and the Existence of Limit Cycles in a Simple Keynesian Macrodynamic Model of Monetary Policy

Toichiro Asada, Michal Demetrian and Rudolf Zimka

Abstract In this chapter, a simple Keynesian macroeconomic model of monetary policy describing the development of nominal rate of interest, expected rate of inflation, and nominal money supply in the period of deflationary depression, which was introduced by Asada (2011) is investigated rigorously. The normal equilibrium point of the model is derived and its dynamic stability is investigated. Questions concerning the existence of limit cycles are studied analytically. The bifurcation equation is found. The formulae for the calculation of its coefficients are gained. A numerical example is presented by means of numerical simulations.

Keywords Keynesian macrodynamic model · Monetary policy · Dynamic stability · Limit cycle · Numerical simulations

1 Introduction

Minsky's (1982, 1986) financial instability hypothesis implies that the financially dominated capitalist economy is inherently unstable. The real situation during several last decades 'proves' it. For example, Japanese economy fell into the serious deflationary depression in the 1990s and could not get out of it for nearly twenty years, and the big financial crisis, that started in USA by mortgage crisis in 2008, rapidly expanded to the European Union and to other parts in the world.[1] As a reaction especially to the deflationary depression in the Japanese economy, Asada (2011)

[1] See Asada (ed.) (2015) and Wakatabe (2015) for the Japanese depression, and see Krugman (2012) for the subprime mortgage crisis in USA.

T. Asada (✉)
Faculty of Economics, Chuo University, Tokyo, Japan
e-mail: asada@tamacc.chuo-u.ac.jp

M. Demetrian
Faculty of Mathematics and Informatics, Comenius University, Bratislava, Slovakia

R. Zimka
Faculty of Economics, Matej Bel University, Banská, Bystrica, Slovakia

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_9

set up a simple Keynesian macrodynamic model of monetary policy describing the development of nominal rate of interest, expected rate of inflation, and nominal money supply. In Asada (2011), however, analytical treatment is rather sketchy and the numerical simulation is not presented.

This paper studies Asada's (2011) dynamic model rigorously both analytically and numerically. In Sect. 2, the model is introduced and its normal equilibrium point is derived. Section 3 is devoted to the dynamic stability / instability of the normal equilibrium point. In Sect. 4 questions concerning the existence of limit cycles are studied rigorously. The bifurcation equation is derived. The formulae for the calculation of its coefficients are gained. In Sect. 5, a numerical example is presented by means of numerical simulation. Section 6 is devoted to the concluding remarks, which gives an economic interpretation of the reached results and suggests other possibilities of the extension of the model.

2 The Model

The model consists of the following system of equations:

$$Y = Y\left(r - \pi^e, G, \tau\right);\ Y_{r-\pi^e} = \frac{\partial Y}{\partial\left(r - \pi^e\right)} < 0, \quad Y_G = \frac{\partial Y}{\partial G} > 0, \quad Y_\tau = \frac{\partial Y}{\partial \tau} < 0 \tag{1}$$

$$\frac{M}{p} = L\left(Y, r, \pi^e\right);\ L_Y = \frac{\partial L}{\partial Y} > 0, \quad L_r = \frac{\partial L}{\partial r} < 0, \quad L_{\pi^e} = \frac{\partial L}{\partial \pi^e} < 0 \tag{2}$$

$$\pi = \varepsilon\left(Y - \bar{Y}\right) + \pi^e;\ \bar{Y} > 0, \quad \varepsilon > 0 \tag{3}$$

$$\dot{r} = \begin{cases} \alpha(\pi - \bar{\pi}) + \beta(Y - \bar{Y}) & if\ r > 0 \\ \max[0, \alpha(\pi - \bar{\pi}) + \beta(Y - \bar{Y})] & if\ r = 0 \end{cases} \tag{4}$$

$$\dot{\pi}^e = \gamma[\theta(\bar{\pi} - \pi^e) + (1 - \theta)(\pi - \pi^e)];\ \gamma > 0, \quad 0 \le \theta \le 1, \tag{5}$$

where the meaning of the symbols is as follows. Y—real national income (real output) > 0, $\bar{Y}$—'natural' output level corresponding to the natural rate of unemployment (fixed) > 0, G—real government expenditure (fixed)> 0, τ—marginal tax rate (fixed, $0 < \tau < 1$), M—nominal money supply > 0, p—price level > 0, $\pi = \frac{\dot{p}}{p}$—rate of inflation, π^e—expected rate of inflation, $\bar{\pi}$—target rate of inflation, r—nominal rate of interest ≥ 0, $r - \pi^e$—expected real rate of interest.

Equation (1) is the reduced form of the IS equation, which corresponds to the equilibrium condition of the goods market. In this formulation, capital accumulation effect is neglected, so that this is a 'short run' model in the sense of Keynes (1936).

Equation (2) is the LM equation, which corresponds to the equilibrium condition of the money market. It is convenient to rewrite this equation as follows. Differentiating this equation with respect to time, we get the following equivalent 'dynamic' expression of the LM equation

$$\mu = \pi + \eta_y \frac{\dot{Y}}{Y} - \eta_r \frac{\dot{r}}{r} - \eta_\pi \frac{\dot{\pi}^e}{\pi^e};\ \mu = \frac{\dot{M}}{M}, \tag{6}$$

where $\eta_y = \frac{(\frac{\partial L}{\partial Y})}{(\frac{L}{Y})} > 0$, $\eta_r = -\frac{(\frac{\partial L}{\partial r})}{(\frac{L}{r})} > 0$, and $\eta_\pi = -\frac{(\frac{\partial L}{\partial \pi^e})}{(\frac{L}{\pi^e})} > 0$ are elasticities of the real money demand with respect to changes of the real national income, nominal rate of interest and the expected rate of inflation, respectively.

Equation (3) is the quite conventional 'expectation-augmented Philips curve'.

Equation (4) describes the monetary policy rule of the central bank in the spirit of the 'Taylor rule'. It is assumed that the central bank chooses the nominal rate of interest r as a policy variable, and the central bank raises or reduces r according to the predetermined policy rule that is specified by Eq. (4). In this equation, the 'nonnegative constraint', which means that the nominal rate of interest cannot become negative, is explicitly considered, and it is assumed that two policy parameters α and β are positive. We can consider that this monetary policy rule means a kind of 'inflation targeting' as well as 'employment targeting', because it means that the central bank aims at the realization of the 'target rate of inflation' $\bar{\pi}$ that is announced by the central bank as well as the realization of the 'natural level of output' $\bar{Y}$.

Equation (5) is a formalization of the inflation expectation formation, which is the mixture of the 'forward looking' and 'backward looking' or 'adaptive' expectations. If the public strongly believes that the actual rate of inflation will be governed by the target rate of inflation that is announced by the central bank in the long run, we shall have in a limit case $\theta = 1$. On the other hand, if the public does not believe the announcement by the central bank at all or the central bank does not announce the target rate of inflation, then we shall have as a limit case $\theta = 0$. Hence, we can consider that the value of the parameter θ reflects the 'degree of credibility' of the central bank's announcement.

Substituting Eqs. (1) and (3) into Eqs. (4)–(6) we receive

$$\dot{r} = \begin{cases} f_1(r, \pi^e; \alpha, \beta, \varepsilon, G, \tau) & if\ r > 0 \\ \max[0,\ f_1(r, \pi^e; \alpha, \beta, \varepsilon, G, \tau)] & if\ r = 0 \end{cases} \tag{7}$$

$$\dot{\pi}^e = f_2(r, \pi^e; \gamma, \theta, \varepsilon, G, \tau) \tag{8}$$

$$\mu = \varepsilon\left[Y\left(r - \pi^e, G, \tau\right) - \bar{Y}\right] + \pi^e + \eta_y \frac{Y_{r-\pi^e}\left(\dot{r} - \dot{\pi}^e\right)}{Y(r - \pi^e, G, \tau)} - \eta_r \frac{\dot{r}}{r} - \eta_\pi \frac{\dot{\pi}^e}{\pi^e} \tag{9}$$

where

$$\begin{aligned} f_1\left(r, \pi^e; \alpha, \beta, \varepsilon, G, \tau\right) &= \alpha\left\{\varepsilon\left[Y\left(r - \pi^e, G, \tau\right) - \bar{Y}\right] + \pi^e - \bar{\pi}\right\} \\ &\quad + \beta\left[Y\left(r - \pi^e, G, \tau\right) - \bar{Y}\right], \\ f_2\left(r, \pi^e; \gamma, \theta, \varepsilon, G, \tau\right) &= \gamma\left\{\theta\left(\bar{\pi} - \pi^e\right) + (1 - \theta)\,\varepsilon\left[Y\left(r - \pi^e, G, \tau\right) - \bar{Y}\right]\right\}. \end{aligned}$$

The system of Eqs. (7)–(9) determines the dynamics of three variables (r, π^e, μ). We can see that this system is a decomposable system in the sense that the dynamics of r and π^e, which is determined by Eqs. (7) and (8), does not depend on equation (9).

Equation (9) only plays the role to determine the growth rate of money supply μ. Therefore, we need only to analyze the two-dimensional system of Eqs. (7) and (8).

The normal equilibrium point $E = (r^*, \pi^{e*})$ of this system is determined by the relations $\dot{r} = 0$, $\dot{\pi} = 0$, $Y = \bar{Y}$. If we neglect the nonnegative constraint of r, then from the structure of the functions $f_1(r, \pi^e; \alpha, \beta, \varepsilon, G, \tau)$ and $f_2(r, \pi^e; \gamma, \theta, \varepsilon, G, \tau)$ we receive

$$Y\left(r^* - \bar{\pi}, G, \tau\right) = \bar{Y}, \tag{10}$$

$$\pi^{e*} = \pi^* = \bar{\pi}. \tag{11}$$

Substituting (10) and (11) into Eq. (9) and considering $\dot{r} = 0$, $\dot{\pi}^e = 0$, we obtain the equilibrium value of μ

$$\mu^* = \bar{\pi}. \tag{12}$$

Equation (10) means that the 'natural' output level is realized at the normal equilibrium point. Equation (11) means that the expected rate of inflation is realized and the realized rate of inflation is equal to the target rate of inflation at the normal equilibrium point. Equation (12) implies that the growth rate of the nominal money supply at the normal equilibrium point is equal to the target rate of inflation. This means that the target rate of inflation that is set by the central bank determines the equilibrium growth rate of money supply and not the other way round in this model.

The nominal rate of interest at the normal equilibrium point r^* is determined as follows. First, the equilibrium real rate of interest ρ^* is determined by the equation $Y(\rho^*, G, \tau) = \bar{Y}$. Solving this equation with respect to ρ^*, we have

$$\rho^* = \rho^*(G, \tau); \frac{\partial \rho^*}{\partial G} > 0, \frac{\partial \rho^*}{\partial \tau} < 0. \tag{13}$$

Then, we have

$$r^* = \rho^*(G, \tau) + \bar{\pi}. \tag{14}$$

If r^* in (14) is negative, the economically meaningful normal equilibrium point does not exist. The condition $r^* > 0$ is equivalent to the condition

$$\bar{\pi} > -\rho^*(G, \tau). \tag{15}$$

If G is sufficiently small and/or τ is sufficiently large, the equilibrium real rate of interest ρ^* may become negative. In this case, the inequality (15) may not be satisfied so that the normal equilibrium point need not exist if the target rate of inflation $\bar{\pi}$ is not sufficiently large. Needless to say, the deflationary biased central bank that selects non-positive $\bar{\pi}$ can easily fail to satisfy the inequality (15). From now on, we assume that the inequality (15) is satisfied so that the equilibrium nominal rate of interest r^* is positive.

3 Dynamic Stability of the Normal Equilibrium Point

For further investigation it is suitable to transform the equilibrium $E = (r^*, \pi^{e*})$ of system of Eqs. (7) and (8) into the origin $E_0 = \left(x_1^* = 0, x_2^* = 0\right)$ by the shifting $x_1 = r - r^*$, $x_2 = \pi^e - \pi^{e*}$. We receive system

$$
\begin{aligned}
\dot{x}_1 &= \alpha\{\varepsilon[Y(x_1 - x_2 + r^* - \pi^{e*}, G, \tau) - \bar{Y}] + x_2 + \pi^{e*} - \bar{\pi}\} \\
&\quad + \beta\left[Y\left(x_1 - x_2 + r^* - \pi^{e*}, G, \tau\right) - \bar{Y}\right] \equiv F_1\left(x_1, x_2; \alpha, \beta, \varepsilon\right) \\
x_2 &= \gamma\{\theta(\bar{\pi} - x_2 - \pi^{e*}) + (1 - \theta)\varepsilon[Y(x_1 - x_2 + r^* - \pi^{e*}, G, \tau) - \bar{Y}]\} \\
&\equiv F_2(x_1, x_2; \gamma, \varepsilon, \theta)
\end{aligned}
\tag{16}
$$

Further on we investigate the impact of the changes in the values of the credibility parameter θ on the behavior of solutions of system (16) in a neighborhood of its equilibrium point $E_0 = (0, 0)$. The Jacobian matrix of system (16) at the equilibrium E_0 is

$$
J\left(E_0, \theta\right) = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}, \tag{17}
$$

where

$$
\begin{aligned}
F_{11} &= \frac{\partial F_1}{\partial x_1} = \alpha\varepsilon Y_{r-\pi^e} + \beta Y_{r-\pi^e} = \left(\alpha\varepsilon + \beta\right) Y_{r-\pi^e} < 0, \\
F_{12} &= \frac{\partial F_1}{\partial x_2} = -\alpha\varepsilon Y_{r-\pi^e} + \alpha - \beta Y_{r-\pi^e} = -\left(\alpha\varepsilon + \beta\right) Y_{r-\pi^e} + \alpha > 0, \\
F_{21} &= \frac{\partial F_2}{\partial x_1} = \gamma\left(1 - \theta\right)\varepsilon Y_{r-\pi^e} < 0, \\
F_{22} &= \frac{\partial F_2}{\partial x_2} = \gamma\left[-\theta - \left(1 - \theta\right)\varepsilon Y_{r-\pi^e}\right].
\end{aligned}
$$

The eigenvalues of Jacobian (17) are the roots of its characteristic equation $\lambda^2 - TrJ\lambda + \det J = 0$, which are given by the formula $\lambda_{1,2} = \frac{TrJ \pm \sqrt{(TrJ)^2 - 4\det J}}{2}$, where

$$
TrJ = \left[\alpha\varepsilon + \beta - \gamma\left(1 - \theta\right)\varepsilon\right] Y_{r-\pi^e} - \gamma\theta, \quad \det J = -\gamma\left(\theta\beta + \alpha\varepsilon\right) Y_{r-\pi^e} > 0.
$$

Conditions for pure imaginary eigenvalues of (17) are:

1. $TrJ = \left[\alpha\varepsilon + \beta - \gamma\left(1 - \theta\right)\varepsilon\right] Y_{r-\pi^e} - \gamma\theta = 0. \quad (18)$

2. $\det J > 0$.

From (18) we receive

$$
\beta = \left(\frac{1}{Y_{r-\pi^e}} - \varepsilon\right)\gamma\theta + \left(\gamma - \alpha\right)\varepsilon. \tag{19}
$$

Expression (19) is the equation of a line with respect to parameter variables θ and β. Its intersection $\bar{\theta}$ with θ-axis is given by the formula

$$\bar{\theta} = \frac{(\gamma - \alpha)}{\gamma} \frac{-\varepsilon Y_{r-\pi^e}}{1 - \varepsilon Y_{r-\pi^e}} < 1.$$

We recall that all parameters $\alpha, \beta, \gamma, \varepsilon$, and θ in the model are considered to be positive. As the determinant of the Jacobian is always positive, we get that the real parts of the eigenvalues $\lambda_{1,2}$ of Jacobian (17) are negative for the values of the parameter β lying above line (17), and positive for lying below this line. Therefore we can state, taking into account that the qualitative properties of the solutions are not changed by the shift of the equilibrium$E = (r^*, \pi^{e*})$ into the origin, the following theorem on the stability of the equilibrium.

Theorem 1 *1. If* $0 < \alpha < \gamma$ *and* $0 < \theta < \bar{\theta}, \bar{\theta} = \frac{(\gamma-\alpha)}{\gamma} \frac{-\varepsilon Y_{r-\pi^e}}{1-\varepsilon Y_{r-\pi^e}}$, *then the equilibrium* $E = (r^*, \pi^{e*})$ *is*

(a) asymptotically stable for $\beta > \left(\frac{1}{Y_{r-\pi^e}} - \varepsilon\right)\gamma\theta + (\gamma - \alpha)\varepsilon$,

(b) unstable for $\beta < \left(\frac{1}{Y_{r-\pi^e}} - \varepsilon\right)\gamma\theta + (\gamma - \alpha)\varepsilon$.

2. If $0 < \alpha < \gamma$ *and* $\bar{\theta} \leq \theta < 1$, *then the equilibrium* $E = (r^*, \pi^{e*})$ *is asymptotically stable for all* β *and* ε.

3. If $\gamma \leq \alpha$, *then the equilibrium* $E = (r^*, \pi^{e*})$ *is asymptotically stable for all* β, ε *and*

$$0 < \theta < 1.$$

4 Existence of the Cycles Around the Equilibrium Point

At the points (θ, β) lying on the segment

$$\beta = \left(\frac{1}{Y_{r-\pi^e}} - \varepsilon\right)\gamma\theta + (\gamma - \alpha)\varepsilon, \quad 0 < \theta < \bar{\theta},$$
$$\bar{\theta} = \frac{(\gamma - \alpha)}{\gamma} \frac{-\varepsilon Y_{r-\pi^e}}{1 - \varepsilon Y_{r-\pi^e}} < 1, \quad \gamma - \alpha > 0$$

there is $TrJ = [\alpha\varepsilon + \beta - \gamma(1-\theta)\varepsilon] Y_{r-\pi^e} - \gamma\theta = 0$ and $\det J = -\gamma(\theta\beta + \alpha\varepsilon) Y_{r-\pi^e} > 0$.

Consider an arbitrary credibility parameter $\theta = \theta_0$, $0 < \theta_0 < \bar{\theta}$. Then at the pair (θ_0, β_0), $\beta_0 = \left(\frac{1}{Y_{r-\pi^e}} - \varepsilon\right)\gamma\theta_0 + (\gamma - \alpha)\varepsilon$ there is $\lambda_{1,2}(\theta_0) = \pm i\omega_0(\theta_0)$. We call this pair the critical pair of Jacobian (17). Let us fix parameters $\alpha, \gamma, \varepsilon$ and the critical value β_0 which corresponds to the chosen θ_0. Further on we shall investigate the properties of the solutions of model (16) with respect to parameter θ from a small neighborhood of the critical parameter θ_0. For this purpose it is suitable to shift θ_0 in model (16) to the origin by shifting $\tilde{\theta} = \theta - \theta_0$. We receive

$$\begin{aligned}
\dot{x}_1 &= \alpha\{\varepsilon[Y(x_1 - x_2 + r^* - \pi^{e*}, G, \tau)] + x_2 + \pi^{e*} - \bar{\pi}\} \\
&\quad + \beta\left[Y\left(x_1 - x_2 + r^* - \pi^{e*}, G, \tau\right) - \bar{Y}\right] \equiv X_1(x_1, x_2; \alpha, \beta_0, \varepsilon) \\
\dot{x}_2 &= \gamma\{(\tilde{\theta} + \theta_0)(\bar{\pi} - x_2 - \pi^{e*}) + (1 - \tilde{\theta} - \theta_0)\varepsilon[Y(x_1 - x_2 + r^* - \pi^{e*}, G, \tau - \bar{Y}]\} \\
&\equiv X_2\left(x_1, x_2; \gamma, \tilde{\theta}, \varepsilon\right).
\end{aligned} \tag{20}$$

In shorten form system (20) can be written as

$$\dot{x} = X(x, \tilde{\theta}), \qquad x = (x_1, x_2). \tag{21}$$

The properties of system (21):

1. $X(0, \tilde{\theta}) = 0$.
2. The eigenvalues of Jacobian of (21) at $x = 0, \tilde{\theta} = 0$ are $\lambda_{1,2} = \pm i\omega_0$, and at $\tilde{\theta}$ from a small neighborhood of $\tilde{\theta} = 0$ are $\lambda_{1,2}(\tilde{\theta}) = \delta(\tilde{\theta}) \pm i\omega(\tilde{\theta})$, $\delta(0) = 0$, $\omega(0) = \omega_0$.

Performing Taylor expansion of model (21) at the equilibrium $E_0 = (0, 0)$ with respect to x we get

$$\dot{x} = J(E_0; \tilde{\theta})x + H(\alpha, \tilde{\theta}), \tag{22}$$

where

$$J(E_0; \tilde{\theta}) = \begin{pmatrix} (\alpha\varepsilon + \beta_0) Y_{r-\pi^e} & -(\alpha\varepsilon + \beta_0) Y_{r-\pi^e} + \alpha \\ \gamma(1 - \tilde{\theta} - \theta_0)\varepsilon\pi^e & \gamma\left[-\tilde{\theta} - \theta_0 - (1 - \tilde{\theta} - \theta_0)\varepsilon Y_{r-\pi^e}\right] \end{pmatrix},$$

$$H(x, \tilde{\theta}) = \begin{pmatrix} H_1(x_1, x_2, \tilde{\theta}) \\ H_2(x_1, x_2, \tilde{\theta}) \end{pmatrix},$$

$$
\begin{aligned}
H_1(x_1, x_2, \tilde{\theta}) = &\frac{1}{2!}\left(a^{(2,0)}x_1^2 + 2a^{(1,1)}x_1x_2 + a^{(0,2)}x_2^2\right)\\
&+\frac{1}{3!}\left(a^{(3,0)}x_1^3 + 3a^{(2,1)}x_1^2x_2 + 3a^{(1,2)}x_1x_2^2 + a^{(0,3)}x_2^3\right)\\
&+\frac{1}{4!}\left(a^{(4,0)}x_1^4 + 4a^{(3,1)}x_1^3x_2 + 6a^{(2,2)}x_1^2x_2^2 + 4a^{(1,3)}x_1x_2^3 + a^{(0,4)}x_2^4\right)\\
&+O\left(|\, x\,|^5\right)\\
a^{(p,q)} = &(-1)^q\,(\alpha\varepsilon + \beta)\,\frac{\partial^k Y\left(r^* - \pi^{e*}, G, \tilde{\theta}\right)}{\partial\,(r - \pi^e)^k}, \quad k = p + q,
\end{aligned}
$$

and

$$
\begin{aligned}
H_2(x_1, x_2, \tilde{\theta}) = &\frac{1}{2!}\left(b^{(2,0)}x_1^2 + 2b^{(1,1)}x_1x_2 + b^{(0,2)}x_2^2\right)\\
&+\frac{1}{3!}\left(b^{(3,0)}x_1^3 + 3b^{(2,1)}x_1^2x_2 + 3b^{(1,2)}x_1x_2^2 + b^{(0,3)}x_2^3\right)\\
&+\frac{1}{4!}\left(b^{(4,0)}x_1^4 + 4b^{(3,1)}x_1^3x_2 + 6b^{(2,2)}x_1^2x_2^2 + 4b^{(1,3)}x_1x_2^3 + b^{(0,4)}x_2^4\right)\\
&+O\left(|\, x\,|^5\right),\\
b^{(p,q)} = &(-1)^q\,\gamma(1 - \tilde{\theta} - \theta_0)\varepsilon\frac{\partial^k Y(r^* - \pi^{e*}, G, \tilde{\theta})}{\partial\,(r - \pi^e)^k}, \quad k = p + q.
\end{aligned}
$$

Denoting $A = -\left(\alpha\varepsilon + \beta_0\right) Y_{r-\pi^e}$, $B = \gamma(1 - \tilde{\theta} - \theta_0)\varepsilon Y_{r-\pi^e}$, then Jacobian $J(E_0; \tilde{\theta})$ in (22) can be written in the form

$$
J(E_0; \tilde{\theta}) = \begin{pmatrix} -A & A + \alpha \\ B & -B - \gamma\,(\tau + \theta_0) \end{pmatrix}.
$$

Consider now the matrix of the eigenvectors $\vec{u_1}, \vec{u_2}$ of the Jacobian $J(E_0; \tilde{\theta})$ in the form

$$
M = (\vec{u_1}, \vec{u_2}) = \begin{pmatrix} -(\alpha\varepsilon + \beta_0)Y_{r-\pi^e} + \alpha & -(\alpha\varepsilon + \beta_0)Y_{r-\pi^e} + \alpha \\ -(\alpha\varepsilon + \beta_0)Y_{r-\pi^e} + \lambda_1 & -(\alpha\varepsilon + \beta_0)Y_{r-\pi^e} + \lambda_2 \end{pmatrix} = \begin{pmatrix} A + \alpha & A + \alpha \\ A + \lambda_1 & A + \lambda_2 \end{pmatrix},
$$

and its inverse matrix

$$
M^{-1} = \frac{1}{-2\mathrm{i}\,(A + \alpha)\,\omega(\tilde{\theta})}\begin{pmatrix} A + \lambda_2 & -A - \alpha \\ -A - \lambda_1 & A + \alpha \end{pmatrix}.
$$

In further considerations we utilize the real forms M_r and M_r^{-1} of the matrices M and M^{-1}. They are as follows:

$$
M_r = \begin{pmatrix} 2\,(A + \alpha) & 0 \\ 2A + 2\delta(\tilde{\theta}) & -2\omega(\tilde{\theta}) \end{pmatrix}, \quad M_r^{-1} = \frac{1}{2\,(A + \alpha)\,\omega(\tilde{\theta})}\begin{pmatrix} \omega(\tilde{\theta}) & 0 \\ A + \delta(\tilde{\theta}) & -A - \alpha \end{pmatrix}.
$$

Perform now in (22) the transformation $x = M_r y$, $y = (y_1, y_2)$. We receive

$$\dot{y} = M_r^{-1}(0; \tilde{\theta}) M_r y + M_r^{-1} H(M_r y, \tilde{\theta}),$$

what gives the system

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} \delta(\tilde{\theta}) & -\omega(\tilde{\theta}) \\ \omega(\tilde{\theta}) & \delta(\tilde{\theta}) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \frac{1}{2\,(A+\alpha)\,\omega(\tilde{\theta})} \begin{pmatrix} \omega(\tilde{\theta}) K_1\,(y_1, y_2) \\ \left(A + \delta(\tilde{\theta})\right) K_1\,(y_1, y_2) - (A+\alpha)\, K_2\,(y_1, y_2) \end{pmatrix}, \tag{23}$$

where

$$\begin{aligned} K_1\,(y_1, y_2) = {} & \frac{1}{2!} \left\{ a^{(2,0)}\,[2\,(A+\alpha)\,y_1]^2 + 2a^{(1,1)}\,[2\,(A+\alpha)\,y_1] \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right] \right. \\ & \left. + a^{(0,2)} \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^2 \right\} + \frac{1}{3!} \left\{ a^{(3,0)}\,[2\,(A+\alpha)\,y_1]^3 \right. \\ & + 3a^{(2,1)}\,[2\,(A+\alpha)\,y_1]^2 \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right] + 3a^{(1,2)}\,[2\,(A+\alpha)\,y_1] \\ & \left. \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^2 + a^{(0,3)} \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^3 \right\} \\ & + \frac{1}{4!} \left\{ a^{(4,0)}\,[2\,(A+\alpha)\,y_1]^4 + 4a^{(3,1)}\,[2\,(A+\alpha)\,y_1]^3 \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right] \right. \\ & + 6a^{(2,2)}\,[2\,(A+\alpha)\,y_1]^2 \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^2 + 4a^{(1,3)}\,[2\,(A+\alpha)\,y_1] \\ & \left. \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^3 + a^{(0,4)} \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^4 \right\} + O\left(|\,y\,|^5\right), \end{aligned}$$

$$\begin{aligned} K_2\,(y_1, y_2) = {} & \frac{1}{2!} \left\{ b^{(2,0)}\,[2\,(A+\alpha)\,y_1]^2 + 2b^{(1,1)}\,[2\,(A+\alpha)\,y_1] \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right] \right. \\ & \left. + b^{(0,2)} \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^2 \right\} + \frac{1}{3!} \left\{ b^{(3,0)}\,[2\,(A+\alpha)\,y_1]^3 \right. \\ & + 3b^{(2,1)}\,[2\,(A+\alpha)\,y_1]^2 \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right] + 3b^{(1,2)}\,[2\,(A+\alpha)\,y_1] \\ & \left. \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^2 + b^{(0,3)} \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^3 \right\} \\ & + \frac{1}{4!} \left\{ b^{(4,0)}\,[2\,(A+\alpha)\,y_1]^4 + 4b^{(3,1)}\,[2\,(A+\alpha)\,y_1]^3 \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right] \right. \\ & + 6b^{(2,2)}\,[2\,(A+\alpha)\,y_1]^2 \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^2 + 4b^{(1,3)}\,[2\,(A+\alpha)\,y_1] \\ & \left. \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^3 + b^{(0,4)} \left[2\left(A+\delta(\tilde{\theta})\right) y_1 - 2\omega(\tilde{\theta}) y_2\right]^4 \right\} + O\left(|\,y\,|^5\right). \end{aligned}$$

The normal form of (23) in real domain (see Wiggins (1990) or Bibikov (1979)) is

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} \delta(\tilde{\theta}) & -\omega(\tilde{\theta}) \\ \omega(\tilde{\theta}) & \delta(\tilde{\theta}) \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} (a(\tilde{\theta}) y_1 - b(\tilde{\theta}) y_2)(y_1^2 + y_2^2) + O(|y|^5) \\ (b(\tilde{\theta}) y_1 - a(\tilde{\theta}) y_2)(y_1^2 + y_2^2) + O(|y|^5) \end{pmatrix}. \tag{24}$$

System (24) in polar coordinates $y_1 = ucos\phi$, $y_2 = u \sin\phi$ has the form

$$\begin{aligned} \dot{u} &= \delta(\tilde{\theta})u + a(\tilde{\theta})u^3 + O(u^5) \\ \dot{\phi} &= \omega(\tilde{\theta})u + b(\tilde{\theta})u^2 + O(u^4). \end{aligned} \tag{25}$$

Taylor expand of the coefficients in (25) at $\tilde{\theta} = 0$ gives

$$\begin{aligned} \dot{u} &= \frac{d\delta(0)}{d\tilde{\theta}}\tilde{\theta}u + a(0)u^3 + O(\mu^2 u, \tilde{\theta}u^3, u^5) \\ \dot{\phi} &= \omega(0) + \frac{d\omega(0)}{d\tilde{\theta}}\tilde{\theta} + b(0)u^2 + O(\tilde{\theta}^2, \tilde{\theta}u^2, u^4) \end{aligned} \tag{26}$$

The bifurcation equation of (26) is

$$\dot{u} = u(au^2 + b\tilde{\theta}), \quad a = a(0), \quad b = \frac{d\delta(0)}{d\tilde{\theta}}. \tag{27}$$

Denote

$$\begin{aligned} h^1\left(y_1, y_2, \tilde{\theta}\right) &= \frac{1}{2\left(A+\alpha\right)\omega(\tilde{\theta})}\omega(\tilde{\theta})K_1\left(y_1, y_2\right) \\ h^2(y_1, y_2) &= \frac{1}{2\left(A+\alpha\right)\omega(\tilde{\theta})}\left[\left(A+\delta(\tilde{\theta})\right)K_1(y_1, y_2) - (A+\alpha)K_2(y_1, y_2)\right]. \end{aligned}$$

Then according to Wiggins (1990) the first Lyapunov coefficient $a = a\,(0)$ in bifurcation equation (27) is defined by the formula

$$\begin{aligned} a = &\frac{1}{16\omega_0}\left[h^1_{y_1y_2}\left(h^1_{y_1y_1} + h^1_{y_2y_2}\right) - h^2_{y_1y_2}\left(h^2_{y_1y_1} + h^2_{y_2y_2}\right) - h^1_{y_1y_1}h^2_{y_1y_1} + h^1_{y_2y_2}h^2_{y_2y_2}\right] \\ &+ \frac{1}{16}\left(h^1_{y_1y_1y_1} + h^1_{y_1y_2y_2} + h^2_{y_1y_1y_2} + h^2_{y_2y_2y_2}\right). \end{aligned}$$

In our case the coefficient $b = \frac{d\delta(0)}{d\tilde{\theta}}$ in the bifurcation equation is negative, as $\delta\left(\tilde{\theta}\right) = TrJ\left(E_0, \tilde{\theta}\right) = (\alpha\varepsilon + \beta_0)\,Y_{r-\pi^e} + \gamma\left[-\tau - \theta_0 - \left(1 - \tilde{\theta} - \theta_0\right)\varepsilon Y_{r-\pi^e}\right]$ and

$$\frac{d\delta\left(\mu\right)}{d\tilde{\theta}} = -\gamma + \gamma\varepsilon Y_{r-\pi^e} < 0. \tag{28}$$

On the base of (28) and the results from the theory on Hopf bifurcation (see for example Wiggins (1990)) we can formulate

Theorem 2 *Let for the coefficient a in the bifurcation equation (27) hold:*

1. *if $a < 0$, then there exists a stable limit cycle for every small enough $\tilde{\theta} < 0$.*
2. *if $a > 0$, then there exists an unstable limit cycle for every small enough $\tilde{\theta} > 0$.*

5 Numerical Simulations

Consider model (7)–(8) constructed on the base of the following functions and values: $Y = C + I + G, C = c(Y - T) + C_0, T = \tau Y - T_0, I = \frac{\kappa}{1+e^{r-\pi^e}}$, $Y(r - \pi^e, G, \tau) = \frac{1}{1-c(1-\tau)}\frac{\kappa}{1+e^{r-\pi^e}} + \frac{G+C_0+cT_0}{1-c(1-\tau)}$, $\alpha = \frac{1}{6}, \gamma = 10, \varepsilon = 0.2, c = 0.8, \tau = 0.4$, $G = 40, C_0 = 2.4, T_0 = 2, \bar{\pi} = 0.02, \bar{Y} = 100, \kappa = 8\left(e^{0.03} + 1\right)$.

The equilibrium of this model is $E = (r^* = 0.05, \pi^{e*} = 0.02)$ and the intersection $\bar{\theta}$ of the line $\beta = \left(\frac{1}{Y_{r-\pi^e}} - \varepsilon\right)\gamma\theta + (\gamma - \alpha)\varepsilon$ with θ-axis is $\bar{\theta} = 0.599$. In the following part there are depicted three pairs of solutions of the model with different values of the critical parameters (θ_0, β_0) and the credibility parameter θ. In these

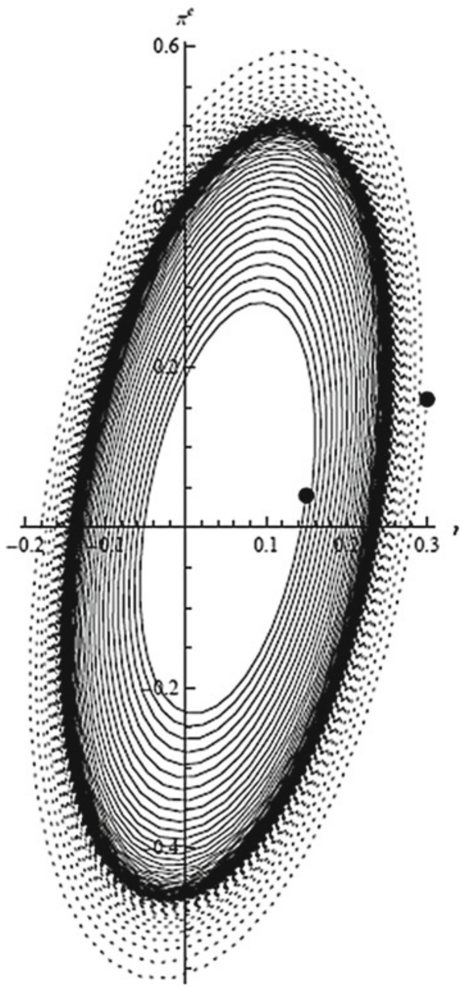


Fig. 1 Without nonnegative constraint: $\bar{\theta} \doteq 0.599$, $\theta_0 = 0.95\,\bar{\theta}$, $\theta = 0.995\theta_0$, initial values $\left(r_0, \pi_0^e\right)$ of solutions: *solid* $\left(r_0 = 3r^*, \pi_0^e = 2\pi^{e*}\right)$, *dotted* $\left(r_0 = 6r^*, \pi_0^e = 8\pi^{e*}\right)$

pairs the first figure depicts solutions without nonnegative constraint, and the second one depicts solutions with nonnegative constraint.

Figures 1 and 2 correspond to the critical pair ($\theta_0 = 0.569$, $\beta_0 = 0.098$) with the value of the credibility parameter $\theta = 0.995\theta_0$ what gives the approximate value $\theta \doteq 0.567$. In Fig. 1 there are depicted two solutions of the model without its nonnegative constraint. We see that their r-components take also negative values. In Fig. 2 there are depicted three solutions of this model with its nonnegative constraint. We see that this constraint controls their courses in the way that their r-components do not

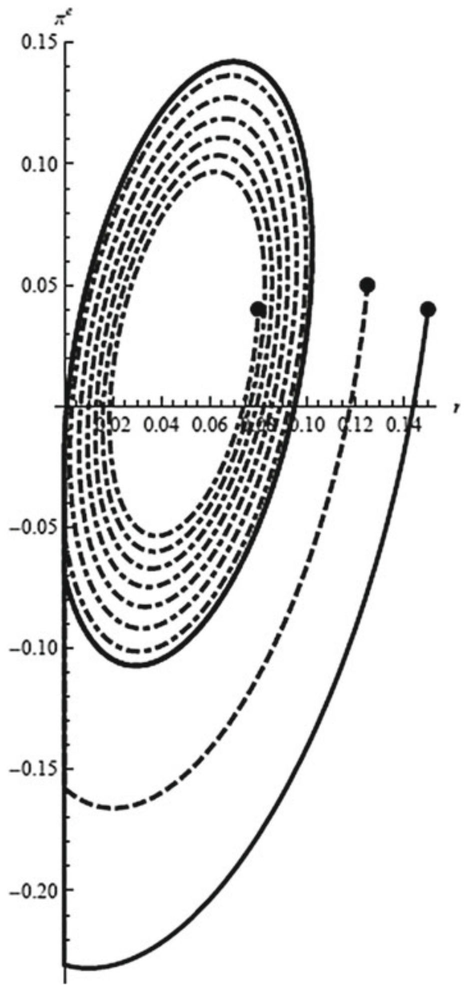


Fig. 2 With nonnegative constraint: $\bar{\theta} \doteq 0.599$, $\theta_0 = 0.95\,\bar{\theta}$, $\theta = 0.995\theta_0$, initial values $\left(r_0, \pi_0^e\right)$ of solutions: *dash-dotted* $\left(r_0 = 1.6r^*, \pi_0^e = 2\pi^{e*}\right)$, *dashed* $\left(r_0 = 2.5r^*, \pi_0^e = 2.5\pi^{e*}\right)$, *solid* $\left(r_0 = 3r^*, \pi_0^e = 2\pi^{e*}\right)$

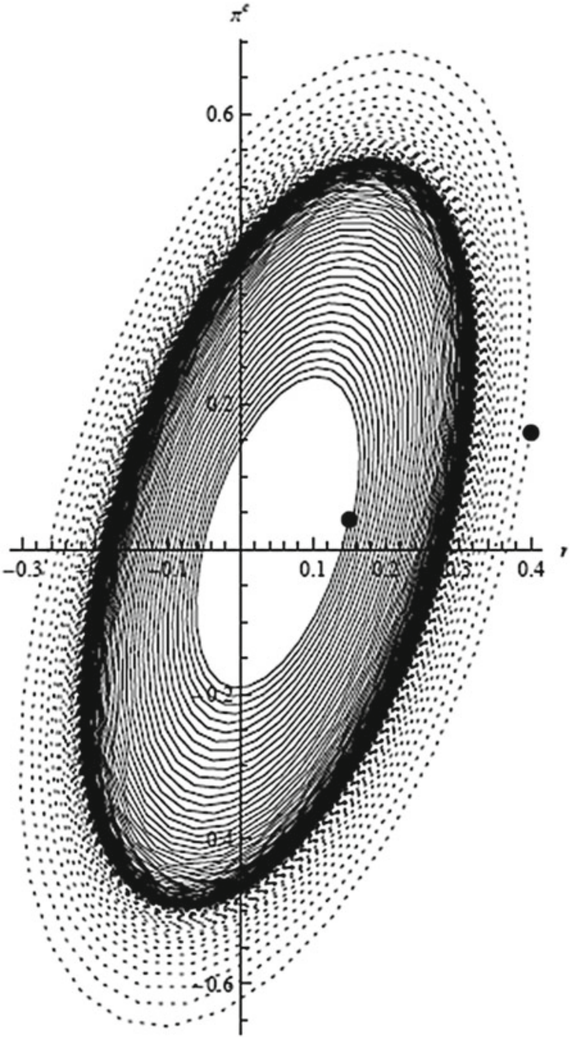


Fig. 3 Without nonnegative constraint: $\bar{\theta} \doteq 0.599$, $\theta_0 = 0.9\,\bar{\theta}$, $\theta = 0.995\theta_0$, initial values $\left(r_0, \pi_0^e\right)$ of solutions: *solid* $\left(r_0 = 3r^*, \pi_0^e = 2\pi^{e*}\right)$, *dotted* $\left(r_0 = 8r^*, \pi_0^e = 8\pi^{e*}\right)$

take negative values, and their courses gradually form a common cycle with different time shifting.

Figures 3 and 4 correspond to the critical pair ($\theta_0 = 0.539$, $\beta_0 = 0.197$) with the value of the credibility parameter $\theta = 0.995\theta_0$, what gives the approximate value $\theta \doteq 0.537$. This value is a little smaller than that one in Fig. 1. In Fig. 3 there are depicted two solutions of the model without its nonnegative constraint. We see that their r-components take also negative values. In Fig. 4 there are depicted three solutions of

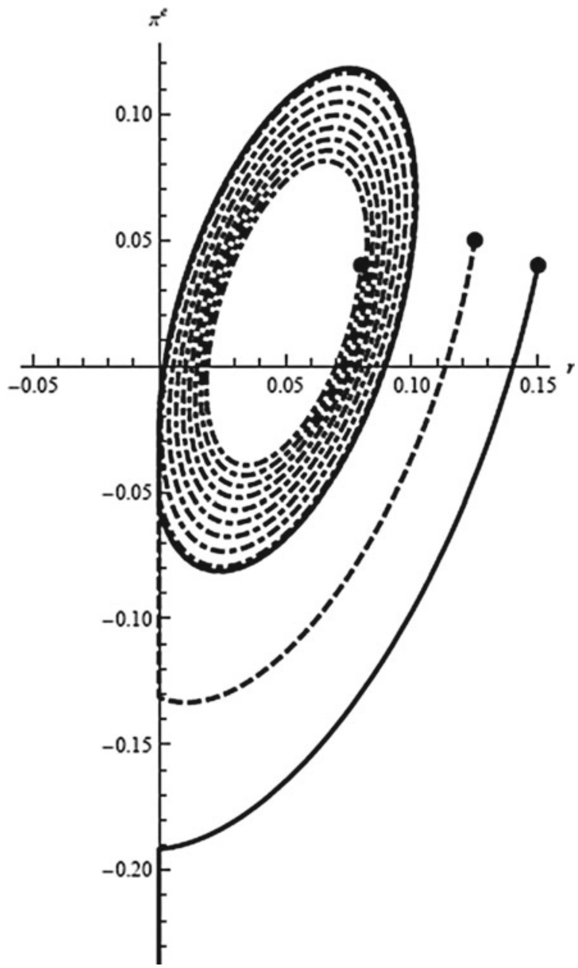


Fig. 4 With nonnegative constraint: $\bar{\theta} \doteq 0.599$, $\theta_0 = 0.9\,\bar{\theta}$, $\theta = 0.995\theta_0$, initial values (r_0, π_0^e) of solutions: *dash-dotted* $(r_0 = 1.6r^*, \pi_0^e = 2\pi^{e*})$, *dashed* $(r_0 = 2.5r^*, \pi_0^e = 2.5\pi^{e*})$, *solid* $(r_0 = 3r^*, \pi_0^e = 2\pi^{e*})$

this model with its nonnegative constraint, having the same initial conditions as the solutions in Fig. 2. We see that the dot-dashed and dashed solutions gradually enter into a common cycle but the solid solution by contrast to the solid solution in Fig. 2 goes to the depression of the expected rate of inflation.

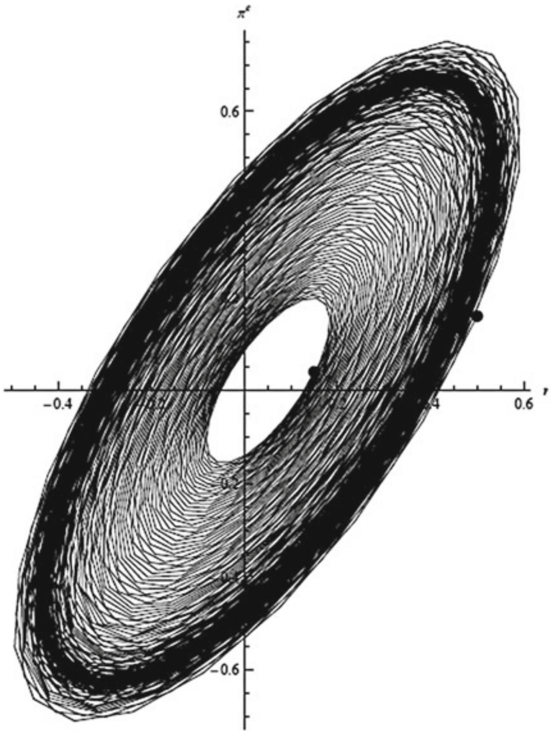


Fig. 5 Without nonnegative constraint: $\bar{\theta} \doteq 0.599$, $\theta_0 = 0.7\,\bar{\theta}$, $\theta = 0.995\theta_0$, initial values $\left(r_0, \pi_0^e\right)$ of solutions: *solid* $\left(r_0 = 3r^*, \pi_0^e = 2\pi^{e*}\right)$, *dotted* $\left(r_0 = 10r^*, \pi_0^e = 8\pi^{e*}\right)$

Figures 5 and 6 correspond to the critical pair ($\theta_0 = 0.420$, $\beta_0 = 0.590$) with the value of the credibility parameter $\theta = 0.995\theta_0$, what gives the approximate value $\theta \doteq 0.418$ which is smaller comparing it with that one in Fig. 3. In Fig. 5 there are depicted two solutions of the model without its nonnegative constraint. We see that their r-components take also negative values. In Fig. 6 there are depicted three solutions of this model with its nonnegative constraint, having the same initial conditions as the solutions in Figs. 2 and 4. We see that the dot-dashed solution gradually form a cycle but the dashed and solid solutions go to the depression of the expected rate of inflation.

Figures 2, 4 and 6 show that the domain of the 'stability' around the equilibrium point $E = (r^*, \pi^{e*})$ of the model which guarantees that solutions starting inside it will not go into the depression of expected rate of inflation is getting smaller with the decrease of the value of the credibility parameter θ.

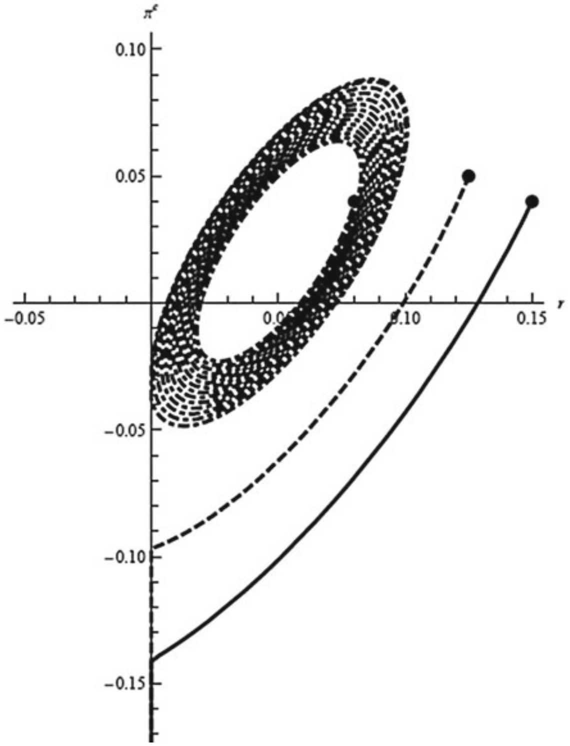


Fig. 6 With nonnegative constraint: $\bar{\theta} \doteq 0.599$, $\theta_0 = 0.7\,\bar{\theta}$, $\theta = 0.995\theta_0$, initial values (r_0, π_0^e) of solutions: *dash-dotted* $(r_0 = 1.6r^*, \pi_0^e = 2\pi^{e*})$, *dashed* $(r_0 = 2.5r^*, \pi_0^e = 2.5\pi^{e*})$, *solid* $(r_0 = 3r^*, \pi_0^e = 2\pi^{e*})$

6 Conclusion

In this paper, a simple two-dimensional Keynesian macrodynamic model of monetary economy describing the development of nominal interest rate and expected rate of inflation is analyzed. Theorem 1 solves the question of the stability of its normal equilibrium point giving conditions on the parameters of the model which guarantee its stability. This theorem implies the following results.

(1) The normal equilibrium point becomes locally stable if the monetary policy of the central bank is sufficiently active (at least one of the monetary policy parameters α or β is sufficiently large) and the central bank's inflation targeting is sufficiently credible (the credibility parameter θ is sufficiently close to 1).
(2) The normal equilibrium point becomes locally unstable if both of the monetary policy parameters α and β are sufficiently small and the credibility parameter θ is sufficiently small.

Theorem 2 gives conditions under which limit cycles around the normal equilibrium point can arise through Hopf bifurcation at the intermediate parameter values. In the presented example the limit cycle is stable. In three pairs of figures it is shown that the nonnegative constraint in the equation for the development of the nominal interest rate can prevent the interest rate to receive negative values during its course. In the presented examples, this nonnegative constraint has a destabilizing effect, and it can trigger the development of the deflationary depression even if the normal equilibrium point is stable in the system without nonnegative constraint. The figures at the same time show that the domain of the 'stability' around the equilibrium point which guarantees that solutions starting inside it will not go into the depression of the expected rate of inflation is getting smaller with the decrease of the value of the credibility parameter θ. These observations are consistent with the performance Japanese economy during the period from the 1990s to the mid 2010s.[2]

Needless to say, the two-dimensional dynamic model in this paper that is based on Asada (2011) is the simplest prototype dynamic model of the monetary policy in the spirit of Keynes (1936) economic theory. More advanced five-dimensional dynamic Keynesian model of monetary and fiscal policy mix with public debt accumulation is studied in Asada and Ouchi (2015).

Acknowledgments This publication was financially supported by the Project: Mobility—enhancing research, science and education at the Matej Bel University, ITMS code: 26110230082, under the Operational Program Education co-financed by the European Social Fund. This research was also financially supported by Chuo University Grant for Special Research, the Japan Society for the Promotion of Science (Grant-in Aid (C) 25380238), and the MEXT-Supported Program for the Strategic Research Foundation at Private University 2013–2017. Needless to say, however, only the authors are responsible for possible remaining errors.

References

Asada T (2011) Central banking and deflationary depression: a Japanese perspective. In: Cappello M, Rizzo C (eds) Central Banking and Globalization. Nova Science Publishers, New York, pp 91–114

Asada T (ed) (2015) The development of economics in Japan: from the inter-war period to the 2000s. Routledge, London

Asada T, Ouchi M (2015) Mathematical formalization of macroeconomic stabilization policy in a high-diensional dynamic model with public debt accumulation. Acta Matheatica Universitatis Comenianae, vol 84, no 2. Bratislava, Slovakia, pp 229–242

Bibikov YN (1979) Local theory of nonlinear analytic ordinary differential equations. Springer, Berlin

Keynes JM (1936) The general theory of employment, interest, and money. Macmillan, London

Krugman P (2012) End this depression now!. W. W. Norton, New York

Minsky HP (1982) Can "it" happen again? Essays on instability and finance. Armonk, New York M. E. Sharpe

Minsky HP (1986) Stabilizing an unstable economy. Yale University Press, New Haven

[2]See Asada (ed.) (2015) and Wakatabe (2015).

Wiggins S (1990) Introduction to applied nonlinear dynamical systems and chaos. Springer, New York

Wakatabe M (2015) Japan's great stagnation and abenomics: lessons for the world. Palgrave Macmillan, New York

Pathology in the Market Economy: Self-fulfilling Process to Chronic Slump

Chronic Slump

Akitaka Dohtani

Abstract In this chapter, we construct an extension of Goodwin's nonlinear accelerator model and detect a possible cause that generates a chronic slump. By introducing a nonlinearity expressing a pessimistic outlook for the future economy in our extended model, we demonstrate that a chronic slump cycle arises from the pessimistic outlook through a self-fulfilling prophecy. In the extended model, income on the cycle is locked in a domain lower than the market equilibrium. This implies that private spending in the model economy fluctuates and is continuously insufficient to make use of the available productive capacity that is estimated at the market equilibrium. The periodic attractor gives a partial description of the recent worldwide chronic slump. Our result shows that the extended Goodwin model provides a partial description of the Krugman's view that explains the recent worldwide slump. Moreover, although booms and slumps come in all sizes, our extended model explains how this is possible.

Keywords Chronic slump · Demand side · Self-fulfilling prophecy · Pessimistic outlook · Asymmetric adaptive expectation formation

1 Introduction

Recently, many countries have experienced slumps. The current depression is not as severe as the Great Depression. However, the recent worldwide slump is critical in the sense that it is chronic, and this chronicity implies a difficulty in recovery: the signs of a serious depression have been observed. Which mechanism is responsible

---

Dedication: This paper is dedicated to the memory of the late Dr. Tatsuji Owase.

A. Dohtani (✉)
Faculty of Economics, University of Toyama, 3190 Gofuku, Toyama 930, Japan
e-mail: doutani@eco.u-toyama.ac.jp

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_10

for generating the recent worldwide chronic slump? Following Krugman's view[1] on the chronic slump, we emphasize the importance of the demand side of the economy. We are interested in the dynamic demand-side model describing Krugman's view. Moreover, the Krugman's view states that the present state of private spending is continuously insufficient to make use of the available productive capacity.[2] Krugman (2008) also asserts that the dynamic notion of the *self-fulfilling prophecy*[3] plays an important role in explaining the chronicity of the recent slump. Regarding the self-fulfilling prophecy, Krugman (2008) focuses on the financial markets. However, we concentrate on the real markets and show that the Krugman's view on the self-fulfilling prophecy holds true for the real markets as well. In a downswing, the self-fulfilling prophecy will often occur in both markets repeatedly. From the perspective of demand-side macroeconomics, we construct a prototype dynamic model expressing a part of the Krugman's view.

The situation that we will describe by the prototype mode is as follows. We will demonstrate that the chronic slump results from a pessimistic outlook on the demand side. We suppose that the Knightian uncertainty[4] arises from a market's loss of confidence, and therefore, the pessimistic outlook spreads.[5] This pessimism often yields the self-fulfilling prophecy. The pessimistic outlook makes the economy inactive. As a result, the inactiveness makes the economic agents believe that the pessimistic outlook is appropriate, and the belief renders the economy even more inactive. This vicious circle (or the self-fulfilling prophecy) continues, triggering an economic avalanche, and a chronic slump emerges. Thus, the pessimistic outlook is a critical barrier to prosperity.

We here make one important remark. The market psychology may change through a "learning", and the pessimistic outlook may change. However, the Knightian uncertainty persists over a long period of time unless the market's loss of confidence is recovered. Consequently, the above vicious circle makes the economic agents believe firmly the validity of pessimistic outlook. Therefore, the pessimistic outlook also persists over a long period of time, and our precondition of argument is robust unless such a loss is recovered.

---

[1] See Krugman (2008, Chap. 10).

[2] Here, the productive capacity is estimated at the market equilibrium.

[3] For insightful arguments on the self-fulfilling prophesy, see Rosser (1991).

[4] Knightian uncertainty applies to situations where we cannot obtain enough information we need in order to set accurate odds. See Knight (1921). For an important relation between the Knightian uncertainty and market psychology, see also Akerlof and Shiller (2009, Chap. 11).

[5] The role of expectation in business cycles has been discussed by many economists from the Keynesian perspective. See, for example, Matthews (1959, Chap. 3.5). Economists have considered expectation to be fickle, and therefore, the corresponding argument lacks clarity. However, owing to the self-fulfilling prophecy, pessimism becomes inflexible and robust in the long run. Thus, the market psychology of pessimism can be considered as a theoretical subject. Akerlof and Shiller (2009) discuss importance of the market psychology from a much wider viewpoint. By using the Michigan Consumer Sentiment Index, Blanchard (1993) pointed out that the loss of confidence can cause a large economic recession.

We present the analytical details as follows. The prototype model constructed in this paper is based on the classical nonlinear-accelerator business cycle model of Goodwin (1951).[6] The Goodwin model is one of well-known demand-side models of the business cycle in the Keynesian tradition.[7] In the model, the sigmoid type of nonlinearity plays the most important role in generating persistent nonlinear fluctuations. Although the Goodwin model has been criticized for the lack of microfoundation, the Keynesian nonlinear business cycle models like Goodwin's nonlinear accelerator model are useful for explaining actual business cycles. We construct an extension of the Goodwin model and detect a possible cause of a chronic slump.[8] Unlike the Goodwin model, we assume that the household distinguishes between short-run and medium-run consumption plans. In the short-run plan, like the Goodwin model, the household determines its consumption depending linearly on its income. On the other hand, in the medium-run plan, the household determines its consumption in proportion to the expected income, which is adjusted by an adaptive expectation rule. We assume that in the case where the actual income is larger than the expected income, the slope of the adjustment function is smaller than that in the converse case. This implies that the household develops a pessimistic outlook for the future economy, and therefore, in the case where the actual income is larger than the expected income, the latter is adjusted merely by a smaller amount than that in the converse case. We show that the introduction of pessimistic adaptive learning into the Goodwin model does generate a chronic slump in which income and expected income are locked in lower domains than the market equilibrium.

All business cycle models in the Keynesian tradition describe a complete recovery from a slump. However, the economic process in an actual chronic slump is not monotonous in the sense that it gradually descends while repeating partial recoveries and slowdowns. In other words, even in the chronic slump, the market economy persistently fluctuates in a low domain of income. To describe this situation, we must construct a nonlinear business cycle model that possesses a periodic path on which income and expected income are lower than those at the market equilibrium. We show that the above extended Goodwin model has such a periodic path. Thus, the extended Goodwin model constructed will be a business cycle model that analytically expresses the above view held by Krugman on the recent chronic slump. Moreover, as stressed in Krugman (1996, p. 68), an important feature of business cycle is that booms and slumps come in all sizes. Slump cycles are a part of such a feature. The extended Goodwin model also gives a theoretical explanation of the feature.

---

[6] Many studies have examined the nonlinear dynamics of the original and extended versions. See, for example, Bothwell (1952), Strotz et al. (1953), Gabisch and Lorenz (1987), Krugman (1996), Owase (1991), and Puu (2003), and many papers on the Goodwin model in Puu and Sushko (2006).

[7] kaldor (1940) constructed another well-known and important nonlinear business cycle model in the Keynesian tradition. The mathematical formulation of the model is given by Chang and Smyth (1971). Our argument holds true for the business cycle model.

[8] For another interesting Keynesian approach, see Varian (1976) and George (1981). This approach employs the catastrophe theory. It also provides the important and useful information on a serious depression. For the catastrophe theory, see Rosser (1991, Chap. 6).

Many models with self-fulfilling features have been proposed.[9] A feature common to these models is that there exist multiple equilibria, which comprise higher and lower equilibria. However, we emphasize that the extended Goodwin model is quite different from these models in the sense that it possesses only a unique equilibrium and all its paths converge to a periodic attractor that is locked in a domain lower than the equilibrium point. From the perspective of Keynesian demand-side economics, we present a new kind of model possessing the self-fulfilling feature.

2 Extension of the Goodwin Model

For constructing the extended Goodwin model, the requirements we impose are as follows:

R.1 The business cycle model is a demand-side model.
R.2 There exists a small periodic path on which income and expected income are constantly lower than their levels at the market equilibrium.

R.1 is the first requirement for following the Krugman's view. In a chronic slump, the market economy does not possess the power of automatic recovery. In this sense, the slump is serious. Therefore, as stated in the Introduction, the economy constantly repeats partial recoveries and slowdowns. To describe such a situation, we require R.2.

R.2 may be stronger than needed. However, R.2 is a convenient requirement for clarifying the meaning of "partial recovery." As stated in the Introduction, R.2 also implies that private spending is continuously insufficient to make use of the available productive capacity.[10] Like the Goodwin model, many business cycle models in the Keynesian tradition possess the power of automatic recovery, and therefore, all the paths fluctuate around the equilibrium point. However, R.2 implies that the model does not possess any power of automatic recovery by itself, and therefore, the income and expected income are lower than their levels at the market equilibrium. The purpose of this section is to construct a prototype dynamic macromodel satisfying R.1 and R.2, which provides an extension of Goodwin's nonlinear business cycle model.

Before constructing the extended Goodwin model, we briefly explain the original nonlinear accelerator model[11] of Goodwin. Throughout this paper, we assume that all functions are continuous. Goodwin's original model is given by

$$\dot{y}_t = \mu\{(c_t + \dot{k}_t) - y_t\}, \tag{2.1a}$$

$$c_t = \alpha y_t + c_0, \text{ and} \tag{2.1b}$$

[9] For the models, see, for example, Krugman (1991) and Murphy et al. (1989).

[10] See also Footnote 2.

[11] It is well-known that the multiplier-accelerator principle plays an important role in explaining business cycles. See Blanchard (1981).

$$\dot{k}_{t+\theta} = \phi(\dot{y}_t), \tag{2.1c}$$

where c is consumption, y is national income, k is capital stock, $\alpha \in [0, 1)$ is the marginal propensity to consume, μ is the adjustment coefficient, and c_0 is a positive constant. A dot over a variable indicates a derivative with respect to time. Goodwin's original model is given by differential-difference equations. Following Goodwin (1951), we transform this system into a system of differential equations. Equations (2.1a) and (2.1b) yield

$$(1-\alpha)y_t = \dot{k}_t - (1/\mu)\dot{y}_t + c_0. \tag{2.2}$$

The linear approximation of $\dot{k}_{t+\theta}$ is given by $\dot{k}_{t+\theta} \approx \dot{k}_t + \theta\ddot{k}_t$. As is typical, using the linear approximation, we replace (2.1c) with

$$\theta\ddot{k}_t = \phi(\dot{y}_t) - \dot{k}_t. \tag{2.3}$$

Let us define

$$x_t = \dot{y}_t \quad \text{and} \quad w_t = y_t - c_0/(1-\alpha).$$

Then, $x_t = \dot{w}_t$. Moreover, Eq. (2.1b) yields $x_t = \mu\{\dot{k}_t - (1-\alpha)y_t + c_0\}$. Therefore, Eq. (2.3) yields

$$\dot{x}_t = \mu\{\ddot{k}_t - (1-\alpha)\dot{y}_t\} = \mu[\{\phi(x_t) - \dot{k}_t\}/\theta - (1-\alpha)x_t].$$

Equation (2.2) yields $\dot{k}_t = (1-\alpha)y_t + (1/\mu)\dot{y}_t - c_0 = (1-\alpha)w_t + (1/\mu)x_t$. Thus, we obtain the following two-dimensional system of differential equations:

$$\Theta_G : \begin{cases} \dot{x}_t = \frac{\mu}{\theta}[\phi(x_t) - \{(1/\mu) + (1-\alpha)\theta\}x_t - (1-\alpha)w_t], \\ \dot{w}_t = x_t. \end{cases}$$

We call this system the Goodwin model. Throughout this paper, the $\phi-$function is supposed to satisfy the following assumptions:

Assumption 1 $\phi(0) = 0$.

Assumption 2 The $\phi-$function is continuously differentiable.

Clearly, Assumption 1 shows that (0, 0) is the equilibrium point in the Goodwin model. On the other hand, Assumption 2 guarantees the existence and uniqueness of solutions in the Goodwin model.[12]

[12] See Guckenheimer and Holmes (1983, p. 3).

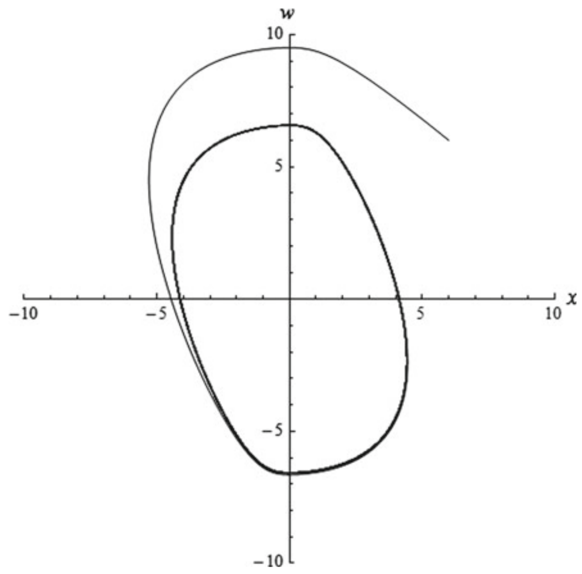


Fig. 1 Typical periodic attractor of the Goodwin model

Goodwin (1951) numerically showed that the ϕ−function of a sigmoid shape yields a limit cycle. System Θ_G is a system of autonomous differential equations of the Rayleigh type.[13] For the Rayleigh-type equation, many mathematical results exist.[14] Therefore, it is not difficult to prove the existence of a limit cycle in System Θ_G under suitable conditions. Since proving this is not the purpose of the present paper, we merely provide a numerical example, where System Θ_G possesses a limit cycle.

Numerical Example 1 We set $\phi(x) = 2\mathrm{Arctan}(1.5x)$, $\mu = 2$, $\alpha = 0.7$, and $\theta = 1$. Clearly, the ϕ−function satisfies Assumptions 1 and 2. The ϕ−function is of a typical sigmoid shape as in Goodwin (1951). Figure 1 shows that System Θ_G possesses a limit cycle. ■

We now extend the Goodwin model. To incorporate the pessimistic outlook for the future economy into the Goodwin model, we consider the long-run consumption plan of the household. Moreover, for this purpose, we replace (2.1b) with

$$c_t = \alpha y_t + \beta y_{et} + c_0, \tag{2.4}$$

where $\beta \in [0, 1]$ and y_e is the expected income. Equation (2.4) states that the consumption plan is decomposed into the short-run plan ($\alpha y_t + c_0$) and the long-run plan (βy_{et}). Throughout this paper, we assume the following:

[13] System Θ_G can also be transformed into a van der Pol-type equation. See Lorenz (1993, Subsection 5.3.2).

[14] See, for example, Sansone and Conti (1964) and Yanqian (1986).

Assumption 3 $1 > \alpha + \beta$.

As proved later, Assumption 3 is utilized to guarantee the existence and uniqueness of the equilibrium point. The expected income is adjusted by

$$\dot{y}_{et} = \psi(y_t - y_{et}). \tag{2.5}$$

Here, we assume the following:

Assumption 4 $\psi(u)u > 0$ for any $u \neq 0$.

We will later discuss the properties of the $\psi-$function that are closely related to the occurrence of a slump cycle. We call the $\psi-$function the adjustment function and Eq. (2.5) the adjustment equation. Given Assumption 3, we define

$$x_t = \dot{y}_t, \tag{2.6a}$$
$$w_t = y_t - c_0/(1-\alpha-\beta), \text{ and} \tag{2.6b}$$
$$z_t = y_{et} - c_0/(1-\alpha-\beta), \tag{2.6c}$$

where w_t denotes the deviation of income from the equilibrium and z_t denotes the deviation of expected income from the equilibrium. Then, $x_t = \dot{w}_t$. In the same way as before, Eq. (2.3) yields

$$\dot{x}_t = \frac{\mu}{\theta}[\phi(x_t) - \{(1/\mu) + (1-\alpha)\theta\}x_t - (1-\alpha)w_t + \beta z_t + \theta\beta\psi(w_t - z_t)].$$

Thus, we obtain the following extension of the Goodwin model:

$$\Theta_{EG}: \begin{cases} \dot{x}_t = \frac{\mu}{\theta}[\phi(x_t) - \{(1/\mu) + (1-\alpha)\theta\}x_t - (1-\alpha)w_t + \beta z_t + \theta\beta\psi(w_t - z_t)], \\ \dot{w}_t = x_t, \\ \dot{z}_t = \psi(w_t - z_t). \end{cases}$$

We call System Θ_{EG} the extended Goodwin model. In the following sections, by introducing a pessimistic outlook into the adjustment function, we consider the dynamic behavior of System Θ_{EG}.

3 Dynamics Resulting from Pessimism

In this section, we demonstrate that the pessimistic outlook held by the household about the future economy causes a chronic slump. The pessimistic outlook is expressed by a nonlinearity incorporated into the adjustment function. Before discussing this, it is convenient to consider the dynamic behavior of System Θ_{EG} with the linear adjustment function. We begin with the verification of simple results on the existence and stability of the equilibrium point. The following lemma is clear.

Lemma 1 *Under Assumptions 1, 3, and 4, System Θ_{EG} possesses a unique equilibrium point (0, 0, 0). In other words, the market equilibrium is uniquely determined and given by* $(y^*, x^*, y_e^*) = (c_0/(1-\alpha-\beta), 0, c_0/(1-\alpha-\beta))$. ■

Proof Direct calculation proves Lemma 1. ■

We now prove the following.

Lemma 2 *We assume that* $0 < \psi'(0)\theta < \mu\phi'(0) - 1 - \mu(1-\alpha)\theta$. *Then, under Assumptions 1–4, the equilibrium point of System Θ_{EG} is unstable.* ■

Proof See Appendix. ■

The linear case provides a direct extension of the Goodwin model. In fact, as will be shown in Numerical Example 2, like the Goodwin model, System Θ_{EG} possesses a similar periodic path that surrounds the equilibrium point. The meaning of "similarity" is clarified using the comparison between systems with nonlinear and linear adjustment functions, which will be presented soon.

Numerical Example 2 We consider the linear adjustment function

$$\psi(u) = \eta_L(u; h) = hu \, (h > 0).$$

We set $\alpha = 0.4$, $\mu = 2$, $\theta = 1$, $\beta = 0.46$, $h = 0.45$, and $\phi(x) = 2\text{Arctan}(1.5x)$. It can be easily verified that these parameters satisfy Assumptions 1–4. Figure 2 describes the projection of a typical path of System Θ_{EG} onto the w–z plane, which converges to a periodic path. The black dot emphasizes the equilibrium point in the

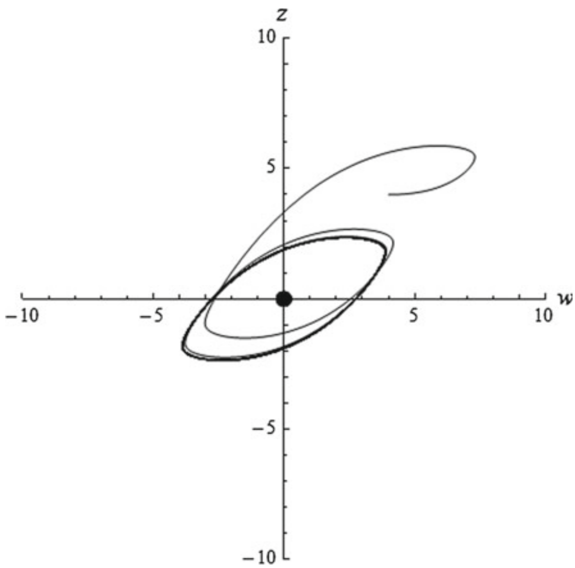


Fig. 2 Typical periodic attractor of the extended Goodwin model with a symmetric (linear) adjustment function

$w-z$ plane. The path in Numerical Example 2 represents the usual business cycles in the sense that the path surrounds the equilibrium point. As in Numerical Example 1, the model economy in Numerical Example 2 possesses the power of automatic and complete recovery from the slump, though the recovery is temporal and the economy repeats a pattern of booms and slumps. ■

The nonlinear factor in Numerical Example 2 is merely incorporated into the $\phi-$function. Therefore, the periodic attractor observed in Numerical Example 2 is generated by the sigmoid nonlinearity of the $\phi-$function. System Θ_{EG} with the linear $\psi-$function generates a periodic attractor that is similar to that of the Goodwin model in the sense that the periodic attractor surrounds the equilibrium point.

Next, we incorporate the nonlinearity of our model into the adjustment function. We define

$$\eta_{NL}(u; a_+, a_-) = \begin{cases} a_+ u & u \geq 0, \\ a_- u & u < 0. \end{cases} \tag{3.1}$$

We now make the following assumption:

Assumption 5 $a_- > a_+$.

We use System Θ_{NEG} to denote System Θ_{EG} wherein $\psi(u) = \eta_{NL}(u; a_+, a_-)$ satisfies Assumption 5. Assumption 5 introduces the asymmetric nonlinearity into the adjustment function. As shown later, Assumption 5 is closely related to the emergence of a chronic slump. In this sense, Assumption 5 plays the most important role in our argument. Here, we explain its economic implication. We assume that the representative household is pessimistic about the future economy. We consider the adjustment function under this assumption. When the actual income exceeds the expected income (i.e., $u = y - y_e > 0$), it is expected that the economy will become more prosperous in the future. However, since the household is pessimistic, it does not have hope for further prosperity. Therefore, the upward adjustment of the expected income is excessively small (in other words, the household is hyperopic). Conversely, when the actual income is lower than the expected income (i.e., $u = y - y_e < 0$), it is expected that the economy will worsen even more in the future. Since the household is pessimistic, it will expect further worsening. Therefore, the downward adjustment of the expected income is excessively large (in other words, the household is myopic). Thus, we see that pessimism about the future economy yields the asymmetric nonlinearity of Assumption 5.

We here make one remark. As showed in the Introduction, through the self-fulfilling prophecy, the pessimistic outlook persists over a long period of time unless the market's loss of confidence is recovered. Consequently, the asymmetric nonlinearity persists over a long period of time. Thus, Assumption 5 is robust.

A viewpoint of the behavioral economics about loss aversion is useful in explaining the asymmetric nonlinearity. We here quote the sentences from Kahneman et al. (1991): Responses to increases and to decreases in prices, for example, might not always be mirror images of each other. The possibility of loss-aversion effects suggests, more generally, that treatments of responses to change in economic variables should routinely separate the cases of favorable and unfavorable changes. Our assumption is consistent with the viewpoint of the behavioral economics. We consider the graph of the nonlinear adjustment function $\psi(u) = \eta_{NL}(u; a_+, a_-)$. Changes to the right (resp. left) side of the origin (i.e., the reference point) are favorable (resp. unfavorable) for households. Thus, from the viewpoint of the behavioral economics, we obtain that the adjustment function habitually possesses the (perhaps weak) nonlinearity. In our model, since we assume pessimism about the future economy, the loss aversion will be reinforced and the asymmetric nonlinearity will be stronger.

Before discussing the dynamics of System Θ_{NEG}, we must confirm the existence and uniqueness of solutions. It should be noted here that the $\psi-$function is not differentiable at $u = 0$. However, it can be easily checked that the $\psi-$function satisfies the Lipschitz condition in R^2. Therefore, from Assumptions 2 and 5, the vector field of System Θ_{NEG} also satisfies the Lipschitz condition. This proves the existence and uniqueness of solutions.[15] Thus, under Assumptions 2 and 5, the solutions of System Θ_{NEG} are determined uniquely. The piecewise linear function (3.1) is written in a very simple form to explain the self-fulfilling process to chronic slump. It is not easy to ascertain whether or not the equilibrium point is stable. Therefore, in this paper, we numerically investigate the dynamics of System Θ_{NEG}. Moreover, as shown from the explanation of Assumption 5, we observe that as the a_+-value decreases or the a_--value increases, the pessimism regarding the future economy becomes strong (in other words, the asymmetry of the adjustment function becomes strong).

Clearly, System Θ_{NEG} with an asymmetric adjustment function possesses the mechanism to produce business cycles, much like the Goodwin model. On the other hand, as stated in the Introduction, in System Θ_{NEG}, the household's pessimistic outlook has a direct influence on the adjustment function of expected income, which is closely related to its outlook for the future economy. Assumption 5 describes this influence. Therefore, given the above, the household's consumption decreases through the reduction in its expected income, and the reduction in consumption makes the model economy inactive. Thus, we can expect the resulting economy to become more inactive than in the Goodwin model. To demonstrate that this intuitive observation is correct, we now consider a typical numerical example of System Θ_{NEG}.

We consider a more pessimistic case than that presented in Numerical Example 2. In Numerical Example 2, we considered the adjustment function $\psi(u) = \eta_L(u; 0.45)$. Therefore, since we consider a more pessimistic case, we set

$$a_+ = 0.1 < 0.45 \quad \text{and} \quad a_- = 0.6 > 0.45. \tag{3.2}$$

We set $\mu = 2$, $\alpha = 0.4$, $\theta = 1$, $\beta = 0.46$, and $\phi(x) = 2\text{Arctan}(1.5x)$. Clearly, these parameters satisfy Assumptions 1–3. It should be noted here that in the symmetric case with $h = a_+ = a_-$, the size of h determines the amplitude of a periodic path that surrounds the equilibrium point. Parts (1) of Figs. 3 and 4 describe the graphs

[15] See Guckenheimer and Holmes (1983, p. 3).

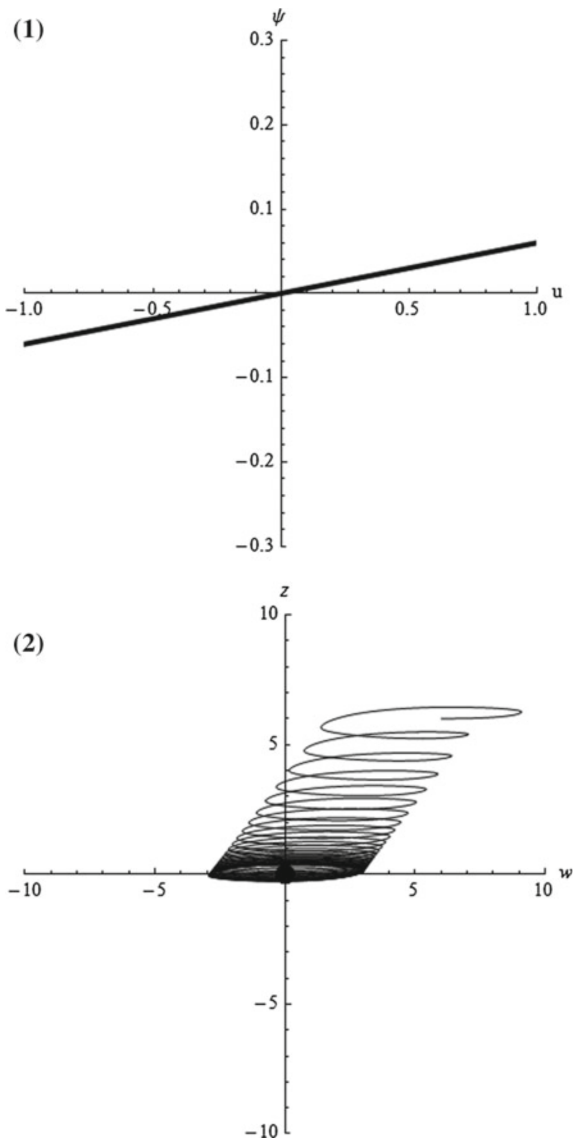


Fig. 3 Periodic attractor of the extended Goodwin model with a small symmetric adjustment coefficient

of the adjustment functions of the systems with $h = a_+ = a_- = 0.1$ and $h = a_+ = a_- = 0.6$, respectively. Parts (2) of Figs. 3 and 4 describe the projections of typical periodic paths of the systems with $h = a_+ = a_- = 0.1$ and $h = a_+ = a_- = 0.6$ onto the $w{-}z$ plane, respectively. In the asymmetric case with (3.2) (i.e., in the mixture of these two cases), different dynamic behavior occurs. Figure 5 shows it. The thick black line of Part (1) of Fig. 5 describes the piecewise linear graph of the adjustment function of System Θ_{NEG}. The piecewise linear graph is the mixture of the graphs of

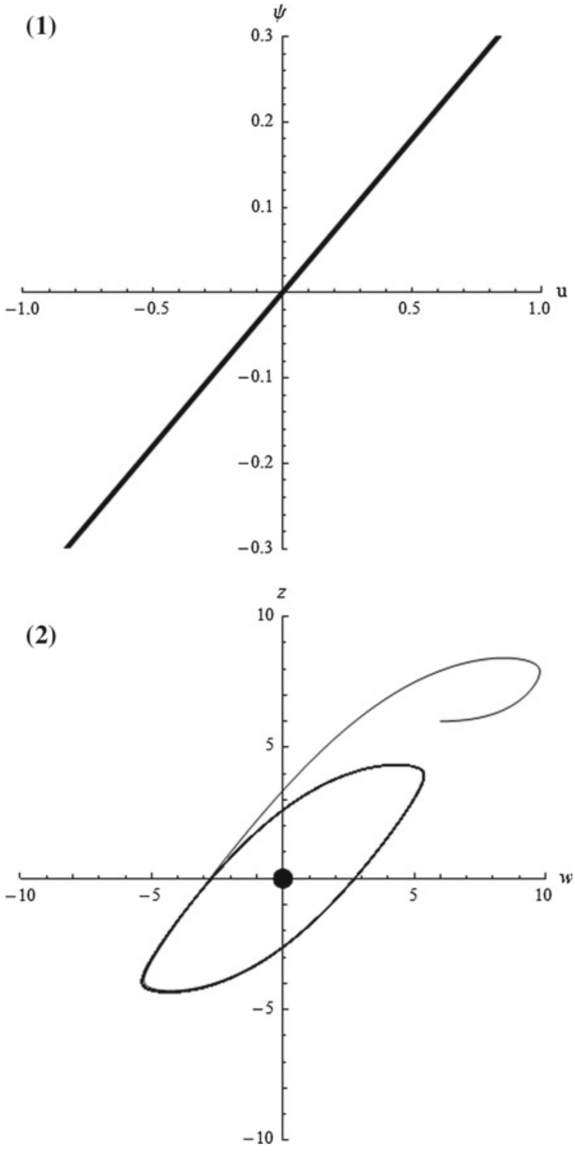
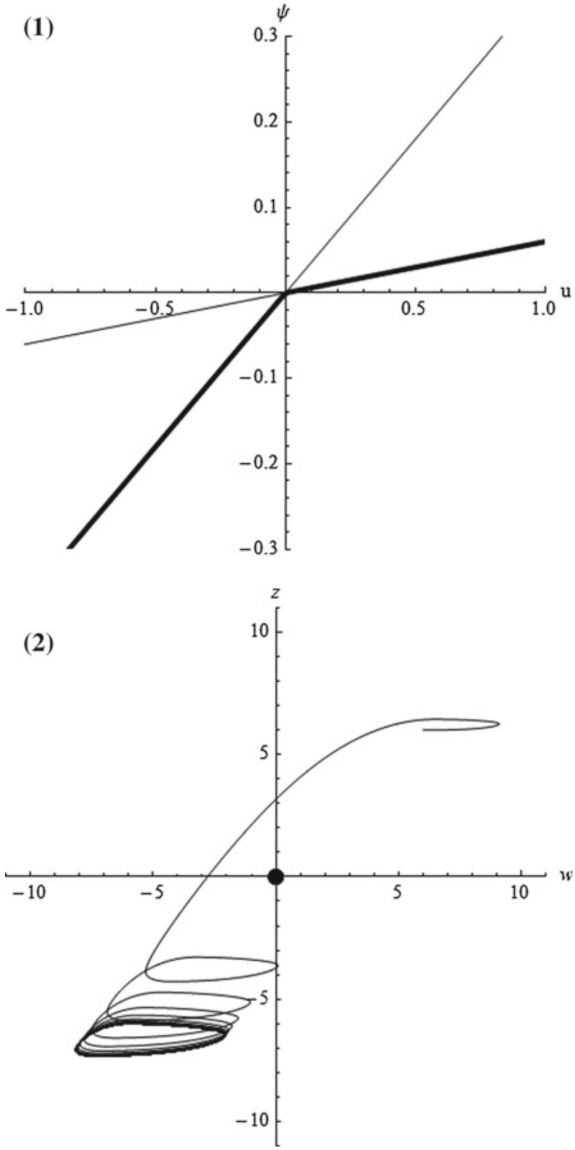


Fig. 4 Periodic attractor of the extended Goodwin model with a large symmetric adjustment coefficient

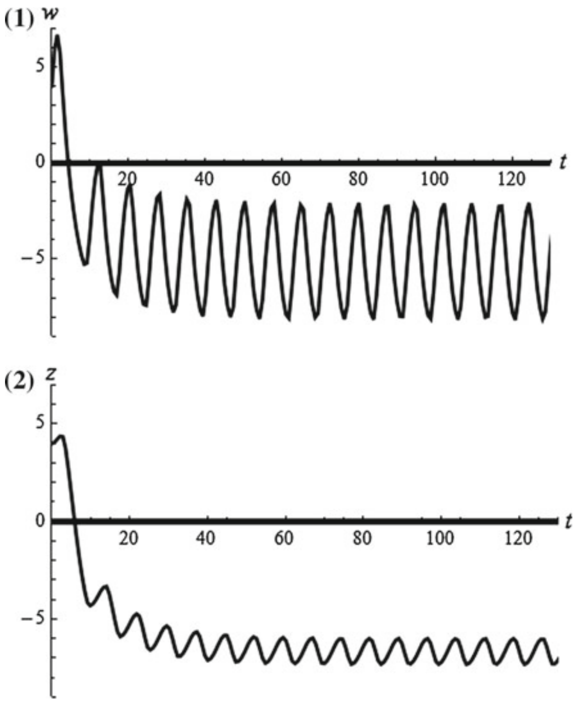
Parts (1) of Figs. 3 and 4. Part (2) of Fig. 5 describes the projection of a typical path of System Θ_{NEG} onto the $w-z$ plane, which converges to a periodic path. In Parts (2) of Figs. 3, 4 and 5, the black dot emphasizes the equilibrium point in the $w-z$ plane. The black curves in Parts (1) and (2) of Fig. 6 describe the time series of the deviations of income and expected income of the path of Fig. 5. In (1) and (2), thick straight lines emphasize the time series of the equilibrium income and equilibrium expected

Fig. 5 Periodic attractor of the extended Goodwin model with an asymmetric adjustment coefficient



income paths, respectively. As compared to the periodic path of Figs. 3 and 4, that of Fig. 5 appears in a domain lower than the equilibrium point (0, 0). In this sense, in a chronic slump, private spending is continuously insufficient to make use of the available productive capacity. See the Introduction. Thus, the periodic path of Fig. 5

Fig. 6 Time series of the path of Fig. 5



describes the situation in which the economy constantly repeats partial recoveries and slowdowns. We call such a periodic path a *slump cycle*.[16]

As stated before, the periodic paths of the Goodwin model and System Θ_{EG} with the linear adjustment function surround the equilibrium point. On the other hand, Fig. 5 show that in the slump cycle of System Θ_{NEG}, the income and expected income are locked in domains lower than the equilibrium point. Thus, we note that the System Θ_{NEG} (i.e., System Θ_{EG} with the asymmetric adjustment function) satisfies R.1 and R.2. We now observe the dynamic behavior of the expected income. Figure 5 also shows that given the loss of symmetry due to the pessimistic outlook, the expected income is locked in a "narrow" domain lower than the equilibrium point. This shows that the household is convinced of the pessimistic outlook for the future economy. Thus, it becomes difficult that the household escapes from a pessimistic outlook.

We consider the effect of the intensity of pessimism about the future economy (i.e., the degree of asymmetry of the adjustment function) on the location and the amplitude of the emerging slump cycle. We set $a_- = 0.4$ and $a_+ = 0.4i$, where i

[16]Needless to say, the periodic paths of System Θ_{NMG} with asymmetric adjustment functions are not necessarily lower than the equilibrium. Therefore, the notion of a slump cycle is restrictive. In the case where the asymmetry of the adjustment function is sufficiently strong, the periodic path becomes lower than the equilibrium. For this point, see Fig. 7 to be given later. However, such a notion is useful in making our argument clear-cut. The occurrence of slump cycles is the most interesting feature of the present paper.

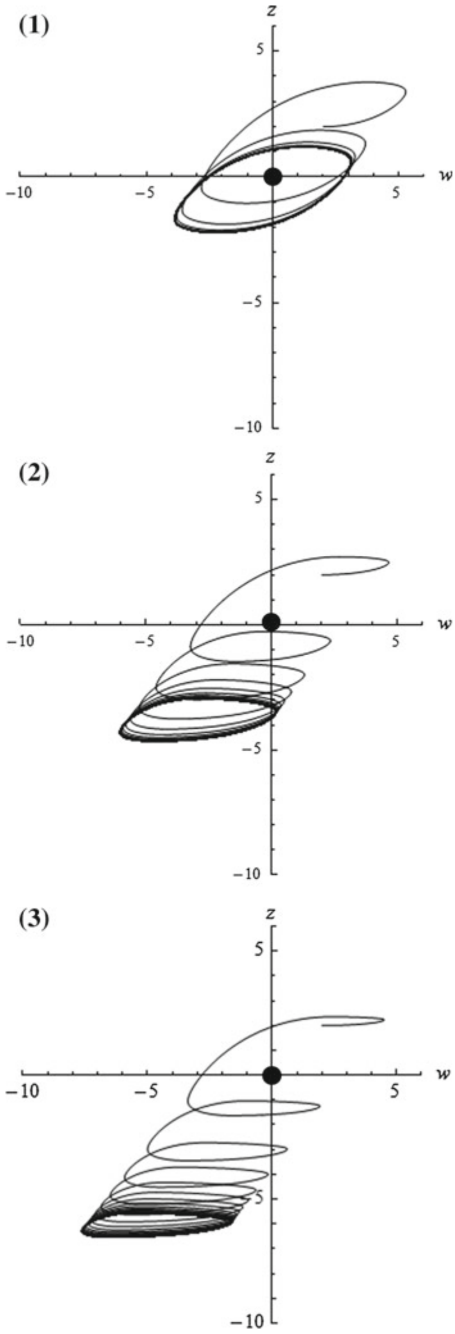


Fig. 7 Periodic attractor goes away from the equilibrium point as the degree of asymmetry of adjustment coefficient (the intensity of pessimism) becomes large

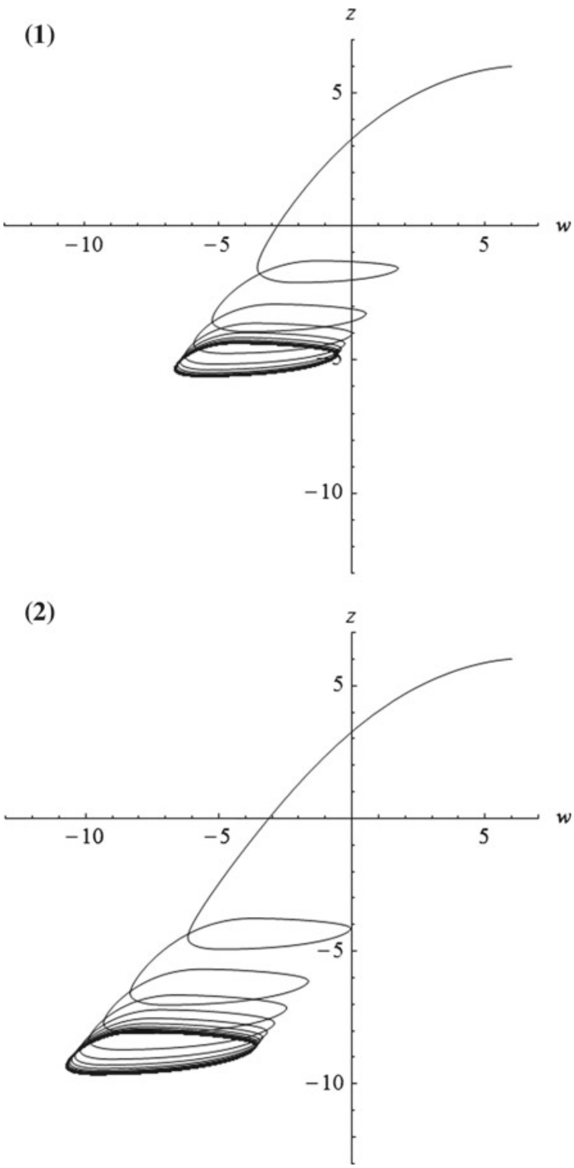


Fig. 8 Periodic attractor goes away from the equilibrium point as the propensity to consume becomes large

represents the intensity of pessimism. Assumption 5 yields $0 < i < 1$. Parts (1)–(3) of Fig. 7 describe the projections of the paths for $i = 0.9$, $i = 0.4$, and $i = 0.2$ onto the $w-z$ plane, respectively. The parameters of Fig. 7 apart from $a_{\pm}$ are the same as in Fig. 5. In Fig. 7, the black dots emphasize the equilibrium point in the $w-z$ plane. Parts (1)–(3) describe how the attractor changes as the intensity of pessimism (i.e., the degree of asymmetry) increases. Figure 7 shows that the intensity of pessimism has a strong effect on the location of the emerging slump cycle. This indicates that as the intensity of pessimism increases (i.e., the parameter a_- increases or the parameter a_+ decreases), the maximum point of a business cycle decreases. Moreover, the figure reveals that the intensity of pessimism has a weak effect on the amplitude of the emerging slump cycle. As the intensity of pessimism increases, the amplitude of the $z-$value in the slump cycle decreases slightly. This indicates that as the household becomes more pessimistic, the outlook for the future economy (expressed by the expected income) becomes more inflexible in a low domain.

System Θ_{NEG} possesses two propensities to consume: the propensities concerning income and expected income. We here numerically see the relation between the seriousness of slump and the propensity to consume concerning income. See Fig. 8. Parts 1 and 2 of Fig. 8 describe the typical dynamic behavior in the case where we set $\mu = 2$, $\theta = 1$, $\beta = 0.43$, $a_+ = 0.1$, $a_- = 0.5$, and $\phi(x) = 2\text{Arctan}(1.5x)$. In Parts 1 and 2 of Fig. 8, we set $\alpha = 0.4$ and $\alpha = 0.47$, respectively. Figure 8 shows that as the propensity to consume concerning income becomes larger, the slump cycle becomes more severe. In other words, comparing with the case where the propensity is small, the occurrence of pessimism in the converse case makes the slump cycle more severe. Since we can obtain the same result on the propensity to consume concerning expected income, we omit the argument.

4 Chronic Slump and Local Stability

Our main result in Sect. 3 is that the maintaining of pessimism can yield the chronic slump through self-fulfilling prophecy. However, the maintaining of pessimism does not always yield the chronic slump. It should be noted that in Sect. 3 we assumed the instability of equilibrium. This assumption is essential to the occurrence of chronic slump. In this section, we make clear this point.

In order to see it, we need the smooth adjustment function in the sense that ψ' is continuous. The $\psi-$function in Sect. 3 is nonsmooth. Therefore, throughout this section, we consider the following smooth adjustment function:

$$\psi(u) = \psi_{m,n,d}(u) = \varphi_{m,n,d}(u) \cdot u = m\{n - d\text{Arctan}(gu)\} \cdot u, \tag{4.1}$$

where m, n, and h are positive constants. The $\varphi_{m,n,d}-$function represents the adjustment coefficient that depends on $u = y_t - y_{et}$. We work under the assumption:

Assumption 6 $n/d > \pi/2$.

The black curve of Fig. 9 describes the graph of the $\varphi_{m,n,d}-$function. In Fig. 9, we set $m = 1$, $n = 0.5$, $d = 0.3$, and $g = 3.4$. The adjustment coefficient function satisfies the following properties:

Lemma 3 *$\varphi'_{m,n,d}(u) < 0$ and $u\varphi''_{m,n,d}(u) > 0$, for any $u \in (-\infty, 0) \cup (0, +\infty)$. $\lim_{t\to+\infty} \varphi_{m,n,d}(u) = m(n - d\pi/2) > 0$ and $\lim_{t\to-\infty} \varphi_{m,n,d}(u) = m(n + d\pi/2)$.* ■

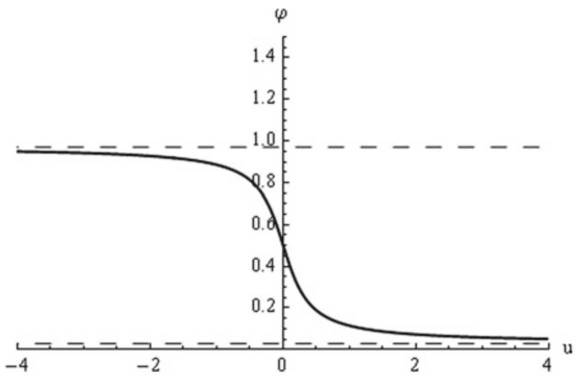


Fig. 9 Adjustment coefficient function

Proof See Appendix. ∎

Thus, we consider the asymmetric adjustment coefficient. In the case where the adjustment coefficient function takes the form of (4.1), the adjustment function satisfies the following properties:

Lemma 4 *we have* $\psi_{m,n,d}''(u) > 0$, $\lim_{u\to+\infty} \psi_{m,n,d}'(u) = m(n - d\pi/2) > 0$, $\lim_{u\to-\infty} \psi_{m,n,d}'(u) = m(n + d\pi/2)$, *and* $\psi_{m,n,d}'(u) > 0$, *for any* $u \neq 0$. ∎

Proof See Appendix. ∎

From Lemma 4, we see that the form of the smooth $\psi_{m,n,d}$−function is almost the same as that of the adjustment function in Sect. 3. The difference between them is in the continuity of the derivative. Figure 10 describes a typical graph of the adjustment function with $m = 1$, $n = 0.5$, $d = 0.3$, and $g = 3.4$.
Moreover, the $\psi_{m,n,d}$−function possesses the following important property:

Lemma 5 *If* $d > s > 0$, *we have* $\psi_{m,n,s}(u) > \psi_{m,n,d}(u)$ *for any* $u \neq 0$.

Proof See Appendix. ∎

See Fig. 11. In Fig. 11, we set $m = 1$, $n = 0.87$, $g = 3$, $d = 0$, $d = 0.12$, $d = 0.3$, and $d = 0.45$. Figure 11 describes that the degree of flexion is larger as the value of d is larger. Thus, the property of Lemma 5 shows that as the parameter d is larger, the degree of pessimism becomes larger. Especially, the line with $d = 0$ describes the usual symmetric adjustment function. Thus, the parameter represents the degree of pessimism. Here, we have the following result:

Lemma 6 *The stability of equilibrium does not depend on the degree of* d. ∎

Proof See Appendix. ∎

Fig. 10 Adjustment function

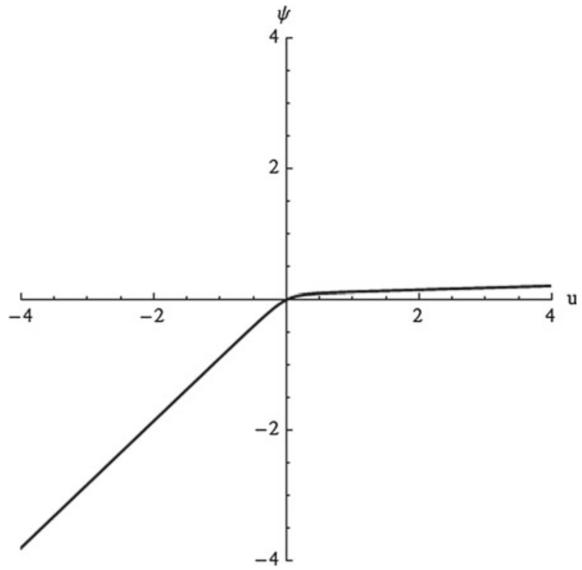
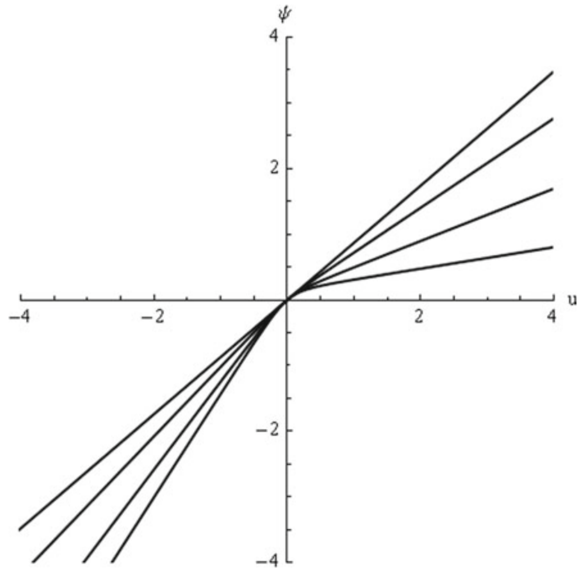


Fig. 11 The degree of flexion (the degree of pessimism) becomes large as the parameter *d* becomes large



Lemma 6 gives us an important message. In order to explain it, in the following, System Θ_{EG} with $d = 0$ and System Θ_{EG} with $d > 0$ are called the nonpessimistic and the pessimistic Goodwin models, respectively. Lemma 6 shows that if the nonpessimistic Goodwin model is locally stable, the pessimistic Goodwin model is locally stable independently of the degree of pessimism.

We numerically show that if the equilibrium is asymptotically stable, the (persistent) chronic slump does not occur. We set $\alpha = 0.4$, $\mu = 2$, $\theta = 1$, $\beta = 0.46$, $m = 1$, $n = 0.5$, $d = 0.3$, $g = 3$, $\phi(x) = \text{Arctan}(qx)$. Under the setting, we have

$$\psi'(0) = \varphi_{m,n,d}(0) = mn = 0.5.$$

Therefore, since $q = \phi'(0)$, we have

$$\lambda_1 + \lambda_2 + \lambda_3 = \left\{ \frac{\mu q - 1 - \mu(1-\alpha)\theta}{\theta} - \psi'(0) \right\} = 2q - 2.7, \tag{4.1a}$$

$$\begin{aligned} \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 &= \frac{\mu\{1 - \alpha - \theta\beta\psi'(0)\} - \{\mu q - 1 - \mu(1-\alpha)\theta\}\psi'(0)}{\theta} \\ &= (1.2 - 0.92\psi') - (2q - 2.2)\psi' = 1.84 - q, \end{aligned} \tag{4.1b}$$

$$\lambda_1\lambda_2\lambda_3 = -\frac{\mu\psi'(0)(1 - \alpha - \beta)}{\theta} = -0.14. \tag{4.1c}$$

For $\lambda_k (k \in \{1, 2, 3\})$, see the proof of Lemma 2 in Appendix. We here define

$$\begin{aligned} -(\lambda_1 + \lambda_2 + \lambda_3)(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1) + \lambda_1\lambda_2\lambda_3 &= (2q - 2.7)(q - 1.84) - 0.14 \\ &= 2q^2 - 6.38q + 4.828 \equiv \Pi(q). \end{aligned}$$

Then, the solutions of $\Pi(q) = 0$ are given by

$$q_+ \equiv \frac{6.38 + \sqrt{6.38^2 - 4 \times 2 \times 4.828}}{4} > 1.84, \tag{4.2a}$$

$$q_- \equiv \frac{6.38 - \sqrt{6.38^2 - 4 \times 2 \times 4.828}}{4} < 2.7/2. \tag{4.2b}$$

We here prove the following lemma:

Lemma 7 *If $q < q_-$, the equilibrium is asymptotically stable. Moreover, if $q_- < q$, the equilibrium is unstable.* ■

Proof See Appendix. ■

Lemma 7 shows that the stability of the equilibrium point depends on q. That is, it depends on the form of the investment function. See Fig. 12 that shows typical paths of the pessimistic Goodwin models. Parts 1 and 2 of Fig. 12 describe paths of the pessimistic Goodwin models with $q = 1.9 > q_+$ and $q = 1.11 < q_-$, respectively. It should be noted here that the adjustment functions of Parts 1 and 2 are the same. Thus, we see that even if the pessimism is not recovered, stabilizing the equilibrium recovers the chronic slump. The recovery from pessimism often requires a lot of time. Therefore, the observation in this section suggests that not only recovering the pessimism but also stabilizing the equilibrium are necessary for the recovery from the chronic slump.

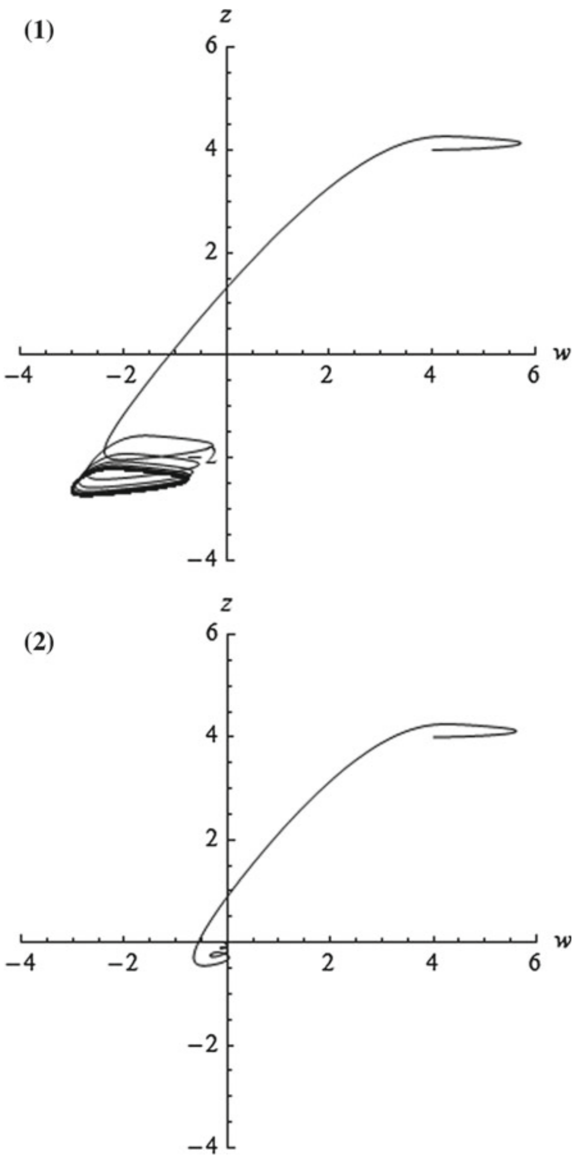


Fig. 12 Paths of the extended Goodwin model: *1* with the unstable equilibrium point; *2* with the stable equilibrium point

Finally, we make one remark. Although stabilizing the equilibrium is necessary for the recovery from the chronic slump, the pessimism makes the recovery slower. Figure 13 describes it. In Fig. 13, we set $\alpha = 0.4$, $\mu = 2$, $\theta = 1$, $\beta = 0.46$, $m = 1$, $n = 0.5$, $q = 1.1$, and $g = 3$. Since $q = 1.1 < q_{-}$, we see from Lemma 7 that the equilibrium point of the extended Goodwin model is asymptotically stable. Dashed curves of Parts 1 and 2 of Fig. 13 describe typical time series of the deviations of

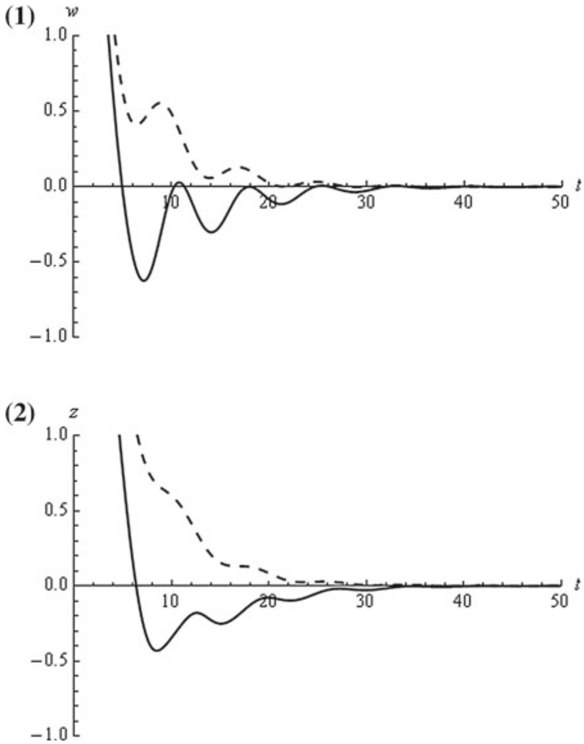


Fig. 13 Paths in the case where the extended Goodwin model with asymmetric adjustment coefficient is stable

income and expected income in the nonpessimistic Goodwin models with $d = 0$. On the other hand, black curves of Parts 1 and 2 of Fig. 13 describe typical time series of the deviations of income and expected income of the pessimistic Goodwin model with $d = 0.3$.

5 Conclusions and Final Remarks

From the Keynesian perspective, we constructed a prototype dynamic model expressing a part of the Krugman's view (Krugman 2008) concerning the recent chronic slump that has spread across the world. We constructed an extension of Goodwin's nonlinear accelerator model, and attempted to show that the pessimistic outlook of the household is an important cause of the chronic slump. Unlike the Goodwin model, the representative household distinguishes between the short-run and long-run consumption plans. The short-run plan is the same as that in the Goodwin model. On the other hand, in the long-run plan, the household determines its consumption in proportion to the expected income, which is adaptively adjusted. We assumed that the household possesses a pessimistic outlook; according to this outlook, the upward adjustment of the expected income is excessively small (in other words, the household is hyperopic). Conversely, the downward adjustment of the expected income is excessively large (in other words, the household is myopic). This assumption introduces an asymmetric nonlinearity into the adjustment function. We also observed that the assumption is related to the result about loss aversion in the behavioral economics. We demonstrated that the asymmetric nonlinearity plays an important role in generating a chronic slump.

An intuitively explanation of our result is as follows. First, we considered the case where the extended Goodwin model is completely unstable. The asymmetric nonlinearity implies that pessimism makes an upturn difficult but makes a downturn easy. Through this mechanism, the model economy spirals downward and falls into a chronic slump. Moreover, in the process, the model economy constantly repeats partial recoveries. But, income and expected income are locked in a domain lower than the market equilibrium. Thus, in the extended Goodwin model, the model economy in the chronic slump cannot continuously achieve the potential ability to produce, which is estimated at the market equilibrium. Thus, we revealed a way in which local instability and a pessimistic outlook cause a chronic slump.

Another important feature of the extended Goodwin model with a pessimistic outlook is that

> *the model economy goes into chronic slump from everywhere, regardless of initial economic conditions.*

The reason is that, in the cases where we numerically investigated, any slump cycle is globally stable. Immediately after the collapse of the bubble economy, the Japanese economy from 1991 through 2002 experienced a chronic slump. The above feature may explain such a transition from a bubble economy to a chronic slump economy. This feature does not appear in the models with multiple equilibria (for example, stable higher, unstable middle, and stable lower equilibria), because, in any model with multiple equilibria, the destination of a path depends on the initial condition of the path.

Next, we considered the case where the extended Goodwin model is stable. We numerically showed that even if the pessimistic outlook is not improved, the economy converges to the equilibrium and therefore, it recovers from the slump, though the recovery time may depend on the strength of stability. Thus, we conclude that the chronic slump results from the instability of equilibrium and the pessimistic outlook about future economy.

From the consideration in this paper, we presented two ways of recovering from the chronic slump: the recovery from the pessimistic outlook and the stabilization of economy. It often takes a long time to recover from the pessimistic outlook. In such a case, it will be effective to carry out a stabilizing policy.

As stated in Introduction, an important feature of business cycle is that booms and slumps come in all sizes. This paper proved that, according to the measure of pessimism, slumps come in all sizes. Moreover, by the same argument as that in this paper, we can prove that, according to the measure of optimism, booms come in all sizes. Thus, the extended Goodwin model also gives a theoretical explanation of the feature.

Acknowledgments The author thanks Toichiro Asada for his helpful comments and suggestions.

Appendix

In this appendix, we prove Lemmas 2–7.

Proof of Lemma 2 We use J to denote the Jacobian matrix of System Θ_{EG}. The characteristic equation of J is given by

$$\begin{aligned}\Lambda(\lambda) &\equiv \det(\lambda I - J)\\ &= \det\begin{bmatrix} \lambda - \dfrac{\mu\phi'(0) - 1 - \mu(1-\alpha)\theta}{\theta} & \dfrac{\mu\{1-\alpha-\theta\beta\psi'(0)\}}{\theta} & \dfrac{\mu(\theta\beta\psi'(0)-\beta)}{\theta} \\ -1 & \lambda & 0 \\ 0 & -\psi'(0) & \lambda+\psi'(0) \end{bmatrix}\\ &= \lambda^3 - \left\{\frac{\mu\phi'(0) - 1 - \mu(1-\alpha)\theta}{\theta} - \psi'(0)\right\}\lambda^2\\ &\quad + \frac{\mu\{1-\alpha-\theta\beta\psi'(0)\} - \{\mu\phi'(0) - 1 - \mu(1-\alpha)\theta\}\psi'(0)}{\theta}\lambda + \frac{\mu\psi'(0)(1-\alpha-\beta)}{\theta}.\end{aligned}$$

We use λ_k $(k \in \{1, 2, 3\})$ to denote the eigenvalue of J. Assumption 3 yields $\Lambda(0) > 0$. Therefore, at least one eigenvalue is negative. Without loss of generality, we suppose that $\lambda_3 < 0$. Assumption 3 gives that $\lambda_1 \cdot \lambda_2 \cdot \lambda_3 = -\mu\psi'(0)(1-\alpha-\beta)/\theta < 0$. Thus, we have $\lambda_1 \cdot \lambda_2 > 0$. Therefore, if λ_1 and λ_2 are real numbers, then λ_1 and λ_2 must be simultaneously positive or negative. The assumption of Lemma 2 shows

$$\lambda_1 + \lambda_2 + \lambda_3 = \frac{\mu\phi'(0) - 1 - \mu(1-\alpha)\theta}{\theta} - \psi'(0) > 0. \tag{A.1}$$

Hence, $\lambda_1 > 0$ and $\lambda_2 > 0$. Moreover, we observe from (A.1) that if λ_1 and λ_2 are complex conjugates, then $\mathrm{Re}\,\lambda_1 > 0$ and $\mathrm{Re}\,\lambda_2 > 0$. Thus, we complete the proof.

Proof of Lemma 3 We have

$\varphi'_{m,n,d}(u) = -mgd/(g^2u^2+1) < 0, \quad u\varphi''_{m,n,d}(u) = mdg^3u^2/(g^2u^2+1)^2 > 0$, and $\lim_{n\to\pm\infty}\mathrm{Arc}\tan(u) = \pm\pi/2$.

The proof follows directly from this fact. ■

Proof of Lemma 4 We have

$$\psi_{m,n,d}{}'(u) = m\{n - d\,\mathrm{Arc}\tan(gu) - dgu/(g^2u^2+1)\}. \tag{A.2}$$

Therefore, since $\lim_{n\to\pm\infty} \mathrm{Arc}\tan(u) = \pm\pi/2$, we see from (A.2) and Assumption 6 that $\lim_{u\to\pm\infty} \psi_{m,n,d}{}'(u) = m(n \pm d\pi/2) > 0$. Moreover, we have

$$\psi_{m,n,d}{}''(u) = -m\left\{\frac{dg}{g^2u^2+1} + \frac{dg(g^2u^2+1) - 2dg^3u^2}{(g^2u^2+1)^2}\right\} = -\frac{2mdg}{(g^2u^2+1)^2} < 0. \tag{A.3}$$

We now prove $\psi_{m,n,d}{}'(u) > 0$. It follows from (A.3) and Assumption 6 that

$$\psi_{m,n,d}{}'(u) > \lim_{u\to+\infty} \psi_{m,n,d}{}'(u) = m(n - d\pi/2) > 0.$$

Thus, we complete the proof. ■

Proof of Lemma 5 Since $\mathrm{Arc}\tan(u) > 0$ (< 0) for any $u > 0$ (< 0), we have $u\mathrm{Arc}\ \tan(u) > 0$ for any $u \neq 0$. Therefore, we have

$$\psi_{m,n,s}(u) - \psi_{m,n,d}(u) = u\mathrm{Arc}\tan(gu)\cdot m(d-s) > 0 \text{ for any } u \neq 0.$$

This completes the proof. ■

Proof of Lemma 6 Since we have $\psi_{m,n,d}{}'(0) = mn$, $\Lambda(\lambda)$ does not depend on d. This completes the proof. ■

Before proving Lemma 7, we prove the following three sublemmas.

Sublemma 1 Let α, β, and γ be solutions of a cubic equation. We assume $\alpha + \beta + \gamma \geq 0$ and $\alpha\beta\gamma < 0$. Then, one of the real parts of the solutions is positive. ■

Proof of Sublemma 1 Since $\alpha\beta\gamma < 0$, one of α, β, and γ must be a negative real number. Without loss of generality, we assume $\gamma < 0$. If α and β are real numbers, α and β are positive. We assume that α and β are not real numbers. Then, α and β are given as $\alpha = \xi + \omega i$ and $\beta = \xi - \omega i$, where ξ and ω are real numbers and $i = \sqrt{-1}$. From the assumption, we have $0 \leq \alpha + \beta + \gamma = 2\xi + \gamma$. Since $\gamma < 0$, we have $\xi > 0$. This completes the proof. ■

Sublemma 2 Let α, β, and γ be solutions of a cubic equation. We assume that, $\alpha + \beta + \gamma < 0$, $\alpha\beta\gamma < 0$, and $-(\alpha + \beta + \gamma)(\alpha\beta + \beta\gamma + \gamma\alpha) + \alpha\beta\gamma < 0$. Then, one of the real parts of the solutions is positive. ■

Proof of Sublemma 2 Since $\alpha\beta\gamma < 0$, one of α, β, and γ must be a negative real number. Without loss of generality, we assume $\gamma < 0$. We assume that α and β are not real numbers. Then, α and β are given as $\alpha = \xi + \omega i$ and $\beta = \xi - \omega i$, so that

$$0 > -(\alpha + \beta + \gamma)(\alpha\beta + \beta\gamma + \gamma\alpha) + \alpha\beta\gamma = 2[-(2\xi + \gamma)\gamma - (\xi^2 + \omega^2)]\xi.$$

Since $\alpha + \beta + \gamma = 2\xi + \gamma < 0$, we have $\xi > 0$. Therefore, if α and β are not real numbers, the proof completes. We next assume that α and β are real numbers. Then

α and β must be simultaneously positive or negative. We assume that α and β are negative. Then, we have

$$\begin{aligned} 0 &> -(\alpha+\beta+\gamma)(\alpha\beta+\beta\gamma+\gamma\alpha)+\alpha\beta\gamma \\ &= -(\alpha^2\beta+\alpha\beta^2+\beta^2\gamma+\beta\gamma^2+\gamma^2\alpha+\gamma\alpha^2+2\alpha\beta\gamma) > 0. \end{aligned}$$

This contradicts to the assumption. Therefore, we see that α and β are positive. This completes the proof. ■

Sublemma 3 Let α, β, and γ be solutions of a cubic equation. A set of necessary and sufficient conditions for all the real parts of the solutions to be negative are given by

$$\begin{aligned} &\alpha+\beta+\gamma < 0, \quad \alpha\beta\gamma < 0, \quad \alpha\beta+\beta\gamma+\gamma\alpha > 0, \text{ and} \\ &-(\alpha+\beta+\gamma)(\alpha\beta+\beta\gamma+\gamma\alpha)+\alpha\beta\gamma > 0. \end{aligned}$$ ■

Proof of Sublemma 3 See Gandolfo (1996, Sect. 16.4). ■

We now prove Lemma 7.

Proof of Sublemma 7 Sublemma 3 yields that a set of necessary and sufficient conditions for all the real parts of the solutions to be negative are given by

$$\begin{aligned} &\lambda_1+\lambda_2+\lambda_3 = 2q-2.7 < 0, \quad \lambda_1\lambda_2+\lambda_2\lambda_3+\lambda_3\lambda_1 = 1.84-q > 0, \text{ and} \\ &\Pi(q) = 2q^2-6.38q+4.828 > 0. \end{aligned}$$

Therefore, the necessary and sufficient condition for all the real parts of the solutions to be negative is given by $q < q_-$. This proves the first half. We now prove the latter half. Sublemmas 1 and 2 show that a set of sufficient conditions for one of the real parts of the solutions to be positive are given by

$$\lambda_1+\lambda_2+\lambda_3 = 2q-2.7 \geq 0 \text{ or} \tag{A.4a}$$
$$\lambda_1+\lambda_2+\lambda_3 = 2q-2.7 < 0 \quad \text{and} \quad \Pi(q) = 2q^2-6.38q+4.828 < 0. \tag{A.4b}$$

Noting $q_- < 2.7/2$, (A.4) implies that $q > 2.7/2$ or $q_- < q < 2.7/2$. Thus, we have a sufficient condition for one of the real parts of the solutions to be positive is given by $q_- < q$. This proves the latter half. ■

References

Akerlof GA, Shiller RJ (2009) Animal spirits: how human psychology drives the economy, and why it matters for global capitalism. Princeton University Press, Princeton
Blanchard O (1981) What is left of the multiplier accelerator? Am Econ Rev 71:150–154

Blanchard O (1993) Consumption and the recession of 1990–1991. Am Econ Rev 83:270–274
Bothwell FE (1952) The method of equivalent linearization. Econometrica 20:269–283
Chang WW, Smyth DJ (1971) The existence and persistence of cycles in a non-linear model: Kaldor's 1940 model re-examined. Rev Econ Stud 38:37–44
Gabisch G, Lorenz H-W (1987) Business cycle theory: a survey of methods and concepts. Springer, Berlin
Gandolfo G (1996) Economic dynamics. third, completely revised and enlarged. Springer, Berlin
George D (1981) Equilibrium and catastrophes in economics. Scott J Polit Econ 28:43–61
Goodwin RM (1951) The nonlinear accelerator and the persistence of business cycles. Econometrica 19:1–17
Guckenheimer J, Holmes P (1983) Nonlinear oscillations, dynamical systems, and bifurcations of vector fields. Springer, New York
Kahneman D, Knetsch JL, Thaler RH (1991) Anomalies: the endowment effect, loss aversion, and status quo bias. J Econ Perspect 5:193–206
Kaldor N (1940) A model of the trade cycle. Econ J 50:78–92
Knight FH (1921) Risk, uncertainty and profit. Century Press, New York
Krugman P (1991) History versus expectations. Q J Econ 106:651–667
Krugman P (1996) The self-organizing economy. Blackwell, Oxford
Krugman P (2008) The return of depression economics and the crisis of 2008, 2nd edn. Penguin Books, London
Lorenz H-W (1993) Nonlinear dynamical economics and chaotic motion, 2nd edn. Springer, New York
Matthews RCO (1959) The trade cycle. Cambridge University Press, Cambridge
Murphy KM, Shleifer A, Vishny RW (1989) Industrialization and the big push. J Polit Econ 97:1003–1026
Owase T (1991) Nonlinear dynamical systems and economic fluctuations: a brief historical survey. IEICE Trans 74:1393–1400
Puu T (2003) Attractors, bifurcations and chaos: nonlinear phenomena in economics. Springer, Berlin
Puu T, Sushko I (2006) Business cycle dynamics: models and tools. Springer, Berlin
Rosser JB Jr (1991) From catastrophe to chaos: a general theory of economic discontinuities. Kluwer Academic Publishers, Boston
Sansone G, Conti R (1964) Non-linear Differential Equations. Pergamon Press, Oxford
Strotz RH, McAnulty JC, Naines JB (1953) Goodwin's nonlinear theory of the business cycle: an electro-analog solution. Econometrica 21:390–411
Varian HR (1979) Catastrophe theory and the business cycle. Econ Inq 17:14–28
Yanqian Y (1986) Theory of limit cycles. Translations of mathematical monographs, vol 66. American Mathematical Society, Providence

Delay Kaldor–Kalecki Model Revisited

Akio Matsumoto and Ferenc Szidarovszky

Abstract This chapter studies the dynamics of the Kaldor–Kalecki model of national income and capital stock. The investment function is assumed to have not only a Kaldorian characteristics, namely, a S-shaped form but also a Kaleckian characteristics, that is, a gestation delay between "investment decision" and "investment implementation." We divide the analysis into two parts. In the first part, we assume that the time period under consideration is short enough so that the capital stock is not affected by the flow of investment and then examine the delay effect on dynamics of national income. In the second part, taking the capital accumulation into account, we draw attention to how the delay affects cyclic dynamics observed in the nondelay Kaldor–Kalecki model. It is demonstrated that the investment delay quantitatively affects the dynamic behavior but not qualitatively.

1 Introduction

In real economy, macroeconomic variables such as national income, capital accumulation, interest rate, etc., exhibit cyclic fluctuations. As a natural consequence, it has been the main interest in studying macro economic dynamics to detect endogenous sources of such cyclic behavior. Since investment is considered to be the key

The authors highly appreciate the financial supports from the MEXT-Supported Program for the Strategic Research Foundation at Private Universities 2013–2017 and the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 24530202, 25380238 and 26380316). The usual disclaimers apply.

A. Matsumoto (✉)
Department of Economics, Institute of Further
Development of Dynamic Economic Research, Chuo University,
742-1, Higashi-Nakano, Hachioji, Tokyo 192-0393, Japan
e-mail: akiom@tamacc.chuo-u.ac.jp

F. Szidarovszky
Department of Applied Mathematics, University of Pécs,
Ifjúság u. 6, Pécs 7624, Hungary
e-mail: szidarka@gmail.com

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_11

factor to cyclic dynamics for the evolution of national economy, a lot of efforts has been devoted to studying investment determinations. As early as in 1930s, a few years before the publication of the Keynes' General Theory, Kalecki (1935) introduces an idea of the consumption function and the multiplier in his statistic analysis of macro economy. Furthermore, regarding investment in the dynamic analysis, he assumes a lag between "investment order" and "investment installation" and call it a gestation lag of investment.[1] Adopting a linear investment function, he constructs a macro dynamic model as a delay differential equation of retarded type and shows that it gives rise to a cycle due to the delay. Kaldor (1940), on the other hand, studies the evolution of production and capital formation and believes that nonlinearities of behavioral equations could be a clincher in such cyclic oscillations. He builds a 2D model of national income and capital with nonlinear investment and saving functions. Its basic movements consist of two sorts. One is the movements along the nonlinear functions and the other is shifts of these functions according to capital accumulations. So far, it has also been confirmed in various ways that the Kaldor model is capable of generating cyclic behavior when nonlinearities become strong enough. Indeed, Ichimura (1955) reduces the model to the Liénard equation, Chang and Smyth (1971) rigorously show the existence of a limit cycle by applying the Poincaré-Bendixson theorem and so does Lorenz (1993) by the Hopf bifurcation theorem. Furthermore, Grasman and Wentzel (1994) show multistability in the Kaldor model, that is, the coexistence of stable and unstable cycles when the equilibrium is locally stable. "Nonlinearity" and "delay" are now treated as two of the main ingredients for endogenous cycles.

More than a half century after Kaldor (1940), Krawiec and Szydlowski (1999) combine the Kaldor model with the Kaleckian time delay in investment to build a delay business cycle model. Their model is often called the (delay) *Kaldor–Kalecki* model. To emphasize the role of delay, they assume a linear investment function as in the Kalecki model and then show the occurrence of a limit cycle with respect to time delay. This model has been developed in various directions. Having an *S*-shaped investment function that reacts to delay output, Zhang and Wei (2004) investigate the local and global existence and the directions of Hopf bifurcation. Kaddar and Talibi Alaoui (2008) further introduce a time delay into capital stock of the investment function and establish sufficient conditions for the local existence of Hopf bifurcation. Wang and Wu (2009) rigorously examine the model with delay in both output and capital stock, applying the center manifold theory with the normal forms and show emergence of periodic solutions. Zhou and Li (2009) assume different gestation delays in output and capital stock and show the occurrence of Hopf bifurcation. In the analysis of these models, the time delay is treated as a bifurcation parameter. Recently Manjunath et al. (2014) exhibit that an exogenous and nondimensional parameter can be a source of cyclic dynamics of the delay Kaldor–Kalecki model.

[1]Since "lag" and "delay" do not have distinctive different meanings, we use these words interchangeably. In particular, we mainly use "delay" in this study.

We will revisit the original Kaldor–Kalecki model with one delay in output and present a new characterization of cyclic oscillations. As mentioned above, in the existing literature, Hopf bifurcation occurs due to the nonlinearity of the investment function with time delay in output (as well as capital stock). This finding indicates that the delay model may explain cyclic dynamic behavior of economic variables. In the existing literature, however, almost no attention is given to a delay effect in the short-run in which capital stock is constant. Furthermore, although the multistability of the Kaldor model is confirmed, it has not been discussed in the delay Kaldor–Kalecki model. We concern these issues, in particular, we are interested in how nonlinearity and delay are responsible for the birth of limit cycles in order to complement the existing literature. For this purpose, we first recapitulate the 2D Kaldor–Kalecki model and specify the investment function. Then we proceed to the analysis of delay dynamics with two steps. At the first step, we investigate dynamics of national income, keeping the stock of capital fixed. As is well known, no cyclic behavior arises in the Kaldor model without capital accumulation. We focus attention on the effect caused by delay on such stable dynamics. Since investment and savings are short term, eliminating capital or fixing capital implies short-run dynamic analysis. At the second step, we consider delay dynamics of the national income and capital accumulation. As is seen above, the nonlinear Kaldor model without delay can generate not only a single limit cycle but also multiple limit cycles while the linear delay Kaldor–Kalecki model also give rise to a limit cycle. A natural question we rise is *how the delay affects the multiple cyclic behavior in the long-run*.

This paper is organized as follows. In Sect. 2, the basic elements of the Kaldor–Kalecki model is rebuilt. In Sect. 3, short-run dynamics is examined under the investment delay. In Sect. 4, after reviewing the Kaldor model without delay quickly, we analytically and numerically detect the delay effect on cyclic dynamics of the Kaldor–Kalecki model from the long-run point of view in the sense that capital accumulation is explicitly taken into account. Section 5 contains concluding remarks and further research directions. The color versions of the figures of this chapter can be found in the e-version of this book.

2 Delay Dynamic Model

Kaldor (1940) relates investment to the level in income (i.e., profit principle) and extends it by proposing a sigmoidal, instead of linear, investment function. A brief description of his model is given by two equations,

$$\begin{aligned}\dot{Y}(t) &= \alpha\left[\Phi(Y(t), K(t)) - S(Y(t))\right] \\ \dot{K}(t) &= \Phi(Y(t), K(t)) - \delta K(t).\end{aligned} \tag{1}$$

where $Y(t)$ is national income at time t, $K(t)$ denotes capital, $\Phi(Y(t), K(t))$ is an investment function, $S(Y(t))$ is a savings function and the parameters α and δ denote the adjustment coefficient and the rate of depreciation. The first equation describes the national income adjustment process and the second describes capital accumulation process. As will be reviewed soon, the Kaldor model is able to generate endogenous limit cycles. There are two ingredients for the birth of the cycles, the nonlinearity of investment in Y and the dependency of investment in K. These two prevent Y and K for global divergence when the equilibrium is locally unstable. Using our notation and following his spirit, a Kaleckian investment function can be presented by

$$I(t) = \Phi(Y(t-\theta), K(t))$$

where θ denotes the gestation delay.[2] Replacing the investment function in the second equation of system (2) with Kaleckian function yields the Kaldor–Kalecki model of the income and capital stock,

$$\begin{aligned} \dot{Y}(t) &= \alpha\left[\Phi(Y(t), K(t)) - S(Y(t))\right], \\ \dot{K}(t) &= \Phi(Y(t-\theta), K(t)) - \delta K(t). \end{aligned} \tag{2}$$

We plan to analyze dynamics generated by (2) with two steps. At first step, we confine attention to short-run dynamics. To this end, we impose two assumptions:

Assumption 1 The time period under consideration is short enough so that the capital stock is not affected by the flow of investment.

Assumption 2 The investment function in the first equation of system (2) contains the delay.

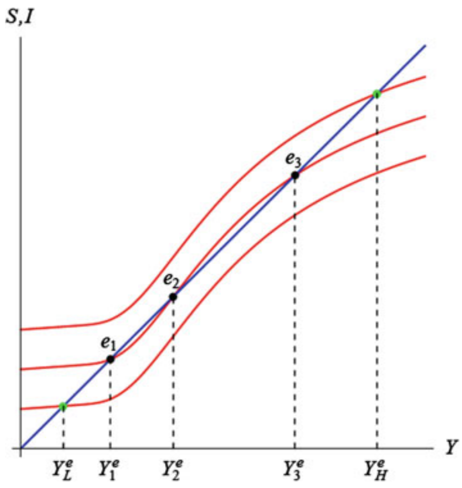
As a consequence of these modifications, the second equation concerning the evolution of capital is eliminated and the variable K in the investment function also disappears. Thus the modified Kaldor–Kalecki model can be presented by a one-dimensional nonlinear differential equation with one time delay,

$$\dot{Y}(t) = \alpha\left\{\varphi[Y(t-\theta)] - S[Y(t)]\right\} \tag{3}$$

After examining short-run dynamics, we proceed to the second step in which the dynamic analysis of $Y(t)$ and $K(t)$ generated by system (2) is investigated. It is considered to be long-run dynamics in the sense that the national income as well as the capital stocks evolve over time. We will numerically confirm the existing analytical results and then consider the effects of the delay on them.

[2] There are several extensions of this delay investment function. Kaddar and Talibi Alaoui (2008) introduce time delay also in capital stock in capital accumulation equation (i.e., $\Phi(Y(t-\theta), K(t-\theta))$). Zhou and Li (2009) assume that the investment function in the capital accumulation depends on the income and the capital stock at different gestation periods (i.e., $\Phi(Y(t-\theta_1), K(t-\theta_2))$ with $\theta_1 \neq \theta_2$).

Fig. 1 Determination of equilibrium with different values of $\bar{K}$



3 Short-Run Dynamics

We first make the following two assumptions for the sake of analytical simplicity.

Assumption 3 The saving function is linear and has no autonomous savings,

$$S(Y) = sY,\ 0 < s < 1.$$

Assumption 4 The investment function has the S-shaped form,

$$\varphi(Y) = A * 2^{-\frac{1}{(CY+D)^2}} + BY - \bar{K}.$$

Parameter $\bar{K}$ in $\varphi(Y)$ is positive implying the fixed level of the capital stock.[3] At a stationary state of Eq. (3), two conditions, $\dot{Y}(t) = 0$ and $Y(t) = Y(t-\theta) = Y^e$ for all $t \geq 0$, are satisfied. Economically, these conditions can be restated as investment is equal to saving at the equilibrium. In Fig. 1, we superimpose the linear saving function, sY with $s = 0.282$, on the three sigmoid investment functions with $A = 35,\ B = 0.02,\ C = 0.01,\ D = 0.00001$ and three different values of $\bar{K}$ (i.e., $\bar{K} = 5, 10, 15$). The equilibrium state is determined by the intersection of these curves.

When the fixed value of $\bar{K}$ is small, we will have a high level of investment and thus a high short-run equilibrium level Y_H^e of national income. As K increases, the investment curve shifts downward. Due to the S-shaped form, there is a case where the saving curve crosses the investment curve three times. These intersections are denoted by the black dotes and their x-coordinates are the corresponding equilibrium levels

[3]This is a simplified version of the investment function adopted in Lorenz (1987). It is replaced with the full version in the latter half of this chapter.

of national income, $Y_1^e < Y_2^e < Y_3^e$. A further increase of $\bar{K}$ shifts the investment curve downward enough resulting in only one intersection denoted by the lower green dotted point and the corresponding national income is Y_L^e. Investment is small here and thus the equilibrium level is also small.

We draw attention to stability of the three equilibrium points obtained under the middle value of $\bar{K}$. Let $Y_{\delta i} = Y - Y_i^e$ for $i = 1, 2, 3$. The linear approximation of Eq. (3) is

$$\dot{Y}_{\delta i}(t) = \alpha\eta_i Y_{\delta i}(t-\theta) - \alpha s Y_{\delta i}(t) \tag{4}$$

where $\eta_i = \varphi'(Y_i^e) > 0$ denotes the slope of the investment curve at the equilibrium income Y_i^e. Substituting an exponential solution, $e^{\lambda t}u$, yields the characteristic equation,

$$\lambda + \alpha s - \alpha\eta_i e^{\lambda\theta} = 0. \tag{5}$$

In the absence of time delay (i.e., $\theta = 0$), we simply have,

$$\lambda_i = \alpha(\eta_i - s). \tag{6}$$

If $\eta_i < s$ or $\lambda_i < 0$, then the equilibrium income Y_i^e is stable. Since investment is greater or less than savings in the left or right of the intersection, any nonequilibrium Y approaches its equilibrium level by the multiplier. Similarly it is locally unstable if $\eta_i > s$ or $\lambda_i > 0$. The sign of $\eta_i - s$ depends on the slopes of the investment and saving curves evaluated at the equilibrium point. Apparently, among three equilibrium values, $\eta_1 - s < 0$ at e_1, $\eta_2 - s > 0$ at e_2 and $\eta_3 - s < 0$ at e_3. The middle one (i.e., Y_2^e) is locally unstable and the remaining high and low ones (i.e., Y_1^e and Y_3^e) are locally stable. It is also clear that the unique equilibrium Y_L^e or Y_H^e is locally stable. As a benchmark, we summarize these results as follows.

Theorem 1 *When Eq. (4) with $\theta = 0$ has three equilibria, then the middle one is locally unstable while both of the larger and smaller ones are locally asymptotically stable.*

Kaldor (1940) focuses on the unstable equilibrium. The main key to it is the capital accumulation that shifts the short-term investment function over time which gives rise to a cycle. On the other hand, we focus on the stable equilibrium and consider whether the delay affects its stability. Hence we return to Eq. (5) with $\theta > 0$. It should be noticed that $\lambda = 0$ does not solve this equation unless $s = \eta_i$. So if stability of Y_i^e switches at $\theta = \bar{\theta}$, then Eq. (5) must have a pair of pure imaginary roots there. Since roots of a real function always come in conjugate pairs, we assume $\lambda = i\omega$ with $\omega > 0$. Substitution of this root divides Eq. (5) into the real and imaginary parts,

$$\begin{aligned} \alpha s - \alpha\eta_i \cos\theta\omega &= 0, \\ \omega + \alpha\eta_i \sin\theta\omega &= 0. \end{aligned} \tag{7}$$

Moving the first term in each equation to the right, squaring the resultant equations and adding them together yield

$$\omega^2 = \alpha^2(\eta_i + s)(\eta_i - s) \tag{8}$$

where the first two factors on the right-hand side are positive. If $\eta_i - s > 0$, then there is a positive ω and stability switch can occur. The condition, $\eta_i > s$, could be possible only at the middle equilibrium point Y_2^e. However, since it is already shown to be locally unstable for $\theta = 0$, Y_2^e still remains unstable for any $\theta > 0$. On the other hand, $\eta_i < s$ holds at the other equilibrium points. If $\eta_i - s \leq 0$, then there is no $\omega > 0$ implying that stability switch does not occur. Summarizing these results yields the following:

Theorem 2 *For any equilibria of the delay Eq. (3), no stability switch occurs for any positive value of the delay.*

Theorem 2 implies that the delay does not affect asymptotic dynamics of the delay model (3). In spite of this result, we show that it really matters in transient dynamics. In particular, following Beddington and May (1975), we show that the delay increases the real parts of the eigenvalues for a stable equilibrium point and decreases the magnitude for an unstable equilibrium point. We proceed to illustrate these effects of the investment delay in the three equilibria case. We assume that $\lambda = x + iy$ with $y \geq 0$ and substitute it into Eq. (5), After arranging the terms, we obtain

$$x + iy = -\alpha s + \alpha\eta_i e^{-x\theta}\cos y\theta + i\left(-\alpha\eta_i e^{-x\theta}\sin y\theta\right).$$

Comparing both sides finds that the real and imaginary parts are

$$\begin{aligned} x &= -\alpha s + \alpha\eta_i e^{-x\theta}\cos y\theta, \\ y &= -\alpha\eta_i e^{-x\theta}\sin y\theta, \end{aligned} \tag{9}$$

from which we derive the form of the real part depending on y and θ,

$$x = -\alpha s - y\cot y\theta. \tag{10}$$

Moving the first term in the right-hand side of the first equation in (9) to the left-hand side, squaring both sides of the resultant equation and adding the square of the second equations to it yield the imaginary part depending on x and θ,

$$y = \sqrt{(\alpha\eta_i)^2 e^{-2x\theta} - (x + \alpha s)^2}. \tag{11}$$

Solving Eq. (9) with $y = 0$ for x yields the real part functional for θ. The equation is

$$x + \alpha s = \alpha\eta_i e^{-x\theta} \tag{12}$$

which clearly has a unique real solution for x that is denoted by $x(\theta)$. For $\theta = 0$,

$$x_i(0) = \alpha(\eta_i - s)$$

which is identical with Eq. (6). By implicitly differentiating Eq. (12), we have that

$$\frac{dx(\theta)}{d\theta} = \alpha\eta_i e^{-\theta x(\theta)} \left(-\frac{dx(\theta)}{d\theta}\theta - x(\theta) \right)$$

implying that

$$\frac{dx(\theta)}{d\theta} = -x(\theta) \frac{\alpha\eta_i e^{-\theta x(\theta)}}{1 + \alpha\eta_i \theta e^{-\theta x(\theta)}}.$$

So if $x(\theta) > 0$, then $dx(\theta)/d\theta < 0$ implying that $x(\theta)$ decreases, and if $x(\theta) < 0$, then $dx(\theta)/d\theta > 0$ implying that $x(\theta)$ increases. In both cases $|x(\theta)|$ decreases. Hence we have Theorem 3.

Theorem 3 *Larger delays result in the smaller absolute value of the real parts and thus slow down convergence speed to stable equilibrium.*

4 Long-Run Dynamics

4.1 Kaldor Model

We review the original Kaldor model (1). So far, in this model, it is demonstrated that the nonlinearity of investment functions leads to the two remarkable results. One is the existence of a stable limit cycle shown by Chang and Smyth (1971) with applying the Poincáre-Bendixson theorem when the equilibrium is locally unstable. The other is the coexistence of a stable limit cycle and a unstable limit cycle by Grasman and Wentzel (1994) with the use of the Hopf bifurcation theorem when the equilibrium is locally stable. Figure 2 graphically confirms these results with the following form of the separable investment function,[4]

$$\Phi(Y, K) = 25 * 2^{-\frac{1}{(0.015Y+0.00001)^2}} + 0.05Y + 5\left(\frac{320}{K}\right)^3.$$

and $\alpha = 3$. The positive sloping dashed curve is the $\dot{K} = 0$ locus and the convex–concave-dashed curve is the $\dot{Y} = 0$ locus. The intersection of these curves determines the stationary equilibrium point denoted by $(Y^e, K^e) \simeq (324.32, 54.05)$. In Fig. 2a where we take $s = 0.3$, the stationary point is locally unstable as we will see shortly and two trajectories starting in a neighborhood of the stationary point explosively oscillate and approach the limit cycle. In Fig. 2b, s is decreased to 0.282 and all other parameters are kept fixed. The equilibrium point becomes stable, which is enclosed by an unstable inner limit cycle which is, in turn, enclosed by an outer stable limit

[4] Lorenz (1987) uses this form of the function to show the occurrence of chaotic motion in a multisector Kaldorian business cycle model. Grasman and Wentzel (1994) also use this form.

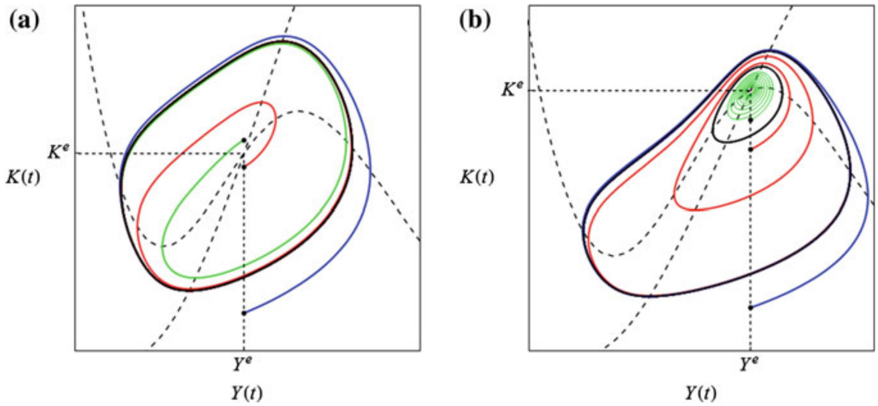


Fig. 2 Limit cycles in the Kaldor model. **a** One limit cycle, **b** Two limit cycles

cycle. A green trajectory starting inside of the inner limit cycle converges to the equilibrium point while both the red trajectory starting outside of the inner limit cycle and the blue trajectory starting outside of the outer limit cycle approaches the outer cycle. The black dots denote the initial points in simulations.

Kaldorian nonlinear dynamics is often examined with respect to the value of the adjustment coefficient α.[5] Figure 2 indicates that Kaldorian dynamics has also strong sensitivity to the value of s. Two different dynamics illustrated in Fig. 2 imply that there is a threshold value of s and the changing of the parameter through this value causes a qualitative change in the nature of dynamics. A bifurcation diagram gives good insights into what is happening to evolution of the equilibrium point as the value of the parameter is changed. In Fig. 3a, the value of s is increased from 0.25 to 0.38 with an increment 1/10000. For each value of s, the dynamic system (1) runs for $t \in [0, 500]$ and the local maximum and minimum values of the specified trajectory for $400 \leq t \leq 500$ are plotted. If the diagram shows one point against s, it implies that the equilibrium is stable and the trajectory converges to it. If it shows two points, then the trajectory has one maximum and one minimum point, implying an emergence of a limit cycle. Figure 3a roughly indicates that the equilibrium point is locally stable for smaller values of s, loses its stability bifurcating to a limit cycle for medium values and then regains stability for larger values. Figure 3b is an enlargement of Fig. 3a around the two threshold values, s_α and s_β. If $s < s_\alpha$, the Kaldor system generates a stable equilibrium. As s passes s_β, the trajectories of the system converges to a stable limit cycle with the radius equal to the distance between the upper and lower red branches. On the other hand, for $s \in [s_\alpha, s_\beta] \simeq [0.281, 0, 283]$,[6] cyclic dynamics can emerge. Notice that the dotted vertical lime at $s = 0.3$ crosses the bifurcation diagram twice in Fig. 3b and the corresponding limit cycle is depicted in

[5]See, for example, Lorenz (1993).

[6]The value of s_a is numerically obtained and the value of s_β is analytically determined as will be seen.

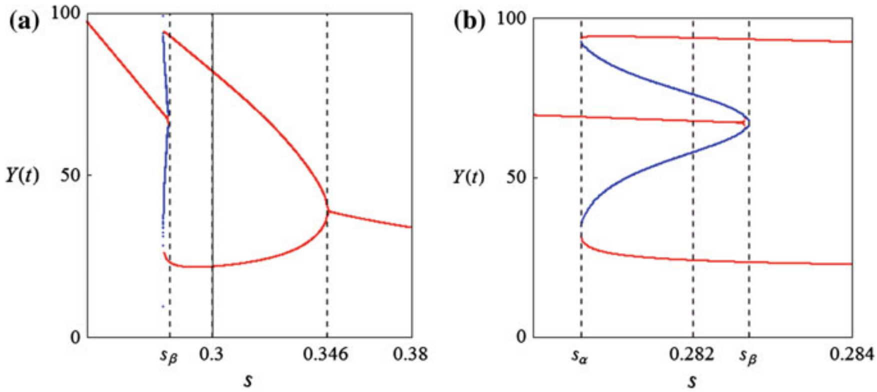


Fig. 3 Bifurcation diagrams with respect to s. **a** $0.25 \leq s \leq 0.38$, **b** $0.28 \leq s \leq 0.284$

Fig. 2a. Similarly, the dotted vertical line at $s = 0.282$ crosses the blue curve twice and the red curve three times in Fig. 3b and the subcritical Hopf bifurcation leads to multistability, that is, the coexistence of a stable equilibrium, an unstable limit cycle and a stable limit cycle as illustrated in Fig. 2b. A distance between the upper and lower blue branches corresponds to the radius of the unstable limit cycle.

4.2 Kaldor–Kalecki Model

We now turn attention to the Kaldor–Kalecki model (2) that has the same stationary equilibrium point, (Y^e, K^e), as the Kaldor model (1). As is already mentioned, Krawiec and Szydlowski (1999) show the existence of limit cycle in the Kaldor–Kalecki model. In this section, we revisit this property and compare the results in the nondelay model with the results in the delay model to find how the delay affect dynamics.

Let $Y_\delta = Y - Y^e$ and $K_\delta = K - K^e$. By linearizing (2) at the equilibrium point, we have

$$\begin{aligned}\dot{Y}_\delta(t) &= \alpha\left[(\eta - s)Y_\delta(t) - \beta K_\delta(t)\right],\\ \dot{K}_\delta(t) &= \eta Y_\delta(t-\theta) - (\beta+\delta)K_\delta(t),\end{aligned} \tag{13}$$

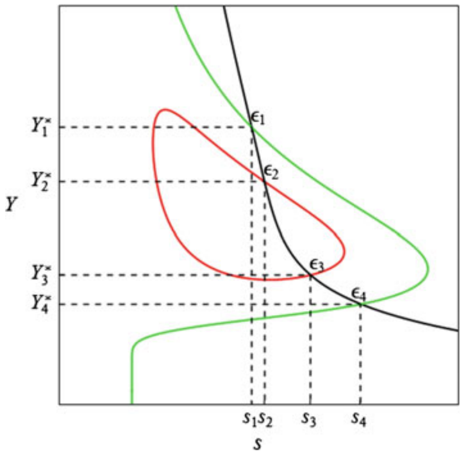
where

$$\eta = \frac{\partial \Phi}{\partial Y} = 0.05 + \frac{0.52 \times 2^{-\frac{1}{(0.015Y^e+0.00001)^2}}}{(0.015Y^e + 0.00001)^3}$$

and

$$\beta = -\frac{\partial \Phi}{\partial K} = -\frac{15 \times (320)^3}{(K^e)^4}.$$

Fig. 4 Stability and instability regions



Notice that the values of β and η depend on the the point where they are evaluated, although their dependency is not explicitly expressed in the following. Suppose exponential solutions, $Y_\delta(t) = e^{\lambda t}u$ and $K_\delta(t) = e^{\lambda t}v$. Then the characteristic equation is written as

$$\lambda^2 + a\lambda + b + ce^{-\lambda\theta} = 0 \tag{14}$$

where

$$a = (\beta + \delta) - \alpha(\eta - s),$$

$$b = -\alpha(\eta - s)(\beta + \delta)$$

and

$$c = \alpha\beta\eta.$$

We first examine the nondelay case (i.e., $\theta = 0$) in which the corresponding characteristic equation is reduced to

$$\lambda^2 + a\lambda + b + c = 0.$$

The necessary and sufficient conditions for the roots of the quadratic equation to be negative if real and to have negative real parts if complex are $a > 0$ and $b + c > 0$. Under the specified values of the parameters, $b + c > 0$ always. In Fig. 4, the negative sloping black curve is the locus of (s, Y^e) and the closed red curve is the locus of $a = 0$. The black locus crosses the red curve four times at $s = s_i$ for $i = 1, 2, 3, 4$.[7] It is verified that $a < 0$ inside and $a > 0$ outside.[8] Hence, the stationary equilibrium

[7]Notice that s_2 is identical with s_β. See footnote 6.

[8]We use the green curve later when the delay model is examined.

obtained along the black curve is locally stable for $s < s_2$ and $s > s_3$ and locally unstable for $s_2 < s < s_3$ where

$$s_1 \simeq 0.265,\ \ s_2 \simeq 0.283,\ \ s_3 \simeq 0.346 \text{ and } s_4 \simeq 0.416.$$

We now return to Eq. (14) and examine the delay case (i.e., $\theta > 0$). Suppose that $\lambda = i\omega$ with $\omega > 0$ is a solution of the equation for some $\theta > 0$. Substituting it into the equation and separating the real and imaginary parts present

$$\begin{aligned} -\omega^2 + b + c\cos\omega\theta = 0, \\ \omega a - c\sin\omega\theta = 0. \end{aligned} \tag{15}$$

Thus

$$c^2 = (\omega^2 - b)^2 + (\omega a)^2.$$

Hence

$$\omega^4 - (2b - a^2)\omega^2 + b^2 - c^2 = 0$$

and its roots are

$$\omega_{\pm}^2 = \frac{2b - a^2 \pm \sqrt{(2b - a^2)^2 - 4(b^2 - c^2)}}{2}$$

where

$$2b - a^2 = -\left[(\beta + \delta)^2 + \alpha^2(\eta - s)^2\right] < 0.$$

If $b^2 - c^2 < 0$, then $\omega_+ > 0$ and there is only one imaginary solution, $\lambda = i\omega_+$. On the other hand, if $b^2 - c^2 \geq 0$, then both roots are negative or complex so no imaginary solution exist. In Fig. 4, the green curve is the locus of $b^2 - c^2 = 0$. It is verified that $b^2 - c^2 > 0$ in the left side of the green curve and $b^2 - c^2 < 0$ in the right side. Thus stability of Y^e along the black curve is affected by the value of s. Along the black curve, stability of Y^e (as well as stability of K^e) can be switched to instability for $s_1 < s < s_2$ and $s_3 < s < s_4$ for some value of θ while Y^e is locally unstable for $s_2 < s < s_3$ regardless of the value of θ. In case of stability switch, solving the first equation of Eq. (15) for θ yields the partition curve[9]

$$\theta = \frac{1}{\omega_+}\cos^{-1}\left(\frac{\omega_+^2 - b}{c}\right) \tag{16}$$

that divides the parameter region into two subregions, stability is reserved in one subregion and stability is lost in the other subregion. Figure 5 presents the division of the (s, θ) plane. For $\theta = 0$ (i.e., no-delay case), as it is already confirmed, stability is lost for $s_2 < s < s_3$. As θ increases and becomes positive, we have two results.

[9]Solving the second equation yields the same partition curve in a different form.

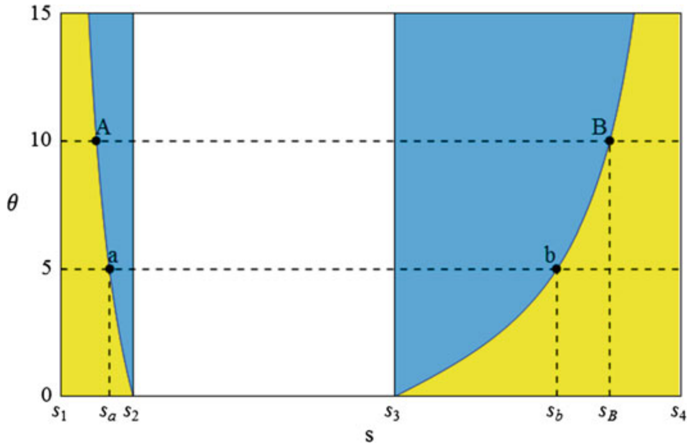


Fig. 5 Partition of the (s, θ) plane

One is that as far as $s \in [s_2, s_3]$, the equilibrium is locally unstable regardless of the values of θ (i.e., in the white region of Fig. 5). The other is that the instability interval of s becomes larger. Stability is preserved in the yellow region and lost in the blue region. The boundary of these regions is the partition curve described by Eq. (16) that is downward-sloping for $s \in [s_1, s_2]$ and upward-sloping for $s \in [s_3, s_4]$. The blue regions are the enlarged instability regions due to the positive delay.

Figure 6 gives the bifurcation diagrams with respect to s and reveals the effects caused by increasing θ on dynamics with respect to s from a different view point. The red curve is for $\theta = 0$ and is identical with the one given in Fig. 3a although multistability phenomenon occurred around s_2 is omitted for the sake of graphical simplicity. The blue curve is for $\theta = 5$ where s is increased along the dotted line at $\theta = 5$ in Fig. 5 where the dotted line crosses the partition curves at points a and b. Let us denote the s-values of the intersections by $s_a \simeq 0.277$ and $s_b \simeq 0.386$. Stability is lost for $s = s_a$ and regained for $s = s_b$. The similar bifurcation cascade is obtained for $\theta = 10$ and described by the green curve. The dotted line at $\theta = 10$ crosses the partition curves at points A and B whose s-values are $s_A \simeq 0.274$ and $s_B \simeq 0.399$. [10] Stability is lost for $s = s_A$ and regained for $s = s_B$. Since qualitatively different dynamics arises according to $s < s_2$ or $s > s_2$, we first consider dynamics for $s > s_2$. In Fig. 6a where the bifurcation diagrams are expanding as θ increases, we observe the following:

(i) the equilibrium point bifurcates to a limit cycle when s crosses the downward-sloping partition curve;
(ii) the amplitude (or ups and downs) of the cycle becomes larger as delay becomes larger as illustrated by the expansion of the bifurcation diagrams;

[10] In Fig. 6, s_a and s_A are not labeled to avoid confusion.

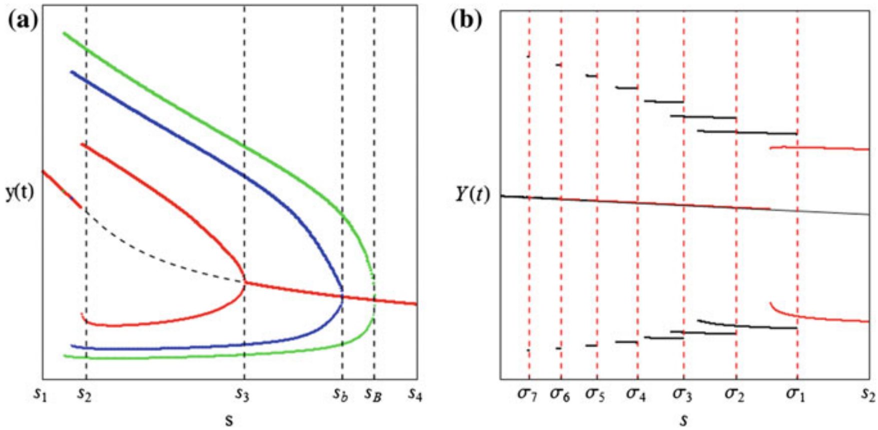


Fig. 6 Bifurcation diagrams with respect to s. **a** Shift due to delay, **b** Multistability

(iii) the stability-regain value of s increases as θ becomes larger, implying that the larger delay has the stronger destabilizing effect by expanding the instability region more.

We turn attention to dynamics for $s < s_2$ to find that it is harder to generate multistability as θ becomes larger. In Fig. 6b the red curve describes the bifurcation diagrams for $s < s_2$, which is the same as Fig. 3b without the blue curves. We increase the value of θ from 1 to 7 and denote the corresponding threshold values of s determined by Eq. (16) as σ_k for $k = 1, 2, ..., 7$. Notice that stability is lost for $s = \sigma_k$ for $\theta = k$ since the real parts of the eigenvalues are zero. The black curves ending at the red dotted line at $s = \sigma_k$ imply that multistability occurs for s between the starting point of the black curves and the ending point for $\theta = k$. It is observed that the lengths of the black curves become shorter with larger length of delay. Furthermore, for $\theta = 7$, the black curves almost disappear. Therefore it becomes harder to have multistability as θ increases.

We perform two more numerical simulations to see how the stable equilibrium is destabilized via increasing value of θ (i.e., the delay effect). In the above simulations, we change the value of s with fixed value of θ. In these simulations, we increase the value of θ, taking the value of s fixed. In the first example, we take $\theta = 5$ and $s \simeq 0.277 \in [s_1, s_2]$ that is obtained via Eq. (16). For $\theta = 0$, the equilibrium is locally asymptotically stable. When θ arrives at the downward-sloping partition curve at $\theta = 5$, stability is lost and further increased θ induces the equilibrium to bifurcate to a limit cycle. Figure 7a shows a bifurcation diagram with respect to θ. It is seen first that the red curve jumps to a limit cycle at $\theta_1 = 5$ via a subcritical Hopf bifurcation. It is further seen that the blue curve appears for $\theta_0 \simeq 3.45$, implying the occurrence of multistability for $\theta \in [\theta_0, \theta_1]$. It is also verified that the occurrence of multistability becomes harder as the value of θ increases. In the second example, we change the value of s to $s \simeq 0.386 \in [s_3, s_4]$. The increasing θ arrives at the upward-

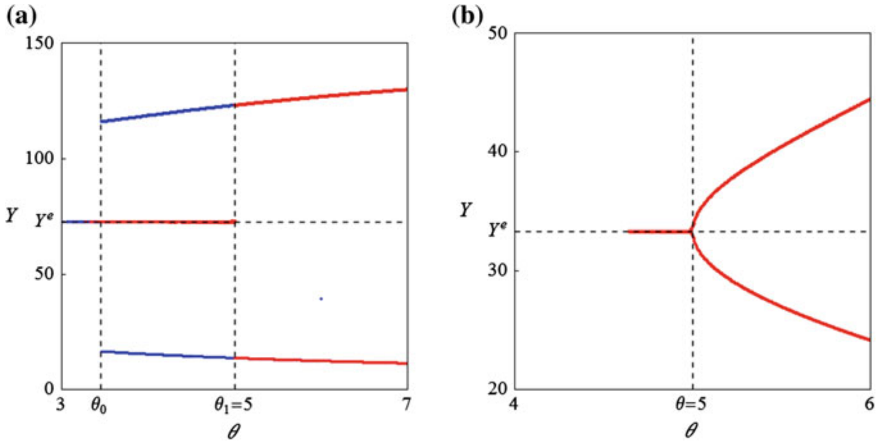


Fig. 7 Bifurcation diagrams with respect to θ. **a** $s \simeq 0.277$, **b** $s \simeq 0.386$

sloping partition curve for $\theta_1 = 5$. Figure 7b indicates that the stable equilibrium loses stability at $\theta = \theta_1$ and bifurcates to a limit cycle via supercritical Hopf bifurcation. Summarizing the delay effects obtained above gives the following results:

(i) The equilibrium point bifurcates to a limit cycle via supercritical Hopf bifurcation when increasing θ crosses the upward-sloping partition curve and via subcritical Hopf bifurcation when it crosses the downward-sloping partition curve;
(ii) Multistability can happen with respect to delay.

5 Concluding Remark

We investigate the Kaldor–Kalecki model in which the investment function has a S-shaped form and a gestation lag of investment between "investment decision" and "investment installation." The main analysis can be divided into two parts. In the first part, with a constant level of the capital stock, short-run dynamics of national income is examined and two results are obtained. First, the delay does not affect asymptotical dynamics in the sense that no stability switch occurs for any values of the delay. Second, the convergence speed gets faster as the delay becomes larger. In the second part, evolution of national income and the capital accumulation are simultaneously examined. Two nonlinear phenomenon, the birth of a limit cycle and coexistence of stable and unstable limit cycles around the stabile equilibrium point, both of which can emerge without delay, are preserved even if the delay is introduced. However, it is numerically confirmed that the amplitude of trajectory's fluctuations becomes larger as the delay becomes larger. It can be concluded that the delay affects the long-run as well as short-run dynamics quantitatively but not qualitatively.

References

Beddington J, May R (1975) Time delays are not necessarily destabilizing. Math Biosci 27:109–117
Chang W, Smyth D (1971) The existence and persistence of cycles in a nonlinear model: Kaldor's 1940 model re-examined. Rev Econ Stud 38:37–44
Grasman J, Wentzel J (1994) Co-existence of a limit cycle and an equilibrium in Kaldor's business cycle model and its consequences. J Econ Behav Organ 24:369–377
Ichimura S (1955) Toward a general nonlinear macrodynamics theory of economic fluctuations. In: Kurihira K (ed) Post-keynesian economics. George Allen & Unwin, London, pp 192–226
Kaddar A, Talibi Alaoui H (2008) Hopf bifurcation analysis in a delayed Kaldor-Kalecki model of business cycle. Nonlinear Anal Model Control 13:439–449
Kaldor N (1940) A model of the trade cycle. Econ J 50:78–92
Kalecki M (1935) A macrodynamic theory of business cycles. Econometrica 3:327–344
Krawiec A, Szydlowski M (1999) The Kaldor-Kalecki business cycle model. Ann Oper Res 89:89–100
Lorenz H-W (1993) Nonlinear dynamical economics and chaotic motion, second, revised and enlarged. Springer, Berlin
Lorenz H-W (1987) Strange attractors in a multisector business cycle model. J Econ Behav Organ 8:397–411
Manjunath S, Ghosh D, Faina G (2014) A Kaldor-Kalecki model of business cycles: stability and limit cycles, IEEEXplore, pp 2650–2656
Wang L, Wu X (2009) Bifurcation analysis of a Kaldor-Kalecki model of business cycle with time delay. Electron J Quant Theory Differ Equ Spec 27:1–20
Zhang C, Wei J (2004) Stability and bifurcation analysis in a kind of business cycle model with delay. Chaos Solitions Fractals 22:883–896
Zhou L, Li Y (2009) A dynamic IS-LM business cycle model with two time delays in capital accumulation equation. J Comput Appl Math 228:182–187

Two Time Lags in the Public Sector: Macroeconomic Stability and Complex Behaviors

Eiji Tsuzuki

Abstract This chapter develops a macroeconomic model that considers two time lags in the public sector—a government expenditure lag and a tax collection lag—and examines the effects of these lags on local stability of the steady state. According to previous studies, a sufficiently large expenditure lag causes economic instability. However, we show that a tax collection lag can have a stabilizing effect on the steady state. In addition, we develop an analysis of global dynamics to demonstrate that an increase in a tax collection lag can yield to complex behaviors.

Keywords Keynesian macrodynamic model · Fiscal policy lag · Delay differential equation · Stability analysis

JEL Classification E12 · E30 · E62

1 Introduction

Recently, many studies have examined the effects of time lags on macroeconomic stability using traditional Keynesian models. For instance, Sportelli and Cesare (2005) introduce a tax collection lag into the dynamic IS-LM model developed by Schinasi (1981) and Sasakura (1994), which is a traditional Keynesian model, and examine the local and global dynamics of the system. The standard dynamic IS-LM model with no policy lag comprises three equations that represent the goods market, monetary market, and budget constraints of the consolidated government. These equations form an ODE (ordinary differential equations) system. The introduction of a time

E. Tsuzuki (✉)
Faculty of Economics, Nanzan University, Nagoya 466-8673, Japan
e-mail: tsuzukie5@gmail.com

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_12

lag transforms this system from an ODE to a DDE (delay differential equations) system.[1]

Generally, models with a time lag can be categorized into two types: fixed lag model and distributed lag model. Fanti and Manfredi (2007) develop a dynamic IS-LM model with a distributed tax collection lag, whereas Sportelli and Cesare (2005) analyze the case of a fixed lag. Both these studies demonstrate that a time lag evidently causes complex behaviors, including chaos, and that a traditional fiscal policy is likely to be ineffective. Moreover, Matsumoto and Szidarovszky (2013) compare the case of a fixed lag with that of a distributed lag in tax collections. They demonstrate that a larger stable region can be established in the case of a fixed lag compared with a distributed lag.

Another type of traditional Keynesian macrodynamic model that incorporates a capital accumulation equation in place of the disequilibrium adjustment function of the monetary market, which is often termed the Kaldorian model, has been proposed. This model originated from Kaldor (1940) and its primary characteristic is found in the assumption of an S-shaped configuration of the investment function. Chang and Smyth (1971) reconstruct the Kaldorian model to form an ODE system. Asada and Yoshida (2001) introduce a fixed government expenditure lag into the model proposed by Chang and Smyth (1971) and show that an increase in the responsiveness of a fiscal policy could lead to economic instability.

Further, Gabisch and Lorenz (1989) propose a hybrid model of the standard dynamic IS-LM model and the Kaldorian model, which involves both functions of capital accumulation and disequilibrium adjustment in the monetary market. Cai (2005) and Neamţu et al. (2007) introduce a fixed capital accumulation lag and a fixed tax collection lag, respectively, into this hybrid model and comprehensively discuss the occurrence of a Hopf bifurcation.

Moreover, Zhou and Li (2009) and Sportelli et al. (2014) propose macrodynamic models with two fixed time lags. Zhou and Li (2009) develop Cai's (2005) model to include two capital accumulation lags. In addition, Sportelli et al. (2014) present a dynamic IS-LM model with two time lags in the public sector: a government expenditure lag and a tax collection lag. These studies demonstrate that the steady states fluctuate between stability and instability as a certain lag increases.

Unlike in Sportelli et al. (2014), this study uses the Kaldorian macrodynamic model to investigate the interaction of two time lags in the public sector. Therefore, our model can be considered as introducing a tax collection lag into Asada and Yoshida's (2001) model. We examine two cases where a fiscal policy is active and where it is passive. An active fiscal policy strongly responds to the national income, whereas a passive fiscal policy is less responsive to the national income. In addition, we perform a stability analysis employing a mathematical method developed by Gu et al. (2005). This method enables us to present an exact figure of a stability crossing

[1] Schinasi (1981) does not consider disequilibrium of the monetary market. Sasakura (1994) develops Schinasi's (1981) model by introducing a disequilibrium adjustment function of the monetary market. Sasakura's (1994) model is now used as a benchmark of the dynamic IS-LM model.

curve—a curve that separates stable and unstable regions on a parameter plane. Few studies have employed this method for economic analysis.[2]

This study proceeds as follows: Sect. 2 presents a dynamic system that represents a model economy. Section 3 examines the local dynamics around the steady state. Subsequently, Sect. 4 examines the global dynamics. Section 5 presents our conclusion.

2 The Model

2.1 *Dynamic System*

The model economy comprises the following equations:

$$\dot{Y}(t) = \alpha[C(t) + I(t) + G(t) - Y(t)];\ \alpha > 0, \tag{1}$$

$$C(t) = c[Y(t) - T(t)] + \bar{C};\ 0 < c < 1;\ \bar{C} > 0, \tag{2}$$

$$T(t) = \tau Y(t - \theta_2) - \bar{T};\ 0 < \tau < 1;\ \bar{T} \geq 0, \tag{3}$$

$$I(t) = I(Y(t), K(t), r(t));\ I_Y > 0;\ I_K < 0;\ I_r < 0, \tag{4}$$

$$\dot{K}(t) = I(Y(t), K(t), r(t)), \tag{5}$$

$$G(t) = \beta[\bar{Y} - Y(t - \theta_1)] + \bar{G};\ \beta > 0;\ \bar{Y} > 0;\ \bar{G} > 0, \tag{6}$$

$$M(t)/P(t) = L(Y(t), r(t));\ L_Y > 0;\ L_r < 0, \tag{7}$$

$$M(t) = P(t)\gamma[\bar{Y} - Y(t)] + \bar{M};\ \gamma > 0;\ \bar{M} > 0, \tag{8}$$

$$P(t) = P(Y(t));\ P_Y > 0, \tag{9}$$

where Y = real national income (output); C = real private consumption; I = real private investment; G = real government expenditure; T = real income tax; K = real capital stock; M = nominal money supply; P = price level; r = nominal interest rate; α = adjustment speed of the goods market; c = marginal propensity to consume; $\bar{C}$ = base consumption; τ = marginal tax rate; $\bar{T}$ = real subsidy; β = responsiveness

[2]We shall refer other Keynesian macrodynamic models that consider a time lag as follows. The time-to-build model developed by Kalecki (1935) is the basis of economic models with a fixed time lag. Szydlowski (2002, 2003) develops this model into models with economic growth. Moreover, Yoshida and Asada (2007) examine the effects of a lag in government expenditure (where they examine both distributed and fixed lags) using the so-called Keynes–Goodwin model. Further, Asada and Matsumoto (2014) introduce a distributed lag of monetary policy implementation into the Keynesian equilibrium model proposed by Asada (2010). Asada's (2010) model comprises a monetary policy rule and an expectation adjustment function. A fixed lag version of this model is proposed by Tsuduki (2015). Furthermore, Matsumoto and Szidarovszky (2014) develop a nonlinear multiplier-accelerator model with investment and consumption lags. Finally, Bellman and Cooke (1963) provide a helpful introductory textbook of delay differential equations (i.e., differential-difference equations).

of government expenditure to national income (i.e., activeness level of the fiscal policy); $\bar{Y}$ = target level of real national income; $\bar{G}$ = target level of real government expenditure; γ = responsiveness of nominal money supply to national income (i.e., activeness level of the monetary policy); $\bar{M}$ = target level of nominal money supply; t = time; θ_1 = government expenditure lag; and θ_2 = tax collection lag.

Equations (1) and (2) represent a disequilibrium adjustment function of the goods market and a consumption function, respectively. Equation (3) is a tax collection function that represents income tax T as a function of past national income $Y(t-\theta_2)$. It may be more general to formulate T as a function not only of a past income but also of the present income denoted by $Y(t)$. However, this change does not affect the nature of our argument; hence, we simply assume that T is a function only of $Y(t-\theta_2)$. Equations (4) and (5) represent an investment function and a capital accumulation function, respectively. For simplicity, we assume that capital depreciation does not exist. Equation (6) represents a fiscal policy reaction function with a government expenditure lag. Equation (7) represents the monetary market equilibrium condition, where the left-hand side denotes real money balance and the right-hand side denotes a demand function for money. In this study, we ensure that the adjustment of the monetary market is rapid, and therefore, the balance of demand and supply of this market is always maintained. Equation (8) represents a monetary policy reaction function. Finally, Eq. (9) represents an aggregate supply function, by which the price level is determined.

In the case of no tax collection lag (i.e., $\theta_2 = 0$), the system compounded from Eqs. (1)–(9) essentially becomes similar to that of Asada and Yoshida (2001). However, the existence of a positive θ_2 significantly complicates the dynamic property of the system, thereby resulting in a major change in the economic implication of time lags.

2.2 Summarizing the Equations

In this section, we summarize Eqs. (1)–(9) in a two-dimensional dynamic system. Substituting Eqs. (8) and (9) into (7) and solving for r, we obtain

$$r(t) = r(Y(t)), \tag{10}$$

where $r_Y = -(\gamma P^2 + P_Y \bar{M} + P^2 L_Y)/P^2 L_r > 0$.

Substituting Eq. (3) into (2) and substituting Eq. (10) into (4), we obtain

$$C(t) = cY(t) - c\tau Y(t-\theta_2) + \bar{C} + c\bar{T}, \tag{11}$$

$$I(t) = I(Y(t), K(t), r(Y(t))). \tag{12}$$

Finally, substituting Eqs. (6), (11), and (12) into (1) and substituting Eq. (12) into (5) yields the following system of differential equations with two time lags:

$$\begin{aligned}\dot{Y}(t) &= \alpha[I(Y(t), K(t), r(Y(t))) - (1-c)Y(t) - \beta Y(t-\theta_1) - c\tau Y(t-\theta_2) \\ &\quad + \bar{C} + c\bar{T} + \beta\bar{Y} + \bar{G}], \\ \dot{K}(t) &= I(Y(t), K(t), r(Y(t))).\end{aligned} \tag{13}$$

2.3 Linearization

To analyze the local dynamics of System (13), we linearize the system around the steady state (Y^*, K^*) and obtain

$$\begin{aligned}\dot{\hat{Y}}(t) &= \alpha[\{A_1 - (1-c)\}\hat{Y}(t) - \beta\hat{Y}(t-\theta_1) - c\tau\hat{Y}(t-\theta_2) + I_K\hat{K}(t)], \\ \dot{\hat{K}}(t) &= A_1\hat{Y}(t) + I_K\hat{K}(t),\end{aligned} \tag{14}$$

where $\hat{Y}(t) = Y(t) - Y^*$, $\hat{K}(t) = K(t) - K^*$, and $A_1 = I_Y + I_r r_Y$. By necessity, the coefficients of these equations are evaluated at the steady state.

Assuming the exponential functions $\hat{Y}(t) = C_1 e^{\lambda t}$ and $\hat{K}(t) = C_2 e^{\lambda t}$ (where C_1 and C_2 are arbitrary constants, and λ denotes the eigenvalue) as the solutions of the above system and substituting these into System (14), we obtain

$$\begin{bmatrix}\lambda - \alpha\{A_1 - (1-c)\} + \alpha\beta e^{-\theta_1\lambda} + \alpha c\tau e^{-\theta_2\lambda} & -\alpha I_K \\ -A_1 & \lambda - I_K\end{bmatrix}\begin{bmatrix}\hat{Y}(t) \\ \hat{K}(t)\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix}.$$

For nontrivial solutions to exist for this system, the determinant of the left-hand side matrix, denoted by $\Delta(\lambda)$, must equal zero; i.e.,

$$\begin{aligned}\Delta(\lambda) &= \lambda^2 - [I_K + \alpha\{A_1 - (1-c)\}]\lambda - \alpha(1-c)I_K \\ &\quad + \alpha\beta(\lambda - I_K)e^{-\theta_1\lambda} + \alpha c\tau(\lambda - I_K)e^{-\theta_2\lambda} = 0 \\ &= p_0(\lambda) + p_1(\lambda)e^{-\theta_1\lambda} + p_2(\lambda)e^{-\theta_2\lambda} = 0,\end{aligned} \tag{15}$$

where

$$\begin{aligned}p_0(\lambda) &= \lambda^2 + b_1\lambda + b_2, \\ b_1 &= -[I_K + \alpha\{A_1 - (1-c)\}], \\ b_2 &= -\alpha(1-c)I_K, \\ p_1(\lambda) &= \alpha\beta(\lambda - I_K), \\ p_2(\lambda) &= \alpha c\tau(\lambda - I_K).\end{aligned}$$

Equation (15) is a characteristic equation of System (14). The significant feature of this equation is the existence of the exponential terms ($e^{-\theta_1\lambda}$ and $e^{-\theta_2\lambda}$).

First, we examine the case with no time lags. When $\theta_1 = \theta_2 = 0$, Eq. (15) can be rewritten as follows:

$$\Delta(\lambda) = \lambda^2 + (b_1 + \alpha(\beta + c\tau))\lambda + b_2 - \alpha I_K(\beta + c\tau) = 0, \tag{16}$$

which is an ordinary quadratic equation of λ.

Thus, we can state that if $b_1 + \alpha(\beta + c\tau) > 0$ (i.e., the coefficient of λ from Eq. 16 is positive), the real parts of the roots of Eq. (16) are negative.[3] In contrast, if $b_1 + \alpha(\beta + c\tau) < 0$, then the real parts of the roots are positive. Therefore, if $b_1 + \alpha(\beta + c\tau) > 0$, the steady state is locally stable, and if $b_1 + \alpha(\beta + c\tau) < 0$, it is unstable.

In the discussion below, we assume the following condition:

Assumption 2.1 $b_1 + \alpha(\beta + c\tau) > 0$.

This assumption implies that if a lag does not exist in the public sector, an economy is stable. Under this assumption, we analyze the effects of the lags (θ_1, θ_2) on local stability.

3 Local Dynamics

The following analysis is performed based on the technique developed by Gu et al. (2005).

3.1 Preconditions

First, to apply the technique of Gu et al. (2005), some preconditions should be checked. According to their study, Eq. (15) should satisfy the following conditions:

(I) $\deg(p_0(\lambda)) \geq \max\{\deg(p_1(\lambda)), \deg(p_2(\lambda))\}$;
(II) $\Delta(0) \neq 0$;
(III) a solution common to all three polynomials $p_0(\lambda) = 0$, $p_1(\lambda) = 0$, and $p_2(\lambda) = 0$ does not exist;
(IV) $\lim_{\lambda \to \infty}(|p_1(\lambda)/p_0(\lambda)| + |p_2(\lambda)/p_0(\lambda)|) < 1$.

In our system, Condition (I) is satisfied by $2 > \max\{1, 1\}$. Condition (II) is also satisfied by $\Delta(0) = \alpha I_K[-(1 - c) - \beta - c\tau] > 0$. Concerning Condition (III), we can check as follows: substituting I_K into $p_1(\lambda)$ and $p_2(\lambda)$, we obtain $p_1(I_K) = p_2(I_K) = 0$. However, $p_0(I_K) = -\alpha A_1 I_K \neq 0$. Hence, Condition (III) is satisfied.

[3] See Chap. 18 in Gandolfo (2010) for details of the relationship between the roots and coefficients of a quadric equation.

Finally, Condition (IV) is satisfied by $\lim_{\lambda\to\infty}(|p_1(\lambda)/p_0(\lambda)| + |p_2(\lambda)/p_0(\lambda)|) = 0$.

Now, we examine the effects of lags (θ_1, θ_2) on the stability of the steady state. The analysis proceeds as follows:

(1) We characterize the points at which the local dynamics can change, i.e., the points at which the pure imaginary roots appear.[4] These points are referred to as the crossing points.
(2) We depict the sets of the crossing points (which we refer to as the crossing curves) on the θ_1-θ_2 plane using numerical simulation.
(3) We reveal the directions of changes in the signs of the real parts that occur when lags (θ_1, θ_2) cross the crossing curves.

3.2 Crossing Points

Dividing Eq. (15) by $p_0(\lambda)$, we obtain

$$1 + a_1(\lambda)e^{-\theta_1\lambda} + a_2(\lambda)e^{-\theta_2\lambda} = 0, \tag{17}$$

where

$$a_1(\lambda) = \frac{p_1(\lambda)}{p_0(\lambda)} = \frac{\alpha\beta(\lambda - I_K)}{\lambda^2 + b_1\lambda + b_2}, \tag{18}$$

$$a_2(\lambda) = \frac{p_2(\lambda)}{p_0(\lambda)} = \frac{\alpha c\tau(\lambda - I_K)}{\lambda^2 + b_1\lambda + b_2}. \tag{19}$$

Moreover, we denote a pure imaginary root as $\lambda = vi$ (where v = imaginary part $\neq 0$ and $i = \sqrt{-1}$). Then, the values of v that satisfy Eq. (17) can be characterized by the following lemma:

Lemma 3.1 (Gu et al. 2005, Proposition 3.1) *For each v satisfying $p_0(vi) \neq 0$, $\lambda = vi$ is a solution of $\Delta(\lambda) = 0$ for some $(\theta_1, \theta_2) \in \mathbb{R}^2_+$ if and only if*

$$|a_1(iv)| + |a_2(iv)| \geq 1, \tag{20}$$

$$-1 \leq |a_1(iv)| - |a_2(iv)| \leq 1. \tag{21}$$

We denote the set of $v > 0$ that satisfy conditions (20) and (21) as Ω, which is termed as the crossing set.[5] For any given $v \in \Omega$, the sets (θ_1, θ_2) satisfying Eq. (17) (each of which corresponds to a crossing point) must satisfy the following relationships (Fig. 1).

[4] It is ensured from precondition (III) that a zero real root cannot be a root.

[5] Pure imaginary roots are always conjugated. Therefore, we can assume $v > 0$ without a loss of generality.

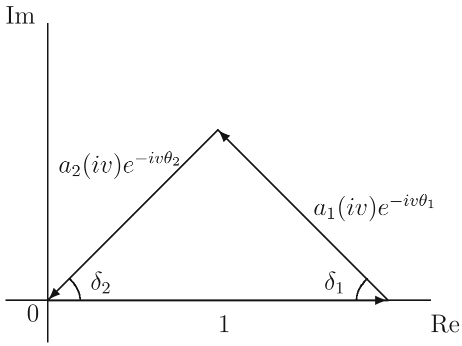


Fig. 1 Triangle formed by 1, $|a_1(iv)|$, and $|a_2(iv)|$ on the complex plane

$$\mp\delta_1 = \arg(a_1(iv)e^{-iv\theta_1}) + 2m\pi; \quad m = 0, 1, 2, \ldots, \tag{22}$$
$$\pm\delta_2 = \arg(a_2(iv)e^{-iv\theta_2}) + 2n\pi; \quad n = 0, 1, 2, \ldots, \tag{23}$$

where $\delta_1, \delta_2 \in [0, \pi]$.

Incidentally, on the complex plane, a multiplication of amplitudes becomes a sum of parts; therefore, we obtain

$$\arg(a_1(iv)e^{-iv\theta_1}) = \arg(a_1(iv)) - v\theta_1, \tag{24}$$
$$\arg(a_2(iv)e^{-iv\theta_2}) = \arg(a_2(iv)) - v\theta_2. \tag{25}$$

Figure 1 also demonstrates that the following relationships hold:

$$\arg(a_1(iv)) = \tan^{-1}\left(\frac{\mathrm{Im}(a_1(iv))}{\mathrm{Re}(a_1(iv))}\right), \tag{26}$$
$$\arg(a_2(iv)) = \tan^{-1}\left(\frac{\mathrm{Im}(a_2(iv))}{\mathrm{Re}(a_2(iv))}\right). \tag{27}$$

Moreover, after some manipulation, Eqs. (18) and (19) derive the following expression:

$$\frac{\mathrm{Im}(a_1(iv))}{\mathrm{Re}(a_1(iv))} = \frac{\mathrm{Im}(a_2(iv))}{\mathrm{Re}(a_2(iv))} = \frac{b_1 v I_K + v(b_2 - v^2)}{b_1 v^2 - I_K(b_2 - v^2)}. \tag{28}$$

Thus, using Eqs. (24)–(28), Eqs. (22) and (23) can be rewritten as follows:

$$\theta_1 = \frac{\tan^{-1}(\frac{b_1 v I_K + v(b_2 - v^2)}{b_1 v^2 - I_K(b_2 - v^2)}) \pm \delta_1 + 2m\pi}{v}, \tag{29}$$
$$\theta_2 = \frac{\tan^{-1}(\frac{b_1 v I_K + v(b_2 - v^2)}{b_1 v^2 - I_K(b_2 - v^2)}) \mp \delta_2 + 2n\pi}{v}, \tag{30}$$

where the interior angles of the triangle denoted by δ_1 and δ_2 are given by the cosine theorem as follows:

$$\delta_1 = \cos^{-1}\left(\frac{1+|a_1(iv)|^2-|a_2(iv)|^2}{2|a_1(iv)|}\right)$$
$$= \cos^{-1}\left(\frac{(b_2-v^2)^2+(b_1v)^2+(\alpha\beta I_K)^2+(\alpha\beta v)^2-(\alpha c\tau I_K)^2-(\alpha c\tau v)^2}{2\sqrt{(\alpha\beta I_K)^2+(\alpha\beta v)^2}\sqrt{(b_2-v^2)^2+(b_1v)^2}}\right),$$
$$\delta_2 = \cos^{-1}\left(\frac{1+|a_2(iv)|^2-|a_1(iv)|^2}{2|a_2(iv)|}\right)$$
$$= \cos^{-1}\left(\frac{(b_2-v^2)^2+(b_1v)^2-(\alpha\beta I_K)^2-(\alpha\beta v)^2+(\alpha c\tau I_K)^2+(\alpha c\tau v)^2}{2\sqrt{(\alpha c\tau I_K)^2+(\alpha c\tau v)^2}\sqrt{(b_2-v^2)^2+(b_1v)^2}}\right).$$

Equations (29) and (30) characterize the sets of the crossing points $(\theta_1, \theta_2) \in \mathbb{R}^2_+$. Depending on the signs of δ_1 and δ_2, we can define two types of crossing points, denoted by $L_1(m,n)$ and $L_2(m,n)$, as follows:

$$L_1(m,n): \quad \theta_1 = \frac{\tan^{-1}\left(\frac{b_1vI_K+v(b_2-v^2)}{b_1v^2-I_K(b_2-v^2)}\right)+\delta_1+2m\pi}{v}, \quad \theta_2 = \frac{\tan^{-1}\left(\frac{b_1vI_K+v(b_2-v^2)}{b_1v^2-I_K(b_2-v^2)}\right)-\delta_2+2n\pi}{v},$$
$$L_2(m,n): \quad \theta_1 = \frac{\tan^{-1}\left(\frac{b_1vI_K+v(b_2-v^2)}{b_1v^2-I_K(b_2-v^2)}\right)-\delta_1+2m\pi}{v}, \quad \theta_2 = \frac{\tan^{-1}\left(\frac{b_1vI_K+v(b_2-v^2)}{b_1v^2-I_K(b_2-v^2)}\right)+\delta_2+2n\pi}{v}.$$

In the next section, based on the study of Asada and Yoshida (2001), we illustrate the examples of $L_1(m,n)$ and $L_2(m,n)$ using numerical simulations.

3.3 Numerical Simulations

Following Asada and Yoshida's (2001) study, we assume the investment function as follows:

$$I(Y(t), K(t), r(Y(t))) = \frac{400}{1+12e^{-0.1(Y(t)-400)}} - 0.01\sqrt{Y(t)} - 0.5K(t) - 10\gamma(\sqrt{Y(t)} - \sqrt{\bar{Y}}).$$

Further, we set the parameter values as follows: $\alpha = 0.9$; $c = 0.625$; $\tau = 0.2$; $\bar{Y} = 400$; $\bar{C} + c\bar{T} + \bar{G} = 200$; and $\gamma = 8.6$. Under these specifications, the steady-state values of System (13) are given by $(Y^*, K^*) = (400, 61.138)$.

In the following discussion, we compare two cases: the case of an active fiscal policy with that of a passive fiscal policy.

3.3.1 Example 1

When $\beta = 4.1$, which represents a relatively active fiscal policy, the crossing set Ω is given by $v \in (3.6506, 3.8716)$ (Fig. 2). For $v \in \Omega$, we can depict $L_1(m, n)$ and $L_2(m, n)$ as shown in Fig. 3, where $m = 0, 1, 2$ and $n = 0, 1, 2$. The dotted curves represent $L_1(m, n)$, and the solid curves represent $L_2(m, n)$. These curves are referred to as the crossing curves.

3.3.2 Example 2

When $\beta = 0.1$, which represents a passive fiscal policy, the crossing set Ω is given by $v \in (0.2636, 0.5120)$ (Fig. 4). In this case, the crossing curves $L_1(m, n)$ and $L_2(m, n)$ can be depicted for $v \in \Omega$ as shown in Fig. 5. The starting points of both curves $L_1(m, n)$ and $L_2(m, n)$ (i.e., the points corresponding to $v = 0.2636$) are given by the upper connecting points of the circles.

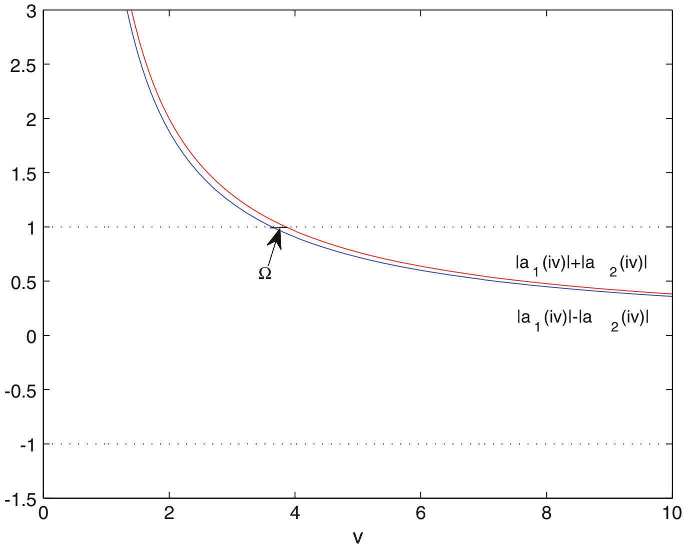


Fig. 2 Crossing set Ω ($\beta = 4.1$)

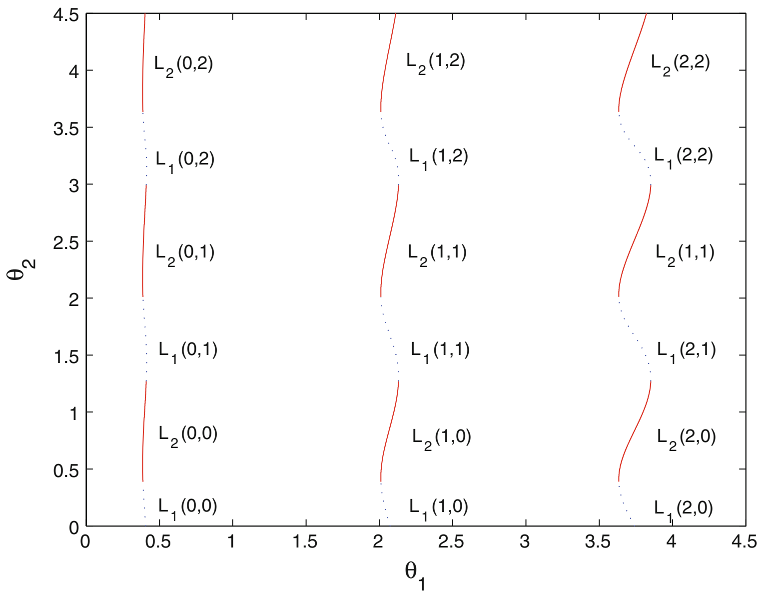


Fig. 3 Crossing curves ($\beta = 4.1$)

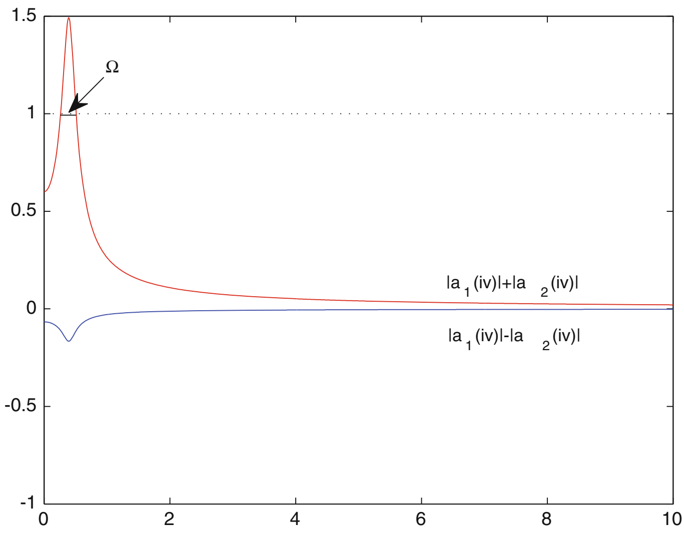


Fig. 4 Crossing set Ω ($\beta = 0.1$)

Next, we examine how the real parts of the roots change when lags (θ_1, θ_2) cross the crossing curves.

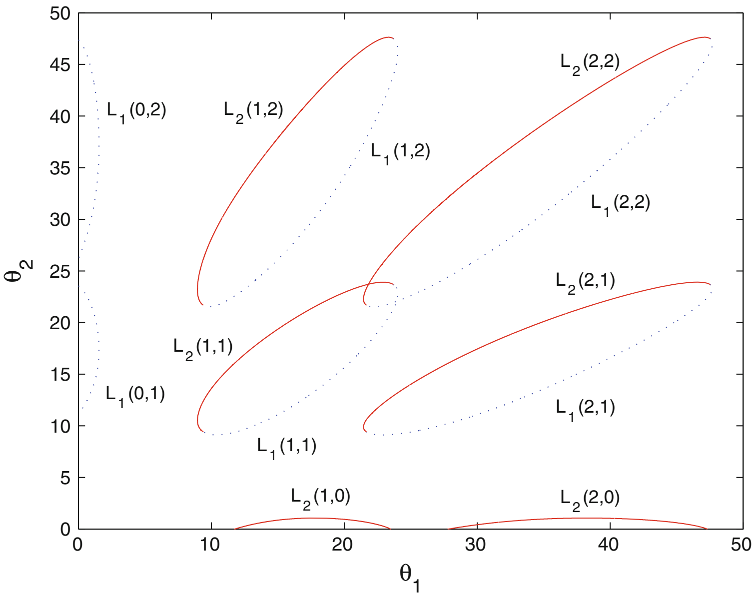


Fig. 5 Crossing curves ($\beta = 0.1$)

3.4 Direction of Crossing

We reveal the direction in which the roots cross the imaginary axis when the value of θ_1 increases. It is determined by the sign of $d\text{Re}\lambda/d\theta_1|_{\lambda=iv}$ ($where\ v \in \Omega$). If $d\text{Re}\lambda/d\theta_1|_{\lambda=iv} > 0$, the roots cross the imaginary axis from left to right with an increase in θ_1 (which indicates destabilization). In contrast, if $d\text{Re}\lambda/d\theta_1|_{\lambda=iv} < 0$, the roots cross the imaginary axis from right to left with an increase in θ_1 (which indicates stabilization). For convenience of calculation, we observe the sign of $\text{Re}(d\lambda/d\theta_1)^{-1}|_{\lambda=iv}$ instead of that of $d\text{Re}\lambda/d\theta_1|_{\lambda=iv}$.

Differentiating Eq. (17) with respect to θ_1, we obtain

$$\left[a_1'(\lambda)e^{-\theta_1\lambda} - a_1(\lambda)e^{-\theta_1\lambda}\theta_1 + a_2'(\lambda)e^{-\theta_2\lambda} - a_2(\lambda)e^{-\theta_2}\theta_2\right]\frac{d\lambda}{d\theta_1} = a_1(\lambda)e^{-\theta_1\lambda}\lambda,$$

or equivalently

$$\left(\frac{d\lambda}{d\theta_1}\right)^{-1} = \frac{a_1'(\lambda)e^{-\theta_1\lambda} + a_2'(\lambda)e^{-\theta_2\lambda} - a_2(\lambda)e^{-\theta_2\lambda}\theta_2}{a_1(\lambda)e^{-\theta_1\lambda}\lambda} - \frac{\theta_1}{\lambda}, \tag{31}$$

where

$$a_1'(\lambda) = \frac{\alpha\beta p_0(\lambda) - \alpha\beta(\lambda - I_K)(2\lambda + b_1)}{p_0(\lambda)^2},$$
$$a_2'(\lambda) = \frac{\alpha c\tau p_0(\lambda) - \alpha c\tau(\lambda - I_K)(2\lambda + b_1)}{p_0(\lambda)^2}.$$

3.4.1 Example 1

Suppose that $\beta = 4.1$. In this case, describing the real part of Eq. (31) as a function of $v \in \Omega$, we can derive Fig. 6, where the dotted curves are the functions evaluated on curve $L_1(m, n)$, and the solid curves are the functions evaluated on curve $L_2(m, n)$.

Figure 6 shows that $\mathrm{Re}(d\lambda/d\theta_1)^{-1}|_{\lambda=iv} > 0$ holds for all cases in Fig. 3. Therefore, at least two imaginary roots with positive real parts emerge when θ_1 crosses the crossing curves from left to right.

Now, a curve formed by connecting curves $L_j(0, n)$ (where $j = 1, 2; n = 0, 1, 2$) is termed as m_0 (an enlarged representation of this curve is proposed in Fig. 7). Then, we can make the following proposition:

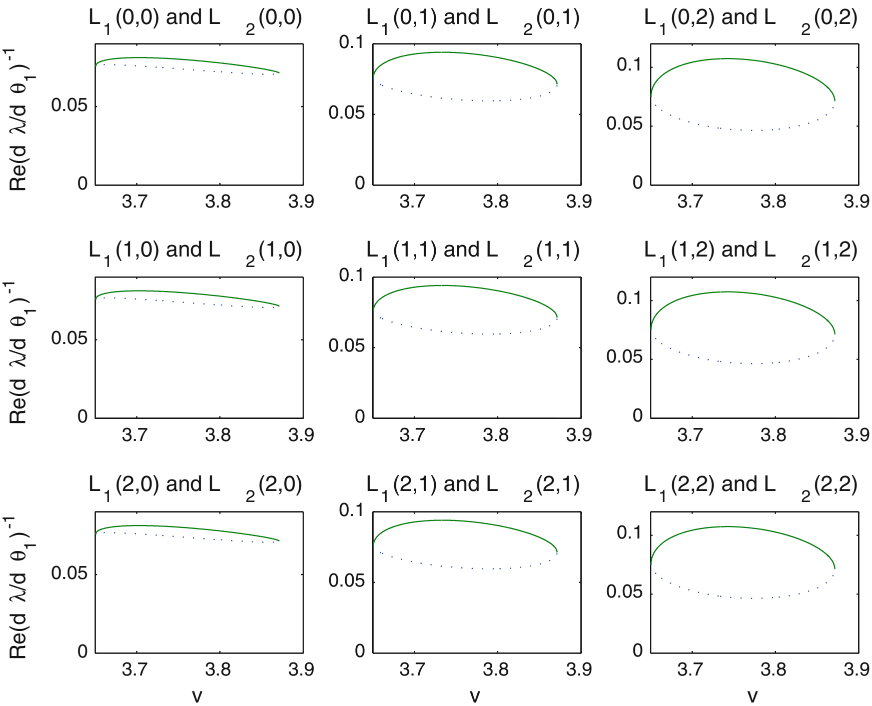


Fig. 6 Direction of crossing ($\beta = 4.1$)

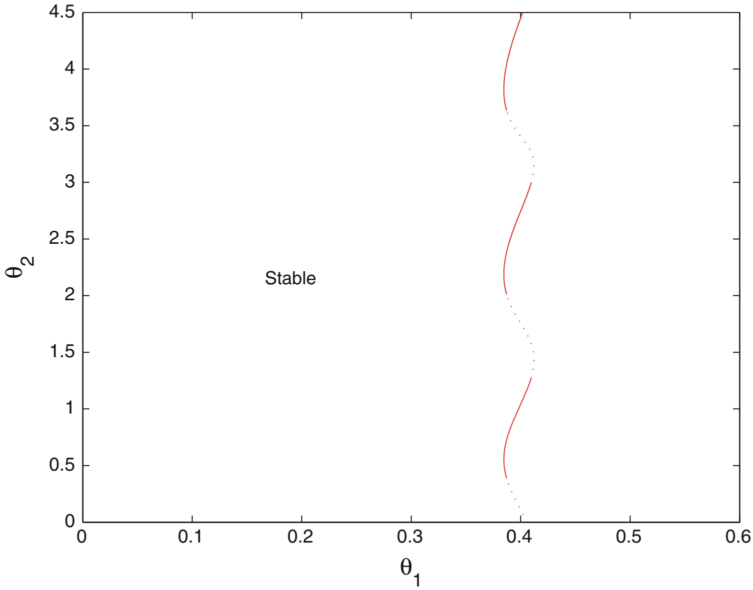


Fig. 7 Curve $m0$

Proposition 3.1 *For lags* (θ_1, θ_2) *lying to the left of curve* $m0$*, the steady state is locally stable. However, for lags* (θ_1, θ_2) *lying to the right of curve* $m0$*, the steady state is unstable.*

Based on this proposition, we can state the following: In the case of $\theta_1 < 0.384$, the steady state is locally stable irrespective of the value of θ_2, i.e., if a government expenditure lag is sufficiently small, a tax collection lag does not affect economic stability. Moreover, in the case of $\theta_1 \in (0.384, 0.412)$, the steady state fluctuates between stability and instability as θ_2 increases. Thus, a tax collection lag can contribute toward stabilizing an economy.

3.4.2 Example 2

When $\beta = 0.1$, the direction of crossing is determined by Fig. 8. Figures 5 and 8 demonstrate the following proposition:

Proposition 3.2 *In Fig. 5, the regions enclosed within curves* $L_1(m, n)$ *and* $L_2(m, n)$ *(i.e., regions inside the circles) are unstable, whereas the others are stable.*

Comparing the case of a passive policy ($\beta = 0.1$) with that of an active policy ($\beta = 4.1$) within an economically meaningful region of (θ_1, θ_2) (i.e., θ_1 and θ_2 take values between 0 and 3), the former achieves a larger stable region. This suggests that an active policy stance may increase economic instability. This result cannot

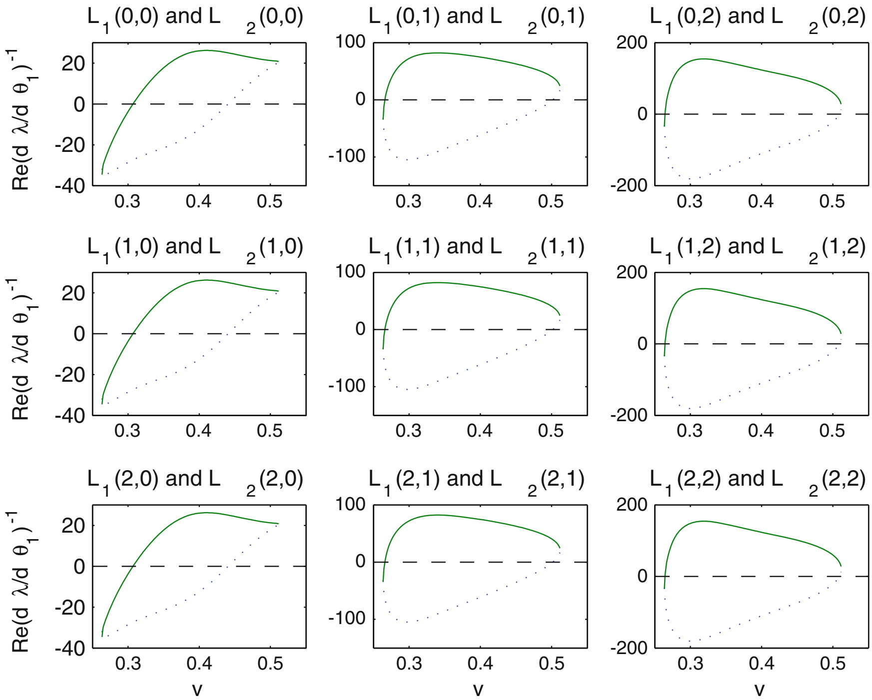


Fig. 8 Direction of crossing ($\beta = 0.1$)

be derived from a model without a time lag. Furthermore, as indicated by Fig. 5, in the case of $\beta = 0.1$, the steady state fluctuates between stability and instability with increases in not only θ_2 but also θ_1. Therefore, not only tax collection but also government expenditure lags can contribute toward stabilizing an economy.

4 Global Dynamics

Thus far, we analyzed the local dynamics of System (13) with regard to the steady state. In this section, we illustrate phase diagrams to visually confirm the result established in the previous section and provide an example of global dynamics of the system.

We set the same parameter values as those in the previous section and assume that $\beta = 4.1$ (This section only examines the case with an active fiscal policy.). Further, we assume $\theta_1 = 0.4$. As indicated by Fig. 7, if θ_2 is sufficiently small (i.e., $\theta_2 \leq 0.038$), the steady state is locally stable. However, if $\theta_2 > 0.038$, then the dynamics of the solutions change depending on the value of θ_2 (Fig. 9). When $\theta_2 = 0.7$, a stable cycle exists and the solutions starting from the initial values of $(Y(0), K(0)) = (390, 55)$

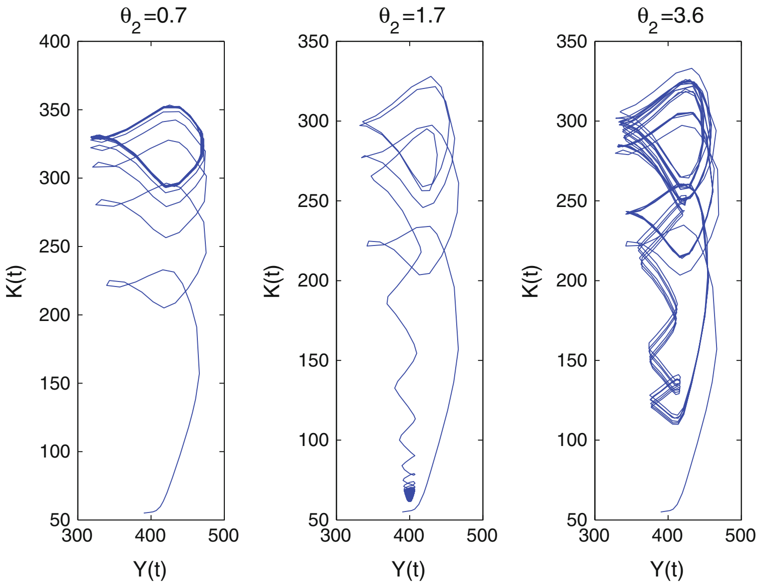


Fig. 9 $\theta_1 = 0.4$

converge to the cycle. When $\theta_2 = 1.7$, the steady state becomes locally stable again, and the solutions converge to the steady state. Moreover, when $\theta_2 = 3.6$, a strange-shaped attractor emerges, and the solutions exhibit chaotic behaviors.

This example demonstrates that while an increase in a tax collection lag contributes toward local stability, it can cause globally complex behaviors.

5 Conclusion

In this study, we developed the Kaldorian model with government expenditure and tax collection lags and examined the effects of these lags on local stability using numerical simulations. In addition, we also examined global dynamics.

As shown by Asada (1987), under a fiscal policy without a lag, the steady state is locally stable as long as the government is sufficiently active. However, Asada and Yoshida (2001) show that under a policy with a sufficiently large expenditure lag, the steady state becomes unstable even if the government is sufficiently active. This study showed that under a policy with government expenditure and tax collection lags, a policy lag can have a stabilizing effect on the steady state.

Under an active policy stance, if a government expenditure lag exceeds a certain threshold level, then the steady state becomes unstable. This result is similar to that in Asada and Yoshida's (2001) study. However, we further demonstrated that in the neighborhood of the threshold, certain positive values of a tax collection lag can achieve local stability. Therefore, a tax collection lag can contribute toward economic stability.

Similarly, under a passive policy stance, both tax collection as well as government expenditure lags can contribute to stabilizing an economy.

We also demonstrated that in an unstable parameter region, limit cycles and complex behaviors can emerge. Therefore, while an increase in a tax collection lag contributes toward local stability, it can cause globally complex behaviors.

According to Friedman (1948), policy lags are classified into three types: recognition, implementation, and diffusion lag. Unlike recognition and diffusion lag, implementation lag can be considered as adjustable to some extent. Therefore, this study suggests that an adjustment of the timing of policy implementation can be a means to achieve stabilization.

References

Asada T (1987) Government finance and wealth effect in a Kaldorian cycle model. J Econ 47(2):143–166

Asada T (2010) Central banking and deflationary depression: a Japanese perspective. In: Cappello M, Rizzo C (eds) Central banking and globalization. Nova Science Publishers, New York, pp 91–114

Asada T, Akio M (2014) Monetary policy in a delay dynamic Keynesian model. In: Annual of the institute of economic research, vol 45. Chuo University, pp 143–160 (in Japanese)

Asada T, Yoshida H (2001) Stability, instability and complex behavior in macrodynamic models with policy lag. Discrete Dyn Nat Soc 5(4):281–295

Bellman RE, Cooke KL (1963) Differential-difference equations. Academic Press, New York

Cai J (2005) Hopf bifurcation in the IS-LM business cycle model with time delay. Electron J Differ Equ 2005(15):1–6

Chang WW, Smyth DJ (1971) The existence and persistence of cycles in a non-linear model: Kaldor's 1940 model re-examined. Rev Econ Stud 38(1):37–44

Fanti L, Manfredi P (2007) Chaotic business cycles and fiscal policy: an IS-LM model with distributed tax collection lags. Chaos Solitons Fractals 32(2):736–744

Friedman M (1948) A monetary and fiscal framework for economic stability. Am Econ Rev 38(3):245–264

Gabisch G, Lorenz HW (1989) Business cycle theory: a survey of methods and concepts. Springer, New York

Gandolfo G (2010) Economic dynamics, 4th edn. Springer, Berlin

Gu K, Niculescu S-I, Chen J (2005) On stability crossing curves for general systems with two delays. J Math Anal Appl 311(1):231–253

Kaldor N (1940) A model of the trade cycle. Econ J 50(197):78–92

Kalecki M (1935) A macrodynamic theory of business cycles. Econ J Econ Soc 3(3):327–344

Matsumoto A, Ferenc S (2013) Dynamics in delay IS-LM model with tax collections. In: IERCU discussion paper No.208. Chuo University

Matsumoto A, Ferenc S (2014) Nonlinear multiplier-accelerator model with investment and consumption delays. In: IERCU Discussion Paper No.227. Chuo University

Neamţu M, Opriş D, Chilarescu C (2007) Hopf bifurcation in a dynamic IS-LM model with time delay. Chaos Solitons Fractals 34(2):519–530

Sasakura K (1994) On the dynamic behavior of Schinasi's business cycle model. J Macroecon 16(3):423–444

Schinasi GJ (1981) A nonlinear dynamic model of short run fluctuations. Rev Econ Stud 48(4): 649–656

Sportelli M, De Cesare L (2005) A dynamic IS-LM model with delayed taxation revenues. Chaos Solitons Fractals 25(1):233–244

Sportelli M, De Cesare L, Binetti MT (2014) A dynamic IS-LM model with two time delays in the public sector. Appl Math Comput 243:728–739

Szydlowski M (2002) Time-to-build in dynamics of economic models I: Kalecki's model. Chaos Solitons Fractals 14(5):697–703

Szydlowski M (2003) Time to build in dynamics of economic models II: models of economic growth. Chaos Solitons Fractals 18(2):355–364

Tsuduki E (2015) Stability in macrodynamic models with a policy lag: Old- vs New Keynesian economics, mimeo

Yoshida H, Asada T (2007) Dynamic analysis of policy lag in a Keynes-Goodwin model: stability, instability, cycles and chaos. J Econ Behav Organ 62(3):441–469

Zhou L, Li Y (2009) A dynamic IS-LM business cycle model with two time delays in capital accumulation equation. J Comput Appl Math 228(1):182–187

Liquidity Shock, Animal Spirits and Bank Runs

Huang Weihong and Huang Qiao

Abstract Since the end of WWII, economists generally believed that the phenomenon of bank runs had died away. However, sufficient evidences occurred in last two decades suggest the revival of bank run, supported by the facts that numerous banking panics occurred repeatedly with the traditional and new styles during the Asian Financial Crisis in 1997, the financial crisis in 2007, and the recent European debt crises. Therefore, classical economic and financial theories about bank runs need to be challenged. In this chapter, a bottom-up behavioral model is built to model bank runs. It is shown with agent-based modeling approach that animal spirits of the depositors together with endogenous optimism and pessimism can cause liquidity shock. In particular, bank run will be trigged if the animal sprit reaches to a tipping point. The model is tested with a set of empirical data, which shows that the effect of animal spirits is significant in reality. The findings in this research may shed some light on central bank's monetary policy.

Keywords Bank runs · Liquidity shock · Animal spirits · Tipping point · Agent-based modeling

1 Introduction

Liquidity is regarded as the lifeline of commercial banks. It is not only a vital aspect for the safety of a single commercial bank, but also significant to the stability of the whole country and the global economy. Nowadays, although the consequence of liquidity shock, i.e., bank runs, has been superficially treated as a solved problem, not all credit institutions are covered by certain clinker-built system of defense. For example, during the Asian Financial Crisis in 1997, liquidity crisis occurred due to bank runs in many countries, which led to the bankruptcy of many commercial banks and the financial crisis in many economies. In addition, the shadow banking

H. Weihong (✉) · H. Qiao (✉)
Division of Economics, Nanyang Technological University, Singapore, Singapore
e-mail: AWHHuang@ntu.edu.sg

H. Qiao
e-mail: qhuang004@e.ntu.edu.sg

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_13

system has raised wide new liquidity concerns during the U.S. subprime mortgage crisis and the global financial crisis. At the same time, numerous traditional financial institutions have begun to withdraw deposits from core financial institutions. Up to now, a lot of investors are still suffering from the panic of this turmoil though the economy has gradually recovered. Over the last two decades, these non-bank banks have grown up and begin to play a role similar as the commercial banks, investment banks and hedge funds. There can also be a run on shadow banks just as what happened to traditional banks in history, especially in the emerging market such as China where supervisions are not foolproof. Hence, the classical economic theory about bank runs need to be modified, and a theory involving the non-economic motivations that can explain the tipping point of the emergence phenomenon before bank failure are in dire need.

There are two main factors that may cause a bank failure: solvency shock and liquidity shock. We will focus on the latter in this paper. Although many factors influence the liquidity of banks, there are two main perspectives. The first is the structure of a bank's assets and liabilities, which is based on the identity: Assets = Liabilities + Capital. When the depositors suspect that the bank cannot repay their deposits as the market value of assets is below liability, bank runs would happen if there is no additional liquidity provided. What's worse, when one of the representative banks bankrupts, people would be afraid of the bankruptcy of other banks. Depositors will panic and start withdrawing money from other banks even if those banks are well-run. The second main factor is the policy of central bank. In the case of Northern Rock, before the breakout of a bank run, the Bank of England announced that it would provide emergency liquidity support. However, it was just after the central bank had the announcement that the depositors started queuing outside the bank. This revealed to us that the central bank could lose its capacity to play as the lender of last resort if the contagion has already been formed. Under such circumstance, it might aggravate the bank run phenomenon even more. In other words, both perspectives suggest that animal spirits might play an essential role in the phenomenon of bank runs.

When such crisis breaks out, human beings tend to fall into a state of chaos. As their willpower weakens, or surrounded by negative signals, two aspects of animal spirit – emotion or emotional drives might play an essential role in their decision-making. Their confidence and stability in their action dramatically reduces after passing a critical value of mind state, and this threshold can be defined as the tipping point. In short, such striking change is driven by the animal spirits, a concept ever laid at the heart of Keynes's explanation for Great Depression, but was abandoned after the blooming of New Classical Economics. Although the reason of bank runs is still controversial, there is no doubt that the public lost confidence in the market though some of the banks are still well-managed, and all the people run on banks when the contagion of panic reaches a critical point. Therefore, the discovery and study of such tipping point is beneficial for us to effectively predict and prevent the occurrence of bank runs.

The optimal bankruptcy rules in monetary conditions are pioneered by Shubik and Wilson (1977), have pointed out that the theory, as the severity of the penalties is varied, considered constraint of credit reappear in a variety of borrowing constraints in modern-equilibrium models. After that, based on Samulson's pure consumption loans model, Bryant (1980) introduced an overlapping-generations model of borrowing and lending alone, without considering such factors as reserves, risky assets, or deposit insurance. In a word, this theory provides the basic strategy of game theory that has been cited in many papers about bank runs. The milestone of bank run theory is that of Diamond and Dybvig (1983) who have proposed the now-classic model based on Nash equilibrium, a model that demonstrates the reason why banks choose to issue deposits that are more liquid than their own assets and to understand the reason why banks are subject to runs. They pointed out the undesirable equilibrium in which all depositors panic and withdraw immediately. After that, the mainstream of academia explains and rationalizes their theory until now. The cost of illiquidity can be avoided with a fragile capital structure in bank (Diamond and Rajan 2001a, b), however, when the bank failures appear, it will shrink the common pool of liquidity, creating, or exacerbating aggregate liquidity shortages and lead to a contagion of failure (Diamond and Rajan 2005).

Uhlig (2010) points out that the new kind of bank runs appears in 2008 financial crisis, and many financial institutions withdraw deposits from core financial institutions. What's more, different from the traditional view, he figures out that the appropriate perspective of bank run might come from insolvency rather than illiquidity. The financial crisis reminds many other economists of this old-fashioned story, for example, Calvo (2012) also claims that the factors that determines the liquidity motivate the financial innovation and crash, and the policy might increase the asset process and steady-state output, however, get reversed as liquidity is destroyed. What's more, as the previous model is impossible to describe the new and complex bank-run phenomenon and the impact on the whole economy, the DSGE model emerges and tries to figure out the real cause. Andre Gerali et al. (2010) set up a DSGE model to explain the credit and banking system in Euro Area, but they just calibrate with the real data without clear explanation of the bank runs occurrence. Based on the idea of Diamond and Rajan (2001a, b), different attitudes to risk in the transmission of shocks (Angeloni and Faia 2009; Wickens 2014) have been mentioned, and different kinds of exogenous shocks are discussed. Hafstead and Smith (2012) introducesin the DSGE framework the interbank lending, which is crucial in nowadays banking system, and they focus on how to adjust the costs and benefits of bank intermediation to smooth the business cycle. Recently, the discipline of social network develops quite fast, and interdisciplinary model is applied in the simulation of the spread of bank-run panic. An epidemiologic model is used by Toivanen (2013), showing that central banks' interventions reduce contagion only slightly. Although some of the agent-based models have appeared (for example, Dias et al. 2015), until now, most of them concentrate on the network of the contagion spread. In a nutshell, no one has explained the shock from a reasonable endogenous perspective, let alone give the explanation on when the chaos would happen and when it would form the trend to be bank run. Therefore, it is of great significance to find the potential stimulus and critical point of bank runs.

The term Animal Spirits was introduced by Hume (1739) to trace the origins of human decision making as major field of human nature. However, we could hardly see any research of economics related to this concept until it is revisited by Keynes (1936). He suggested that human sentiments or feelings (e.g. optimism or pessimism) would drive the aggregate economy. Animal spirits, instead of just employing mathematics and strong economic analysis, would lead people to act without shyness and doubt, or even they would not react at all. This concept lays at the core of Keynes's explanation for the Great Depression, however, during 1970s, the New Classical Economics arose and criticized that animal spirits should be not considered at all. During the waived period of animal spirits, the valuable development came from Loewenstein and O'Donoghue (2004). They divided animal behavior into deliberative processes that assess options with a broad, goal-based perspective within the standard economic conception and affective processes that encompass emotions and emotional drives. It is noticed that the affective system is inherently myopic, while the deliberative system cares about both short-term and long-term outcomes. Furthermore, Adaptive learning mechanism, which consists of two kind of rules: fundamentalist and chartist rules, has often been applied in the behavioral economic literature (Brock and Hommes 1997; De Grauwe and Grimaldi 2006; Branch and Evans 2006). Pfajfar and Zakelj (2009) find out the experimental evidence in support of these two rules for inflation forecasts in a new Keynesian model.

Recently, the study of animal spirits was renewed by Akerlof and Shiller (2009) and began to raise greater concern. Different from seeking to minimize asmuch as possible in pure economic motivation and rationality, they explain the deviations that actually do occur and that can be observed. Besides, they generally account for the impact of animal spirits in the bank runs and point out that the dramatic loss of confidence is indispensable to cause bank runs inevitably even though some insurance measures have been set up. Paul De Grauwe (2012) brings animal spirits into the DSGE framework, and it is the first time that this concept is combined with new Keynesian model. He breaks the assumption of rational expectation that gives an endogenous shock with non-normal distribution, and points out that when the sample is large enough, the uncertainty is transformed into measurable risk, which means that the deterministic trend is observable. Therefore, with the development of animal spirits theory, considering the bank run phenomenon from the perspective of endogenous shock becomes possible.

In 2000, Gladwell and Malcolm define tipping points as the moment of critical mass, the threshold, the boiling point. To be specific, it refers to one dramatic moment in an epidemic when everything can change all at once. Silver (2008) has pointed out that a Tipping-Point state is defined as a state that would be most likely to alter the outcome if it were decided differently. Besides, the latest and exhaustive interpretation has been given by Lamberson (2012). They define tipping points as discontinuities in the relationship between present conditions and future states of the system. In other words, it is regarded as the action or outcome that dramatically reduces uncertainty about the future. The authors divide tipping points into two categorical distinctions, which describe the tipping point by scientific division. Beyond the pure theory, Solé (2011) has showed various models with the feature of tipping point in the phase transitions. Such process focuses on the properties of the interaction structure, which would undergo a dramatic shift when passing the tipping point.

In the bank-run phenomenon, the spread of panic would increase dramatically when the surroundings is in chaos. Namely, some stimuli would tip the contagion, and bank runs is nearly formed above the tipping point. This implies the reason that why the economic behavior would always be explained by the unstable equilibrium, and actually in that case some forces in nature are driving us strikingly to a deterministic tendency.

There are two disparate fields in studying networks: graph theory and social science. In the social science, the set of contacts of an individual is their 'neighborhood', and the size of this neighborhood is the individual's 'degree' (Keeling and Eames 2005). In random networks, the spatial position of individuals is irrelevant, and connections are formed at random (Bollobás 2001). In the most simplified version of the random network, each individual has a fixed number of neighborhood through which infection can spread. In fact, the growing random network is more reliable, and the social science develop the mode by simulating the methodology of graph theory. The random graphs were first introduced by Erdös and Rényi (1959), Erdös and Rényi (1960), who have given the probable structure of a random graph. In the development of their theory, they have found that there exists a probability distribution function with a regular sharp threshold. However, they could not find the precise condition of the critical value that the network would tip in their theory, and there are not the delicate trends of human nature considered in their model. After that, Schelling (1978) published Micromotives and Macrobehavior, where he is interested to find that a small and ostensible meaningless decisions and actions made by individuals often lead to dramatically unintended results for a large group. Similarly, Miller and Page (2004) set up a Standing Ovation Problem (SOP) that explains peer effect, which is the part of everyone's common experience. It stresses some of the key themes caused in social systems, such as social learning, diffusion, heterogeneity, incentives, and networks. Even so, both the Schelling's Segregation Model and the Standard Ovation Problem explaining the same individual behavior would lead to the unintended results at large group with the different condition. In addition, the agent-based modeling used in these models allow a variety of assumptions about information transmission, and it allows the researcher to easily alter the

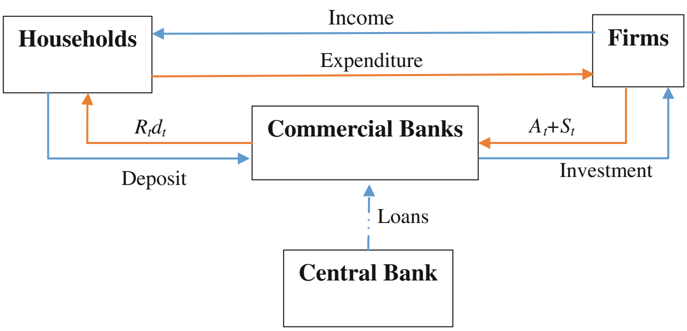


Fig. 1 Circulation of commercial banks

assumed behaviors and parameters in an effort to identify the key factors driving the results. As the great merit of agent-based modeling, many scholars nowadays attempt to simulate the social network with the epidemic models. For the bank runs phenomenon, the driver of the dramatic chaos is driven by the animal spirits in nature, so the model should base on the condition and result of animal spirits. In a word, it is crucial to find out the tipping point that animal spirits would react prominently, and we should know which factor leading to this tip, which is direct or contextual.

2 The Generic Model

The economy is considered with entrepreneurs in firms, depositors in households and the commercial banks. The central bank acts as the lender of last resort, but we assume it is inactive to figure out the effect of animal spirits in this paper. Refer to Diamond and Rajan (2001a, b) and Ignazio Angeloni and Faia (2009), a present "liquidation value" of the assets equals to a fundamental value A_t plus a value would be shocked by animal spirits S_t. Thus the assets is $A_t + S_t$, and the interval of S_t is defined as $1/2z$. z is the shock by animal spirits, which is between 0 and 1. As the Fig. 1 shown below, the circulation of banking system should satisfy the function $A_t + S_t \geq R_t d_t$, otherwise the depositors might run on the banks. R_t is the interest rate of deposit, and d_t is the deposit ratio.

Further, as the commercial bank has the professional knowledge about the project, it will have an advantage in figuring out the project value before it is completed compared with other lenders. Therefore,other lenders' liquidation value can be set as $\lambda(A_t + S_t)$ where $0 < \lambda < 1$. In conclusion, there are two cases for the bank-run happened (Table 1).

Table 1 Cases for the bank runs

Category	Scenario	Range between the return of investment and saving	Returns of depositors and banks
Case 1	The depositors must run on bank as the outcome of the project is lower than the gross deposits (including interest)	$A_t + S_t < R_t d_t$	Depositors: $\frac{(1+\lambda)\times(A_t+S_t)}{2}$ Bank: $\frac{(1-\lambda)\times(A_t+S_t)}{2}$
Case 2	The other lenders figure out the outcome of project is lower than the gross deposit, but the calculation of the specific commercial bank is not	$\lambda(A_t + S_t) < R_t d_t < A_t + S_t$	Depositors: $\frac{A_t+S_t+R_t d_t}{2}$ Bank: $\frac{A_t+S_t-R_t d_t}{2}$

In the first case, depositors alone would get $\lambda \times (A_t + S_t)$ by expectation, and the rest $(1 - \lambda) \times (A_t + S_t)$ share between depositor and the bank. The excess return is divided into half for each. The returns receive from the project investment is lower than gross deposit, which means the bank face the illiquidity risk. For the second case, the depositors run or not based on whether they believe the extraction of bank. At this time, the outcome of project seems uncertain for outsider. If the outcome of project calculated by other lenders is higher than the gross return by deposit, that is $\lambda(A_t + S_t) > R_t d_t$, the depositor will not run the bank in this situation.

Because the interval is larger than z, the expected value of total payment to the depositors is maximized as follows:

$$\max_{d_t} \quad \tilde{E}p_t = \frac{1}{2z}\int_{-z}^{R_t d_t - A_t} \frac{(1+\lambda)\times(A_t+S_t)}{2} dS_t + \frac{1}{2z}\int_{R_t d_t - A_t}^{z} \frac{A_t + S_t + R_t d_t}{2} dS_t$$

$$\text{s.t.}\lambda\frac{A_t + z}{R_t} < d_t < \frac{A_t + z}{R_t}$$

$$\Rightarrow d_t = \frac{1}{R_t}\frac{A_t + z}{2 - \lambda}$$

The probability of a run occurring can be written as

$$B = \frac{1}{2z}\int_{-z}^{R_t d_t - A_t} dS_t = \frac{1}{2}\left(\frac{R_t d_t - A_t}{z} + \frac{z}{z}\right) = \frac{1}{2} - \frac{A_t(1-\lambda) - z}{2z(2-\lambda)}$$

The probability B ranges between 0 and 1, impacted by the return of bank, the expectation of depositors and the shock in the market. The higher the expectation λ and shock z with the lower return A_t, the higher probability of the bank runs happens. The problem here is to figure out the shock z in the market, and an endogenous shock aroused by animal spirits is discussed in the following section.

3 Shock Aroused by Animal Spirits

While the mainstream model assumes that agents are able to understand thewhole world with the rational expectations, and here replaced by homo sapiens, the agents use the simple rules to portray thecognitive limitation and willingness to learn(e.g. Kirman 2006; Brazier et al. 2008; De Grauwe 2012). In the other word, we will use-heuristics (Tversky and Kahneman 1974) to decide the economic behavior with the adaptive learning mechanism, which can be simplified as two kinds of rules: One is the basic justifications underlying rational expectations, and the other is contrasting a great deal with the rational expectations forecasting rule. This is the basic mechanism that is applied for figure out the shock coming from animal spirits. What's more, each agent can use different forecasting rules. Therefore, there will be heterogeneity among agents. While rational expectations assume that agents understand the complex structure with understanding the only one "Truth" and allows builders of rational expectations models to focus on just one "representative agent", this heterogeneity creates interactions between agents and lead to a dynamics that is absent from rational expectations models, which is more close to our real life. Here, two types of behavior rules (heuristics) could be assumed as follows: the first is "fundamentalist rule" that Agents estimate the future with uncertainty and only have too pessimistic and pessimistic biased estimates of it. The second is "chartist rule" that does not presuppose that agents have biased estimates. It is ambiguous for the future and predict by the previous observation.Two rules can be listed as

(i) The fundamentalist rule is defined by
The optimistic fundamentalist rule $\tilde{E}^{fo}S_{t+1} = b$
The pessimistic fundamentalist rule $\tilde{E}^{fp}S_{t+1} = -b$
(ii) Thechartist rule is defined by $\tilde{E}^{c}S_{t+1} = S_{t-1}$

The rules are simple in the sense that do not need the agent to understand the whole picture, and just use the information they can understand. The experimental evidence has supported these two rules for inflation forecasts in a new Keynesian model (e.g. Duffy 2006). Although the agents could not figure out the whole picture of the market, they will have the adaptive learning ability that continually try to learn from their mistakes and choice the better rule they think, and generally, the market forecast is based on the weighted average of these two forecasts.This is the mechanism of animal spirits' effect, which is driven by the memory and the willingness to learn. The shock in next period is written as $\tilde{E}S_{t+1}$that

$$\tilde{E}S_{t+1} = \beta_{f,t}\tilde{E}^{f}S_{t+1} + \beta_{c,t}\tilde{E}^{c}S_{t+1} = \pm\beta_{f,t} \times b + \beta_{c,t} \times S_{t-1}, \qquad \text{where} \qquad \beta_{f,t} + \beta_{c,t} = 1$$

$\tilde{E}S_{t+1}$ is the weighted sum of two kinds of rule, so we should figure out the weight value firstly. Apply the discrete choice theory (Anderson et al. 1992), under pure rationality agents would choose the fundamentalist ruleif $U_{f,t} > U_{c,t}$, and vice versa. $U_{f,t}/U_{c,t}$ is forecast utility of the fundamentalist/chartist rule separately. Yet, psychologists have get the result that when we make an alternative decision we are also influenced by our state of mind that is impacted by recent emotional experiences, environment, etc. Thus, we can derive the new probability of choosing the fundamentalist as

$$\beta_{f,t} = P(U_{f,t} + \varepsilon_{f,t} > U_{c,t} + \varepsilon_{c,t})$$

$\varepsilon_{f,t}$and $\varepsilon_{c,t}$are the component that will aggrandize the animal spirits, and a more precise expression can specify as the logistically distributed (Anderson et al. 1992, p. 35):

$$\beta_{f,t} = \frac{\exp(\gamma U_{fo,t})}{\exp(\gamma U_{fo,t}) + \exp(\gamma U_{fp,t}) + \exp(\gamma U_{c,t})}$$
$$\beta_{f,t} = \frac{\exp(\gamma U_{fp,t})}{\exp(\gamma U_{fo,t}) + \exp(\gamma U_{fp,t}) + \exp(\gamma U_{c,t})}$$
$$\beta_{c,t} = \frac{\exp(\gamma U_{c,t})}{\exp(\gamma U_{fo,t}) + \exp(\gamma U_{fp,t}) + \exp(\gamma U_{c,t})}$$

If the $U_{f,t}/U_{c,t}$ improves greater thanthe outcomes of chartist/fundamentalist rule, the people prefer the fundamentalist/chartist rule. The parameter γ refers to willingness to learn from the past performance, and it shows as the intensity of choice in economic behavior. It is related to the variance of $\varepsilon_{f,t}$ and $\varepsilon_{c,t}$. When the variance is large, γ approaches 0. In this situation, agents choose the rule randomly and the probability of fundamentalist/ extrapolator equals to 0.5 as the time goes on. The variance becomes small when γ increases, and the utility is then fully deterministic. In the other words, the willingness to learn is 0 when $\gamma = 0$, and it will increase with the size of γ. In this adaptive learning mechanism, only those rules that pass the fitness test remain in place, or kick out.

$$U_{fo,t} = -\sum_{k=0}^{\infty} w_k (S_{t-k-1} - \tilde{E}_{fo,t-k-2} S_{t-k-1})^2$$
$$U_{fp,t} = -\sum_{k=0}^{\infty} w_k (S_{t-k-1} - \tilde{E}_{fp,t-k-2} S_{t-k-1})^2$$
$$U_{c,t} = -\sum_{k=0}^{\infty} w_k (S_{t-k-1} - \tilde{E}_{c,t-k-2} S_{t-k-1})^2$$

For $U_{f,t}/U_{c,t}$, the geometrically declining weights is introduced, because the agents would forget and give lower weight to errors made far in the past compared with errors made recently. Therefore, they are the negative value of geometrically declining weights (w_k) times mean squared forecasting errors of the fundamentalist and chartist rules respectively.To be specified, the $w_k = (1-\rho)\rho^k (0 \le \rho \le 1)$ is defined, and the equation can be written as follows

$$U_{fo,t} = \rho U_{fo,t-1} - (1-\rho)(S_{t-1} - \tilde{E}_{fo,t-2} S_{t-1})^2$$
$$U_{fp,t} = \rho U_{fp,t-1} - (1-\rho)(S_{t-1} - \tilde{E}_{fp,t-2} S_{t-1})^2$$
$$U_{c,t} = \rho U_{c,t-1} - (1-\rho)(S_{t-1} - \tilde{E}_{c,t-2} S_{t-1})^2$$

When $\rho = 0$, there is no memory, which means that onlyperformance in last period impacts on the forecasting rule, and the memory increases with the growth of ρ. When $\rho = 1$, the agent has infinite memory and the weight allocated to each period is the same and approaches to 0. At this time, the agents forget nothing, but the impact of new information can be also ignored. The animal spirit might disappear in this situation.

4 Tipping Point Simulation

Based on the random social network, thepossibility shocked by animal-spirit effects can be simulated by heterogeneous agent-based modeling. As the rumor between depositors becomes more frequent under the bad economic circumstance, the information in links would increase convexly (Fig. 2a). The vertical axis of *d*means the degree (information number) of each depositor, and the horizontal axis of *n*represents the probability of group panic. At the beginning, when the panic group is small, the neighbor is almost zero as almost no one will care about the bad situation and each one is isolated from others. If the group becomes larger, the depositors start to concern more information, which would increase infinitely when the major group focuses on it. The links between depositors increase convexly, because human being would become very sensitive to changes in the low probability and be impacted by the dramatic loss of confidence. Based on the prospect theory, there are overweighting of small probability and the underweighting of large probability. When the expectation begins to deviate the reference point, people would become sensitive observably, and they react much more strongly than the rational necessity. In Fig. 2b, the probability of rational withdrawal of individual depositor is expressed as the horizontal axis, and the probability of reality is the vertical axis. During the bad period, the panic people are likely to withdraw their money when they believe other people were also withdrawing. As a result, they withdraw the money at high level in reality, while the rational choice still closes to zero. Further, as people become less sensitive to changes in probability that moves away from the reference point, the willing of withdrawal will decrease gradually. After this change driven by animal spirits, the new equilibrium in standard economic would appear.

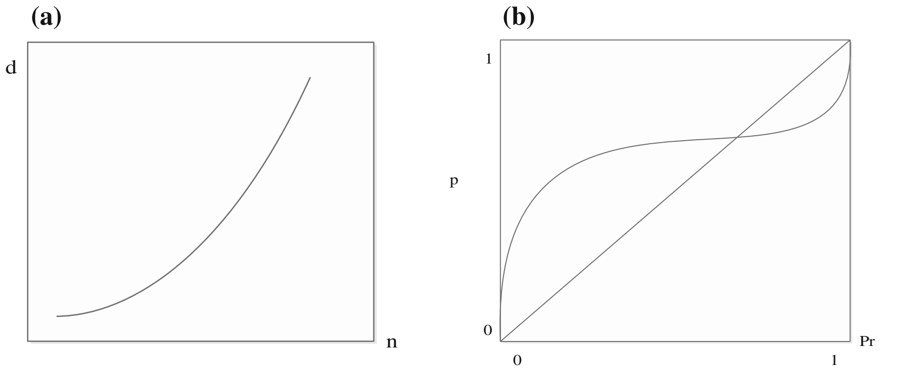


Fig. 2 **a** The degrees each depositor when the population of panic is increasing; **b** the relationship between real probability of individual depositor withdraw rationally and the probability of individual withdraw in fact

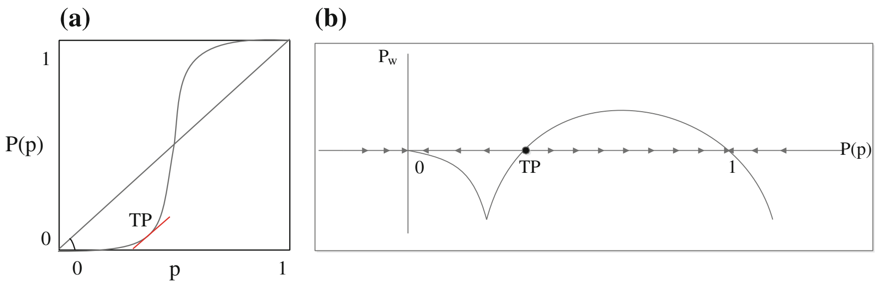


Fig. 3 Tipping point of bank runs

In reality, individual withdrawal gradually changes in the normal situation, however, there exists the extreme scenario when passing a critical value that leads to a discontinuous jump to the whole withdrawing behavior. The moment of jumping is the tipping point (TP), which appears due to the fact that animal spirits has played an essential role in human decision making. What we can see in Fig. 3a is that the probability of individual depositor (p) is the horizontal axis, and the probability of the whole depositors (P(p)) showing as vertical axis is increasing slowly and then becomes drastic. Bank run phenomenon forms when the P(p) gets 1 rapidly. Regarded as the equilibrium problem in Fig. 3b, the probability 0 and 1 of P(p) are two stable equilibrium similar to the Nash equilibrium in Diamond and Dybvig model. However, when the P(p)is larger than TP, it will tend to probability 1, or when the p is less than TP, the tendency to 0 would happen. In other words, the TP is an unstable at the lower p. However, when the change arrives at the tipping point, the increase equilibrium, and the probability of withdraw becomes positive above TP. Thus, it is necessary for us to find out the tipping point and then recognize the tendency of bank runs.

Now it is ensured that after passing the tipping point, the group withdrawal dramatically grows to cause a social bank run phenomenon impacted by the extremely enlarging of animal spirits (Fig. 4a), and the simulation result of agent-based modeling is demonstrated with different assumptions. To simplify, there are 4 connections for each agent that have the same probability to deliver the information, and 39507, the amount of withdrawing agents is selected arbitrarily. Initially, the serious consequence is not considered because this kind of phenomenon seldom happened before, like bank runs occurred in the early 20th. We show this result in Fig. 4b, and the tipping point is around 59 % in this situation. What's more, the animal spirits rises when story of crisis imprints a profound impression to public, and then the tipping point becomes lower (Fig. 4c). Driven by the animal spirits, the tipping point, around 40 % is just one possibility that becomes much stronger, and the real tipping point is based on the emotional change of human being. Nevertheless, the insurance mechanism is necessary which can delay the tipping point. Even when the bank run occurs, it can help keep the confidence among some depositors. As is seen from the Fig. 4d, the tipping points become higher, with several fluctuations before the process of

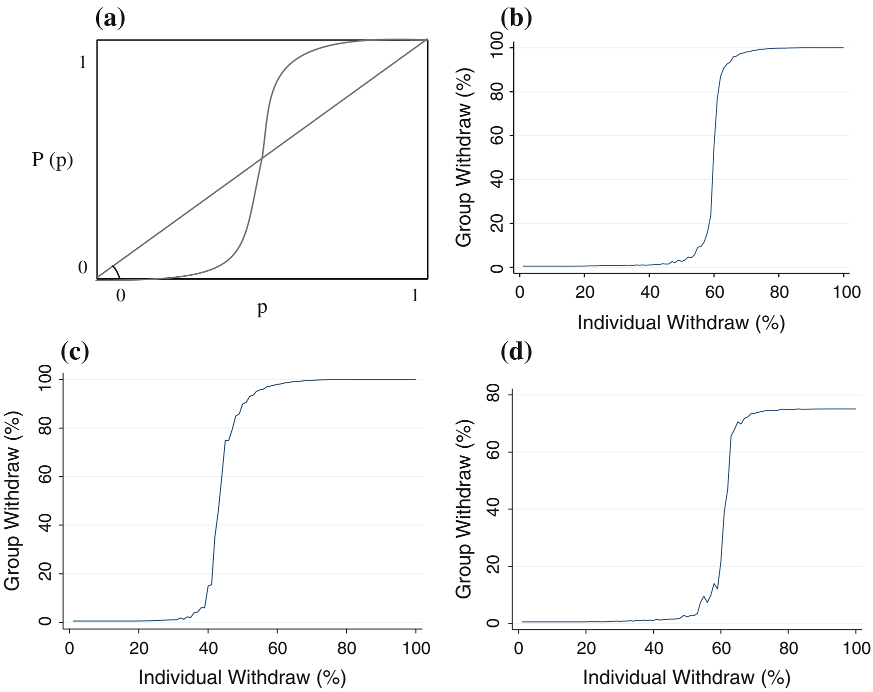


Fig. 4 Simulation with different conditions

dramatic change. Hence, it can leave the time to stop the tips in this scenario. Furthermore, there is no more than 80 % group withdrawal, which implies that more effective insurance can control the occurrence of bank-run in a low degree.

5 Data Analysis

From the discussion above, for the equation of bank run probability $B = \frac{1}{2} - \frac{A_t(1-\lambda)-z}{2z(2-\lambda)}$, the US real data is used in this paper. TheKBW Bank Total Return Index (KBW BTRI), which is abbreviated as BI below, is set as the return of bank A_t, and the Michigan Consumer Sentiment Index (CSI) is regarded as the ratio of other lenders' value λ. It is the monthly data from Jan. 1993 to April 2015 (Fig. 5).

The number of bank failure is regarded as the proxy of bank runs tendency. Influenced by the saving and loan crisis during the 1980–1990s, a lot of banks failed during the two decades. Fig. 6a is the real data, and there are 2362 banks closed between year 1980 and year 1995. After that, US entered a boom time that only 50 bank failed until 2006. However, the breakout of financial crisis revisits this kind of old story thatthe illiquidity risk in bank strikingly increases. The result of the model

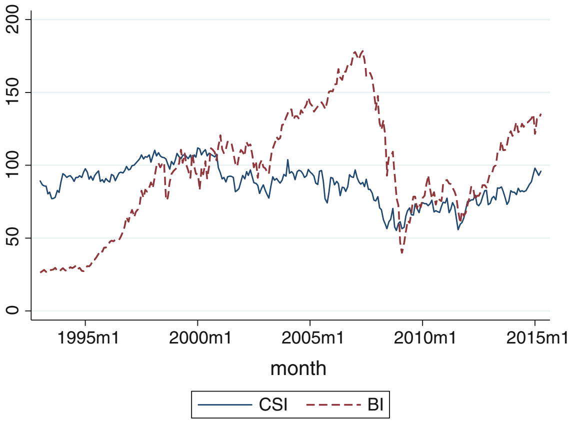


Fig. 5 CSI and BI

is shown in Fig. 6b. Coinciding with the real data, after the peak around year 1989, the risk got down gradually and the strong shock with animal spirits effect still existed until year 1995. The probability of bank runs became a bit lower than 50 % after that year. Yet, between the end of 1990 s and the beginning of 20th century, a small number of bank run cases still appeared because the adjustmentwas implemented before the boom time. Since year 2000, the bank crisis got down till the eruption of subprime crisis. Obviously, there is a discontinuous jump of the bank run risk just as what the real story tells us in financial crisis. Behind this result, the shock aroused by animal spirits might guide the central bank effectively, and the analysis is expressed in the following.

As mentioned in the model section, animal spirits are determined by the memory and the intensity of choice. When $\rho = 0$, there is no memory, and the memory increases with the growth of ρ. When $\rho = 1$, the agents forget nothing, theycan

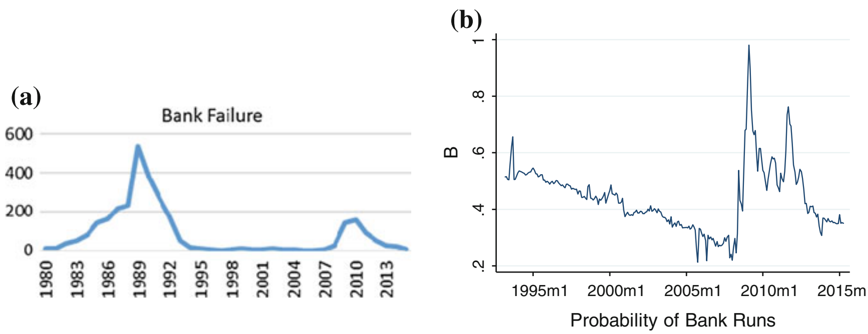
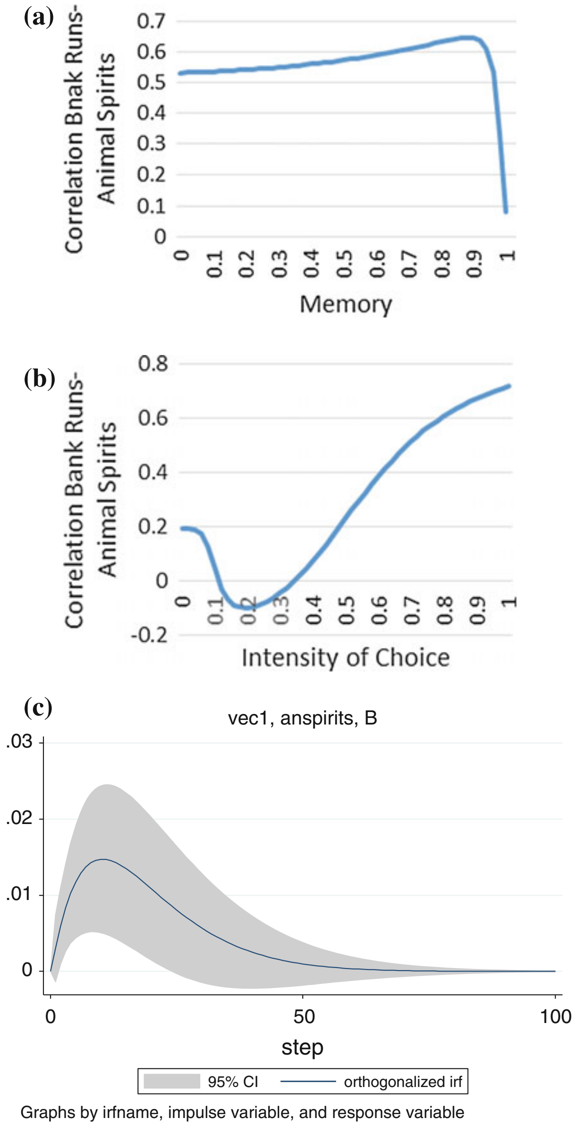


Fig. 6 Real data and modeling result of bank runs

remember all the things with the same weight. The animal spirit might disappear in this situation as they don't have any cognitive limitation. Computing the correlation between bank runs and animal spirits for consecutive values of ρ, the animal spirits arise when they try to remind the history, however, after ρ approaches to 0.95, it will dramatically drop to almost zero (Fig. 7a). This proves that the agents should have the cognitive limitation that forgets something to generate the animal spirits.

Fig. 7 Correlation between bank runs and animal spirits



The parameter γ refers to the intensity of choice that comes from the willingness to learn from the past performance. Whenγ approaches 0, agents choose the rule randomly and the probability to be fundamentalist (or extrapolator) is exactly 0.5 as the time goes on. When γ increases, the willingness to learn will increase. From Fig. 7b, it shows that the correlation between bank runs and animal spirits would get down to a negative value first, which means people will take the irrational behavior as they don't want to learn from their mistake any more when γ is around 0.1 and 0.3.In our data when the economic environment is good enough, γ is low. In the contrast, the γ turns to be quite high when the recession would happen. At this time, the animal spirits arise, which means that the agents need a minimum level of willingness to learn for animal spirits to emergence and to influence the fluctuation of banking system. So far, it is ensured that the cognitive limitation and the willingness to learn would shock the banking system. What's more, considering the longitudinal effect of animal spirits by taking the impulse response function in Fig. 7c, the effect is great and lasting for 100 steps. The agents need a short time to react, so the effect is zero but increases fast at the beginning. Around the10th step the effect becomes largest and then declines gradually for a long time. Overall, it shows that the shock of animal spirits is quite significant and long lasting.

In a good environment, the economic activities are frequent because it is easy to earn money, and the animal spirits are low. Normally, the economic world is working in line with the law. Thus, from the Fig. 8a, it is easy to find that the major distribution is allocated at the lower level of animal spirits. Yet, just a few values are between 0.4 and 0.6, because people become very prudent when the economic environment is not good or bad, and they do not try to switch their choice or learn from the history. And then there is a small peak from 0.6. During the crisis, people would like to swift their wealth to a safer way, so the activities are also relatively frequent. They will try to learn from the aftermath of the history and they will take the extreme reaction when this kind of contagion forms and spreads. From Fig. 8b, there is an increase with the increasing rate from the value around 0.5 of bank run probability. What's more, the

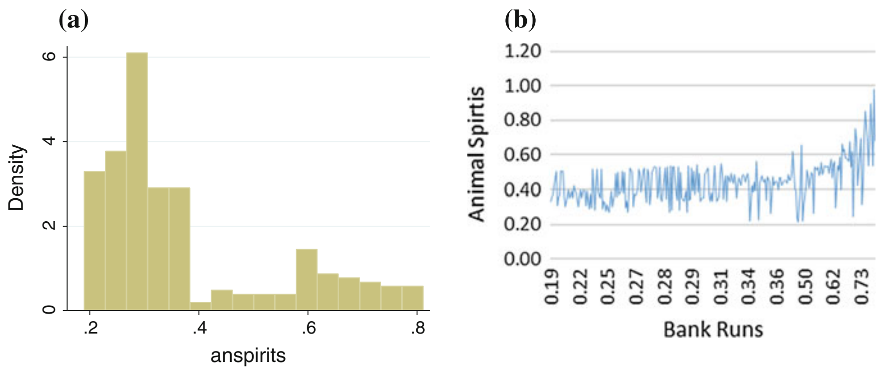


Fig. 8 Distribution and Trend of Animal Spirits

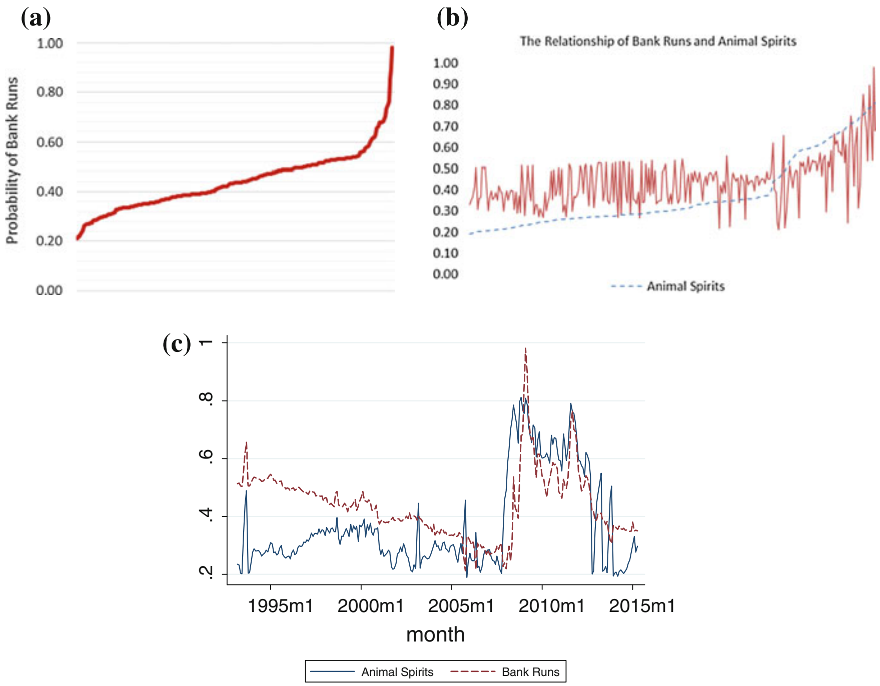


Fig. 9 Bank Runs and Animal Spirits

value becomes very oscillatory until getting the higher level. Therefore, the tipping point here for animal spirits is around 0.5.

When all the probability of bank runs is plotted with the ascending order in Fig. 9a, it is clear that there is a discontinuous jump. What's more, before the jumping the probability is concave shape that increases with the decreasing rate, but after the jumping the shape becomes convex that increases extremely fast. In Fig. 9b, taking the animal spirits with ascending order with the corresponding value of bank runs, the animal spirits jump when the value is around 0.3 andthen the banking system becomes chaotic at first. At this time, the central bank would like to solve the problem, and the banking system turns to stable for a while. However, as the real financial situation exacerbate and the animal spirits approach to 0.5, the trend of bank runs turns into deterministic that to be a higher level, and the so called bank runs phenomenon is formed in this case.

Overall, it is clear that when the shock of animal spirits dramatically increases, the bank runs phenomenon would erupt. The result of real data is shown in Fig. 9c. Before the value of bank runs jumps, the animal spirits dramatically increase ahead during the year 2007. Luckily, before the tipping point of animal spirits appears, there might be some inflection point that makes the economic statementbecome chaotic, and this is the signal that the central bank should implements or swift the monetary policy for preventing the following crisis.

6 Discussion

The effect of animal spirits shows us that small changes could have big, and desirable, consequences. From the result of agent-based simulation, we can see that the policy intervention might be effective if the policy can be actualized before the tip of animal spirits. Mostly, human beings are naive, and preferences often fluctuate as a result of purely present-bias effect (Loewenstein 1996). When the crisis occurs, people normally pay attention to the recent depression, instead of the gain in the future, and then the tips of bank runs might easily emerge. On the other side, sophistication effect means sophisticates are fully aware of any self-control problems they might have in the future, and this awareness can influence behavior now. Thus, sophisticates are influenced by the sophistication effect in addition to the present-bias effect. However, a person with projection bias understands the qualitative nature of change in her preferences but underestimates the magnitude (Loewenstein 2000). Thus, while people believe the crisis gradually recovers nowadays, they are still too prudent for the risk free investment. Further, Loewenstein and O'Donoghue (2004) model the behavior motivated by animal spirits as the interaction between deliberative system (assesses options with a broad, goal-based perspective) and affective system (experiences emotions, such as anger and fear, and motivational drives, such as hunger and sex). The deliberative system may expandwillpower to exert influence on the myopic affective system. However, it is a limited cognitive resource that may have an impact on exercising self-control. Individuals have a finite amount of willpower, which is used for all kinds of tasks. Ego depletion theory postulates that willpower (or ego) depletes when using it, but is replenished after a while (like a muscle). Exercising self-control now impacts self-restraint later, e.g., individuals can maintain their diet discipline for some time, but not over long time (Baumeister et al. 1998). To solve the self-control problem, the potential benefits of commitments suggested by Gul and Pesendorfer (2001) and Thaler and Benartzi (2004), and this result supports Diamond and Dybvig (1983). Therefore, a good commitment can control the shocks aroused by animal spirits to avoid this kind of disaster, and this has explained the current situation in China that the banking system is very stable due to the strong commitment from central bank. Yet, although China can be regarded as the symbol with strong commitment, as the inclusive financial system strikingly developing, the mechanism to figure out the probability of crisis breakout is still necessary and attractive, let alone in the open system for most of countries. Hence, this study pays attention on the threshold of time that central bank should implement a stronger intervention for banking system, and discovering the animal-spirit effect is quite remarkable.

7 Conclusion

With the thriving of the complex banking system and its unsound supervision, the risk of bank runs should be considered again. This paper attempts an exploratory study about how the shock aroused by animal spirits can tip the bank run occurrence based on the adaptive learning mechanism. The effect of animal spirit does not come from the irrational expectation, instead, it is from the fact that the agents are eager to learn from the mistake they ever have with the bounded memory ability. Yet, those agents could not capture all the information in the complex market, and they only use the information they can understand driven by both the optimal/pessimistic and extrapolated motivation. Thus, such kind of animal spirit lays at the core of the banking crisis, and it will dramatically rise when the economic environment deteriorates. The distribution of animal spirits shows that economic activities are frequent in both good and bad situation, and the shock caused by animal spirits explains that when the tipping point is passed, the financial contagion takes places even if thereis intervening policy. On the other hand, the result of agent-based simulation reveals that the preventative policy is helpful in preventing the arrival of tipping point, and central bank should implement rescue policies before that point. The tipping point and long run effect of animal spirits demonstrates the reason why crisis is difficult to recover in a short time once it is formed. This has explained why the bank run phenomenon still occurs in Northern Rock even after the announcement of the central bank, and why the public is still scared when we talk about financial crisis.

Overall, this behavior model with endogenous waves of optimism and pessimism accounts for the micro-behavior with the animal spirits —the shock caused by cognitive limitation and willingness to learn —in the bank run phenomenon, and the signal from this kind of shock is beneficial for central bank's monetary policy.

References

Akerlof GA, Shiller RJ (2009) Animal spirits : how human psychology drives the economy, and why it matters for global capitalism. Princeton University Press, Princeton

Anderson SP et al (1992) Discrete choice theory of product differentiation. MIT press, Cambridge

Angeloni I, Faia E (2009) A tale of two policies: prudential regulation and monetary policy with fragile banks, Kiel working paper

Baumeister RF et al (1998) Ego depletion: is the active self a limited resource? J Pers Soc Psychol 74(5):1252

Bollobás B (2001) Random graphs / Béla Bollobás. Cambridge University Press, New York, p 2001

Branch WA, Evans GW (2006) A simple recursive forecasting model. Econ Lett 91(2):158–166

Brazier A et al (2008) The danger of inflating expectations of macroeconomic stability: heuristic switching in an overlapping-generations monetary model. Int J Cent Bank 4(2):219–254

Brock WA, Hommes CH (1997) A rational route to randomness. Econ: J Econ Soc 1:1059–1095

Bryant J (1980) A model of reserves, bank runs, and deposit insurance. J Bank Financ 4(4):335–344

Calvo G (2012) Financial crises and liquidity shocks a bank-run perspective. Eur Econ Rev 56(3):317–326

De Grauwe P (2012) Lectures on behavioral macroeconomics. Princeton University Press, Princeton

De Grauwe P, Grimaldi M (2006) Exchange rate puzzles: a tale of switching attractors. Eur Econ Rev 50(1):1–33
Diamond DW, Dybvig PH (1983) Bank runs, deposit insurance, and liquidity. J Polit Econ 401–419
Diamond DW, Rajan RG (2001a) Banks, short-term debt and financial crises: theory, policy implications and applications. In: Carnegie-Rochester Conference Series on Public Policy, Elsevier
Diamond DW, Rajan RG (2001b) Banks and liquidity. Am. Econ. Rev 422–425
Diamond DW, Rajan RG (2005) Liquidity shortages and banking crises. J Financ 60(2):615–647
Dias A et al (2015) An Agent Based Propagation Model of Bank Failures. Advances in Artificial Economics, Springer, pp 119–130
Duffy J (2006) Agent-based models and human subject experiments. Hand Comput Econ 2:949–1011
Erdös P, Rényi A (1959) On random graphs I. Publ Math Debrecen 6:290–297
Erdös P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5(17–61):43
Gerali A et al (2010) Credit and banking in a DSGE model of the Euro area. J Money, Credit Bank 42(s1):107–141
Gul F, Pesendorfer W (2001) Temptation and self-control. Econometrica 1403–1435
Hafstead M, Smith J (2012) Financial shocks, bank intermediation, and monetary policy in a DSGE model. Unpublished Manucript
Hume D (1739). A treatise of human nature, McLean, Va. : IndyPublish.com, [2002]
Keeling MJ, Eames KTD (2005) Networks and epidemic models. J R Soc Interface 2(4):295–307
Keynes JM (1936) The General Theory of Employment, Interest, and Money. Kessinger Publishing, Whitefish (2010)
Kirman A (2006) Heterogeneity in economics. J Econ Interact Coord 1(1):89–117
Lamberson P, Page SE (2012) Tipping points. Q J Polit Sci 7(2):175–208
Loewenstein G (1996) Out of control: visceral influences on behavior. Organ Behav Hum Decis Proces 65(3):272–292
Loewenstein G (2000) Emotions in economic theory and economic behavior. Am Econ Rev 90(2):426–432
Loewenstein G, O'Donoghue T (2004) Animal spirits: affective and deliberative processes in economic behavior
Miller JH, Page SE (2004) The standing ovation problem. Complexity 9(5):8–16
Schelling TC (1978) Micromotives and macrobehavior. Norton, New York
Pfajfar D, Zakelj B (2009) Experimental evidence on inflation expectation formation. Working paper, Tilburg University
Shubik M, Wilson C (1977) The optimal bankruptcy rule in a trading economy using fiat money. J Econ 37(3):337–354
Silver N (2008) Tipping Point v2.0
Solé RV (2011) Phase transitions. Princeton University Press, Princeton
Thaler RH, Benartzi S (2004) Save more tomorrowTM: using behavioral economics to increase employee saving. J Polit Econ 112(S1):S164–S187
Toivanen M (2013) Contagion in the interbank network: an epidemiological approach. Bank of Finland Research Discussion Paper(19)
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. science 185(4157):1124–1131
Uhlig H (2010) A model of a systemic bank run. J Monet Econ 57(1):78–96
WickensM(2014) How useful are DSGE macroeconomic models for forecasting? Open Econ Rev 25(1):171–193

A Mathematical Note on Stabilization Policy and Dynamic Inefficiency

Masahiro Yabuta

Abstract This chapter aims at discussing an essential idea concerning the economic stabilization policy and its outcome from a dynamics perspective., focusing on a theoretical discussion of the consequences of the stabilization policy. In the current economic situation, not only the fragility of the financial system but also the erratic fluctuation of the economy resulting from operational issues have become apparent. Among financial matters, such as the expansion of the budget deficit, the importance of a stabilization policy via the policy instruments is increasing. In this context, the paper explores the discipline of the stabilization system.

Keywords Stabilization policy · Dynamic inefficiency · Consumption and investment · Policy instruments · Income distribution

1 Introduction

The purpose of this paper is to discuss an essential idea concerning the economic stabilization policy and its outcome from a dynamics perspective. The paper focuses on a theoretical discussion of the consequences of the stabilization policy. In the current economic situation, not only the fragility of the financial system but also the erratic fluctuation of the economy resulting from operational issues have become apparent. Among financial matters, such as the expansion of the budget deficit, the importance of a stabilization policy via the policy instruments is increasing. In this context, the paper explores the discipline of the stabilization system.

This paper is organized as follows: In Sect. 2, we present the basic concept of economic stabilization that will be discussed in this document. In Sect. 3, an optimal stabilization policy is discussed. Then, Sect. 4 analyzes the issue of dynamic inefficiency.

M. Yabuta (✉)
Chuo university, Hachioji, Japan
e-mail: yabuta@tamacc.chuo-u.ac.jp

A. Matsumoto et al. (eds.), *Essays in Economic Dynamics*,
DOI 10.1007/978-981-10-1521-2_14

2 Formulation of the Stabilization Policy

The following sections investigate the policy instruments such as fiscal expenditure and monetary supply to make the economy stable. Yabuta (1993) analyzed the unstable property of the economic growth model, and Chiang (1992), Holbrook (1972) and Scarth (1979) have examined issues of the instrument instability. The economy that this chapter analyzes using foundational research of Harrod (1973) is based on the discussion of the relevance of a myopic stabilization policy and the optimal policy plan when faced with instability. First of all, to clarify the characteristics of the analytical framework, the following single differential equation is assumed:

$$\dot{x} = \phi(x, u), \tag{1}$$

where x is a macroeconomic variable, such as production and employment, and u represents the policy variable to be controlled, such as government spending and high-powered money. For the purpose of analysis, the function ϕ is assumed to be partially differentiable and the following is assumed:

$$\phi_x > 0, \phi_u > 0. \tag{2}$$

Because the first sign of (2) implies that an increase of x leads to an additional increase of x itself, the macroeconomic system shown by (1) is unstable. The second inequality of (2) means that a rise in policy measures, such as government spending and money supply leads to an increase in macroeconomic variables, such as production and employment. In this context, for the second partial derivatives concerning (1), the following is assumed:

$$\phi_{xx} < 0, \phi_{uu} < 0, \quad \phi_{ux} = \phi_{xu} = 0. \tag{3}$$

Equations (1)–(3) are justified when introducing the investment function of the Harrod type. The Harrod model (1939), usually, is characterized by an instability property with a "knife-edge" shown by the following:

$$\dot{g} = h(\delta(g), u), \partial h/\partial \delta > 0, d\delta/dg > 0, \tag{4}$$

where g is the rate of capital accumulation and δ is the capital utilization rate. The rise in the rate of capital accumulation leads to excess demand in the market to increase the rate of capital utilization. Typically, companies that face a shortage of capital equipment expand new investment, resulting in a rise in the capital accumulation rate. Equation (4) implies $\partial \dot{g}/\partial g = h_\delta \delta' > 0$, meaning the instability of the economy, which is known as the "balance on the blade of the knife (knife-edge equilibrium)".

In the following section, the paper focuses on a structural formula concerning the instability of the economy given in (1). When the private economy itself has instability in (1), the paper looks for the way that leads to the overall stability of the entire economy, in addition to the use of government functions. Removing the overall economic instability through a variety of policy measures and leading the economy to a stable equilibrium is expected by the government. In this case, a stabilization policy has the formula to adjust a policy measure u in (1):

$$\dot{u} = \varphi(x, u), \tag{5}$$

where φ is assumed to have continuous partial derivatives.[1] To clarify the nature of the equilibrium (x^*, u^*) given by (1) and (5), the Olech theorem, which provides sufficient conditions for global stability (Desai 1973, is applied. For convenience of analysis, the equilibrium is re-evaluated at the $(x - x^*)$ and $(u - u^*)$, leading to a new equilibrium evaluated at (0, 0). Further, for simplicity, until now, the same variable symbols x and u apply.

The Jacobi matrix J of (1) and (5) is

$$J = \begin{bmatrix} \phi_x & \phi_u \\ \varphi_x & \varphi_u \end{bmatrix}. \tag{6}$$

According to the Olech theorem, the conditions under which the equilibrium (0, 0) becomes globally stable are as follows[2]:

$$\textit{Trace of } J = \phi_x + \varphi_u < 0, \text{ everywhere} \tag{7a}$$

$$\textit{Determinant of } J = \phi_x \varphi_u - \phi_u \varphi_x > 0, \text{ everywhere} \tag{7b}$$

and

$$\phi_x \varphi_u \neq 0, \text{ everywhere or } \phi_u \varphi_x \neq 0, \text{ everywhere.} \tag{7c}$$

Equation (7c) implies that whenever a policy variable u changes, the economy shown by x should change. Equation (7a), (7b), and (2) together provide the following:

$$\varphi_u < -\phi_x < 0, \tag{8a}$$

$$\varphi_x < \frac{\phi_x}{\phi_u} \varphi_u < 0. \tag{8b}$$

[1] In this context, note that it does not mention the automatic stabilizer of the economy (built-in stabilizer). The built-in stabilizer, such as the progressive taxation system, is known as one of the effective means of stabilizing the economy. In the case of introducing it, for example, a formula $u = u(x)$ leads to the stability conditions concerning (1) to be expressed as follows:

$$\frac{d\dot{x}}{dx} = \phi_x + \phi_u u'(x) < 0 \ \textit{or} \ u'(x) < -\frac{\phi_x}{\phi_u} < 0.$$

[2] It is notable that the conditions (7a)–(7c) does not assure the non-negativity of the variables, meaning that these conditions are the ones for semi-global stability rather than global stability from the economic perspective.

Equation (8a) and (8b) show the rule that the government should follow to make the system of (1) and (5) stable by adjusting u as a policy stabilizer. This rule means that an adjustment of the policy measure should be sufficiently large to compare the adjustment in the economy. For example, if the government's stimulus policy leads to economic recovery, the rule means that the reversal of or reduction in government spending is needed authoritatively.

The following phase diagrams help us to understand the adjustment process of u and x visually. For the system of (1) and (5), the slope of the curve of $\phi(x, u) = \varphi(x, u) = 0$ is given by the following:

$$0 > \frac{du}{dx}\,|_{\dot{x}=0} = -\frac{\phi_x}{\phi_u} > \frac{du}{dx}\,|_{\dot{u}=0} = -\frac{\varphi_x}{\varphi_u}. \tag{9}$$

On the other hand, the characteristic equation of the system and its discriminant is as follows:

$$\Delta = (\phi_x + \varphi_u)^2 - 4(\phi_x\varphi_u - \phi_u\varphi_x). \tag{10}$$

Equation (7a) and (7b) together do not give a clear sign condition of (10). If Δ is positive (or zero), two of the roots of the characteristic equation are each a negative of different real solutions (or a multiple root). In this case, the equilibrium becomes a stable node (node). This is shown in Fig. 1a. On the contrary, if Δ is negative, two of the roots become a conjugate complex solution, and the equilibrium point becomes a stable spiral point. Figure 1b depicts this case.

It is evident that the economy can not eliminate instability without the government adjusting appropriate policy measures. To stabilize the unstable economy as depicted in the Harrod model, the government should appropriately adjust the policy variable u in the form of (5), while satisfying (8a) and (8b). In cases where the economy

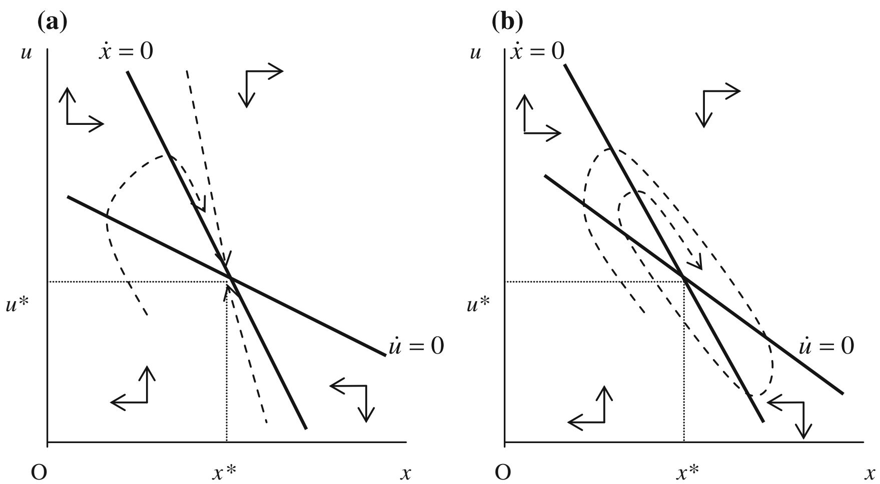


Fig. 1 Stabilization policy and policy variable. **a** Stable node. **b** Stable spiral point

is of an unstable nature, we can define such a short-term stabilization policy of the government as a myopic stabilization policy (MSP). In this regard, to define a set of MSPs for economic system (1) will be as follows:

$$\boldsymbol{M} = \{\text{MSP} \mid \dot{u} = \varphi(x, u) \textit{ with } (6) \textit{ and } (7) \ (\textit{or } (6) \textit{ and } (8))\},$$

where $\boldsymbol{M}$ is not null, and the following conditions will be assured:

Proposition 1 *The government can stabilize the economy by adopting a certain MSP* $\in \boldsymbol{M}$.

Proposition 2 *Assume that the autonomous system (1) is a neoclassical type of stable economy with* $\phi_x < 0$ *and* $\varphi_u > 0$. *In this case, the government can stabilize the economy by adopting a certain MSP* $\in \boldsymbol{M}$.

These propositions suggest that whenever the government suitably adopts an MSP, it can manage the economy so as to attain stability. In this sense, especially concerning Proposition 2, an MSP belonging to $\boldsymbol{M}$ can be referred to as too cautious a policy (or an overcautious policy).

3 Optimal Stabilization Policy

Following Phillips (1954), consider the following typical optimization policy:

$$\textit{Minimize} \int_0^\infty \{(x - x*)^2 + \beta(u - u*)^2\} e^{-\rho t} dt \textit{ subject to } (1) \textit{ with } (2). \quad (11)$$

A fundamental analytical framework for the stabilization policy was given by Blancherd and Fisher (1996) and Kamien and Schwartz (1991). Here, the variables with an asterisk, such as x^* and u^*, show a target value.[3] Any deviation from these target values reduces the welfare of the society. In this optimal control problem, x is a state variable and u is a control variable.

The Hamiltonian concerning (11) is as follows:

$$H = \{x^2 + \beta u^2\} + \lambda\phi(x, u), \quad (12)$$

where λ is the adjoint variable and shows the shadow price of x. From the maximum principle in (12) and the variation of x in (1), the following necessary conditions are assured:

$$\frac{\partial H}{\partial u} = 2\beta u + \lambda\phi_u = 0, \quad (13)$$

[3]In the equilibrium, these target values are adjusted to be zero.

and

$$\dot{\lambda} = -\frac{\partial H}{\partial x} + \rho\lambda = -2x + (\rho - \phi_x)\lambda_D. \tag{14}$$

It is clear that $\partial^2 H/\partial u^2 = 2\beta\{1 - u\phi_{uu}/\phi_u\}$ is positive from (2), meaning that H is minimized. Differentiating (13) with respect to time, an adjustment equation concerning the policy variable u is attained as follows:

$$\dot{u} = \chi(x, u) = \frac{1}{1+\varepsilon}\left[\frac{x}{\beta}\phi_u + (\rho - \phi_x)u\right], \tag{15}$$

where $\varepsilon = -\phi_{uu}u/\phi_u > 0$. Consider the system of (1) and (15). In the equilibrium of this system, $\phi(0, 0) = \chi(0, 0) = 0$ is assured. The Jacobi matrix evaluated at (0, 0), J_0, becomes as follows:

$$Jo = \begin{bmatrix} \phi_x & \phi_u \\ \chi_x & \chi_u \end{bmatrix}, \tag{16}$$

where

$$\chi_x = \frac{\phi_u}{(1+\varepsilon)\beta}, \tag{17}$$

and

$$\chi_u = \frac{\rho - \phi_x}{1+\varepsilon}. \tag{18}$$

In (16), considering (17) and (18) leads to the following:

$$\textit{Trace of Jo } = \phi_x + \chi_u = \frac{1}{1+\varepsilon}(\rho + \varepsilon\phi_x) > 0 \tag{19a}$$

$$\textit{Determinant of Jo } = \phi_x\chi_u - \phi_u\chi_x = \frac{1}{1+\varepsilon}\left[\phi_x(\rho - \phi_x) - \frac{\phi_u^2}{\beta}\right] \equiv \frac{1}{1+\varepsilon}F(\phi_x) \tag{19b}$$

Because the trace of J_0 is positive, the equilibrium can be neither a stable node nor a stable spiral point. Here, only a saddle point can be converged to the equilibrium along a dynamic stable path. It is necessary for the equilibrium to be a saddle point that $F(\phi_x)$ of the second term of the right-hand side of (19b) is negative. Consider the discriminant, Δ_F, of a quadratic function $F(\phi_x)$ of ϕ_x. Note that all values are evaluated at the equilibrium point (0, 0). Also, confirm the following equations to focus on the sign conditions concerning $F(\phi_x)$:

$$F(0) = F(\rho) = -\phi_u^2/\beta < 0, \tag{20a}$$

$$F\prime(\rho/2) = 0, \tag{20b}$$

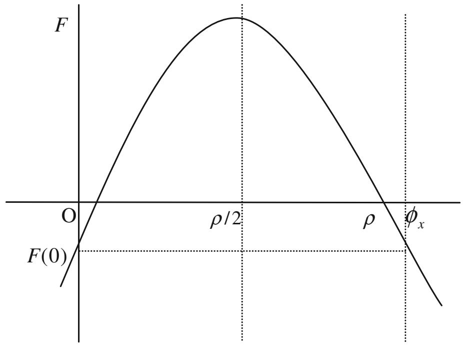


Fig. 2 Optimization and saddle point domain

$$\Delta_F = \rho^2 - 4\phi_u^2/\beta. \tag{20c}$$

Taking (20a), (20b), and (20c) into consideration, $F(\phi_x)$ can be depicted as in Fig. 2.

Let the economy have the inherently instability shown by $\phi_x > 0$. Even if an optimal stabilization policy is taken, controlling the policy variables along the path to the saddle point is not always possible at any time. The case that such control is not possible is shown by the area where F is above the horizontal axis ϕ_x depicted in Fig. 2. The magnitude of $\phi_x (= \partial\dot{x}/\partial x)$ represents the rate of change of the economy itself induced by changes in the variable. For example, a change in the unit of investment causes the investment to change itself through varying income and employment. Therefore, it is notable that whether or not controlling the policy variable can successfully lead to a stable saddle point is dependent on the individual nature of the economy in advance. As the neoclassical model assumed, as far as the economy is inherently stable, as shown by $\phi_x < 0$, the equilibrium becomes a saddle point, and it is possible for the government to choose the optimal path toward a steady state.

On the other hand, when the economy is unstable with $\phi_x > 0$, some additional conditions to make $F(\phi_x)$ negative are needed. First, in the case of $\Delta_F < 0$, F is always negative regardless of the restrictions on ϕ_x. An additional condition is given by the following:

$$\frac{\sqrt{\beta}\rho}{2} < \phi_u. \tag{21}$$

Equation (21) is satisfied easily when the impact of policy variables on the private economy is (i.e., the larger ϕ_u gets) large enough or a weight of welfare related to the target policy variable β and the social discount rate is small enough. Even if (21) does not hold true, once the followings inequalities are satisfied, it is possible for the government to perform the stabilization policy:

$$0 < \phi_x < \frac{-\rho - \sqrt{\Delta_F}}{2} \text{ or } \frac{-\rho + \sqrt{\Delta_F}}{2} < \phi_x. \tag{22}$$

As already discussed, when the economy is inherently unstable, whether a stable path toward a stable saddle point exists or not is dependent on the ranges that ϕ_x and ϕ_u take. Unlike the case of MSP in (4), the government must continuously adjust and control the policy variable u according to (15), whose essential components are given by (1). In this sense, the optimal stabilization policy is closely related to the stability conditions of the private economy. Unfortunately, there exists a case where the government can never realize the optimal control policy that leads to the steady state.

While the private economy is inherently unstable, if an optimum stability control policy exists, it is referred to as the optimal stabilization policy (OSP). Then, a set of OSPs related to (1) characterized by (2) is as follows:

$$\boldsymbol{O} = \{\mathrm{OSP} \mid \dot{u} = \chi(x, u) \ \textit{with} \ (22) \ \textit{and negative} \ F(\phi_x) \ \textit{in} \ (23)\}.$$

To summarize the above discussion, the following propositions on the set $\boldsymbol{O}$ will hold:

Proposition 3 *$\boldsymbol{M} \cap \boldsymbol{O}$ is empty. Hence, an MSP is not in OSP, and it is impossible for an OSP to be in MSP.*

Proposition 4 *The government can only pursue an optimum stabilization path only if $F(\phi_x)$ in (19b) is negative. It is possible for the set $\boldsymbol{O}$ to be empty.*

Proposition 5 *Unlike the Harrodian economy with instability, let's assume that the economy is inherently stable, with $\phi_x < 0$ and $\varphi_u > 0$. In this case, $\boldsymbol{O}$ is not an empty set, and it is possible for the economy to follow a stable optimized path toward a steady state.*

A comparison of $\boldsymbol{O}$ and $\boldsymbol{M}$ seems to have significant implications for the following points. As far as the goal of economic policy is concerned, full employment, improvement of income, and price stability are often discussed. In reality, it is necessary to implement a countercyclical policy against economic fluctuation. In this regard, (5) assumes that the government carries out an economic policy post-correspondingly to the reality of economic trends, but (15) shows a case in which the government can control the economy when it confirms the policy goal. The former is a policy related to $\boldsymbol{M}$, but the latter is involved in $\boldsymbol{O}$. Hence, a stabilization policy that has these two properties at the same time does not exist. The understanding itself that stabilization policies must be carried out by either formulation defines our analytical method.

4 The Issue of Dynamic Inefficiency

The macroeconomic policies for stabilizing economic fluctuations induced by the private sector were examined in the previous sections. We found that the relationship between the stability of the private economy and the effectiveness of the stabilization policy of the government is the problem. This section focuses on a private sector economy from a strategic optimization perspective. One of the modern economic challenges is in the relationship of labor and capital, therefore on the income distribution of wages and profits. In fact, even in Abenomics in Japan, which puts the monetary policy within the axis, salary increases are considered to be a meaningful measure to stabilize the private economy, which means that at the same time, a wage hike means the reallocation of profits.

The problem of income distribution clearly includes the intertemporal decision-making problem because the way the current income is distributed determines future income distribution. It is evident that a larger profit distribution due to less wages paid results in rapid capital accumulation, leading to the expansion of future income levels as a consequence. Hence, considering the intertemporal relationship between investments and savings is significant for the analysis from a long-term perspective. Here, this study considers somewhat older topics that Lancaster (1973) and later scholars have developed.

The framework of optimal control to be used here includes two players, the workers and the companies (or the capitalist) in a dynamic differential game. The problem, rather than governing, is of finding the optimal decisions for each economic entity. As mentioned earlier, there is a dilemma with regard to income distribution. First, workers, even though they gain enhanced current consumption acquired from higher wages, they might reduce consumption of the future. Whether or not future consumption becomes significant depends on corporate savings and its accumulation behavior. On the other hand, while companies seem to be able to control future production, they do not appear to be able to control income distribution. In other words, the outcome distribution that each entity can receive is dependent on the decisions of other economic agents. Such a conflict over current and future consumption is understood as a dynamic battle (an active conflict of the capitalist economy); therefore, this framework can be formulated as a differential game—a differential game between workers and capitalists).[4]

First of all, we try to set up a basic model on the basis of Lancaster (1973). The production function is assumed to be of the AK type:

$$Y = aK, \tag{23}$$

where Y is income and K is capital stock. On the demand side, the demand for the product is given by the following:

$$Y = Cw + Cc + I, \tag{24}$$

[4]The analytical framework here depends on Lancaster (1973). Also, there are analyses such as Pohjola (1983, 1984) and Zeeuw (1992). In Pohjola (1984), in response to the previous study, incentives, both institutional and consensus rules of the cooperation policy in capitalism, are analyzed, and the fact that the presence of the Nash bargaining solution in the Lancaster model is proven to affect the bargaining solution with the nature of the threat optimum strategy has been analyzed. Moreover, Pohjola (1983) conducted a comparative study of the Nash and Stackelberg solutions, where dynamic biological inefficiencies that were dealt with in this section are likely to be reduced by the Stackelberg solution. It also proved that the Stackelberg game is in a state of stalemate because both workers and capitalists never act as leaders.

where C_{w} is the consumption by workers and C_{c} is the consumption by firms. In this relation, let μ be the labor share and ν be the investment rate of firms. Then, investment and consumption are assumed to be shown as follows:

$$Cw = \mu Y, c \leq \mu \leq \quad b, (1/2 \leq b) \tag{25}$$

$$Cc = (1-\mu)(1-\nu)Y, 0 \leq \nu \leq 1 \tag{26}$$

$$I = (1-\mu)\nu Y. \tag{27}$$

The accumulation process of capital is as follows:

$$\dot{K} = I. \tag{28}$$

The workers' problem to maximize their consumption is given by the following:

$$\max J_1 = \int_0^T aK\mu dt \ subject\ to\ \dot{K} = aK(1-\mu)\nu, \tag{29a}$$

while the firms' problem becomes

$$\max J_2 = \int_0^T aK(1-\mu)\nu dt \ subject\ to\ \dot{K} = aK(1-\mu)\nu. \tag{29b}$$

It is assumed that workers try to control μ to pursue (29a), whereas firms would control ν so as to meet (29b). Here, we apply the maximum principle to an optimal control problem related to (29a) and (29b). We define Hamiltonian for workers, H_1, and firms, H_2, respectively, with each adjoint variable $\lambda_i (i = 1, 2)$ as (30a) and (30b). Then, as far as the workers' problem is concerned, the following hold true:

$$H_1 = aK\mu + \lambda_1[aK(1-\eta)\nu], \tag{30a}$$

$$\partial H_1/\partial\mu = aK(1-\lambda_1\nu), \tag{31a}$$

$$\dot{\lambda}_1 = -\partial H_1/\partial K = -a[\mu - \lambda_1(1-\mu)\nu]. \tag{32a}$$

Moreover, the related transversality condition is given by the following

$$\lambda_1(T) = 0. \tag{33a}$$

Then,

$$\mu = cif\lambda_1\nu > 1, \mu = bif\lambda_1\nu < 1 \tag{34a}$$

will be confirmed from (31a).

Similarly, the following are assured as far as the firms' problem (29b) is concerned:

$$H_2 = aK(1-\mu)\nu + \lambda_2 aK(1-\eta)\nu, \tag{30b}$$

$$\partial H_2/\partial \nu = aK(1-\nu)(\lambda_2 - 1), \tag{31b}$$

$$\dot{\lambda}_2 = -\partial H_2/\partial K = -a(1-\mu)[1 + (\lambda_2 - 1)\nu], \tag{32b}$$

$$\lambda_2(T) = 0. \tag{33b}$$

Hence, the following will hold true:

$$\nu = 0 \, if \, \lambda_2 < 1, \nu = 1 if \lambda_2 > 1. \tag{34b}$$

The transversality condition implies $\mu = b$ and $\nu = 0$ for any t ($t_0 \leqq t \leqq T$), leading to the following:

$$\dot{K} = 0, \dot{\lambda}_1 = -ab, \dot{\lambda}_2 = -a(1-b). \tag{35}$$

In (35), because the time variation of λ_2 is equal to $a(1-b)$, it is clear that for t_0 to be equal to $T - 1/(a(1-b))$, $\lambda_2(t_0) = 1$, and $\lambda_1(t_0) = b/(1-b)$. Hence, because $b \geqq 1/2$, it is assured that $\lambda_1(t_0) > 1$, and $(\mu, \nu) = (c, 1)$ for $t \in [0, t_0]$. In any event, ultimately, $(\mu, \nu) = (b, 0)$ becomes the control to be performed.

Apart from the optimal control behavior of the economic agents in such a conflict between workers and firms, there would be an optimal control behavior that should be considered from a social perspective. Here, the sum of the welfare of both entities is envisaged, i.e., $J_1 + J_2$ is regarded as a social welfare function to be maximized. In this framework, a social planner problem is as follows:

$$\max J_1 + J_2 = \int_0^T aK(1-\sigma)dt \textit{ subject to } \dot{K} = aK\sigma, \tag{29s}$$

where $\sigma = (1-\mu)\nu$, and $0 \leqq \sigma \leqq 1 - c$. It is necessary for a social planner to control the social investment rate σ so as to meet (36). Hamiltonian Hs is defined as follows:

$$H_s = aK(1-\sigma) + \lambda_s aK\sigma. \tag{30s}$$

Hence, the followings will be

$$\partial H_s/\partial \sigma = -aK(1-\lambda_s) \tag{31s}$$

$$\dot{\lambda}_s = -\partial H_s/\partial K = a[1 - \sigma(1-\lambda_s)]. \tag{32s}$$

Moreover, the transversality condition becomes the following:

$$\lambda_s(T) = 0. \tag{33s}$$

From this, for any t in $t^* \leqq t \leqq T$ consisting of $\lambda_s < 1$, it can be seen that $\sigma = 0$ becomes the control to be performed. Furthermore, in this case, since $\dot{K} = 0$, $\dot{\lambda}_s = -a$, and for $t^* = T - 1/a$, $\lambda_s\ (t^*) = 1$ is satisfied. Thus, for any t in $0 \leqq t \leqq t^*$, $\lambda_s > 1$ and $\sigma = 1 - c$ are also satisfied. In this way, a social planner, even initially taking the maximum value of the investment rate (1 − c), should drive a policy to convert to zero at some point in time.

The problem is whether or not divergence exists between the consequence brought about by the social optimization program and that brought about by the individual optimizing behavior of each entity (let's call this "the capitalist program"). To clarify this point, let's focus on the time when the policy change is needed, t_0 and t^*, respectively. Clearly, the following holds true:

$$t^* = T - 1/a > T - 1(a(1-b)) = t_0. \tag{36}$$

Therefore, it is justified to classify the policy periods into (i) $t \in [0, t_0]$, (ii) $t \in [t_0, t^*]$, and (iii) $t \in [t^*, T]$. Table 1 summarizes the differences in variables, such as consumption and investment in each period, wherein, C is the representative of the total consumption. As in Table 1, the difference between the two programs occurs after period (ii).

The total welfare, W_0, attained through period (ii) and period (iii), is given by the following:

$$W_0 = aK_0(T - t_0) = K_0/(1-b), \tag{37}$$

while the welfare by a social optimum program, W^*, considering (36), is given as follows:

$$\begin{aligned} W^* &= \int_{t_0}^{t^*} caK_0 \exp[a(1-c)(t-t_0)]dt + aK_0 \exp[a(1-c)(t^*-t_0)](T-t^*) \\ &= [\exp\{b(1-c)/(1-b)\}/(1-c) - c/(1-c)]K_0. \end{aligned} \tag{38}$$

The first term on the right side of (38) is a welfare measure implemented in period (ii), and the second term refers to the welfare measure in period (iii).

Exp$(x) > 1 + x$ holds true for $x > 0$, and let $x = b(1-c)/(1-b)$, meaning that Exp $(b(1-c)/(1-b)) > 1 + b(1-c)/(1-b)$ holds. Upon dividing this by $1 - c$

Table 1 Capitalist program and social optimization program

	Capitalist program	Social optimization program
(i)	$\dot{K} = a(1-c)K$ $C = caK$	$\dot{K} = a(1-c)K$ $C = caK$
(ii)	$\dot{K} = 0$ $C = aK, K = K_0$	$\dot{K} = a(1-c)K$ $C = caK$
(iii)	$\dot{K} = 0$ $C = aK, K = K_0$	$\dot{K} = 0$ $C = aK, K = K^* > K_0$

and multiplying by K_0, the expression $\{b(1-c)/(1-b)\}K_0/(1-c) > [1/(1-c) + b/(1-b)]K_0$ is satisfied. Then, considering both (37) and (38), $W^* > [1 + b/(1-b)]K_0 = W_0$ holds true, meaning that the capitalism program can never achieve the welfare to be attained in the social optimum program. This is related to the issue of the capital accumulation system in period (ii) and is caused by the separation between workers and firms for making decisions of consumption and investment.

Of course, the result of the inefficiency of the capitalism program might come from the Lancaster type of analytical system. It is notable that essentially, the relationship of conflict is caused by a relationship over income distribution. Hence, the assumption that it is possible to control the rate of distribution μ in an optimizing way for the benefit of the entire group of workers might be simply meaningful in the normative hypothesis. On the other hand, along this analytical method, it is also possible to envisage a workers' organizer. This, named as the workers' program, would be written as follows:

$$\max J_w = \int_0^T aK\omega dt \text{ subject to } \dot{K} = aK(1-\omega), c \leq \omega \leq 1, \tag{29w}$$

where ω represents the consumption rate that the workers can facilitate. Solving this optimal control problem leads to the following:

$$H_w = aK\omega + \lambda_w[aK(1-\omega)], \tag{30w}$$

$$\partial H_w/\partial\omega = aK(1-\lambda_w), \tag{31w}$$

$$\dot{\lambda}_w = -\partial H_w/\partial K = -a[\omega - \lambda_w(1-\omega)], \tag{32w}$$

and the transversality condition is given by the following:

$$\lambda_w(T) = 0. \tag{33w}$$

From (33w), $\omega = 1$ for $1 > \lambda_w$. Here, it is clear that $\lambda_w(t_w) = 1$ holds for any $t, t_w \leqq t \leqq T$ and $t_w = T - 1/a$. After all, a workers' program will bring an entirely equivalent outcome as the optimal social program seen before. This result might imply that the noncooperative game between both classes (leading to a Nash solution) brings a prisoner's dilemma solution and that a cooperative game shows the possibility of a Pareto improvement. Unfortunately, the analysis mentioned above does not indicate that the direction of the cooperation itself between both classes transcends class.

References

Blancherd OJ, Fisher S (1996) Lectures on macroeconomics. The MIT Press, Cambridge
Chiang AC (1992) Elements of dynamic optimization. McGraw- Hill, New York
Desai M (1973) Growth cycles and inflation in a model of the class struggle. J Econ Theory 6:527–545
Harrod RF (1973) Economic dynamics. Macmillan Press, London
Holbrook R (1972) Optimal economic policy and the problem of instrument instability. Am Econ Rev 62–1(2):57–65
Kamien MI, Schwartz NL (1991) Dynamic optimization, 2nd edn. North Holland Publishing, Amsterdam
Lancaster K (1973) The dynamic inefficiency of capitalism. J Polit Econ 81–5:1092–1109
Phillips AW (1954) Stabilization policy in a closed economy. Econ J 64:290–323
Pohjola M (1983) Nash and Stackelberg solutions in a differential game model of capitalism. J Econ Dyn Control 6:173–786
Pohjola M (1984) Threats and bargaining in capitalism a differential game view. J Econ Dyn Control 8–3:291–302
Romer PM (1990) Endogenous technical change. J Polit Econ 98:71–102
Scarth WM (1979) Bond financed fiscal policy and the problem of instrument. J Macroecon, pp 107–117
Yabuta M (1993) Economic growth models with trade unions; NAIRU and union behavior. J Macroecon 15(2):381–400
Zeeuw A (1992) Note on Nash and Stackelberg solutions in a differential game model of capitalism. J Econ Dyn Control 16–1:139–145