A Self-Switching Markov Approach to the Analysis of the Business Cycle

by

Fred Engst

A dissertation submitted to the

Graduate School - New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for a degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Professor Hiroki Tsurumi

and approved by

New Brunswick, New Jersey

October, 1997

UMI Number: 9814069

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

Abstract of the Dissertation

A Self-Switching Markov Approach to the Analysis of the Business Cycle

by FRED ENGST

Dissertation Director:

Professor Hiroki Tsurumi

This dissertation explores a new class of time series models that is more compatible with endogenous business cycle theories than the existing models in contemporary economic literature. It is more compatible in the sense that it can capture more readily the endogenous component of business cycle fluctuations, and it can model the business cycle as a limit cycle. This goal is achieved by nesting both the threshold autoregressive (TAR) model and the Markov-switching (MS) model into a class of self-switching Markov (SSM) models.

The motivations for pursuing an endogenous business cycle modeling strategy are reviewed first. In this section, I evaluate a variety of factors that provide nonlinear feedback to an economy, including the dual roles of competition, innovation, the impact of changing the optimal scale of production, the

role of money, asymmetry of information, and many other factors that can affect the dynamics of an economy. Next I review the evidence of time series nonlinearity, and models of endogenous business cycles.

After the formulation of a SSM model, I compare it to a number of benchmark time series models, including James Hamilton's (1989) fixed transition probability (FTP) MS model, Andrew Filardo's (1994) time-varying transition probability (TVTP) MS model, as well as Howell Tong's (1990) TAR models.

In contrast to MS models that assume regimes are exogenously determined, I find the endogenous information to be significant in predicting the switch in regime. Compared to TAR models that rely on discrete thresholds and delay factors, the SSM approach also improves likelihoods significantly.

When the SSM model is applied to the monthly changes in the U.S. unemployment rate, I identify many estimates that exhibit stable and persistent limit cycles of diverse periodicity, up to 20 months, in forecasts or simulations.

Finally, I observe that the endogenous switching parameters to be statistically significant in all models. This finding lends empirical support to endogenous business cycle theories. The challenge remains, however, to model the business cycle as a low frequency (five to ten years) limit cycle.

Acknowledgments and Dedication

I thank my committee members: Hiroki Tsurumi, Willi Semmler, Bruce Mizrach, Bartholomew Moore, and Peter Ireland; for their patience and encouragement, as well as invaluable inputs. I am grateful to Dorothy Rinaldi, the graduate secretary in the economic department, and all staffs in the department. They created an environment that made my graduate study possible.

Thanks also go to Richard Smith. Without his help in editing an earlier draft, this dissertation would have been less comprehensible.

I am thankful for the support and encouragement from my friends and family. I am especially indebted to my parents Joan Hinton and Sid Engst who sparked my interest in economics while I was growing up in China, and who supported me in time of difficultly.

I dedicate this dissertaton to my daughters Katie and Gina, with love.

Table of contents

List of Tables

List of Illustrations

1. Introduction

A vast body of economic literature has been devoted to the study of business cycles, and yet the profession's understanding of the cause for cycles is still limited. Some of the most basic questions are not well understood. For example, "Is there a systematic tendency for *cycles?*". In other words, is there a "natural" rhythm in a market economy? If there is, is the cycle or rhythm generated mainly from exogenous or endogenous sources? These questions are fundamental to a better understanding of the nature of the business cycle.

To address these questions, economists have built numerous models of the economy. Imbedded within these business cycle models, however, implicit or otherwise, are the philosophical world views of the model builders. The view that a market economy is best characterized by balance and equanimity despite fierce competition is implicit in reactive models that passively respond to outside influence. This dominant world view in the economic literature sees the economy as an orderly, balanced system; a system that is always in a state of market-clearing equilibrium. In the dynamic version of this world view, not only do all markets clear at every moment of time, the economy is also assumed to be on a path towards a steady-state equilibrium. The steady-state is either assumed to be stable, or the instability is largely ignored. The focus of this literature is more on *how* rather than *why* an economy fluctuates. The question of *why* is relegated to factors external to the economy.

This combination of a short-run market-clearing and long-run steady-state assumptions together with agent optimizations are the foundations of the new classical school of dynamic general equilibrium models. By assuming infinite speed of price adjustment, these models make optimization synonymous with equilibrium. Thomas Cooley states:

> Most economists now accept as incontrovertible the notion that theories of the business cycle should be consistent with the long-term observations about economic growth and with the principles of *competitive equilibrium* theory (1995, xv, emphasis added).

This new classical school attributes the causes of cycles to external shocks, such as supply interruptions, technological innovations, government interference, and wars. Otherwise, they believe that the market economy will function smoothly. Their focuses are on the propagation mechanism of the economy from external shocks, while ignoring the possibility that impulses might be internally generated.

In contrast, the view that a market economy is best characterized by turbulence and upsets is implied in proactive models with an internal dynamic. According to this literature, not all markets clear at the same instantaneous speed. For example, the financial market clears faster than the product and labor markets.

From this nonmarket-clearing perspective, the fact that there might exist a price that a market can clear does not mean that the market knows what that price is. Only through non-zero excess demand could that market-clearing price be revealed, especially when a new product is introduced. Furthermore, according to this literature, the market might not be converging towards the steady-state, especially when the steady-state is locally unstable. This nonmarket-clearing literature focuses more on the question of *why* than *how* an economy fluctuates. This focus leads to endogenous theories of business cycles and the study of the out-of-equilibrium adjustment mechanisms.

Some critics in this literature see the infatuation with steady-state equilibrium among some new classical business cycle model builders as missing the point. "Everything that matters and is of interest to us happens *be-*

cause the system is not in equilibrium." For example, in meteorology, "The true equilibrium state, in the absence of heat input from the sun, is at a temperature where all life comes to a stop!" (Blatt 1983, 5).

From the endogenous and disequilibrium school perspective, the economy that perhaps came close to being in a steady-state equilibrium was the medieval economy. To compete in a market economy, firms do not seek balance in the market. Each firm innovates to create an imbalance that forces the others to follow. The process of market competition, or what Joseph Schumpeter termed "creative destruction" (Schumpeter, 1939), is one of the real driving forces that moves the economy forward.

The insistence on eternal market-clearing equilibrium is defensible only in a linear world (Puu 1993, 3); for it is an unattainable modeling strategy otherwise. In a nonlinear world, however, it is no cause for concern. On the contrary, builders of endogenous models are attracted to locally unstable and yet globally stable systems; for they make their systems more complex and interesting. Many of them prefer the dynamical system approach to the study of business cycles for its ability to capture, to a greater degree, the internal dynamic of a complex system. They believe that the observed business cycle can be better understood as the result of interactions of many conflicting forces at play, such as the conflict between capital and labor, between innovation and imitation, and between financial and real sectors of the economy. It is these conflicts that provide the economy with energy that causes it to fluctuate.

Critics of the new classical school, such as Victor Zarnowitz, contend that exogenous school models — and the Real Business Cycle (RBC) models in particular — have no theoretical explanations for what might cause the economy to fluctuate (1992, 46). Instead, causes are mystified as external (technological) shocks. Zarnowitz notes: "It is very unlikely, however, that any nationwide technological decline occurred in the recessions of 1949, 1954, 1958, 1960, 1979 . . . without anyone having noted it at the time . . . ." (p. 8)[1].

For many economists of the endogenous school, the conditions for the existence of an Arrow-Debreu general equilibrium are either too stringent, not realistic, or simply irrelevant to the study of the modern business cycle. For example, as pointed out by Joseph Stiglitz (1994, 148), in a technologically changing world, there are no buyers for goods that have yet to be invented, thus the market could not be complete. Furthermore, in this perfectly competitive world, there is no incentive to innovate (Stiglitz 1994, Ch. 8).

The nonmarket-clearing endogenous school — with a long history, running from Marx through Keynes and his followers[2], — attributes the causes of the business cycle to factors internal to the market economy. This school does not believe that the market economy has an inherent tendency to function smoothly; it sees cyclical phenomena as a demonstration of the internal dynamics of a complex economic system; and it believes that there is a "natural" rhythm to a market economy.

Many endogenous school economists have used tools developed in nonlinear dynamical systems as the basis for their business cycle theories[3]. Instead of relying on external shocks, these economists often employ limit cycle analysis in modeling cyclical phenomena.

By adhering to linear time series models, however, more than half a century of empirical time series model builders, since G. Yule's 1927 model of

---

[1] The response by RBC model builders, such as Hansen and Prescott (1993), is that the cause of those recessions can be traced to changes in government regulations. This is an example of the view that government interference causes cycle in an otherwise smooth functioning market economy.

[2] See, for example, surveys by Zarnowitz (1992, ch.2) and Mullineux (1990).

[3] See, for example, Goodwin 1951, 1967; Lorenz 1993; Rosser 1991; Medio 1992; Klein and Preston 1969, and others.

the Walfer's sunspot numbers, have implicitly or explicitly supported the exogenous school. Theoretical as well as practical difficulties in building nonlinear time series models have prevented empirical researchers from doing otherwise.

The typical "Frisch-Slutsky" formulation is to separate the impulse from the propagation mechanism through lags. This formulation made it possible for a linear model to mimic the observed business cycle fluctuations.

Richard Goodwin (1990, 10) charges, however, that the formulation of "Frisch misled a generation of investigators by resolving the problem with exogenous shocks." Since linear models could not generate cycles endogenously, these economists became convinced that the real world economy also could not fluctuate without outside shocks.

Guy Laroque and Guillaume Rabault summarize the last 50 years:

> Deterministic theories of the business cycle were very fashionable in the 1940s, as witness by the work of Goodwin (1951), Hicks (1950) and Kaldor (1940) among others. They were based on strong nonlinearities in investment behavior. Then the data came, and econometrics became popular. The econometricians need random shocks and liked linear models. Adelman and Adelman (1959), studying the Klein Goldberger (1955) model, convinced the profession that the basic features of the cycle could be explained through the lag structure of a linear model. The large scale macroeconometric models of the 1960s therefore, paid little attention to nonlinearities. Eventually, Sims (1980) questioned the use of economic theory to specify lag structures and promoted the vector autoregressive modelling strategy: all nonlinearities disappeared from numbers of macroeconomic empirical studies (1995, 283).

The reason that nonlinearities disappeared, however, is more basic than these two authors indicate. Since neoclassical model builders view the observed business cycle as an outcome of a self-correcting system that passively responds to outside influences, a linear system is sufficient for this purpose[4].

---

[4] Lorenz notes that: "Nonlinear approaches to economic dynamics have been investigated mainly by economists who felt uncomfortable with the classical paradigm of equilibrium economics" (Lorenz 1993, 25).

To show how misleading it is to have a linear model mimic a nonlinear one, John Blatt (1978) regressed a linear model on data generated by a simulated John R. Hicks 1950 nonlinear investment accelerator model with ceilings and floors. The parameters derived from the linear regression falsely manifest a stable system when the true system is far from stable. To distinguish linear and nonlinear time series, Blatt (1980, 1983) contrasted the average slope of many empirical series during ascending and descending phases. He found evidence of asymmetry in the absolute value of the average slopes between the two phases. Blatt then concluded that all linear models of the Frisch type are thus incompatible with observed time series asymmetry.

Andy Mullineux and WenSheng Peng (1993) point out, that there is an inherent limitation to the prevailing linear time series models. Without exogenous shocks, the economy either diverges or converges. The probability of sustained cycles in a linear model is zero. In a nonlinear world, however, local instability need not imply global instability. Internal as well as external factors can lead an economy to experience sustained cycles. Yet, a stable economy need not converge to a steady-state equilibrium point. One approach to model such a world is to model it as a limit cycle.

The concept of a limit cycle, loosely speaking, is simply the idea that a series, free from outside shocks and insensitive to starting values, will wander around within some bounded region, repeatedly traversing its path time after time. Shocks within the bounded region can only alter its regularity, not its basic cyclical characteristic.

Many non-economic time series statisticians take for granted that business cycles should be modeled as limit cycles. "It is our view that if the notion of a *business cycle* is to be taken seriously [,] it must be related to a non-linear economic system, not unlike a limit cycle in a dynamical system" (Tong 1990, 232).

To date, however, few nonlinear time series models that permit limit cycles derive their parameters from real world economic data. It is the objective of this dissertation to expand the arsenal of models that can accommodate endogenous cycles — as a step toward effective empirical modeling of the business cycle.

In the pages that follow I provide some rationales in section 2 for the existence of internal dynamics in an economy. In this section I also review the empirical evidence for time series nonlinearity over the business cycle, and survey endogenous business cycle theories and their implications for time series models. In section 3, I formulate and in section 4, estimate a class of self-switching Markov models of the U.S. business cycle. The estimation provides empirical evidence for modelling the business cycle as an endogenous cycle, perhaps even a limit cycle. Finally I discuss the conclusions of the research and their implications.

2. Business Cycle Dynamics

2.1. Introduction

Theories of the business cycle that rely exclusively on exogenous shocks to generate cycles are unsatisfactory, for the exogeneity of the shocks are unconvincing, especially the shocks to technology. What is considered, for example, an exogenous technological shock, might be largely generated endogenously within the market.

Without obscuring the causes to external factors, one is faced with the challenge of trying to answer the question of *why* an economy fluctuates. This section sets out to examine some internal factors that might contribute to the business cycle; some evidence in time series data that might justify seeking internal causes; and some models that might endogenize economic dynamics.

2.2. Dynamics via Nonlinear Feedback

There are two distinct approaches to analyzing and theorizing a market economy. They differ on whither market clears at all times or not. Peter Flaschel, Reiner Franke, and Willi Semmler (1997) summarize these differences. Market-clearing

> Macrotheory is built on microeconomic principles, using the competitive equilibrium model with given endowments of agents, preferences and technology as a reference model. The equilibrium approach posits that representative agents (e.g., consumers, firms) are rational, have full information, and optimize intertemporally. The paths of prices, wages, and rental rates of capital are usually assumed to be known in advance. The decision-making process is modeled as if decisions are undertaken by an idealized policy maker and as if this were a good approximation of the complex decision making in an industrial society (p. 2).

Nonmarket-clearing macrotheory

> . . . focused on diverse sources of instability and macroeconomic fluctuations. Instabilities are seen to originate in: (*i*) stock-flow relationships (for example, accelerator-multiplier or output-inventory interactions); (*ii*) price dynamics and price expectation dynamics (nominal-real interaction); (*iii*) large demand shocks; (*iv*) the conflict over distributive shares; and (*v*) the financial sector and the financial-real interaction. . . (p. 4).

In the spirit of the nonmarket-clearing approach, I will examine economic factors that can contribute to dynamic interactions. In various ways these factors can generate strong feedback that might prevent the market from clearing instantaneously, or cause a steady-state equilibrium to be unstable.

2.2.1. Dual Roles of Competition

The foundation of new classical analysis is built upon a competitive equilibrium where

> The dominant adjustment mechanisms are price adjustments, which are supposed to equilibrate markets infinitely fast. A perfect foresight path of prices is often assumed so that markets can instantaneously clear. In intertemporal models with typical saddle point properties, prices are assumed to jump in order to bring about convergence to the long-run equilibrium. Product markets are cleared, and imbalances in the labor market are seen to be a result of the choice between leisure-work effort (Flaschel, Franke and Semmler 1997, 2-3).

From a nonmarket-clearing perspective, competition does not necessarily bring an economy to a short-run or a long-run equilibrium[5]. The stability of the equilibrium point and the speed of convergence in a model economy are critical to the outcome, due to the dual roles of competition in a production economy. In contrast to a pure exchange economy, competition is not only a force that equalizes prices and rates of profits between firms, but it also can be a force that destroys any hope of the economy to maintain a state of

[5] The classical economists, especially Marx seems to have a more realistic view of competition than contemporary neoclassical or new classical economists (see, for example, Clarke, 1994, Ch. 2).

equilibrium for long.

In the long-run, firms either imitate or innovate. Both activities take place simultaneously. Before imitators have a chance to equalize the market, innovators might have already undermined the opportunity for tranquillity. It is similar to a cat chasing its tail. The long-run equilibrium can be a moving target due to innovators. If the rate of convergence is low, the economy could endlessly chase after the long-run equilibrium, with little hope of ever reaching it. Thus, imitation is like a centripetal force while innovation is like a centrifugal force. It is the contradiction between these two forces that gives energy and vigor to a market economy (Schumpeter, 1939). The business cycle is but one manifestation of this contradiction.

In the short-run, competition might not lead to a market-clearning equilibrium either. To maintain a state of equilibrium where the markets are always cleared, new classical models, such as RBC models, treat innovations as exogenous shocks, not endogenous inspirations based on market conditions. If innovations are endogenous, the insistence that all markets are always cleared becomes problematic. For a market to clear immediately after each innovation, the market needs to know what the market-clearing price is. This requires perfect foresight, a requirement that is inconsistent with endogenous technological innovations.

Since production is a sequential process, the nonmarket-clearing approach sees a production economy different from a pure exchange economy. In a production economy, not all demands for a new technology, or a new product can be satisfied instantly, and no quantities of goods can be produced instantaneously. Without perfect foresight, the market-clearing equilibrium price in the product market can only be revealed through an iterative process that relies on the signals from the non-zero excess demands. This is one example of the difference in the speed of adjustment between financial and product markets stressed by the nonmarket-clearing approach.

Furthermore, a change in technology is not seen as an economy wide or an industry wide phenomena. New technology takes time to defuse. Between the initial investment and the initial product's introduction into the market place, there is a time lag.

In addition, not all innovations prove to be profitable. It takes time for others to be convinced that they are worth imitating. Thus, for an innovation to succeed, it must withstand the test of time.

The time lag is seen as a factor working against the neoclassical notion of a market-clearing equilibrium in the short-run and the steady-state equilibrium in the long-run. The diffusion process — made evident by the time lag — is seen as capable of preventing the economy from reaching a new equilibrium point instantly. Before the market-clearing price is reached, the economy is in a disequilibrium. Before the diffusion is complete, the economy is not in a steady-state. Since technological change occurs constantly, disequilibrium is viewed as pervasive, while equilibrium is seen as only temporary.

Market-clearing approach, however, sees the diffusion time lag as a result of the cost of adjustment. In RBC models, for example, the time lag, or "time to build" (see, for example, Kydland and Prescott 1982) is seen as the reason for an evolving equilibrium point over time, not as a reflection of the market adjustment towards equilibrium. Moreover, the assumption of perfect foresight allows the market to clear at all times as the economy moves toward the steady-state equilibrium after each shock.

It seems ironic that the RBC approach, while stressing micro-fundations and technological innovations to the study of business cycles, provides no micro-foundation for technological innovations. The RBC approach merges innovators and imitators, and thereby eliminates the motivation for firms to innovate or to imitate. The optimization problem posed by a typical RBC model ignores potential gains from innovation or imitation.

In the real world, however, there is a critical time lag from innovation to imitation (see, for example, Cheng and Dinopoulos 1992). During this period, innovators can enjoy a higher than long-run equilibrium profit (see, for example, Deneckere and Judd 1992).

2.2.2. Over-production

What Karl Marx viewed as the "crisis of over-production"(see, for example, Clarke, 1994) and what Joseph Schumpeter (1939) viewed as "creative destruction" are different formulations for the same phenomenon. Competing modern firms do not plan their production scale according to the equilibrium level of market supply and demand. Instead, each wants to gain market share at their competitor's expense. Over-production is often used as a competitive tool to drive out high cost producers. Consequently, the business cycle is necessary for a vibrant market economy (Cheng and Dinopoulos, 1992). Without it, inefficient producers would not withdraw from the market by themselves.

From a market-clearing perspective, "creative destruction" can be modeled as an equilibrium response to changes in technology or taste. Ricardo J. Caballero and Mohamad L Hammour (1994) studied the "cleansing effect" of a recession in their market-clearing dynamic partial equilibrium model. Since capital destruction is costly, firms in their model can meet the changes in market demand by either creating new capital that embeds more efficient technology or destroying the less efficient old capital.

To maintain market-clearing, though, the changes in technology or demand must be exogenous. When over-production is used as a competitive tool, in the sense that some firms are willing to sustain sort term losses to achieve long term gains, the changes in technology are no longer purely exogenous phenomena; neither is the change in demand. Thus the market might not be cleared during each period.

2.2.3. Variable Returns to Scale

Another factor that potentially can destabilize an economy is the variation in optimal return to scale induced by technological change. Each innovation might represent a structural change that alters the optimum scale of production and changes the optimum number of firms in each industry. This outcome occurs when returns to scale are a nonlinear function of the scale — where output might be increasing at one level and decreasing or constant at another.

For ease of analysis, the returns to scale tend to be fixed in most economic models. This assumption is acceptable as a local approximation of a long-run equilibrium point. It might be invalid, however, if the economy moves too far away from the equilibrium point, — a possibility that is largely ignored in market-clearing models.

A change in optimum scale can be destabilizing, for it can create opportunities for new entries or can force some existing firms to exit. Some RBC models attempt to address the issue of entry and exit together with increasing returns to scale (see, for example, Hornstein 1993, Devereux, Head, and Lapham 1996). But, by assuming infinite speed of price adjustment, market is always cleared, and the returns to scale is essentially fixed.

2.2.4. Factor substitutability

Like many other aspects of the economy, factor substitution requires also the passage of time. Investment capital behaves more like a mass of putty initially: it can be molded into any combination of physical capital and labor. Once the investment is made into physical capital, however, it behaves more like a mass of clay. Remolding becomes costly.

With heterogeneous production and varying capital intensity, the dynamic path of an economy might be unstable. A change in a factor price can lead to a disequilibrium in the short-run. There might not be a steady-state equilibrium in the long-run either, due to the instability of the dynamic path. This outcome is often the case with extensions of Goodwin's 1967 growth models (see, for example, Flaschel, Franke, and Semmler 1997, Ch. 4).

Given the adjustment cost and the narrow range of physical capital-labor substitutability in the short run, technology is perhaps better represented by a fixed proportion Leontief production function. In the long run, Cobb-Douglas is perhaps a better representation, since firms can adjust, over time, to technologies with the most profitable capital-labor ratio. In a imperfectly competitive world, a factor's share can affect the pace and the type of technological innovations and adaptations, so that a factor price is not purely determined by its marginal product[6].

2.2.5. Risk versus Uncertainty

Technological progress introduces an element of uncertainty in a production economy that is different from risk. This difference largely has been ignored by neoclassical models.

Take, for example, the assumption of perfect foresight. In an uncertain world, it is not possible to have contingent contracts—a pure exchange economy concept – for all possible future technological innovations. Unlike the expected value of a lottery that can be calculated from a known distribution, the likely market value of an innovation, as a random variable, has no known (objective) moments. The track record of past innovations provides only a

---

[6] It is perhaps because of the changes in substitutability with time that Flaschel, Franke and Semmler (1997, Ch. 11) dismiss the neoclassical marginalist view in their model.

suggestion — it is only a subjective guide for expectation formation. The shape of the distribution or its parameters cannot be known with certainty.

Similarly, a "rational expectation" could not be objectively rational, since the true distribution is not known. Under true uncertainty, the *ex ante* optimizing surface is flat. The range of behaviors that is consistent with optimization is great. As pointed out by Axel Leijonhufvud (1993), one should only impose *ex post* rationality, for *ex ante* rationality is indeterminate under true uncertainty.

If every agent is capable of forming only subjective expectations, the chance of all their expectations being consistent with each other is nil. Since consistency of expectations among all economic agents in the economy is a necessary condition for the solution of a rational expectation model, true uncertainty presents a formidable challenge to that class of models.

Allan H. Meltzer (1982) formally distinguished the difference between risk and uncertainty. The former involves only a transitory change, while the latter reflects a permanent change. Whereas temporary changes can be captured in the error term of a regression model, structural changes cannot. The latter is what Meltzer means by uncertainty.

In a time series framework, the existence of true uncertainty implies that the ergodicity assumption is invalid or misleading (Davidson 1991, 133). Thus by assuming ergodicity, one assumes away the difference between risk and uncertainty. Since theories of non-ergodic time series are underdeveloped, in practice one is forced to assume ergodicity.

The existence of true uncertainty lends support to the rationality of adaptive expectations, or Keynes' "animal spirits." Elaborating on Meltzer 1982, Mullineux (1990, 41) states: "Under uncertainty, instead of forming expectations independently, agents must take account of the weight of opinion guiding the activities of other agents in the manner of the Keynes (1936, 156) 'beauty contest' example".

In other words, economic agents learn from each other what the true nature of the world is. The learning and consensus-forming activities take time. When the road ahead is foggy, the optimizing behavior of a driver on a highway is to slow down. By analogy, true uncertainty can prevent the fast convergence of the economy to a new long-run equilibrium point.

Furthermore, from a nonmarket-clearing perspective, agents without full information,

> . . . cannot fully optimize since the "cost of optimization" — properly computed — might be very large and, even worse, not known in advance. This often results in the presupposition of bounded rationality or procedural rationality in which agents imitate the behavior of others, follow rules of thumb, and adjust gradually to a changing environment (Flaschel, Franke and Semmler 1997, 4).

Thus, true uncertainty creates more than just a continuum of equilibria, where a purely exogenous event such as a change in the "sun spots" determines the actual outcome. Many authors, including Flaschel, Franke and Semmler (1997) have shown that infinite speed of price and expectation adjustment can lead to instability, not steady-state equilibrium.

2.2.6. Investment Multiplier and Bubble

That the multiplier-accelerator effect of investment spending can cause the equilibrium point of a model economy to be unstable has long been recognized by economists (Hicks 1950 and Kaldor 1940). True uncertainty in the face of technological innovations might be the underlying reason for this instability.

Goodwin (1990, 19) has given an example of how an investment bubble might happen under uncertainty. If innovation occurs only once, the time lag stemming from the diffusion of technology, with its accelerated initial investment in adapting the new technology, can lead to an investment boom.

Later when the diffusion is complete, a bust occurs due to the deceleration of investment. Optimizing behavior does not eliminate the boom or the bust as long as the same innovation does not occur in all firms simultaneously. Only an assumption of perfect foresight could eliminate it.

Differential rates of investment growth between capital goods and consumption goods industries can be another source of instability. Given the uncertainty due to innovations in a production economy, it would be unlikely for the current level of investment spending in the capital goods industry to be consistent with both the following at the same time: the future level of investment spending in the same industry, and the future level of consumption that would have been stimulated by the current spending. This is one source of internal dynamics of many two sector growth models[7].

2.2.7. Destabilizing Money Holding

Historically, the role of money as a store of wealth emerged partially in response to the true uncertainty in the economy. A quantity of money in circulation represents a claim on goods and services in general, and its value is more certain than the holding of a particular good alone. Money holding renders Say's law — supply creates its own demand — invalid in the short run. (In the long run, Say's law is meaningless other than an identity). As a result, this role for money makes the economy less stable than a barter economy.

In a barter economy, by Say's law, every intended investment purchase is also an intended sale. The difference between two intended transactions is zero in every period. In a momentary economy, by contrast, an investor does not have to invest if the expected return in one period is below a threshold.

---

[7] See, for example survey by Nishimura and Sorger 1996 for the market-clearing approach to internal dynamic models. Related also is Marx's idea of the disproportion that happens between the production of investment goods and consumption goods in a capitalist economy. See, for example, Clarke 1994.

The investor can invest the idle money in the next period when the expected return is higher. Thus the total intended sale can be higher or lower than the total intended purchase in each period. The difference reflects the change in intended money holding. This kind of fluctuation in intended investment has long been recognized by economists as the leading cause of business cycles.

Neoclassical models typically assume that savings equals investment. This approach eliminates the role of money as a store of value. The focus instead is on the transaction demand for money. Consequently, Keynes' liquidity trap cannot exist in their linear and perfectly competitive models.

Given the role of money as a store of wealth, changing expectations of the future value of that wealth, or rapid swings in inflationary expectations can create havoc in a monetary economy. Both consumption and investment can be affected by these changes in inflationary expectations. Money, thus, creates an additional source of instability in an economy.

2.2.8. Capital Depreciation and Utilization

Investment spending by firms can be "synchronized" by the business cycle. Automobile sales, for example, typically boom during the recovery phase of a business cycle. The boom, however, plants the seed for a fixed capital replacement cohort 8 to 10 years later, and subsequently, a market saturation and bust in that industry.

Another source of endogenous fluctuation is a phenomenon called "self-organized criticality." Per Bak *et al.* (1993) have shown that instead of small sector fluctuations canceling out when their activities are aggregated, local interactions between production units together with non-convex technology can cause aggregate fluctuations. "Large interactive dynamic systems can 'self-organize' into a critical state." A state that "can actually be an attractor for the dynamical system, toward which the system naturally evolves, and to which it returns after being perturbed by some large external shock"(p. 5). Like an avalanche that can be caused by dropping an additional grain of sand to a sand pile that has grown to a critical shape (a prototypical example of such self-organized criticality), an economy that is close to its capacity can experience great fluctuations when a shock to one firm can amplifies throughout the economy. In other words, the destabilizing impact of an innovation to the economy is higher when the economy is closer to full capacity than at other times.

2.2.9. Asymmetry of Information

Stiglitz has devoted most of his professional life to the study of asymmetry of information. He has argued forcefully in his numerous writings (Stiglitz 1994 and the references therein) that not only is perfect foresight unattainable in the real world, but it is also contrary to the optimizing behavior of agents. If knowledge is valuable in the market place, it will not be freely shared. Instead, there is an incentive to misinform potential competitors. To innovate is to find new and better ways to do something that others have not been able to do. The assumption of perfect foresight denies this possibility and thus deprives agents the incentive to innovate.

Asymmetry of information plays a critical role also in many other aspects of the economy, as pointed out by Stiglitz in his numerous writings. In the financial sector, it leads to credit rationing — when the interest rate alone is an insufficient criterion for investment decision. Banks and creditors need more than just a higher interest rate to screen out low return investment projects; for it can also screen out low risk projects. A debtor knows the risk of his or her default better than a lender. Thus, there might not be an interest rate level for which the desired investment equals the desired savings when information is asymmetrical.

In the labor market, asymmetry of information leads to an efficiency wage that is higher than the neoclassical marginal product wage. A higher wage attracts higher quality workers when monitoring and enforcement cost of labor inputs are non-zero. This situation prevents the labor market from clearing in the neoclassical manner.

Besides preventing the market to clear, asymmetry of information can contribute to the business cycle more directly. Stiglitz (1993) developed a model that makes research and development (R&D) spending pro-cyclical. Since an investment on R&D is not collateralizable, the funding is mostly from an internal source. An economic fluctuation can cause a change in the R&D spending that is internally funded. The change in the R&D spending, in turn, will cause a fluctuation in the volume of technological innovations that affect the growth rate of the economy. This positive feedback can prevent the economy from converging to a steady-state growth path.

In summary, many of the neoclassical assumptions hinder our ability to understand the observed business cycle. For instance, neither is perfect foresight a good representation of the behavior of people in a technologically changing world, nor is instantaneous market-clearing a good approximation of the price formation in a dynamic economy. Steady-state equilibrium is also not a good characterization of the market dynamics.

Instead, a market economy is better understood as having a bounded rationality with delayed market-clearing and a unstable growth path. The adjustment speed towards the equilibrium or the steady-state and the instability of the equilibrium or the steady-state are the central components of a dynamic macroeconomy.

The factors that I have just reviewed underlie these aspects of the economy. They either slow down the speed of adjustment towards a market-clearing equilibrium, or they make the steady-state equilibrium unstable. Both phenomena of nonmarket-clearing and unstable steady-state can exist more likely in a nonlinear world. Thus I now turn to the evidence of nonlinearity in the real-world market economy.

2.3. Evidence of Nonlinearity

A system of nonlinear feedback can manifest itself in nonlinear behavior of time series data. Finding nonlinearities in a variety of macro time series thus provides an empirical foundation for endogenous business cycle models. Without such empirical support, the rationales listed above for endogenizing business cycles amount only to empty theories.

A telltale sign of a nonlinear data series is the asymmetry of its distribution. Since a linear function is skew-symmetric, distributional asymmetry in a time series might be explained either by an asymmetric shock, or by a nonlinear data generating function (Tong 1983, 16). Simon Potter (1994, 320) points out, however, that an asymmetric shock cannot account for an asymmetric response — only a nonlinear data generating function can.

Assuming that distributional asymmetry is from a nonlinear data generating function, a host of nonlinearity tests in the time domain has been developed in the literature[8]. They have tested linearity against either nonlinearities in general, or a specific class of nonlinearity. The power of each test is sensitive to the type of nonlinearity encountered (see, for example, the examination by Teräsvirta 1996).

The task of identifying nonlinearity in a macro time series is hampered

---

[8] De Gooijer and Kumar (1992) surveyed the literature on recent developments in the linearity tests.

by the process of aggregation. For example, the U.S. GDP series is an aggregated series from many component series. One of them is the gross domestic investment series. Suppose the latter series displays clear signs of nonlinearity. The aggregation process, however, could conceivably dilute any signs of nonlinearity in the aggregate series. Despite this difficulty, many economists have found evidence of nonlinearities in many U.S. macro time series.

Salih Neftci (1984) used a three-state Markov chain to calculate the transition probabilities of a series movement from expansion phase to contraction phase. The asymmetry of transition probabilities is seen as a sign of nonlinearity. He found significant evidence of nonlinearity in the U.S. unemployment rate. Although Daniel Sichel (1989) found a probable error in Neftci's test, Philip Rothman (1991) used a modified Neftci test, and confirmed Neftci's earlier findings.

J. Bradford DeLong and Lawrence Summers (1986) conducted a simple skewness test on U.S. macro data, such as GNP and the Industrial Production (IP) index. They wanted to test the claim made by John Keynes (1936) and Wesley Mitchell (1927) that the shape of a downturn is steeper than the shape of an upturn in a business cycle. They found evidence of asymmetry in the unemployment series only. They found no evidence of asymmetry in the GNP or IP series.

Zarnowitz (1992, 258) suspects that DeLong and Summers' conclusion is "premature, being based on uncertain assumptions and evidence" in deriving their standard error. Potter (1994, 315) points out that the test used by DeLong and Summers — the expectations of the third moment — is the weakest of all asymmetry tests.

Timo Teräsvirta and Heather M. Anderson (1993) tested the industrial production index of 13 countries and the combined "Europe" index. They rejected linearity assumptions in most of the series using smooth threshold autoregressive (STAR) models. C. Q. Cao and Ruey. S. Tsay (1993) also found the volatility of the monthly stock returns from 1928 to 1989 in NYSE to be nonlinear.

Both James Hamilton (1989) and Andrew Filardo's (1994) results can be seen as evidence of nonlinearity in the series they have studied (U.S. GNP and post WWII IP index, respectively). Both found asymmetric transition probabilities in the data from recession to recovery and vise versa.

Sichel (1991) estimated a parametric hazard function model and found asymmetric business cycle duration dependency. He found that the length of contraction affects the hazard rate of recovery, but the length of expansion does not affect the hazard rate of contraction. Similarly, Allan Brunner (1992) found evidence of conditional asymmetry in the real GNP growth rate using a semi-nonparametric approach. He found the persistence during an expansion is greater than the persistence during a contraction. Robert Hussey (1992) found an asymmetric conditional variance in employment data — the conditional variance is larger following a contraction than following an expansion.

Supporting earlier findings, Grant McQueen and Steven Thorley (1993) found asymmetric business cycle turning points. They found "round" peaks and "sharp" troughs in growth rate variations. Paul Beaudry and Gary Koop (1993) found that a positive shock to the growth rate is more persistent than a negative shock.

Once the economists start looking for nonlinearity in the data, the evidence seems to appear everywhere. Economists should no longer ignore this real-world nonlinearity when constructing their models.

2.4. *Models of Internal Dynamics*

From the numerous studies cited, clearly the data generating processes determining most macro time series are not linear. To understand the process, many economists have developed various nonlinear models that attempt to highlight key features of a market economy's internal dynamics.

Mullineux and Peng (1993) give an excellent introductory survey of nonlinear endogenous business cycle models in both the Keynesian and equilibrium traditions.

In the Keynesian tradition, they surveyed early models based on Hicks' 1950 model with consumption ceiling and investment floor, Nicholas Kaldor's 1940 nonlinear investment function model, and Goodwin's 1951 nonlinear investment accelerator model. They also surveyed models based on Goodwin's 1967 capital-labor dynamic growth model and H. Rose's 1967 nonlinear Phillips curve model.

A common assumption in the Keynesian tradition is that markets do not clear instantaneously. Deterministic cycles emerge in these models simply from nonlinearities of the models. This approach in modeling business cycles has been criticized in the 1970's by new classical economists as being too adhoc, for the behavior assumptions in those models were not based on solutions to any agent optimization problems. Ever since then, the economic literature has been dominated by the new classical models.

The discovery of low dimensional chaos in last few decades and the evidence of nonlinearity in time series data post new challenges to the new classical tradition. These developments make endogenous cycles more likely, but do "not set well with the idea of strict economic equilibrium" (Benhabib 1992, 3).

Faced with the challenge, many new classical economists have developed sophisticated models to show the possibility of cycles and chaos in their models. The collection in Jess Benhabib (1992) is an attempt in this direction, containing many interesting general equilibrium endogenous cycle models. Mullineux and Peng (1993) surveyed also other models in this tradition, such as Richard H. Day's 1982 nonlinear growth model with chaotic tendencies, and Jean-Michel Grandmont's 1985 and 1986 nonlinear over-lapping generation model with non-neutrality of money[9].

Endogenous cycles, nonetheless, do not occur naturally in the market-clearing equilibrium tradition.

> If we assume complete markets and decreasing returns to scale, however, optimal growth models can display interesting fluctuations only when extreme parameter constellations, for example, extremely high discount rates, are assumed (Flaschel, Franke, and Semmler 1997, 3).

In contrast, endogenous cycles occur quite naturally in the nonmarket-clearing tradition. For example, Goodwin's 1967 predator-prey model of capital-labor dynamic is well known for its limit cycle property. Although Goodwin's model does not start from the principle of agent optimization, it does not follow that his model is incompatible with the principle. Renato Balducci, G. Candela and G. Ricci (1984) developed an optimization model in a non-cooperative differential game setting for which Goodwin's model is a simplified version. Balducci, Candela, and Ricci (1984) showed that even when agents discount future consumptions and prefer consumption smoothing, the Nash-equilibrium is still not globally stable. The time path of the model is a limit cycle, or an irregular motion, depending on the assumptions made.

Many new contributions to nonlinear endogenous business cycle theory in the nonmarket-clearing tradition have appeared since Mullineux and

---

9 Other new contributions to endogenous cycles in the general equilibrium tradition have been surveyed by Nishimura and Sorger (1996).

Peng's 1993 survey. Among them, Duncan Foley (1992) developed a business cycle model based on rationally behaving representative firms facing real liquidity costs, production lags, and a systematic money supply rule. In his model, a stable limit cycle occurs due to the instability of the stationary equilibrium.

Laroque and Rabault (1995) developed a perfect foresight over-lapping generation inventory cycle model that endogenizes the cycle. They also estimated the model using the U.S. GNP data. In their model, the cycles persist due to the asymmetric specification of the growth and recessionary phases of the economy.

In Semmler (1989; 1994), there are many new extensions to the nonlinear business cycle models of Hicks (1950), Michal Kalecki (1935), Kaldor (1940), and Goodwin (1967), as well as nonlinear extensions of Keynesian business cycle models. The emphasis is in the macro instability introduced by financial instability.

Raymond Deneckere and Kenneth Judd (1992) developed a model based on the interactions between innovators and imitators that causes capital investment to follow a limit cycle path or to be chaotic. In this model, the notion of general equilibrium is replaced with the momentary monopoly power of the innovators.

The dominant business cycle models in the contemporary economic literature are the RBC models. They are not endogenous cycle models. Although these general equilibrium models are dynamic, the dynamics are externally imposed through shocks. Geert Rouwenhorst (1991, 242) charges that

> For persistent deviations in output and investment to occur in the neoclassical model it is required that the time series of shocks that hit the economy behave very much like the fluctuations which the model seeks to explain.

Furthermore, RBC models typically assume investment and consumption decisions are made by the same "representative" agents in the economy. Jean-Pierre Danthine and John Donaldson (1995, 228) reveal that this is a key to a successful RBC modeling strategy, i.e., "only one model agent solves an intertemporal problem."

The merging of these two very distinct roles glosses over one of the main contradictions in a market economy. It removes at one stroke the dynamic interplay between firms' profit optimization problem on the one hand, and households' consumption optimization problem on the other. This distinction is neglected because in the neoclassical world,

> . . . capital markets are assumed to be perfect and spending is determined by an equilibrium path. . . . Liquidity and borrowing constraints resulting from imperfect capital markets, which means that one cannot borrow against future income without collaterals, are mainly disregarded (Flaschel, Franke and Semmler 1997, 3).

The assumption of homogeneity is crucial in a dynamic general equilibrium model. When heterogeneity is introduced into a model, either by assuming heterogeneity of agents (such as capital and labor), or by assuming heterogeneity of products (such as capital goods and consumption goods), the maintenance of a state of equilibrium becomes problematic. George Stadler (1994, p. 1771) points out that

> Even introducing a small amount of agent heterogeneity can have destructive consequences. If agents have identical preferences, and differ only in terms of the income they receive, the "representative agent" for such an economy need not be well behaved and the economy can manifest a large number of unstable equilibria.

Without an out-of-equilibrium adjustment mechanism, these types of models are not credible. Furthermore, multi-sector RBC models "[have] the property that aggregate randomness disappears if the number of sectors is made large" (Bak et al. 1993, 2n. 2).

Neoclassical economists are more willing to accept chaos than limit cycle as the basic model of the business cycle. José Scheinkman (1990), for example, argues that a purely deterministic model could not reasonably explain the behavior of aggregate quantities or asset pricing in an actual economy. Deterministic chaos looks random, so models based on chaos can be grouped together with stochastic general equilibrium models.

From a non-neoclassical endogenous business cycle perspective, a limit cycle model is more promising than a chaos model. Limit cycle and business cycle share many features in common. First, we observe neither an ever-increasing growth rate nor an unending contraction in any market economy. There seems to be a bounded region whin which the growth rates of most of the industrialized countries' economy operate. (Even during the Great Depression of the 1930's, the contraction lasted only a few years, not generations). Second, a period of high growth is sure to be followed by a period of low growth or contraction; just as a period of contraction lays the foundation for eventual recovery and higher growth. However, a business cycle is more than a limit cycle: it has stochastic elements that make it irregular.

Chaos, as a business cycle model, is not very credible for theoretical reasons. Michele Boldrin (1994) lists many economic impediments to chaos. Chief among them is the preference for consumption smoothing. It also suffers from a lack of empirical evidence (see for example Brook and Sayers 1992; Ramsey, Sayers, and Rothman 1992; and Sayers 1994).

Nevertheless, one major achievement of the modern endogenous business cycle model builders is that they have shown that cycles can occur in a model economy without violating the principle of agent optimization. The reliance on external shocks is no longer necessary in an optimizing model of the business cycle.

2.5. Conclusion

Given the rationals of dynamic feedback, the evidence of nonlinearity, and theoretical models of endogenous business cycles, empirical researchers are now challenged to develop nonlinear time series models for the study of real world business cycles. The task is to build models that can better accommodate the observed nonlinearity, and to embed the feedback mechanism within the model so that one can test various aspects of a theoretical model. Towards this end, a class of self-switching Markov models is developed next.

3. Internal Dynamics via Self-Switching Markov Models

3.1. *Introduction*

The motivation for developing a self-switching Markov model of the business cycle is mainly based on the desire to have a model that can capture the "natural" rhythm of a market economy. Since linear models cannot generate robust and persistent cycles without shocks, their value for business cycle analysis is limited mainly to questions of *how* an economy fluctuates, not *why*. Furthermore, the response function (the skeleton, the deterministic part, or the non-stochastic part) of a linear model is symmetric, thus incompatible with the time series data exhibiting asymmetric behavior. Nonlinear endogenous models are better suited to study the questions of *why*, have greater explanatory power for the persistence of cycles, and are more compatible with the distributional asymmetry found in the data.

The question of *why* an economy fluctuates can be addressed in a non-linear time series model in various ways. The first step is to show that there is a systematic cycle, such as a limit cycle, in the business cycle. The existence of a limit cycle in the time series data indirectly validates theories of business cycle endogeneity. Secondly, the transition dynamics in a self-switching Markov model can shed some light to the nature of the business cycle. Thirdly, in a higher dimensional multivariate system, the relative importance of each series to the transition dynamics can be studied. The questions regarding macro aggregates, such as the money supply, interest rate, unemployment rate, and the wage share, to the output fluctuation can be addressed in future research.

To build nonlinear models, Flaschel, Franke, and Semmler (1997, 283) used three "essential nonlinearities" in the continuous-time. They are "regime changes"; "viability constraints", i.e., ceilings and floors; and " the time rate of change of variables (derivative control)". The task of employing all three type of nonlinearities in discrete time models can be formidable. Instead I will model mainly the first type of nonlinearity, i.e., threshold and regime switching. Occasionally, I might be able to model the third type, the derivative control, using a pseudo time rate of change as a contributing factor in regime switching.

A useful tool in modeling fluctuations in a time series model is to have a mechanism for feedback. The lag structure in a linear model is one feedback mechanism. The limitation here is that the combined feedback, both positive and negative, must be less than one for the model to be stable, and thus traceable. If the total feedback is greater than one in a linear model, it will always be greater than one. Thus any deviation from the equilibrium point is amplified to eternity.

This limitation need not be the case in a nonlinear model. Here the force of feedback can be strong at times and weak at other times. The model can be globally stable and traceable, behaving like a limit cycle.

3.2. Limit Cycles in Time Series

A limit cycle in discrete time is different from a limit cycle in continuous time. Since the field of nonlinear difference equations is still in its infancy, Howell Tong (1990, 49-50) provides a set of working definitions.

For each integer t >0, let

$$X_t = f(X_{t-1}), X_0 \in R^n, \quad (3\text{-}1)$$

where f is a vector valued function. Let $f^{(j)}$ denote the j[th] iteration of f, that is

$$X_{t+j} = f^{(j)}(X_t) = \underbrace{f(f(\ldots f}_{\text{j of them}}(X_t)\ldots)), \quad (3\text{-}2)$$

For example, $X_2=f(X_1)=f^{(2)}(X_0)=f(f(X_0))$.

The attractor set for f is a compact set A such that the basin

$$B = \left\{X : \lim_{j\to\infty} \inf_{Y\in A} \left\| f^{(j)}(X) - Y \right\| = 0\right\}, \quad (3\text{-}3)$$

has positive Lebesgue measure, where $\|\cdot\|$ is the Euclidean norm, and A is minimal with respect to this property.

A limit cycle is an attractor set A with T points $\{X_1,\ldots,X_T\}$ such that

$$X_{t+1} = f(X_t),\ t = 1..T-1, \text{ and } f(X_T) = X_1. \quad (3\text{-}4)$$

These definitions of Tong's have some implications. At one extreme, where T=1, the attractor set A is simply a limit point (equilibrium point) instead of a limit cycle. At the other extreme where T approaches infinity, the set A is a strange (or chaotic) attractor, in which case the value of the j^{th} iteration of f has a sensitive dependence on initial conditions[10]. If T is in between the two extremes, the set A is then a limit cycle set. For example, many computer generated pseudo random number series are limit cycle series with a very large but finite T. This is one extreme case of a limit cycle that looks random.

3.3. Limit Cycles via Threshold Models

Although there are a wide variety of nonlinear time series models, only a few classes have been developed in the literature that are capable of producing limit cycles. The class of threshold autoregressive (TAR) models introduced by Tong (1983, 1990) is one. Another potential candidate is the class of Markov-switching (MS) models used extensively since Hamilton (1989). Both types of

---

[10] Medio (1992, 46) defined a map f:U->U, $U \in R^n$, which has a sensitive dependence on initial conditions if $\exists\delta>0$ s.t. $\forall x \in U$, and $\forall$ neighborhood N_x, $\exists y \in N_x$, and $j \ge 0$ s.t. $|f^n(x) - f^n(y)| > \delta$.

models are in essence piecewise linear models. The latter type, however, needs modifications to create limit cycles.

Tong's two-regime TAR model has the following form:

$$Y_t = (\alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_t^{(1)})(1 - I(\eta - Y_{t-d})) + (\beta_0 + \sum_{i=1}^{q} \beta_i Y_{t-i} + \varepsilon_t^{(2)}) I(\eta - Y_{t-d}) \text{ ,} \quad (3\text{-}5)$$

where $I(\eta - Y_{t-d}) = \begin{cases} 0 & \text{if } Y_{t-d} \le \eta \\ 1 & \text{if } Y_{t-d} > \eta \end{cases}$, η is the threshold parameter, d is the delay parameter, and $\varepsilon_t^{(i)}$, i=1,2 are white noises. The choice of η and d can be determined from a grid search for the pair that gives the minimum sum of square errors.

If the indicator function I(·) is replaced by a continuous non-decreasing function g(·), the result is a smooth threshold autoregressive (STAR) model. There are a number of candidates for the function g(·): logistic and semi-normal forms are the two most commonly used[11].

In Tong's application of TAR to both the Canadian Lynx and the sunspot numbers, he produced coefficients that exhibit robust limit cycle phenomena in simulations. The key ingredient for the success is the flip-flop nature of the model structures.

For example, in the sunspots model (Tong 1990, 425), the two equations after expanding (3-5) to the equivalent forms are

$$\begin{aligned} Y_t &= 1.89 + 0.86Y_{t-1} + 0.08Y_{t-2} - 0.32Y_{t-3} + 0.16Y_{t-4} - 0.21Y_{t-5} - 0Y_{t-6} + 0.19Y_{t-7} - 0.28Y_{t-8} \\ &\quad + 0.2Y_{t-9} + 0.1Y_{t-10} + \varepsilon_t^{(1)} \qquad \text{if } Y_{t-8} \le 11.93, \quad (3\text{-}6) \\ &= 4.53 + 1.41Y_{t-1} - 0.78Y_{t-2} + \varepsilon_t^{(2)} \qquad \text{if } Y_{t-8} > 11.93, \end{aligned}$$

where $Y_t = 2\{(1+X_t)^{1/2} - 1\}$, X_t is the annual average of daily observations of the sunspots 1700 to 1979, and $\varepsilon_t^{(i)}$ are white noises.

---

[11] Teräsvirta and Anderson (1993) contrasted logistic and semi-normal functional forms in a univariate STAR model.

Table 1 Characteristics of the sunspots model

Regime	Dominant Eigenvalue	Modulus	Equilibrium
1	0.832 ± 0.484i	0.962	8.59
2	0.705 ± 0.532i	0.883	12.24

Since the modulus of the imaginary roots of both equations in Table 1 are less than 1, both equations are stationary and oscillatory. Depending on the initial conditions, the skeleton of the model has 3 stable convergence paths. It either converges to one of the two stationary equilibrium points or it will oscillate for ever in a stable limit cycle.

The reason that the series can converge to a limit cycle is due to the imaginary roots of the two equations. Suppose the series starts out oscillating in the first regime and the oscillation is large enough to have one of its lags greater than the threshold, then the governance of the series will switch to the second regime after some lags. The same is true in the second regime. Each regime counteracts the damping effect of the other regime. The contradictory process generates a stable limit cycle.

In the Canadian Lynx example (Tong 1990, 387), the two equations are

$$\begin{aligned} Y_t &= 0.546+1.032Y_{t-1}-0.173Y_{t-2}+0.171Y_{t-3}-0.431Y_{t-4}+0.332Y_{t-5} \\ &\quad -0.284Y_{t-6}+0.21Y_{t-7}+\varepsilon_t^{(1)} && \text{if } Y_{t-2} \leq 3.116, \quad (3\text{-}7) \\ &= 2.632+1.492Y_{t-1}-1.324Y_{t-2}+\varepsilon_t^{(2)} && \text{if } Y_{t-2} > 3.116. \end{aligned}$$

where Y_t is the $\log_{10}$ of annual Lynx trappings in the Hudson bay from 1821 to 1934.

Table 2 Characteristics of the Lynx model

Regime	Dominant Eigenvalue	Modulus	Equilibrium
1	0.887 ± 0.000i	0.887	3.82
2	0.746 ± 0.876i	1.151	NS[12]

[12] NS stands for non-stationary.

Table 2 implies that the first regime is stationary, while the second regime is explosive and oscillatory. In contrast to the sunspots model, the skeleton of this model converges only to a limit cycle. With any starting value, a simulation of the series will alternate between the stationary and the explosive regimes. Since the threshold of 3.116 is less than the long run equilibrium value of the stationary regime, the series will pass the threshold and switch to the explosive regime in due time. The series will not stay in the explosive regime for long, however. Once the explosive oscillation dips below the threshold, the regime is switched back to the stationary one after two periods.

Inspired by Tong's threshold idea, Potter (1993; 1994, 323) applied a TAR model to the U.S. GNP data. In his model the two equations[13] are

$$Y_t = \begin{cases} -0.808+0.516Y_{t-1}-0.946Y_{t-2}+0.352Y_{t-5}+\varepsilon_t^{(1)} & \text{if } Y_{t-2}\le 0, \\ 0.517+0.299Y_{t-1}+0.189Y_{t-2}-0.143Y_{t-5}+\varepsilon_t^{(2)} & \text{if } Y_{t-2}>0, \end{cases} \qquad (3\text{-}8)$$

where $Y_t = 100(1-L)\log(GNP_t)$, from 1947q1 to 1990q4.

Table 3 Characteristics of the GNP model

Regime	Dominant Eigenvalue	Modulus	Equilibrium
1	0.334 ± 1.047i	1.099	NS
2	0.663 ± 0.334i	0.743	0.789

Contrary to Tong's Lynx model (3-7), the threshold in Potter's GNP model is less than the equilibrium of the second regime, so there are two convergence paths for the skeleton in this model. Depending on the initial condition, the model either converges to the long run equilibrium of the second regime, or it converges to a stable limit cycle (with periodicity equals to 4, based on a simulation of (3-8)).

---

13 There seems to be a misprint in Potter (1994, 323). The coefficients for the 5th lag in both regimes have the wrong signs. Potter (1993, table 3, 25) has the right signs.

3.4. *Threshold versus Markov-Switching*

In contrast to TAR models, MS models in the literature normally are not capable of generating limit cycles. The reason is that the states are exogenous to the system, and their transition probabilities are fixed. Laroque and Rabault's 1995 model is perhaps one exception. There they used a filter similar to that of Hamilton's, and yet the transition probabilities are governed by a process more like that of a TAR model.

One possibility of generating a limit cycle in an MS model is to endogenize the transition probabilities. Generalization of MS models to incorporate time-varying transition probabilities (TVTP) has been studied by Francis Diebold, Joon-Haeng Lee, and Gretchen Weinbach (1994), and by Filardo (1994). The next step is to make the TVTP a function of lagged endogenous variables. Once this is done, the line between a TAR model and an MS model becomes blurred.

Although in both TAR and MS models there are at least two regimes, the approaches differ in their treatment of the regimes. Tong explicitly views the threshold and the regimes as a linearization of a nonlinear process. Each regime is seen as a local approximation of a nonlinear function at a point. An MS model, by contrast, makes no reference to the cause of regime switching. It is exogenous to the model.

The general model that I develop below can encompass both a TAR model and an MS model as special cases. It endogenizes the regime switching as in a threshold model. However, it only tries to infer the probability of being in a particular regime through filtering as in a Markov-switching model. This combined model can be called an autoswitching Markov or a self-switching Markov (SSM) model — a name inspired by Tong's self excited threshold principle.

This combination enriches the theoretical content of an MS model. The regimes are no longer governed by a hidden force. The approach provides a richer environment to study factors that cause a regime to switch.

This combination also enriches the statistical content of a threshold model. It becomes possible to estimate the threshold value and delay factor jointly, without an auxiliary grid search.

I will now define the SSM model formally, and show how it can be constrained to represent either a TAR or an MS model. The related estimation strategies will also be noted.

3.5. A Self-Switching Markov (SSM) Model

3.5.1. The Preliminaries

An n-dimentional multi-state (or regime) vector autoregressive model with r-lags can be defined by the following equation[14] if the states are observable:

$$Y_t = C_{S_t^*} + \Phi_{1,I(S_t^*,1)} Y_{t-1} + \Phi_{2,I(S_t^*,2)} Y_{t-2} + \cdots + \Phi_{r,I(S_t^*,r)} Y_{t-r} + \varepsilon_{t,S_t^*} \qquad (3\text{-}9)$$

where

Y_t is an nx1 vector of endogenous variables,

$C_{S_t^*}$ is an nx1 vector of state dependent constants, where

S_t^* is a state indicator variable taken on values 0 or 1, and

$\Phi_{h,I(S_t^*,h)}$ is an nxn matrix indexed by lag and by an indicator function $I(S_t^*,h)$, where

$I(S_t^*,h)$ takes on values 0, or 1, based on the value of S_t^* and h, h=1,...,r,

and

---

[14] In the notations that follow, unless otherwise noted, k is the number of states; r is the number of lags; indices i, j are from 1 to k; h is from 1 to r; and an '*' next to a variable is a binary state variable.

$$\varepsilon_{t,S_t^*} \sim N(0,\Omega_i), \qquad (3\text{-}10)$$

where

$\Omega_i = \Omega_0$ if S_t^* =0, otherwise $\Omega_i = \Omega_1$.

Because each element in the state vector $\{S_t^*, S_{t-1}^*, ..., S_{t-r}^*\}$ can have two possible values, we can specify the indicator function $I(S_t^*, h)$ of (3-9) in a variety of ways. At one extreme, we can define $I(S_t^*, h)$=0, for all h. In which case, the AR coefficients stay the same in both regimes as in Hamilton's 1989 model. However, we can relax the assumption of uniformity of AR coefficients across states and set $I(S_t^*, h)$=S_t^* for all h, as suggested by Robert McCulloch and Tsay (1994). In this case, the dynamic of the system (3-9) is determined solely by the current state. A TAR model such as Potter's 1994 model is also of this type. At the other extreme, we can define the indicator function $I(S_t^*, h)$=S_{t-h}^*, so that there are 2^{r+1} possible combinations of AR coefficients and constants. The dynamics in (3-9) then become more complex. Similarly, the variance-covariance matrix Ω_i in (3-10), can either be free to differ between two states, or to be constrained.

Table 4 Binary-state to K-state correspondence

S_t	S^*_t	S^*_{t-1}	...	S^*_{t-r}
1	0	0	0	0
2	1	0	0	0
3	0	1	0	0
:	:	:	:	:
k/2	1	1	1	0
1+k/2	0	0	0	1
:	:	:	:	:
k	1	1	1	1

To simplify notation, I will define a new state variable S_t based on S_t^*. Suppose $I(S_t^*, h)$=S^*_{t-h}, then the indicator function equals the value of the state at time t-h. In this case, the value of S_t based on S_t^* is defined according to Table 4. As a consequence, not all transition probabilities are positive.

Although there are multiple states in S_t, there can only be one or two long-run equilibria. If both regimes are stationary, their long-run equilibria are:

$$\mu_i = (I - \Phi_{1.i} - \Phi_{2.i} - \cdots - \Phi_{r.i})^{-1} C_i, \text{ i=0,1.} \qquad (3\text{-}11)$$

If one of the two regimes is non-stationary (let it be regime 1), then at least one modulus of the eigenvalues of the following matrix (see Hamilton 1994, eq. 10.1.10)

$$F = \begin{bmatrix} \Phi_{1.1} & \Phi_{2.1} & \cdots & \Phi_{r-1.1} & \Phi_{r.1} \\ I & 0 & \ldots & 0 & 0 \\ 0 & I & \ldots & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ O & O & \ldots & I & 0 \end{bmatrix} \qquad (3\text{-}12)$$

will be greater than one.

3.5.2. The Regimes and Their Transitions

When the states are not observable, we can weight (3-9) by the probability of each state's occurrence and sum over all k-state probabilities, i.e.,

$$Y_t = \sum_{j=1}^{k} \hat{p}_t^j \left(\Gamma_j X_t + \varepsilon_{t,j^*}\right), \qquad (3\text{-}13)$$

where

$\hat{p}_t^j$ is the expected probability for $S_t = j$,

j^*=0 if j is even, and j^*=1 if j is odd,

$$X_t = [1, Y_{t-1}', Y_{t-2}', \cdots, Y_{t-r}']', \qquad (3\text{-}14)$$

and the system of the autoregressive coefficients

$$\Gamma_1 = [C_{S_t^*=0}, \Phi_{1.I(S_{t-1}^*=0.1)}, \Phi_{2.I(S_{t-2}^*=0.2)}, \cdots, \Phi_{r.I(S_{t-r}^*=0.r)}], \ldots, \qquad (3\text{-}15)$$

$$\Gamma_k = [C_{S_t^*=1}, \Phi_{1,I(S_{t-1}^*=1,1)}, \Phi_{2,I(S_{t-2}^*=1,2)}, \cdots, \Phi_{r,I(S_{t-r}^*=1,r)}].$$

The density of Y_t conditioned on S_t is then

$$f(Y_t \mid S_t = j, X_t) = (2\pi)^{-n/2} \mid \Omega_j^{-1} \mid^{1/2} \exp[(-1/2)(Y_t - \Gamma_j X_t)' \Omega_j^{-1} (Y_t - \Gamma_j X_t)] \quad (3\text{-}16)$$

The marginal density, or the likelihood of an observation is

$$f(Y_t \mid X_t) = \sum_{j=1}^{k} f(Y_t \mid S_t = j, X_t) \hat{p}_t^j . \quad (3\text{-}17)$$

The different specifications for the weighing variable $\hat{p}_t^j$ distinguish the various multi-state models, such as the TAR, STAR, MS, or SSM models.

In a TAR model,

$$\hat{p}_t^j = I(\eta - Y_{t-d}), \text{ j=1, 2,} \quad (3\text{-}18)$$

where

η is the threshold value,

d is the delay factor, and the indicator function

$I(\eta - Y_{t-d})$ takes on value of 0 or 1.

In a STAR model,

$$\hat{p}_t^j = g(\eta - Y_{t-d}), \text{ j=1, 2,} \quad (3\text{-}19)$$

where

$g(\eta - Y_{t-d}) \in [0,1]$, is a continuous non-decreasing function.

In an MS model,

$$\hat{p}_t^j = \sum_{i=1}^{k} p^{ij} p_{t-1}^i , \text{ t>1,} \quad (3\text{-}20)$$

is the projected probability of S_t=j given the information at time t-1, where

$$p^{ij} = p(S_t = j \mid S_{t-1} = i) \quad (3\text{-}21)$$

is a constant of the state transition probability, and

$$p_{t-1}^i = P(Y_{t-1}, S_{t-1} = i \mid X_{t-1}) / f(Y_{t-1} \mid X_{t-1}) , \text{ t>2,} \quad (3\text{-}22)$$

is the filtered probability of $S_{t-1}=i$ given the information at time t-1. In (3-22)

$$P(Y_{t-1}, S_{t-1} = i \mid X_{t-1}) = f(Y_{t-1} \mid S_{t-1} = i, X_{t-1})\hat{p}^{i}_{t-1} \qquad (3\text{-}23)$$

is the joint density, and the likelihood $f(Y_{t-1} \mid X_{t-1})$ is defined in (3-17). At the time t=1, $\hat{p}^{j}_{1}=\rho^{j}$, a constant to be defined later.

In an SSM model,

$$\hat{p}^{j}_{t} = \sum_{i=1}^{k} p^{ij}_{t} p^{i}_{t-1}, \qquad (3\text{-}24)$$

where p^{i}_{t-1} is the same as (3-22) in an MS model, but the state transition probability

$$p^{ij}_{t} = p(S_t = j \mid S_{t-1} = i, X_t) \qquad (3\text{-}25)$$

is an element in the kxk state transition probability matrix

$$P_t = \begin{bmatrix} p^{*00}_t & 1-p^{*00}_t & 0 & 0 & & 0 & 0 \\ 0 & 0 & 1-p^{*11}_t & p^{*11}_t & & \vdots & \vdots \\ \vdots & \vdots & 0 & 0 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & 0 & 0 \\ 0 & 0 & 0 & 0 & & 1-p^{*11}_t & p^{*11}_t \\ p^{*00}_t & 1-p^{*00}_t & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1-p^{*11}_t & p^{*11}_t & & \vdots & \vdots \\ \vdots & \vdots & 0 & 0 & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & & 0 & 0 \\ 0 & 0 & 0 & 0 & & 1-p^{*11}_t & p^{*11}_t \end{bmatrix}, \qquad (3\text{-}26)$$

where $p^{*ii}_t \in P^{*}_t$. The latter is a first order Markov state transition probability matrix, i.e.,

$$P^{*}_t = \begin{bmatrix} p^{*00}_t & 1-p^{*00}_t \\ 1-p^{*11}_t & p^{*11}_t \end{bmatrix}, \qquad (3\text{-}27)$$

where

$$p^{*ii}_t = g_i(\beta_i, X_t, \delta_i),\ i=0, 1, \qquad (3\text{-}28)$$

where

$g_i(\cdot)$ is a function that maps R^{2m-1} to [0,1], where m=nr+1,

δ_i is a scaling factor for state i,

$$\beta_i = [\eta_i, \beta_{i,1}, \beta_{i,2}, \cdots, \beta_{i,r}] \quad (3\text{-}29)$$

is a 1xm row vector[15], where

η_i is a constant,

$\beta_{i,h}$ is a 1xn row vector so that the product

$$\beta_i X_t = \eta_i + \beta_{i,1} Y_{t-1} + \beta_{i,2} Y_{t-2} + \cdots + \beta_{i,r} Y_{t-r} \quad (3\text{-}30)$$

is a scalar.

There are many functional forms for $g_i(\cdot)$ that one can use to satisfy the mapping. Logit is one, and semi-normal is another. In a logit mapping, $g_i(\cdot)$ can be defined as

$$p_t^{*ii} = g_i(\beta_i, X_t, \delta_i) = \left(1 + \exp\left(-\delta_i(\beta_i X_t)\right)\right)^{-1}. \quad (3\text{-}31)$$

In a semi-normal mapping, $g_i(\cdot)$ can be defined as

$$p_t^{*ii} = g_i(\beta_i, X_t, \delta_i) = \exp\left(-\delta_i(\beta_i X_t)^2\right). \quad (3\text{-}32)$$

In either case, only β_i or δ_i is identified, not both[16].

3.5.3. The Initial States

Hamilton (1989) set ρ^j, j=1 to k, to the ergodic probabilities[17] implied by the

---

[15] A double subscript for β identifies the regime and the lag, whereas a single subscript identifies only the regime.

[16] Simulations with both mappings have shown that the semi-normal mapping is less likely to have an explosive trajectory when faced with a short sequence of strong shocks while one regime is not stationary. This is partly due to the symmetry of the mapping. Also there is a focal point (when $\beta_i X_t$=0) for each regime for which (3-32) is one (p_t^{*ii} =1). The combination of these two factors constraints the behavior of the model. The probability of staying in an explosive regime diminishes the farther is the series deviates from the focal point.

[17] Since (3-26) is assumed to be irreducible, and all eigenvalues of (3-26) except one are inside the unit circle, (3-26) represents an ergodic Markov chain. The ergodic initial state probability vector is then the normalized eigenvector of (3-26) associated with the unit eigenvalue (see Hamilton 1994, 681).

transition matrix (3-26). It is possible in his case to do so, for the matrix is a matrix of constants without time subscript. When state transition probabilities are time-varying, however, the ergodic state transition probabilities for the whole sample are not well defined.

In a 2-state model without lags, Diebold, Lee, and Weinbach (1994) proposed treating ρ^j as an additional parameter to estimate. When there are lags, as in (3-9), there may be as many as 2^{r-1}-1 independent initial state probabilities (as defined in Table 4). It is not practical to estimate so many parameters.

An alternative approach is to use an iterative process. For example, let

$$\rho^j = P(S_1 = j \mid Y_T, \ldots Y_1), \tag{3-33}$$

where $P(S_1 = j \mid Y_T, \ldots Y_1)$ is the full sample smoothed inference of the state probabilities (see Hamilton 1994, 694) based on the initial parameter vector, and estimate the model conditioned on this initial set of values. Next, based on the new parameter vector, repeat (3-33) and re-estimate the model. Repeat the process until it converges.

The log likelihood of the model is then simply the summation of the log of (3-17) from t=1 to T, with a special case for the initial state probability at t=1:

$$\begin{aligned} \log f(Y_T, .. Y_1 \mid Y_0, .. Y_{-r+1}) = \log \sum\nolimits_{j=1}^{K} f(Y_1 \mid S_1 = j, X_1)\rho^j + \\ \sum\nolimits_{t=2}^{T} \log \sum\nolimits_{j=1}^{k} f(Y_t \mid S_t = j, X_t)\hat{P}_t^j \end{aligned} . \tag{3-34}$$

3.5.4. The Gradient

To fully appreciate the recursive nature of an SSM model, I have derived the gradient of the model in what follows. To make clear the dependency of the likelihood on the model parameters, first I define the super parameter set

$$\Theta = \{\Gamma_1, \ldots, \Gamma_k, \Omega_0, \Omega_1, \beta_0, \beta_1, \delta_0, \delta_1\}, \tag{3-35}$$

where the subset $\{\Gamma_1, ..., \Gamma_k\}$ is from the structural equation (3-15), or (3-9), the subset $\{\Omega_0, \Omega_1\}$ is from the variance-covariance matrix (3-10), and the subset $\{\beta_0, \beta_1, \delta_0, \delta_1\}$ is from the transition equations (3-29) and (3-28).

Given the definition of the density of an observation (3-17), the log likelihood of the t[th] observation, t >1, conditioned on the past observations and the parameter set Θ is

$$L_t = \log f(Y_t \mid X_t, \Theta) = \log \sum_{j=1}^{k} f(Y_t \mid S_t = j, X_t, \Theta) P(S_t = j \mid X_t, \Theta), \qquad (3\text{-}36)$$

where

$$P(S_t = j \mid X_t, \Theta) = \hat{p}_t^j = \sum_{i=1}^{k} p^{ij} p_{t-1}^i \qquad (3\text{-}37)$$

as is in (3-24).

The recursion is due to the dependency of $P(S_t = j \mid X_t, \Theta)$ on $f(Y_{t-1} \mid X_{t-1}, \Theta)$ in view of (3-22). Since the density of each observation is recursively defined, the gradient of the model parameters with respect to the likelihood function is also recursive in nature. This result can be shown in the following.

Let γ be an active element in the parameter set, i.e.$\gamma \in \Theta$. (Not all elements in this set are distinct. For instance, most of the elements in the subset $\{\Gamma_1, ..., \Gamma_k\}$ could be the same, especially in Hamilton's model (3-61).) The partial derivative of the log likelihood of the t[th] observation with respect to γ is

$$\frac{\partial L_t}{\partial \gamma} = \left(\sum_{j=1}^{k} \hat{P}_t^j \partial f(Y_t^j) / \partial \gamma + \sum_{j=1}^{k} f(Y_t^j) \partial \hat{P}_t^j / \partial \gamma \right) \Big/ f(Y_t \mid X_t), \qquad (3\text{-}38)$$

where the short-hand notation

$$f(Y_t^j) = f(Y_t \mid S_t = j, X_t, \Theta). \qquad (3\text{-}39)$$

Of course, some of the partial derivatives are zero. Evaluating the $\partial f(Y_t^j)/\partial \gamma$ is straight forward in view of (3-16). Complications arise for the $\partial \hat{P}_t^j / \partial \gamma$. From

(3-22) and (3-24), we see that

$$\frac{\partial \hat{p}_t^j}{\partial \gamma} = \frac{\partial \sum_{i=1}^{k} p_t^{ij} p_{t-1}^{i}}{\partial \gamma} = \frac{\partial \sum_{i=1}^{k} p_t^{ij} f(Y_{t-1}^{i}) \hat{p}_{t-1}^{i} \big/ f(Y_{t-1})}{\partial \gamma}, \tag{3-40}$$

where

p_t^{ij} is an element of (3-26), defined by (3-28),

$\hat{P}_{t-1}^{i} = P(S_{t-1} = i \mid X_t, \Theta)$ is the filtered i[th] state probability from (3-22), and

$f(Y_{t-1}) = \sum_{j=1}^{k} f(Y_{t-1}^{j}) \hat{P}_{t-1}^{j}$ is the unconditional density form (3-17).

The last expression in (3-40) is a sum of 4 multiplicative factors, three of which are the recursion of the derivative at time t-1. At time t=1, however, (3-40) equals to zero, because the initial state probabilities are assumed to be independent from the model parameters. When (3-40) equals to zero, the second summations in (3-38) are all zeros as well.

3.5.5. The Estimation

a) Finding a Starting Parameter Vector

Unlike a linear model, the starting parameter vector plays a crucial role in estimation of an SSM model. One reason is that the likelihood is defined only in a subset of the parameter space. For example, if a regime has only one data point, the variance of that regime becomes zero and the log likelihood explodes to infinity. The model is also ill-defined when both regimes are non-stationary. For these reasons, the starting parameter vector should not be arbitrary. I have developed the following algorithm to find a starting parameter vector:

1. Fit a single set of AR coefficients using OLS
2. Create a vector with only binary values based on the signs of the OLS residuals

3. Regress this new vector to generate a set of linear probability coefficients and use it as the starting parameter values for the logit transition probability equation
4. Group sample observations into two sub-groups based on the signs of the OLS residuals
5. Regress each sub-group separately to generate the AR coefficients for the two regimes.

This approach works well for the logistic mapping of the transition probabilities, for the linear probability coefficients are consistent estimators of the true probabilities. The same is not true for the semi-normal mapping. Only ad-hoc methods seems to exist for creating vectors of transition coefficients. One method is to use the two AR coefficients as the transition coefficients. These can lead to undefined likelihoods, however. To ensure the consistency between the focal point of the initial coefficients in the transition equation and the long-run equilibrium of the initial AR coefficients from (3-11), at least for the stationary regimes, I opted to initialize

$$\eta_i = -\mu_i(\beta_{i,1}+\beta_{i,2}+\ldots+\beta_{i,r}), \tag{3-41}$$

so that $\beta_t X_t = 0$ and $p_t^{ii} = 1$ in (3-32), whenever $Y_{t-h} = \mu_i$, h=1 to r.

b) Finding Multiple Local Maxima

The reason that the starting parameter vector plays a crucial role in finding the global maximum of an SSM model is that in a nonlinear model of this type, there might be multiple local maxima as well as saddle points in the likelihood surface. The algorithm above leads only to a local maximum. To find the global maximum, we need to find as many local maxima as possible.

Each local maximum has a domain in the parameter space for which, starting from any points inside the domain, a hill-climbing algorithm, such as the quasi-Newton algorithm, will find the local maximum eventually. However, if one starts from a point outside the domain of a local maximum, convergence is not assured. Thus, to find as many local maxima as possible, we need additional strategies.

In the following, I will introduce 3 strategies for finding a point in the domain of a local maximum. The first strategy is a grid search algorithm that looks for promising starting parameter values from one local maximum to another. In practice, however, only a few alternative domains can be identified by the grid search algorithm, due to the complexity of the likelihood surface. This drawback leads to the need for a second strategy, an algorithm that generates a set of starting parameter vectors randomly. A simple, and yet valuable third strategy is to use one local maximum parameter vector as the starting vector of a new estimation, except that the parameters in the transition probability equations are switched. Since the third strategy is straight forward, I will only explain the first two in greater detail.

i) Grid-Search for Starting Vectors

If one designs a grid in the parameter space and tries to use all the grid points as the starting values in a quasi-Newton estimation process, the computational burden becomes prohibitive. A grid search for all points that have a higher likelihood value than their neighboring points is one alternative strategy to narrow down the set of starting values. The finer the grid, the less the chance of missing the domain of a local maximum. However, the finer the grid, the more chance those starting values will lead to the same set of local maxima. The computational trade-off is still high. Thus, finding a manageable set of starting values is a practical challenge for building a successful SSM model.

The algorithm developed here is a peak-to-peak search algorithm. It searches from one local maximum to the next, without estimating from all the grid points in between. The key is to identify a point that is outside the domain of the current local maximum. Using this algorithm, one can quickly identify the starting parameter vector for a new local maximum without a high computational burden. The trade-off is that one cannot be guaranteed that the global maximum is among the local maxima found.

The algorithm is the following:

First we design a parameter grid with, say, 10 points for each parameter. (Figure 1 represents an example of a model with 5 local maxima and only two parameters.) Next, we estimate a model from any reasonable set of starting values (such as those given by point "A" in Figure 1). Once the local maximum is found (point "B"), we add the estimated parameter values to the parameter grid. Using the local maximum just found as the starting point, say the k^{th} point on the grid, we test for the likelihood value on points along the grid lines for each parameter. If we find a grid point, say, the j^{th} point ($j \neq k$) on the grid that has a likelihood value higher than points j-1 and j+1 for the i^{th} parameter, then we have found a point that is outside the domain of the current local maximum. It could be a point that is closer to the next local maximum (points 1 and 1' in Figure 1). We set the i^{th} parameter value to the j^{th} value on the grid and repeat the process for the next parameter.

Since we are incorporating local maxima into our grid, any peak different from the local maxima already on the grid represents a point within the domain of a new local maximum. If we start our estimation from such a point, we should reach the new local maximum. Further search along the grid for points like 2, 3, and 4, not only helps reduce the time it takes to estimate the model, but also eliminates false starts from such points as 5 and 5' in Figure 1. Those points lie within the domain of the local maximum found previously.

By searching on the grid for the highest point within the domain of a local maximum, we can find a point such as point 7 in Figure 1 that is outside the domain of any local maximum we have already found. We can repeat the process until any highest point found on the grid is a local maximum found earlier.

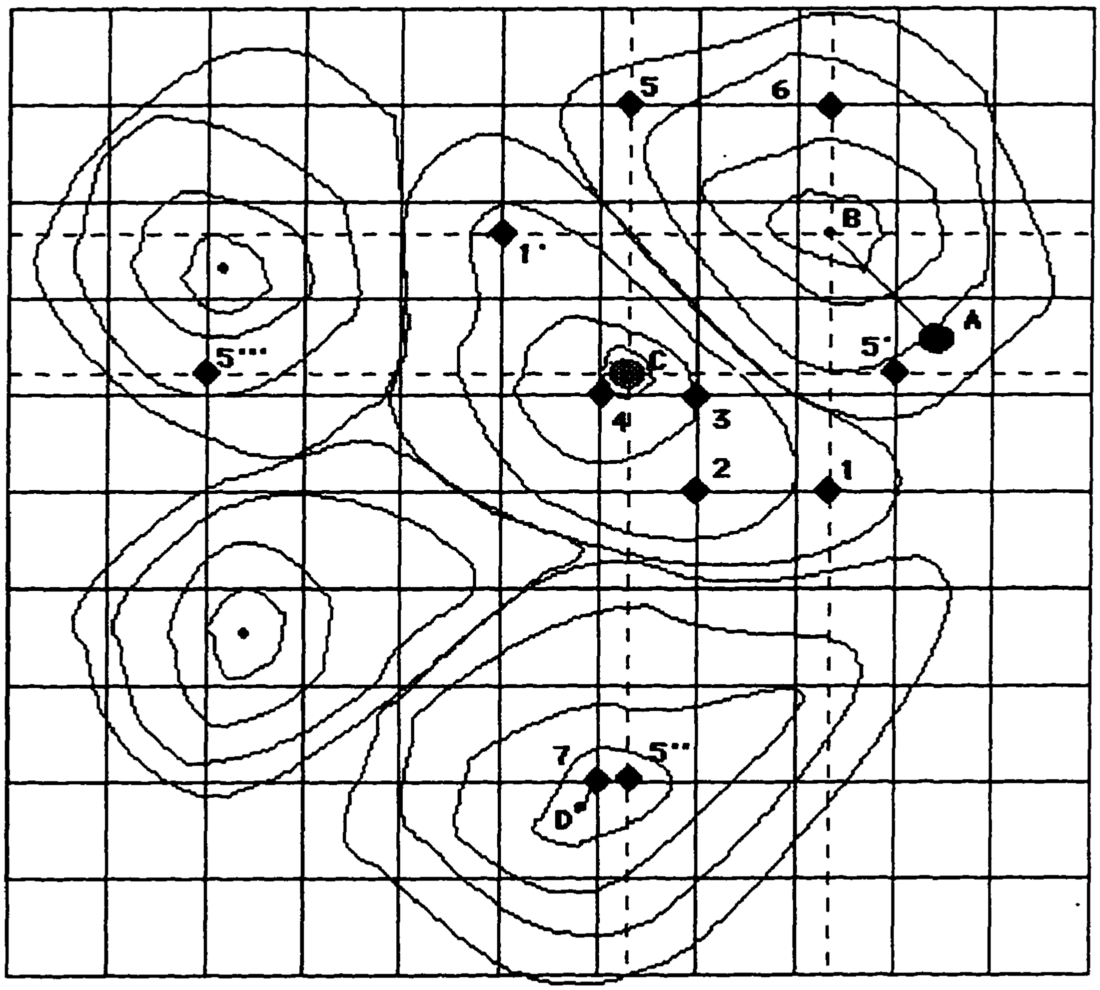


Figure 1 Grid search of local maxima.

ii) Random Generation of Starting Vectors

Since the likelihood is only defined on a subset of points in the parameter space, random starting vectors within bounds often lead to undefined likelihood. To generate starting parameter vector randomly and efficiently, we need to generate within the subset as much as possible. Since the algorithm for "Finding a Starting Parameter Vector" on page 45 was not a recursive algorithm, it can be repeated with each random subset of observations.

An experiment using the yearly Canadian Lynx trapping series is very illuminating. The likelihood is defined on only 50 out of 20,000 random parameter vectors. The reason for this low number is that although each parameter is generated within its reasonable range, the combination of all the parameters leads to an undefined likelihood. Instead, when random subsets of observations were used to generate 100 starting parameter vectors, only a few leads to undefined likelihoods, and yet many distinct local maxima were found.

c) Parameter Transformation

In actual estimation it is necessary to ensure that parameter values do not wander outside their proper domain. In particular, we need to ensure that the variance-covariance matrix in the autoregressive equation is positive definite. The same holds true for the scaling factors in the transition probability equation when they are free parameters. To this end, Cholesky decomposition is used for the variance-covariance matrix, and exponential transformation is used for the scaling factor during estimation.

3.5.6. The Evaluation

For a given model, when there are multiple local maxima, the best model estimate should be the one that has the highest likelihood value. However, due to small sample size, the global maximum in a sub-sample might not be the global maximum in a larger sample. Thus there are alternative criteria for selecting the best estimate of a model.

a) Post-Sample Prediction Error

One selection criterion for the best model estimate is to based the selection on the forecasting power of the estimates. The global maximum that has the best in-sample fit might not be the one that gives the minimum post-sample prediction error.

i) Long Horizon Forecast

Suppose we are interested in forecasting τ-step-ahead (τ > r, the number of lags) at time T, then the expected $Y_{T+\tau}$ at time T conditioned on the $S_{T+\tau}=j$ is

$$\hat{Y}^{j}_{T+\tau|T} = E(Y_{T+\tau} \mid S_{T+\tau} = j, \hat{X}_{T+\tau|T}) = \Gamma_j \hat{X}_{T+\tau|T}, \quad (3\text{-}42)$$

where Γ_j is from (3-15) with $S_{T+\tau}=j$, and

$$\hat{X}_{T+\tau|T} = [1, \hat{Y}'_{T+\tau-1|T}, \hat{Y}'_{T+\tau-2|T}, \cdots, \hat{Y}'_{T+\tau-r|T}]', \quad (3\text{-}43)$$

is known from the forecast for time T+τ-1 of the unconditional expected $\hat{Y}_{T+\tau-1|T}$.

The unconditional expected $\hat{Y}_{T+\tau|T}$ is found by summing over the conditional $\hat{Y}^{j}_{T+\tau|T}$, weighted by the projected state probabilities, i.e.,

$$\hat{Y}_{T+\tau|T} = \sum_{j=1}^{k} \hat{p}^{j}_{T+\tau|T} \cdot \Gamma_j \hat{X}_{T+\tau|T}, \quad (3\text{-}44)$$

where

$$\hat{p}^{j}_{T+\tau|T} = \sum_{i=1}^{k} p^{ij}_{T+\tau|T} \cdot p^{j}_{T+\tau-1|T}, \quad (3\text{-}45)$$

and where

$p^{ij}_{T+\tau|T}$ is an element of (3-26), defined by (3-28), and

$p^{j}_{T+\tau-1|T}$ is the filtered probability of $S_t = j$ forecasted for time T+τ-1. The unconditional density forecast is

$$f(\hat{Y}_{T+\tau|T} \mid \hat{X}_{T+\tau|T}) = \sum_{j=1}^{k} f(\hat{Y}_{T+\tau|T} \mid S_{T+\tau} = j, \hat{X}_{T+\tau|T}) \hat{p}^{j}_{T+\tau|T}, \quad (3\text{-}46)$$

where

$$f(\hat{Y}_{T+\tau|T} \mid S_{T+\tau} = j, \hat{X}_{T+\tau|T}) = (2\pi)^{-n/2} |\Omega^{-1}_{j'}|^{1/2} \exp[(-1/2)(\hat{Y}_{T+\tau|T} - \Gamma_j \hat{X}_{T+\tau|T})' \Omega^{-1}_{j'} (\hat{Y}_{T+\tau|T} - \Gamma_j \hat{X}_{T+\tau|T})]', \quad (3\text{-}47)$$

where j'=0 if j is even, and j'=1 if j is odd. From these quantities, we find the filtered probability of $S_t = j$ forecasted for time T+τ:

$$p^{j}_{T+\tau|T} = f(\hat{Y}_{T+\tau|T} \mid S_{T+\tau} = j, \hat{X}_{T+\tau|T}) \hat{p}^{j}_{T+\tau|T} / f(\hat{Y}_{T+\tau|T} \mid \hat{X}_{T+\tau|T}). \quad (3\text{-}48)$$

This process can be repeated to forecast $\tau+1$ steps ahead.

Suppose that a model is estimated with sample size T when the full sample has T+h data points. Then the h-step mean post-sample prediction error in a univariate case (see Harvy 1990, 189) is

$$\xi_{h,1|T} = \frac{1}{h}\sum_{\tau=1}^{h}\left(\hat{Y}_{T+\tau|T} - Y_{T+\tau} \mid \Theta\right)^2 \qquad (3\text{-}49)$$

where the first subscript for ξ is the forecast horizon, the second subscript is the number of forecasts made, and the third subscript is the data set used to estimate the parameter set Θ. The forecast $\hat{Y}_{T+\tau|T}$ is from (3-44), and $Y_{T+\tau}$ is an observed data point.

This test of the forecast performance might not be very informative when the forecast horizon h is large. The forecast $\hat{Y}_{T+\tau|T}$ might converge rapidly to a single point. The power to detect the change in regimes between different estimates is lost.

Experiences with various models show that if one regime of a model is nonstationary and the logistic mapping is used for the transition equation, then there exist a short sequence of shocks after time T for which the simulation of the model can become nonstationary. It is less likely, however, for the semi-normal mapping to experience the same explosive trajectory.

ii) Simulated Real Time Forecast

An alternative definition of the mean one-step post-sample prediction error is to simulate a real time forecast, and take the average of the one-step-ahead forecast over horizon h, i.e.,

$$\xi_{1,h|T} = \frac{1}{h}\sum_{\tau=1}^{h}\left(\hat{Y}_{T+\tau} - Y_{T+\tau} \mid \Theta\right)^2, \qquad (3\text{-}50)$$

where

$$\hat{Y}_{T+\tau} = \sum_{j=1}^{k} \hat{p}_{T+\tau}^{j} \cdot \Gamma_j X_{T+\tau} \qquad (3\text{-}51)$$

is derived using information up to time T+τ-1, the same as the in-sample expectation of (3-13), except for the change in time subscript.

b) Bootstrap Likelihood

Another selection criterion for the best model estimate is to select the estimate that is least influenced by the sample period. The influence can be measured by a version of bootstrapping. In our case, we test the goodness of fit of each local maximum using a set of random sub-samples of the data and record the likelihood values from such random sub-samples.

Given the recursive nature of the likelihood, the random subsample cannot be generated by randomly selecting one observation at a time. Instead, I have developed an alternative procedure to generate a random sub-sample.

Based on the ergodicity assumption, the bootstrap sample is derived by organizing the original sample into a repeating data series, going from the τ^{th} observation to the T^{th} and back. A data point in the bootstrap sample

$$Y_t \in \{ Y_{\tau}, \ldots, Y_T, Y_{\tau}, \ldots, Y_T, Y_{\tau}, \ldots \}, \qquad (3\text{-}52)$$

where τ is the minimum index of τ*, where τ* is any point for which

$$D_{\tau^*-1} > D_{\tau^*} < D_{\tau^*+1}, \quad \tau^* = r, \ldots T-1, \qquad (3\text{-}53)$$

where

$$D_{\tau^*} = \|Y_T - Y_{\tau^*}\| + \ldots + \|Y_{T-r+1} - Y_{\tau^*-r+1}\|, \qquad (3\text{-}54)$$

so that the last r observations and the first r observations starting from the τ^{th} observation have the minimum Euclidean distance in the neighborhood of τ=r. Once the bootstrap sample is created by repeating the new data series $\{Y_{\tau}, \ldots, Y_T\}$, a sub-sample can be extracted by starting from a random point in the bootstrap sample and selecting consecutive observations of a given size. The randomness here is only in the sense of randomly selecting a slice of the data within the bootstrap sample. Since only one point of the data for every T-τ data points in the bootstrap sample is artificially connected, its influence on the overall likelihood should be minimal.

The log likelihood of (3-34) for a given vector of parameter values and using a random subsample of the bootstrap sample starting from data point α to α+d is

$$\begin{aligned}\log f(Y_{\alpha+d},..Y_{\alpha} \mid Y_{\alpha-1},..Y_{\alpha-r}) = \\ \log\sum_{j=1}^{K} f(Y_{\alpha} \mid S_{\alpha}=j, X_{\alpha})\rho_{\alpha}^{j} \\ +\sum_{t=\alpha+1}^{\alpha+d}\log\sum_{j=1}^{k} f(Y_t \mid S_t=j, X_t)\hat{P}_t^{j},\end{aligned} \tag{3-55}$$

where

$$\rho_{\alpha}^{j} = P(S_{\alpha}=j \mid Y_{\alpha+d},..Y_{\alpha}) \tag{3-56}$$

is the full-sample smoothed inference of the initial state at time α.

3.5.7. Nesting a TAR or an MS Model in an SSM Model

In a univariate case, the system of a 2-state SSM model with logistic mapping has the general form

$$\begin{aligned}&Y_t=c_0+\phi_{0,y_{t-1}}Y_{t-1}+\phi_{0,y_{t-2}}Y_{t-2}+...+\phi_{0,y_{t-r}}Y_{t-r}+\varepsilon_t^{(1)} && \text{if } S_t=0,\\ &Y_t=c_1+\phi_{1,y_{t-1}}Y_{t-1}+\phi_{1,y_{t-2}}Y_{t-2}+...+\phi_{1,y_{t-r}}Y_{t-r}+\varepsilon_t^{(2)} && \text{if } S_t=1,\\ &p_t^{00}=1/\{1+\exp(-\delta_0[\eta_0+\beta_{0,y_{t-1}}Y_{t-1}+...+\beta_{0,y_{t-r}}Y_{t-r}+\beta_{0,z_{t-1}}Z_{t-1}+...+\beta_{0,z_{t-r}}Z_{t-r}])\},\\ &p_t^{11}=1/\{1+\exp(-\delta_1[\eta_1+\beta_{1,y_{t-1}}Y_{t-1}+...+\beta_{1,y_{t-r}}Y_{t-r}+\beta_{1,z_{t-1}}Z_{t-1}+...+\beta_{1,z_{t-r}}Z_{t-r}])\},\end{aligned} \tag{3-57}$$

where r is the maximum lag used, and Z_t is any exogenous variable used, and p_t^{ii} is the time-varying transition probability from state i to state i.

a) Nesting a TAR Model in an SSM Model

Given (3-57), a TAR model can be represented as a constrained SSM model. When $g_i(\cdot)$ in (3-31) becomes an indicator function, the state transition probability matrix (3-27) becomes a matrix with columns of all zeros or all ones. The summation in (3-24) for each $\hat{p}_t^j$, j=1 to 2 becomes a binary variable, as in

(3-18). For example, Potter's 1994 TAR model with a threshold of 0 in (3-8) above, is equivalent to a 2-state SSM model after setting all coefficients in the transition equation to zero, except for $\beta_{0,y_{t-2}}=-1$, $\beta_{1,y_{t-2}}=1$, and δ equal to some large number, say, 10^4. More precisely

$$p_t^{00}=1/\{1+\exp[-10^4\ (0-Y_{t-2})]\},$$
$$p_t^{11}=1/\{1+\exp[-10^4\ (0+Y_{t-2})]\}, \qquad (3\text{-}58)$$

so that $g_i(\cdot)$ in (3-31) becomes an indicator function.

Tong's TAR models can be represented in a similar fashion, e.g.,

$$p_t^{00}=1/\{1+\exp[-10^4\ (\eta- Y_{t-d})]\},$$
$$p_t^{11}=1/\{1+\exp[-10^4\ (-\eta+Y_{t-d})]\}. \qquad (3\text{-}59)$$

where η is the threshold, and d is the delay factor. However, an SSM model that employs the semi-normal mapping cannot nest a TAR model, due to the symmetry of the mapping. Neither can an SSM model nests a STAR model, due to the absent of filtering of state probabilities in a STAR model.

b) Nesting an MS Model in an SSM Model

A Markov-switching model also can be represent as a degenerate self-switching model using either the logistic or the semi-normal mapping. For example, Hamilton's 1989 model of the log difference of GNP is modeled as:

$$\begin{aligned} y_t - \mu_{s_t^*} &= \phi_1(y_{t-1} - \mu_{s_{t-1}^*}) + \phi_2(y_{t-2} - \mu_{s_{t-2}^*}) \\ &+\phi_3(y_{t-3} - \mu_{s_{t-3}^*}) + \phi_4(y_{t-4} - \mu_{s_{t-4}^*}) + \varepsilon_t' \end{aligned} \qquad (3\text{-}60)$$

where the state S_t^*=0 or 1. The equation (3-60) can be written in an equivalent form using the notation of (3-9) with $I(S_t^*, h)$=0 for all h:

$$y_t = c_{s_t} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \phi_4 y_{t-4} + \varepsilon_t, \qquad (3\text{-}61)$$

where c_{s_t} is one of the 32 combinations of $\mu_{s_t^*} - \phi_1\mu_{s_{t-1}^*} - \phi_2\mu_{s_{t-2}^*} - \phi_3\mu_{s_{t-3}^*} - \phi_4\mu_{s_{t-4}^*}$, depending on the value of the state vector $\{S_t^*, S_{t-1}^*, \ldots, S_{t-4}^*\}$.

All the coefficients in the state transition probability equations are constrained to be zero except for η_0 and η_1. The scaling factors δ_i are fixed to 1 as well. The new logistic state transition probability equations are:

$$p^{00}=1/\{1+\exp(-\eta_0)\},$$
$$p^{11}=1/\{1+\exp(-\eta_1)\}. \tag{3-62}$$

It is no longer a self-switching model, for the 32x32 transition probability matrix in (3-26) is independent of time. Instead, it is a fixed transition probability (FTP) model. One can thus test a self-switching model against the null hypothesis of a FTP model represented by equations (3-61) and (3-62), in particular the transition coefficients in (3-57)

3.5.8. Time-Varying Transition versus Self-Switching

The contributions made by both Diebold, Lee, and Weinbach (1994) and Filardo (1994) are in the formulation of the time-varying transition probabilities (TVTP). Instead of a fixed value for p^{ij}, theirs vary with time, similar to (3-28).

For example, Filardo's 1994 model is very similar to Hamilton's above except that Filardo included the Composite Leading Indicators (CLI) as an exogenous regressor in the state transition probability functions. In these equations, the coefficients of the first two lags of CLI, ($\beta_{i,z_{t-h}}$, i=0,1, and h=1,2) are free from constraints. In his model, the transition probability equations are:

$$p_t^{00}=1/\{1+\exp(-[\eta_0+\beta_{0,z_{t-1}}z_{t-1}+\beta_{0,z_{t-2}}z_{t-2}])\},$$
$$p_t^{11}=1/\{1+\exp(-[\eta_1+\beta_{1,z_{t-1}}z_{t-1}+\beta_{1,z_{t-2}}z_{t-2}])\}, \tag{3-63}$$

He thus makes the $g_i(\cdot)$ and the transition probability matrix in (3-26) a function of time.

The difference between Diebold, Lee, and Weinbach's and Filardo's approaches is that Diebold, Lee, and Weinbach used the full sample smoothed inference to define

$$\hat{p}_t^j = P(S_t = j \mid Y_T, \ldots Y_1), \tag{3-64}$$

instead of (3-24).

The only difference between an SSM model and a TVTP model is in the argument used for $g_i(\cdot)$. In a TVTP model,

$$p_t^{*ii} = g_i(\beta_i, Z_t, \delta_i),\ i=0,1, \tag{3-65}$$

where Z_t is a vector of exogenous variables, where as in an SSM model, the vector X_t can include lagged endogenous variables as well as the exogenous variables.

This might seem to be a trivial difference, but the implication is profound. Without endogenizing the transition probabilities, it is not possible to nest a TAR and an MS model into a unified SSM model, as I have shown earlier. Furthermore, it is not possible also to have endogenous, or limit cycles in the model without endogenizing the transition probabilities.

Another drawback of the TVTP model is the presence of exogenous variables. They make forecasting difficult. Since neither Diebold, Lee, and Weinbach's 1994 model nor Filardo's 1994 model has a dynamic specification for the exogenous variables used, forecasting requires extensions to both. In addition, in their model, Diebold, Lee, and Weinbach do not define the full sample smoothed inference of the state probability at time T+τ, τ>0, i.e., $P(S_{T+\tau}=j|Y_T..Y_1)$. A further extension, therefore, is needed for forecasting to be possible.

The situation is different for the SSM and MS models. Since no exogenous variable is needed in either model, forecasting is more straight forward, as I have shown.

One implication of the SSM approach is that a shock to the system has 3 channels to impact the future. The first channel is through the usual linear AR equations. The second channel is through the change in the filtered state probabilities, thus changes the weight that is given to each AR equation. The third channel is through the change in the transition probabilities, which also affects the weight given to each equation.

This third channel is unique for an SSM model. In a TVTP model, only a change in an exogenous variable can affect the transition probability; shocks to an endogenous variable have no impact.

3.6. Conclusion

The advantage of an SSM model is that it provides a unified framework for modeling both an MS model and a TAR model. By constraining various parameters, the framework also can be used to test a variety of hypotheses. An SSM model is also superior to a TVTP model; for it makes an endogenous cycle as well as forecasting possible.

In anticipation of the complexity of the SSM model, I have provided an algorithm for finding a set of estimation starting values, and a set of test statistics to evaluate the estimations. The next step is to apply the model, the algorithm, and the test statistics to the real economic data.

4. SSM Models of the U.S. Business Cycle

4.1. Introduction

If there are different regimes in a business cycle, one important question to ask is to what extent is the regime endogenously determined. This question can be addressed using the SSM model developed in the previous section. By comparing an SSM model that incorporates endogenous information with a fixed transition probability (FTP) model or a time-varying transition probability (TVTP) model that ignores this information, we can gauge the relative importance of business cycle endogeneity.

4.2. SSM Models of the GDP Series

To make a direct comparison between an SSM model and Hamilton's 1989 FTP Markov-switching model, I applied each model to 3 data sets. The results are reported in Table 5. For all models in the table, the AR equation (3-60) is used instead of (3-9). The transition equations for all models are

$$p_t^{00}=1/\{1+\exp(-\eta_0-\beta Y_{t-1})\},$$
$$p_t^{11}=1/\{1+\exp(-\eta_1+\beta[Y_{t-1}+Y_{t-2}])\}. \tag{4-1}$$

The model "FTP1.1" is the replication of Hamilton's original model. The results are not exactly the same as his except for the underlined digits. The differences are due to the different specification of the initial state probabilities $P(S_1=j)$. Hamilton used the ergodic state probabilities as the initial state probabilities, whereas the initial state probabilities for all models in Table 5 are derived using the iterative process described above on page 57. When the ergodic state probabilities are used, the results are identical to those of Michael

Boldin's reported in his Table 1 (1996, 39). The iterative process improved the mean log likelihood from -1.384 to -1.377.

There are some noteworthy observations[18] to be drawn from Table 5.

The model "FTP1.2" is the second local maximum found using Hamilton's model and data. Michael Boldin (1996) found that Hamilton's 1989 estimation is sensitive to the date set used, as is evident from Table 5. Although the log likelihood of the reported Hamilton model is higher than the second local maximum, the bootstrapped likelihood is lower, indicating a less robust estimate.

In Hamilton's model, the dominant coefficients are the 3rd and 4th lag coefficients. In the other estimates, it is the coefficient of the first lag that dominates.

---

[18] Notes on the symbols used in the table:

Except for σ^2, subscript next to a coefficient is the absolute value of the ratio of the estimate and its standard error. σ^2 is not estimated directly. Instead, it is computed from the estimate of σ. The subscript next to σ^2 is the ratio of the estimate of σ and the standard error of σ.

An entry without a subscript is derived from other parameters and not estimated.

"P^{u}" is calculated from (4-1) using the sample mean, so that all models can be compared.

AIC is the Akaike's Information Criterion. AIC= -2*MeanLogL+2*(number of parameters)/T, where T = sample size.

LL = Mean Log Likelihood.

"BootL" is the mean bootstrapped log likelihood. It is based on a version of bootstrapping described in the section on "Bootstrap" on page 53.

"Std BtL" is the standard error of the bootstrap estimate. Both the mean and the standard error of the bootstrapped log likelihoods are based on 100 random extraction of 50% sub-samples.

"SP^{0}" is the average $P(S_t=0|Y_T,...,Y_1)$, t=1,...,T; i.e., the full sample smoothed recession state probability.

"$\approx Cyc^{0}$" is the approximate mean duration of state 0. Each duration of state 0 is the consecutive periods of length d in which

$P(S_t=0|Y_T,...,Y_1)>0.5$, $P(S_{t+1}=0|Y_T,...,Y_1)>0.5$,..., $P(S_{t+d}=0|Y_T,...,Y_1)>0.5$.

"$\approx$Cyc" is the approximate mean cycle length. It is the average length of a full cycle from the start of state 0 back to state 0.

"Err_{NBER}" is the mean square distance from the smoothed recession state probability to the National Bureau of Economic Research (NBER) reference cycle classification.

"Est" is the rank of the estimate out of the number of local maxima found in the model.

Table 5 32-State FTP versus SSM models of GDP

Data[19]	1951q1 to 1984q4			1959q3 to 1997q1		1947q1 to 1997q1	
Model	FTP1.1	SSM1	FTP1.2	FTP2	SSM2	FTP3	SSM3
μ_0	$-0.373_{1.5}$	$-0.545_{2.1}$	$-0.686_{1.0}$	$-1.213_{3.1}$	$-0.718_{3.0}$	$-1.392_{3.3}$	$-0.337_{1.6}$
μ_1	$1.175_{17.}$	$0.976_{9.6}$	$0.938_{3.7}$	$0.811_{8.2}$	$0.866_{11.}$	$0.825_{9.5}$	$0.908_{11.}$
ϕ_1	$0.002_{0.0}$	$0.118_{1.0}$	$0.287_{1.7}$	$0.298_{2.7}$	$0.215_{2.3}$	$0.416_{5.2}$	$0.312_{4.0}$
ϕ_2	$-0.071_{0.5}$	$0.074_{0.6}$	$0.145_{1.1}$	$0.134_{1.1}$	$0.015_{0.1}$	$0.136_{1.5}$	$0.031_{0.4}$
ϕ_3	$-0.254_{2.6}$	$-0.031_{0.3}$	$-0.081_{0.2}$	$-0.032_{0.3}$	$0.017_{0.1}$	$-0.008_{0.0}$	$0.096_{1.2}$
ϕ_4	$-0.212_{2.2}$	$-0.087_{0.8}$	$-0.103_{0.9}$	$-0.022_{0.2}$	$0.044_{0.4}$	$-0.184_{2.6}$	$-0.152_{2.0}$
η_0	$1.150_{2.2}$	$-2.296_{1.6}$	$-0.157_{0.1}$	$-1.339_{0.9}$	$-4.00_{3.0}$	$-13.58_{0.0}$	$-5.021_{2.9}$
β		$-3.052_{2.1}$			$-2.640_{3.0}$		$-3.971_{3.1}$
η_1	$2.209_{5.2}$	$-0.451_{0.5}$	$2.481_{1.1}$	$3.542_{4.8}$	$0.579_{0.7}$	$3.663_{6.0}$	$-0.928_{1.1}$
σ^2	$0.578_{12.}$	$0.730_{12.}$	$0.707_{8.8}$	$0.415_{13.}$	$0.393_{15.}$	$0.582_{18.}$	$0.575_{18.}$
p^{00}	0.759	0.011	0.461	0.208	0.003	0.000	0.000
p^{11}	0.901	0.982	0.923	0.972	0.987	0.975	0.994
AIC	2.890	2.851	2.919	2.302	2.251	2.566	2.546
L L	-1.377	-1.349	-1.391	-1.098	-1.057	-1.237	-1.222
LR/P-v[20]	7.12/0.0076			9.52/0.0020		5.91/0.0151	
BootL	-1.432	-1.357	-1.407	-1.064	-1.103	-1.246	-1.254
Std BtL	0.109	0.084	0.092	0.186	0.186	0.052	0.056
SP^0	0.288	0.168	0.136	0.042	0.087	0.029	0.118
$\approx Cyc^0$	5.5	2.5	1.6	1.25	1.3	1	1
$\approx Cyc$	17	17	23	27	13	24	10
Err_{NBER}	0.070	0.105	0.104	0.106	0.076	0.157	0.107
Est	1/2	1/4	2/2	1/5	1/3	1/2	1/2

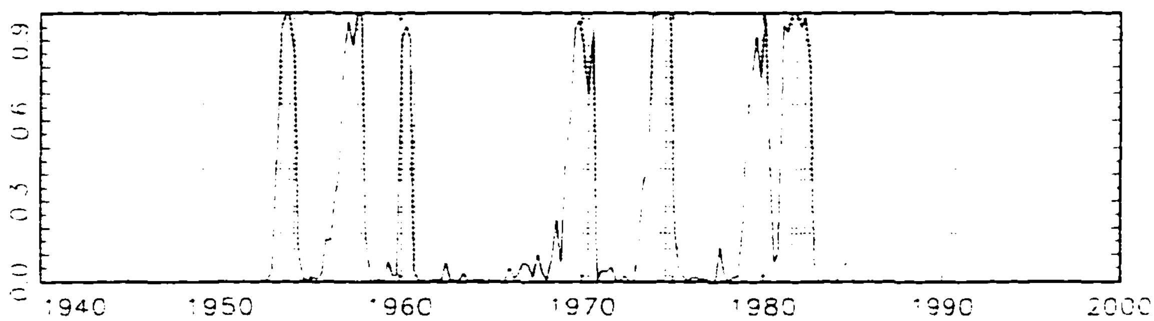
In every data set, SSM has improved the likelihood. However, the boot-

[19] Data sources: The first (GNP 1951q1 to 1984q4 in 1982 dollars, 131 observations) is Hamilton 1989 model's original data set. The second (GDP 1959q3 to 1997q1 in 1992 dollars, 146 observations) is from the St. Louis Federal Reserve's Economic Database (FRED). The third (GDP 1947q1 to 1997q1, 196 observations) is the same as the second after 1959q3. For data before that date, a subseries from FRED (GDP 1947q1 to 1992q1 in 1987 dollars) is used. Since all data series are transformed using log-differences, only a few data points show discrepancies greater than 0.2% between the two basic series.

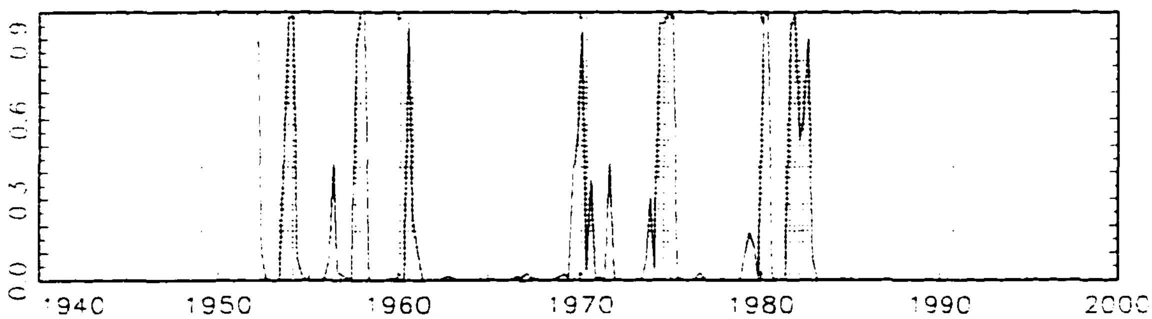
[20] The likelihood ratio and the P-value are between the two adjacent columns of estimates.

strap results shows that the improvement is not robust. In two out of three models, the SSM models have a lower bootstrap likelihood. What is of interest, however, is the improvement in the detection of the 1990-1991 recession in the SSM2 and SSM3 models, as shown in Figure 3 and Figure 4.

For all models in Table 5, the variance is constrained to be the same for both regimes. Otherwise, an estimation of the GDP series can lead to an unreasonably high local likelihood. Since the log-difference GDP series do not show a pronounced conditional difference in variances, the estimation can converge to a situation where only a few observations are fitted almost perfectly into one regime, thus exploding the likelihood[21].



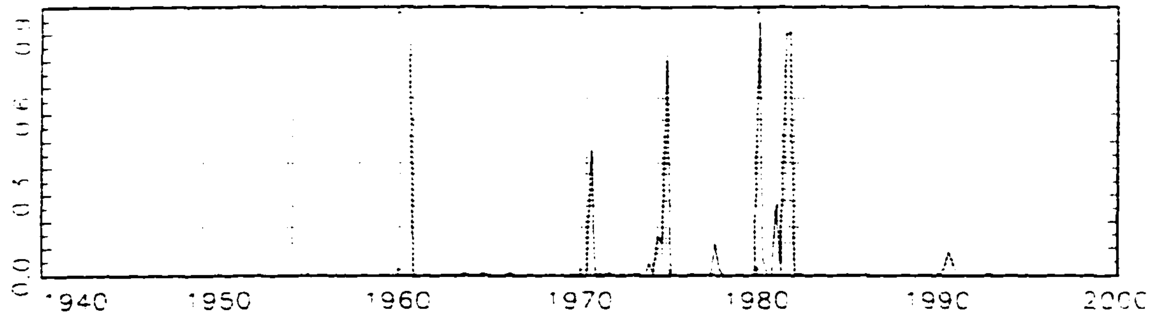
a) 32-State FTP 1.1 Model



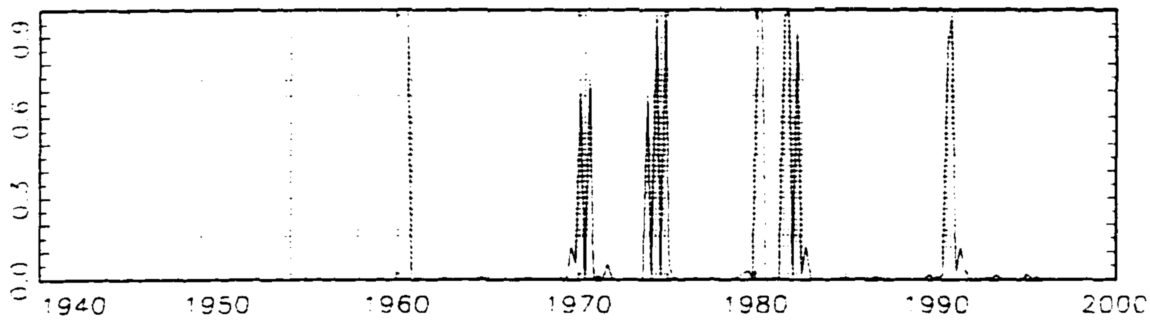
b) 32-State SSM 1 Model

Figure 2 Full-sample recession probabilities of the 32-state GDP models (1)

[21] A TAR model, however, could accommodate two regimes of variances easily, for the number of observations classified in each regime is set by a fixed threshold.

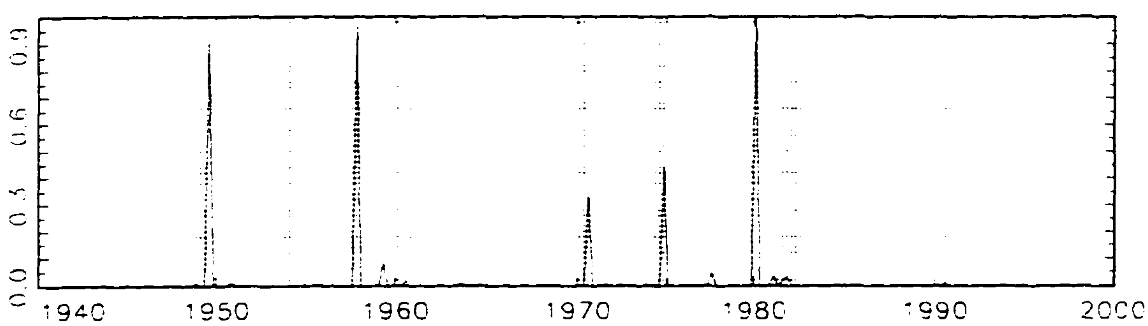


a) 32-State FTP2 Model

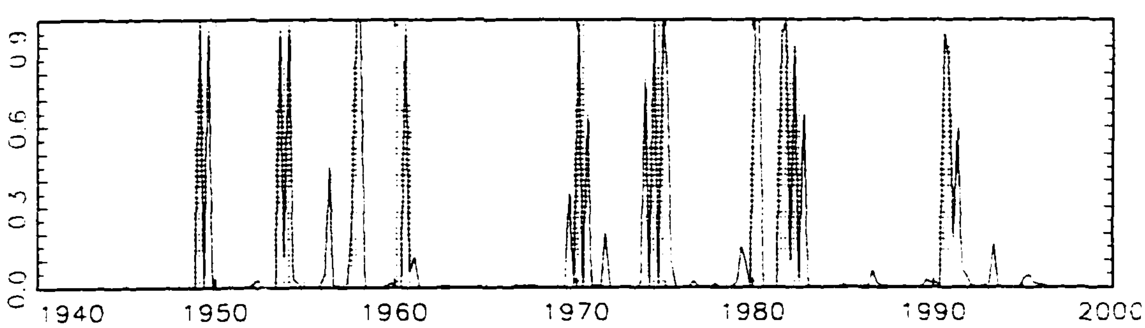


b) 32-State SSM2 Model

Figure 3 Full-sample recession probabilities of the 32-state GDP models (2)



a) 32-State FTP3 Model



b) 32-State SSM3 Model

Figure 4 Full-sample recession probabilities of the 32-state GDP models (3)

Since Hamilton's model allows the two regimes to differ only in levels, not in their AR equations, it might be too restrictive for the purpose of capturing the business cycle dynamics. In Table 6, this restriction is relaxed. Instead, an 8-state AR model in the form of (3-9) is used.

Here the indicator function associated with each lag AR coefficient is defined as $I(S_t^*, 1) = S^*_{t-1}$, and $I(S_t^*, h) = S^*_{t-2}$, $h > 1$. In other words, the intercept and AR coefficients have 8 combinations, based on the values of the state vector $\{S^*_t, S^*_{t-1}, S^*_{t-2}\}$. Furthermore, an asymmetric AR dynamic is employed where one regime has 4 lags while the other has only 2.

This asymmetric and multi-state AR dynamic is suggested by Table 7, where FTP models based on (3-9) with symmetric lags are used for both regimes (the data is the same as the GDP data 1947q1 to 1997q1 used in Table 5). There were multiple local maxima found for each model in Table 7, but only the highest from each model is used to calculate the Akaike's Information Criterion (AIC). In a 4-state model, the intercept follows the current state and all lags follow the state at time t-1, whereas in an 8-state model, all lags greater than 1 follow the state at time t-2. There are two low points for the AIC on lags 2 and 4 in 2-state and 4-state models. In 8-state models, the AIC is lowest on lag 2. The two different low points for AIC suggest an asymmetric AR dynamic. The detection of the change in regime is best using 8 states with 2 lags, for it has a more reasonable cycle length of 32, and best matches the NBER reference cycle dates.

After experimenting with various specifications of the transition equations, the best turned out to be a pair of equations where the transition is a function of the sum of the last two observations. In particular,

$$p_t^{00} = 1/\{1+\exp(-\eta_0 - \beta[Y_{t-1}+Y_{t-2}])\},$$
$$p_t^{11} = 1/\{1+\exp(-\eta_1 + \beta[Y_{t-1}+Y_{t-2}])\}. \qquad (4\text{-}2)$$

In Table 6, the difference between the FTP model and the two SSM models is that in the FTP model, β=0. The likelihood ratio test (LRT) shows that the

SSM1 model is a substantial improvement over the FTP model. Furthermore, the bootstrap likelihood is lower too, indicating a more robust model. The difference between SSM1 and SSM2 in Table 6 is that in SSM2, $\eta_1=-\eta_0$, similar to a STAR model. The low likelihood ratio between SSM1 and SSM2 shows that the regime switching can be modelled as a symmetric switching model.

The new specification in Table 6 improved the FTP model only marginally. However, it improves the SSM model substantially, as is evident from the improvement in both the AIC and the bootstrap likelihood values.

The improvement in the detection of the change in regime is evident in Figure 5, especially for the 1990-1991 recession. Overall, the beginning and the end of a recession period became more pronounced, and more distinct.

In contrast to a linear AR model or an MS model, Table 6 shows that the feedback mechanism of the U.S. economy exists not only in the AR dynamics, but also in the regime switching dynamics. Although the feedback mechanism is not strong enough to sustain a limit cycle, it provided an additional mechanism for the persistence of the business cycle.

Table 6 8-State asymmetric AR GDP models[22]

Data	1947q1 to 1997q1		
Model	FTP	SSM1	SSM2
μ_0	NS	-0.371	-0.402
μ_1	0.740	0.880	0.874
C_0	$-1.343_{3.4}$	$-1.077_{4.1}$	$-1.234_{4.9}$
$\phi_{0,1}$	$-0.249_{1.1}$	$-0.672_{3.6}$	$-0.727_{3.7}$
$\phi_{0,2}$	$-0.611_{2.8}$	$-0.782_{4.6}$	$-0.841_{4.9}$
$\phi_{0,3}$	$0.672_{1.4}$	$-0.473_{2.9}$	$-0.528_{3.2}$
$\phi_{0,4}$	$0.503_{1.2}$	$0.028_{0.1}$	$-0.027_{0.1}$
C_1	$0.328_{3.8}$	$0.503_{4.6}$	$0.491_{4.6}$
$\phi_{1,1}$	$0.366_{5.0}$	$0.316_{4.2}$	$0.338_{4.6}$
$\phi_{1,2}$	$0.191_{2.6}$	$0.113_{1.4}$	$0.100_{1.2}$
η_0	$-13.02_{0.0}$	$-3.286_{2.1}$	$-1.084_{2.6}$
η_1	$3.295_{5.9}$	$-0.265_{0.0}$	[1.084]
β		$-3.410_{3.2}$	$-1.886_{4.5}$
σ^2	$0.549_{18.}$	$0.515_{18.}$	$0.505_{18.}$
p^{00}	0.000	0.000	0.019
p^{11}	0.964	0.993	0.981
AIC	2.563	2.513	2.515
L L	-1.226	-1.195	-1.201
LR/P-v[23]	11.79/0.0006		2.29/0.1302
BootL	-1.226	-1.198	-1.194
Std BtL	0.052	0.055	0.057
SP^0	0.034	0.125	0.168
$\approx Cyc^0$	1	2	2.5
$\approx Cyc$	22	14	19
Err_{NBER}	0.155	0.102	0.106
Est	1/4	1/4	1/4

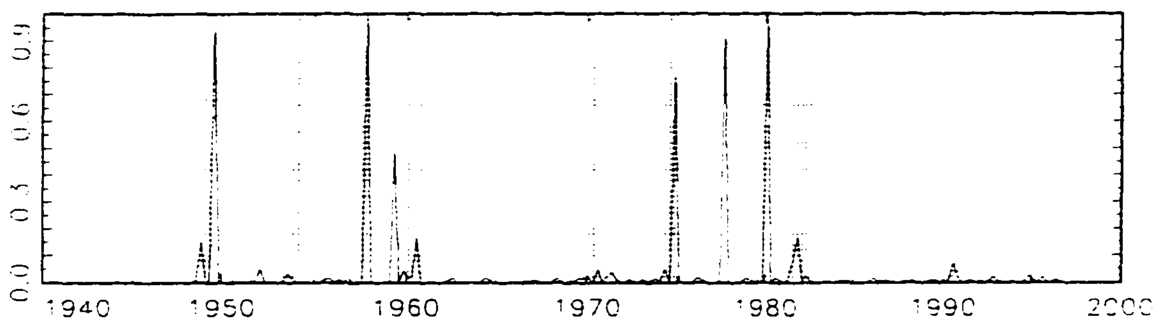
---

22 For an explanation of the symbols use in the table, see footnote 18 on page 42. Additionally, the entry for η_1 for the SSM2 model with a set of "[]" around it indicates that the entry is constraint, other than the sign, to be the same as η_0.

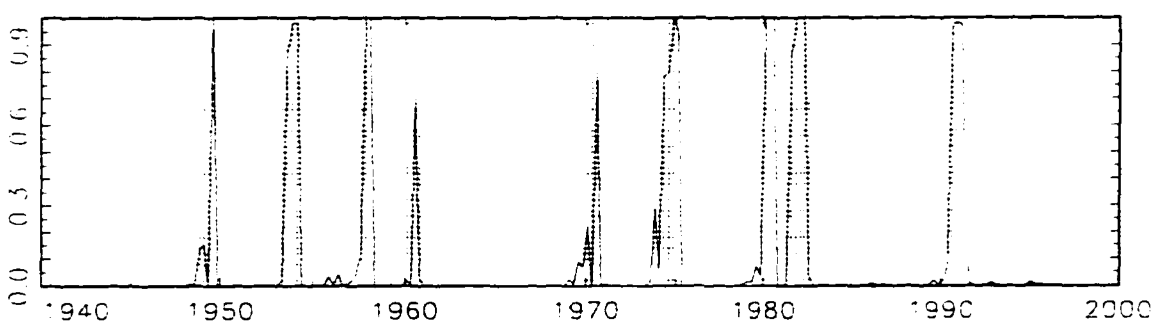
23 In both likelihood ratios the alternative is the SSM1 model.

Table 7 Symmetric AR FTP model exploration[24]

AR Lags	2 states AIC	Cyc Length	Err NBER	4 states AIC	Cyc Length	Err NBER	8 states AIC	Cyc Length	Err NBER
1	2.5682	4	.16	2.5731	3	.24			
2	2.5444	3	.17	2.5454	12	.16	2.5476	32	.14
3	2.5676	3	.17	2.5571	5	.30	2.5494	20	.15
4	2.5285	4	.30	2.5260	3	.31	2.5542	17	.15
5	2.5871	6	.15	2.5506	4	.31	2.6052	2	.29
6	2.6079	5	.15	2.5679	3	.29	2.5891	2	.30



a) 8-State FTP Model



b) 8-State SSM1 Model

Figure 5 Full-sample recession probabilities of the 8-state GDP models

4.3. *SSM Models of the IP Series*

It is not surprising that SSM models above have improved upon the basic Hamilton FTP model in view of Filardo's TVTP model using exogenous information. The natural question to ask, then, is can one improve upon Filardo's model

[24] See footnote 18 on page 42 for an explanation of the column headings.

using endogenous information. Since Filardo's model is based on the index of Industrial Production (IP) and the Composite Leading Indicator (CLI) series, I will address this question using a data set covering the same period and using the same log-difference transformation. The results are reported in Table 8. For all models in the table, the AR equation (3-60) is used instead of (3-9). The transition equations are

$$p_t^{00}=1/\{1+\exp(-[\eta_0+\beta_{y_{t-1}}Y_{t-1}+\beta_{y_{t-2}}Y_{t-2}+\beta_{0,z_{t-1}}Z_{t-1}+\beta_{0,z_{t-2}}Z_{t-2}])\},$$
$$p_t^{11}=1/\{1+\exp(-[\eta_1-\beta_{y_{t-1}}Y_{t-1}-\beta_{y_{t-2}}Y_{t-2}+\beta_{1,z_{t-1}}Z_{t-1}+\beta_{1,z_{t-2}}Z_{t-2}])\}, \qquad (4\text{-}3)$$

where $Y_t = 100(1-L)\mathrm{Log}(IP)$, and $Z_t = 100(1-L)\log(CLI)$.

Since Y_t is not variance-stationary, Filardo divided the pre-1960 sub-sample by the standard deviation of that sub-period. The same method is used for the data in Table 8.

The interpretation of the AR coefficients in Table 8 is thus problematic. We can, however, gauge the relative importance of the lagged values of the IP versus CLI series in regime switching. To this end, a third restricted model SSM2, is developed, where only the lagged IPs are entered into the transition function.

Given the possibility of multiple local maxima, Boldin (1996) suggests starting the restricted model using the unrestricted model's parameter estimates. Since the computational burden of the SSM1 model in Table 8 is very high[25], I could not explore multiple local maxima, and then compare the global maximum found for each model within a reasonable time. Thus, in Table 8, both the TVTP and SSM2 model estimations started from the local maximum found for the SSM1 model. The full sample smoothing of the recession prob-

---

[25] Due to a yet unresolved bug in my gradient programing code, when there are exogenous series in the model, I could only use the numerical gradient method for both the TVTP and SSM1 models in Table 8. The SSM1 model took over 5 hours to converge.

abilities for each model is illustrated in Figure 6.

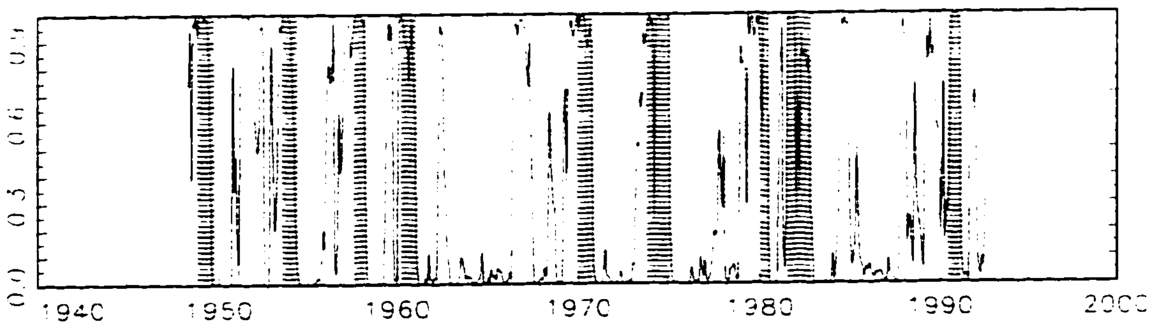
Drops in likelihoods in Table 8 are significant for both restrictive models. This result implies that both the IP and CLI series contain significant additional information regarding the regime shift, at least for the given local maximum.

Table 8 TVTP versus SSM models of IP[26]

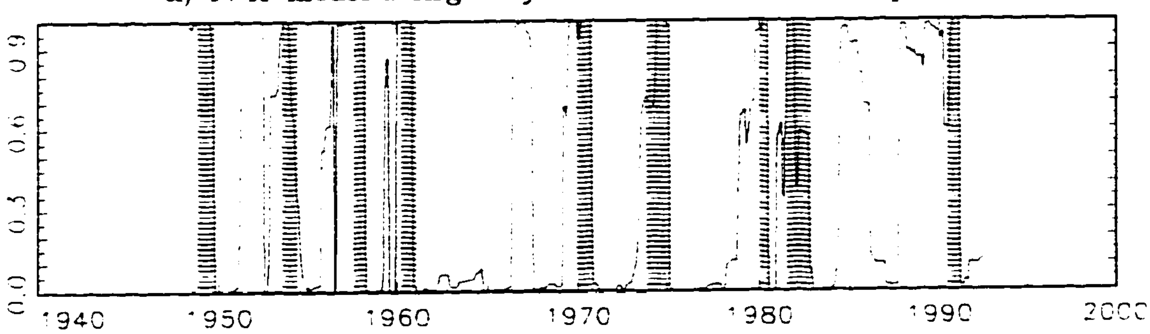
Model	TVTP	SSM1	SSM2
μ_0	$-0.559_{5.2}$	$-0.589_{4.7}$	$-0.092_{1.1}$
μ_1	$0.359_{5.0}$	$0.435_{4.4}$	$2.525_{9.5}$
ϕ_1	$0.237_{4.7}$	$0.260_{5.0}$	$0.422_{8.8}$
ϕ_2	$-0.016_{0.2}$	$0.071_{1.3}$	$0.104_{1.9}$
ϕ_3	$0.053_{1.2}$	$0.025_{0.6}$	$0.085_{1.7}$
ϕ_4	$0.035_{0.8}$	$0.029_{0.6}$	$-0.011_{0.2}$
η_0	$1.025_{1.7}$	$6.877_{2.8}$	$4.850_{8.5}$
$\beta_{0.z_{t-1}}$	$-1.036_{1.0}$	$-1.543_{1.0}$	
$\beta_{0.z_{t-2}}$	$-3.296_{1.9}$	$-6.102_{2.0}$	
$\beta_{y_{t-1}}$		$3.292_{2.3}$	$1.405_{3.9}$
$\beta_{y_{t-2}}$		$1.643_{1.8}$	$-0.233_{0.8}$
η_1	$2.987_{2.5}$	$6.842_{3.2}$	$2.661_{2.9}$
$\beta_{1.z_{t-1}}$	$3.261_{1.5}$	$9.195_{2.4}$	
$\beta_{1.z_{t-2}}$	$7.035_{1.6}$	$2.993_{1.3}$	
σ^2	$0.559_{29.}$	$0.554_{30.}$	$0.524_{30.}$
p^{00}	0.734	0.999	0.934
p^{11}	0.953	0.999	0.992
AIC	2.3980	2.3644	2.3837
L L	-1.175	-1.154	-1.171
LR/P-v[27]	21.88/0.0000		18.33/0.0011
BootL	-1.1737	-1.161	-1.177
Std BtL	0.046	0.038	0.056
SP^0	0.391	0.426	0.036
≈Cyc^0	9	14	2
≈Cyc	15	33	52
Err_{NBER}	0.19	0.24	0.21

26 For an explanation of the symbols use in the table, see footnote 18 on page 42. The CLI data set is from University of Michigan archive through the Internet. It is not the identical to Filardo's data set; for I could not replicate his results exactly.

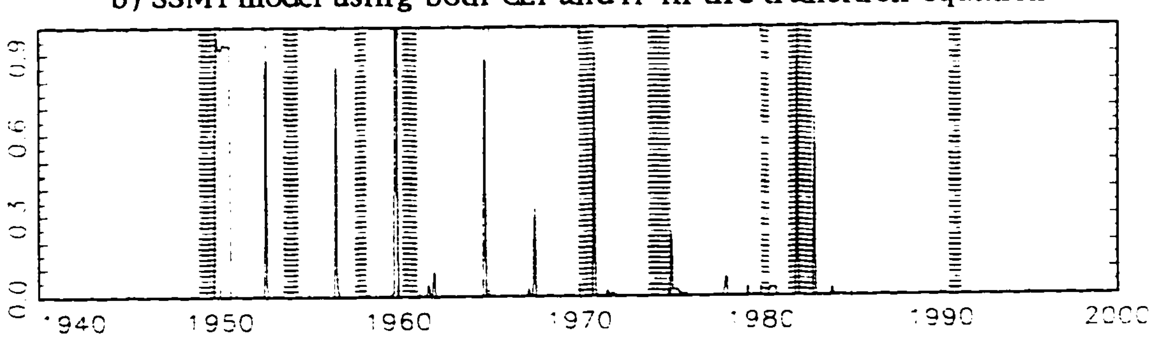
27 For both likelihood ratios, the alternative is the SSM1 model.



a) TVTP model using only CLI in the transition equation



b) SSM1 model using both CLI and IP in the transition equation



c) SSM2 model using only IP in the transition equation

Figure 6 Full-sample recession probabilities of the IP models

4.4. *An SSM Model of the Unemployment Rate*

Although the self-switching models have been shown to be significant in both output series, no estimations so far lead to limit cycle trajectories in their forecasts or simulations. The U.S. unemployment rate series is different. It is known to be highly nonlinear. It offers us an opportunity to model the business cycle as a limit cycle.

The U.S. unemployment rate series show signs of non-stationariety, as evident in the first panel of Figure 7. Until the current recovery, the rate tends to inch upward from one recession to the next. To achieve stationarity, I applied the first differencing to the rate series, as many authors have done (see, for example, Hansen 1997). Panel (b) of Figure 7 shows the rate series after differencing.

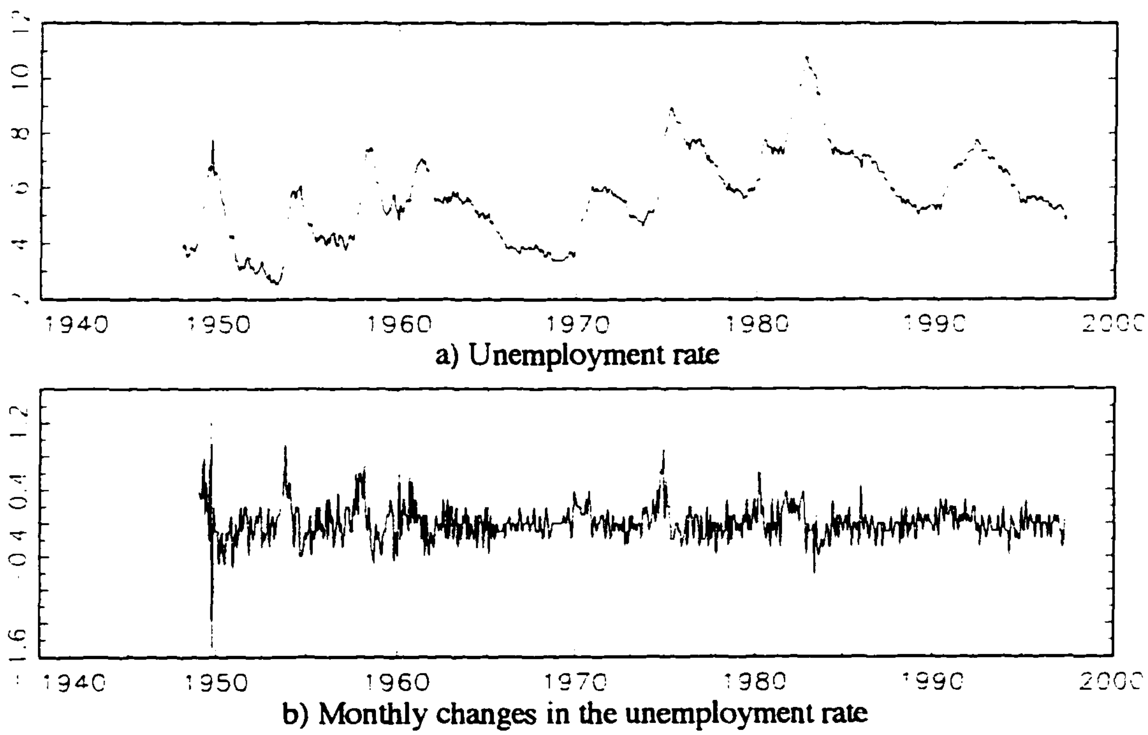


Figure 7 The U.S. unemployment rate and the monthly changes

To narrow down the range of choices for estimating an SSM model, I compared a series of 2-state symmetric FTP models with 2 to 20 lags to determine the optimal lags in the AR structure[28]. Mindful of Tong's (1990, 289) warning against overparameterization based on the lowest AIC, I opted for a lag length of 12 (a one year lag comparable to the 4 lags used for the quarterly data). The AIC value increased after lag 15 and then decreased drastically after that. Once the highest lag length is set, I compared another series of FTP models with asymmetric AR structures to determine the optimal lags for the second regime, while fixing the first AR regime to 12 lags. The lowest AIC for the second regime is at lag 5.

After experimenting with various self-switching parameters using the semi-normal mapping, and applying the initialization (3-41), Table 9 reports the SSM model that seems to be the best in terms of the AIC. The less interesting AR coefficients are reported in Table 10. In this model only the intercept and the coefficient for lag 5 in the transition equations are free, i.e.,:

$$p_t^{00}=\exp(-[\eta_0+\beta_0 Y_{t-5}]^2),$$
$$p_t^{11}=\exp(-[\eta_1+\beta_1 Y_{t-5}]^2). \quad (4\text{-}4)$$

Although there were 19 local maxima found with 50 random starting parameter values for the SSM model, the SSM1 and SSM2 estimates are the most interesting. To contrast the FTP and SSM models within the neighborhood of the same local maximum, the starting values for the two FTP model estimates were based on the SSM1 and SSM2 parameter values reported in Table 9, respectively.

The reason for employing the semi-normal mapping for the models in Table 9 is that the mapping is more stable in simulations, due to it's symmetry, especially in 2-state models. This is why the models in Table 9 are 2-state models. In contrast, some short series of strong shocks can send a model's trajec-

---

[28] The number of local maxima increased drastically in 4-state and 8-state models, making the selection more difficult.

tory based on the logistic mapping into explosive paths.

The possibility of explosive path using the logistic mapping seems to be a particular problem of SSM models. Since an FTP model has no self-switching component, an estimated FTP model is unlikely to experience this problem. Neither should a TAR model with a fix threshold. In numerous simulations, I have found no explosive paths in either FTP or TAR models.

The differences between the two panels in Figure 8 seemed small. They represent the full-sample (1948.03 to 1997.06) recession probabilities of the SSM1 and FTP1 models. The small difference between the two panels is confirmed in Table 9. The recession probabilities are 4.9% and 6.6%, respectively. The one-step forecast seems to be indistinguishable between the two panels in Figure 9 as well.

The dramatic difference does occur, though, in the long-run forecast between the two panels in Figure 9. The SSM1 model has a cyclical pattern with periodicity around 20 month, whereas the FTP1 model converges to a straight line. Forecasting out 500 periods, starting from 1990.05, the SSM1 estimate revealed a limit cycle trajectory, as evident in Figure 10.

The mean post-sample prediction error of the SSM model is inferior to the FTP model in both the estimates reported in Table 9, at least for the local maximum found and the sample forecast period used.

The SSM2 estimate in Table 9 behaves very differently from the SSM1 estimate. Whereas the full-sample recession probabilities from the SSM1 estimate missed many recessions, including the 1991 recession in Figure 8, the SSM2 estimate missed many recoveries in Figure 11.

Although there were many local maxima, the estimations generally fall into 4 categories, as shown in Table 11. Figure 8 of SSM1 estimate and Figure 11 of SSM2 estimate represent the two well behaved categories. The last two categories have the highest number of local maxima, and are the most sensitive to the choice of the starting parameter values. The full-sample smoothing of the recession probabilities from these estimates either show a very high frequency of business cycles, or they show extremely small probability of recession. Of the latter type, two of the estimates have unreasonably high likelihood values.

On the one hand, from Table 11, it seems that the probability of a random starting value leading to an estimate with a limit cycle outcome, similar to SSM1, is fairly high, especially if we exclude the last two categories of estimates. In fact, there were 13 different starting values converged to the same limit cycle estimate. On the other hand, based on the bootstrap likelihood, the SSM2 estimate seems to be a more robust estimate with a higher likelihood value.

Table 9 2-state models of the changes in the U.S. unemployment rate[29]

	SSM1	FTP1	SSM2	FTP2
μ_0	NS	NS	0.081	0.163
μ_1	-0.020	-0.024	-0.036	0.071
η_0	$-0.024_{0.1}$	$0.524_{4.3}$	$-0.025_{0.3}$	$0.180_{4.1}$
β_0	$3.217_{3.1}$		$-0.730_{4.2}$	
η_1	$-0.098_{2.7}$	$0.122_{3.3}$	$0.105_{2.6}$	$0.137_{3.8}$
β_1	$-0.297_{1.8}$		$0.652_{3.1}$	
σ_0^2	$0.076_{5.5}$	$0.117_{6.1}$	$0.082_{18.}$	$0.080_{19.}$
σ_1^2	0.032_{28}	$0.031_{23.}$	$0.019_{21.}$	$0.019_{21.}$
p^{00}	0.999	0.760	0.999	0.968
p^{11}	0.990	0.985	0.989	0.981
AIC	-0.3555	-0.3469	-0.351	-0.3418
L L	0.2284	0.2200	0.2262	0.2174
LR/P-v[30]	8.299/0.0158		8.694/0.0129	
BootL	0.245	0.240	0.248	0.265
Std BtL	0.068	0.071	0.065	0.076
SP^0	0.049	0.066	0.383	0.407
≈Cyc^0	3	4	28	41
≈Cyc	58	62	66	96
Err_{NBER}	0.134	0.113	0.204	0.217
$\xi_{1,86}$[31]	0.0232	0.0231	0.0236	0.0225
$\xi_{86,1}$	0.0807	0.0697	0.0303	0.0252

[29] For an explanation of the symbols use in the table, see footnote 18 on page 42. Data source: *Civilian Unemployment Rate (SA, %)* 1948.01 to 1997.06 from St. Louis Fed's FRED. The estimation used data from 1949.03 to 1990.04, a total of 494 data points. The first 12 data points were lost due to the lags used in the model. The data from 1990.05 to 1997.06, a 15% of the sample was reserved for forecast comparison purpose.

[30] For both likelihood rations, the nulls are the FTP models.

[31] $\xi_{1,86}$ and $\xi_{86,1}$ are the mean one-step ahead forecast error and the long horizon forecast error, respectively, based on parameter estimation of data up to 1990.04.

Table 10 AR coefficients for the models in Table 9.

	SSM1	FTP1	SSM2	FTP2
C_0	$-0.016_{1.8}$	$-0.020_{2.1}$	$-0.051_{5.4}$	$-0.049_{4.7}$
$\phi_{0.1}$	$0.034_{0.8}$	$0.022_{0.5}$	$-0.197_{3.5}$	$-0.195_{3.1}$
$\phi_{0.2}$	$0.135_{3.4}$	$0.120_{3.0}$	$-0.045_{0.8}$	$-0.051_{0.9}$
$\phi_{0.3}$	$0.106_{2.7}$	$0.085_{2.1}$	$-0.103_{1.6}$	$-0.131_{2.0}$
$\phi_{0.4}$	$0.102_{2.5}$	$0.104_{2.6}$	$0.056_{1.2}$	$0.087_{1.5}$
$\phi_{0.5}$	$0.074_{1.8}$	$0.072_{1.8}$	$0.086_{1.7}$	$0.080_{1.5}$
$\phi_{0.6}$	$-0.011_{0.3}$	$-0.001_{0.1}$	$0.020_{0.4}$	$0.036_{0.6}$
$\phi_{0.7}$	$0.019_{0.5}$	$-0.020_{0.5}$	$-0.014_{0.3}$	$-0.000_{0.0}$
$\phi_{0.8}$	$0.040_{1.0}$	$0.050_{1.3}$	$0.150_{2.9}$	$0.149_{2.8}$
$\phi_{0.9}$	$0.006_{0.1}$	$0.011_{0.3}$	$-0.036_{0.7}$	$-0.034_{0.6}$
$\phi_{0.10}$	$0.102_{2.6}$	$-0.103_{2.6}$	$-0.182_{3.8}$	$-0.190_{3.9}$
$\phi_{0.11}$	$0.024_{0.6}$	$0.010_{0.3}$	$0.030_{0.6}$	$0.041_{0.8}$
$\phi_{0.12}$	$-0.162_{4.2}$	$-0.166_{4.3}$	$-0.195_{4.2}$	$-0.196_{4.2}$
C_1	$0.309_{2.2}$	$0.384_{2.1}$	$0.028_{1.2}$	$0.020_{0.9}$
$\phi_{1.1}$	$-0.673_{3.2}$	$-0.529_{1.9}$	$0.021_{0.3}$	$0.043_{0.6}$
$\phi_{1.2}$	$1.321_{4.3}$	$0.795_{2.4}$	$0.330_{4.1}$	$0.321_{4.4}$
$\phi_{1.3}$	$0.517_{1.6}$	$0.306_{0.9}$	$0.232_{2.8}$	$0.217_{2.8}$
$\phi_{1.4}$	$-0.697_{1.9}$	$-0.791_{1.9}$	$-0.036_{0.4}$	$-0.012_{0.2}$
$\phi_{1.5}$	$0.586_{1.2}$	$0.245_{0.6}$	$0.115_{1.3}$	$0.060_{0.8}$

Table 11 Estimation Characteristics

No. of Starting Values	No. of Local Maxima	Full-Sample Graph Looks Like	SP^{0}	Skeleton
15	2	Figure 8	0.05	limit cycle
3	1	Figure 8	0.10	limit point
1	1	Between Figures 8 & 11	0.25	limit point
5	2	Figure 11	0.38	limit point
15	6	Numerous recessions	0.14	limit point
11	7	No recessions	<0.04	limit point

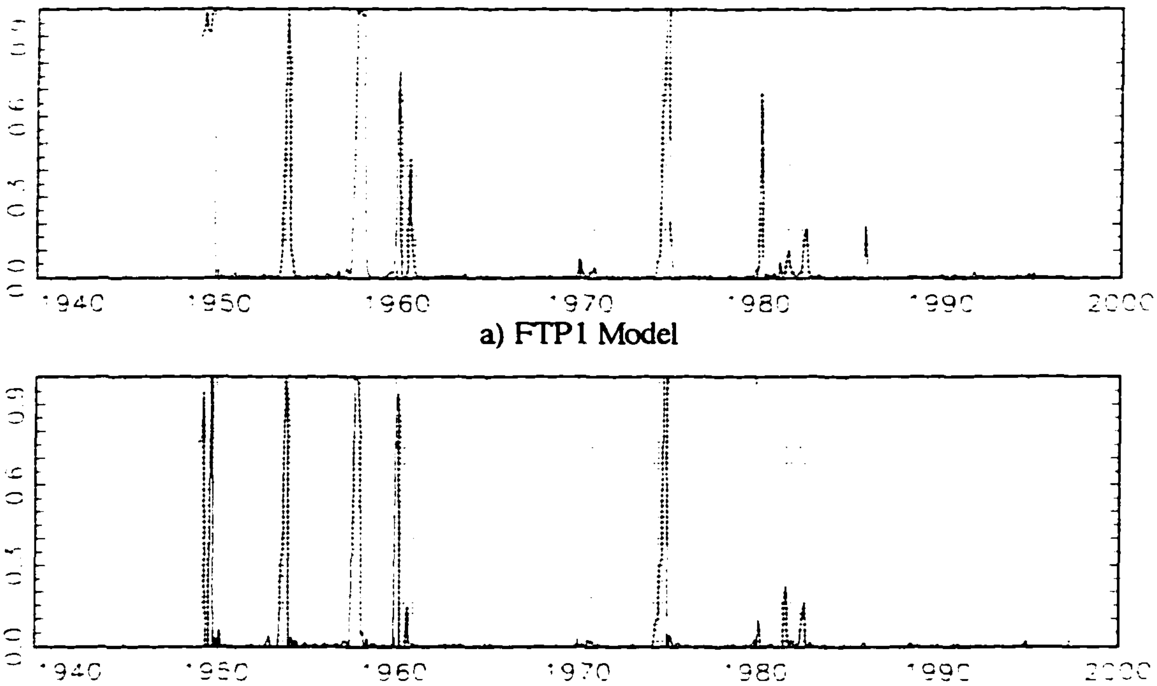


Figure 8 Full-sample recession probabilities of the FTP1 and SSM1 models.

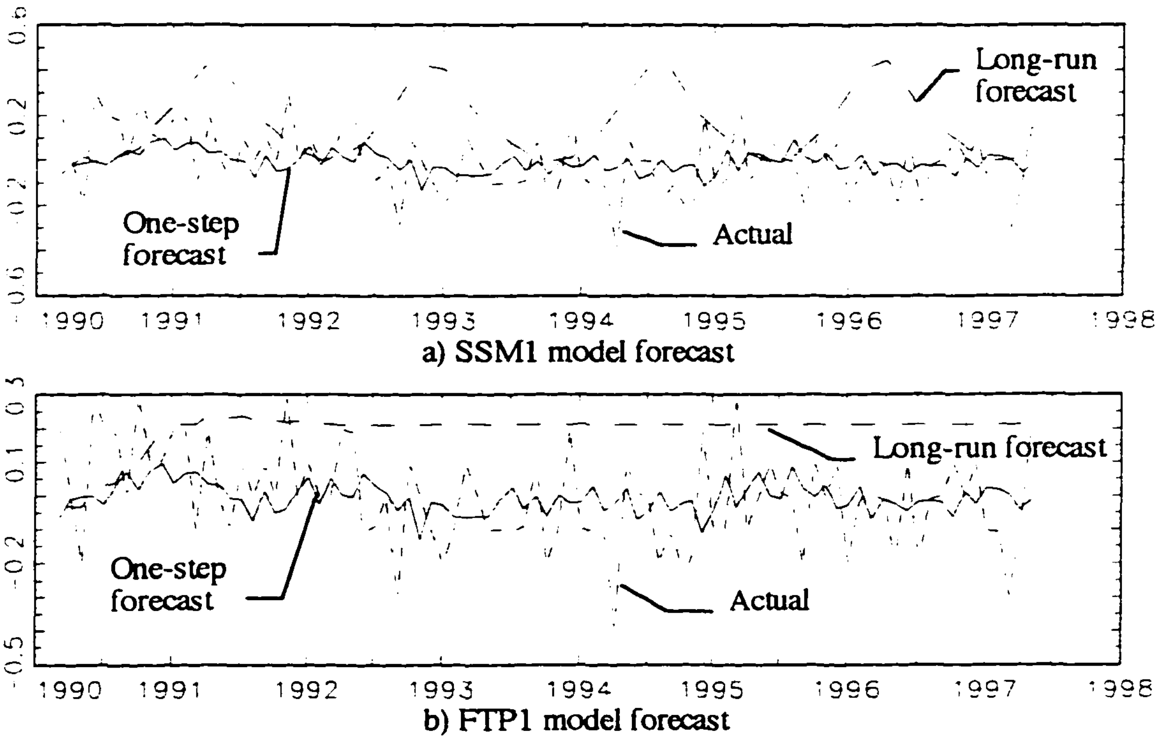


Figure 9 Forecasting based on the SSM1 and FTP1 models

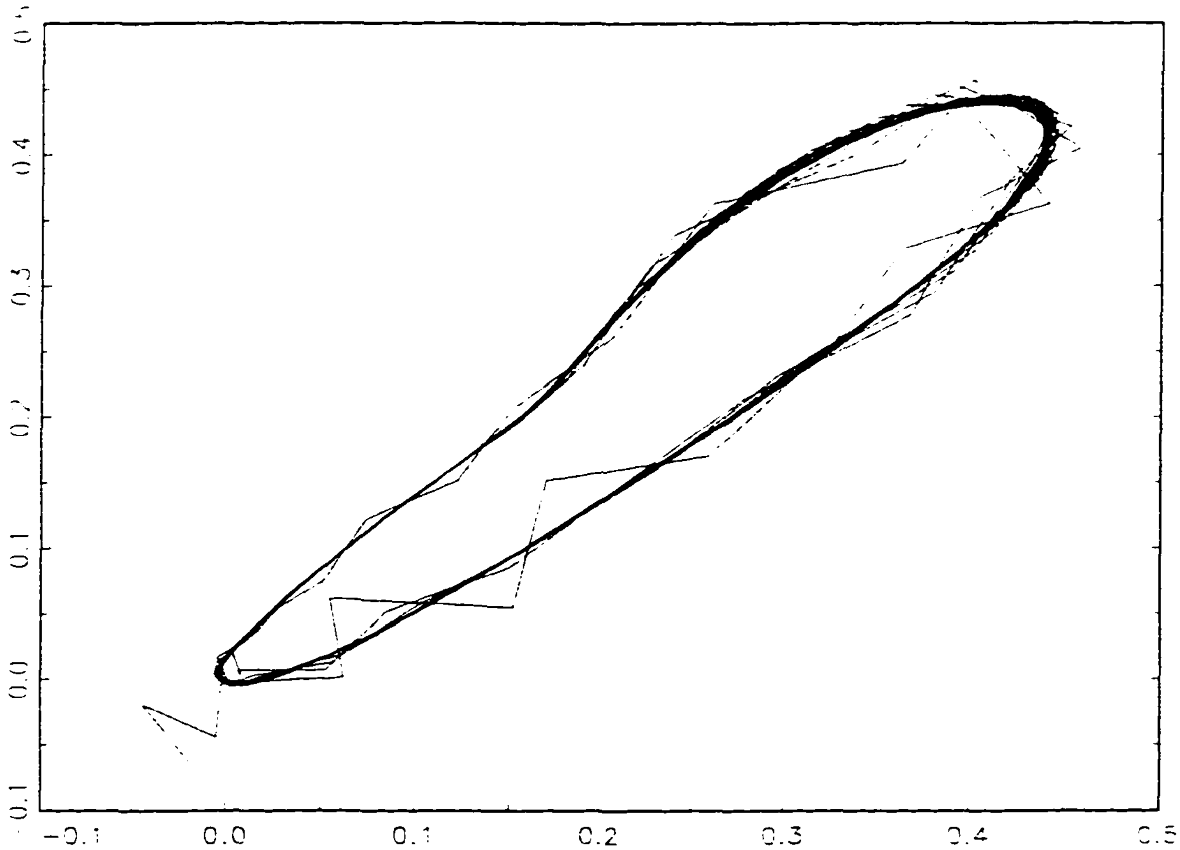
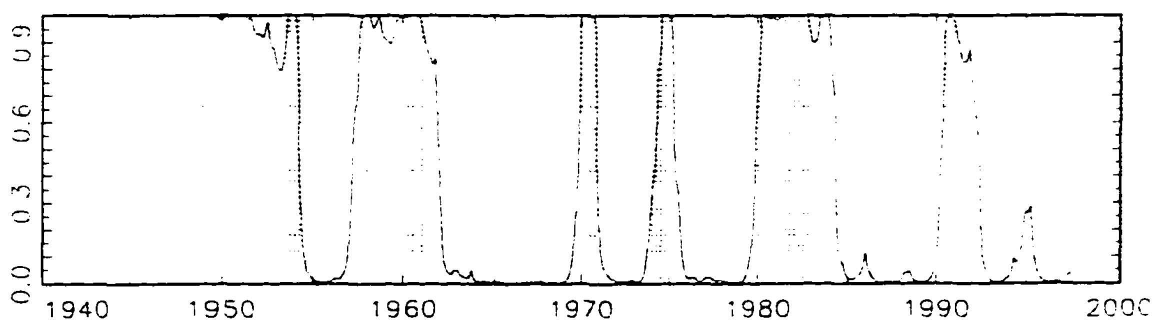
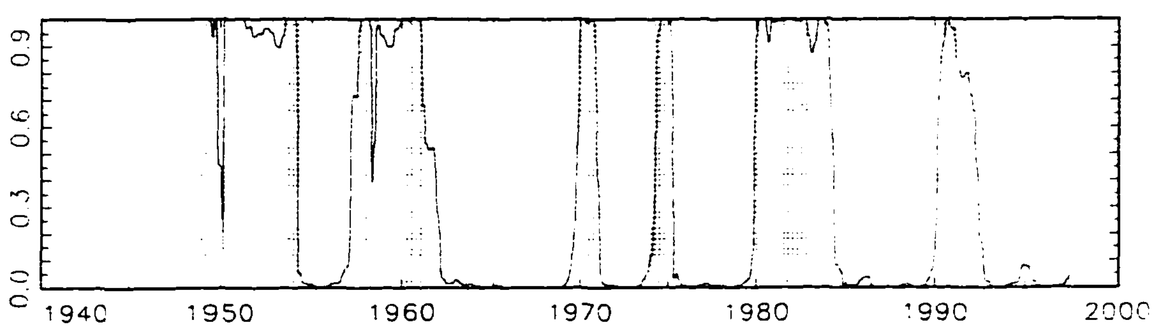


Figure 10 Limit cycle trajectory (phase diagram) of the SSM1 forecast



a) FTP2 model



b) SSM2 model

Figure 11 Full-sample recession probabilities of the SSM2 and FTP2 models.

4.5. Conclusion

SSM models above have shown that self-switching Markov is a successful strategy for embedding nonlinear feedback. It can generate complex dynamics where a Markov switching model fails to.

The SSM models also provide empirical evidence in a variety of settings that endogenous information is significant in predicting regime switching. Not only are the coefficients for the lagged endogenous variables significant, but likelihood ratio test results also show the endogenous information to be important in all models.

The goal of capturing the "natural" rhythm of a market economy is achieved when an SSM model is applied to the monthly changes in the U.S. unemployment rate. Many estimates exhibit stable limit cycles of diverse periodicity in forecasts or simulations. These estimates show that an endogenous cycle (limit cycle) model is consistent with the data.

The advantage of SSM models over TAR models that rely on discrete thresholds and delay factors is also evident from results reported in the Appendix. The SSM approach improves likelihoods significantly in these models as well.

The skeleton of an SSM model behaves very much like a TAR model. It can have multiple paths of convergences, including a limit cycle path. The impulse-response function is thus path dependent. The persistence, short of a limit cycle, can still be much greater than in the linear case.

There are also many challenges facing an SSM model builder. The choice of the transition function affects the model behavior in simulation and forecasting. While the logistic mapping is intuitive and its starting parameter vector can be found consistently, it can easily lead to a globally unstable outcome in simulation. The Semi-normal mapping, by contrast, is more globally stable, and yet no obvious method exists for finding a consistent initial parameter vector.

Another challenge facing an SSM model builder is the tendency for the model to converge, based on the starting parameter values, to a large number of local maxima. Many of which seem spurious.

The performance of the model in terms of the mean post-sample prediction error is also mixed. The SSM model is inferrer to FTP model in the U.S. unemployment rate series of Table 9, but superior to TAR model in the sunspots series of Table 13.

The finding that the endogenous switching parameters are statistically significant lends empirical support to endogenous business cycles theories, despite some of the difficulties encountered in estimating the models. The challenge of modeling a business cycle as a low frequency (five to ten years) limit cycle remains for future research.

5. Appendix: SSM Models on Benchmark Time Series

To evaluate the performance of SSM models, I contrasted them with the TAR models reported in Tong (1990) using a series of benchmark tests. In particular, I contrasted them with Tong's models based on the Canadian Lynx yearly trapping series and the sunspot numbers.

5.1. The Canadian Lynx Yearly Trapping Series

Tong's model (1990, eq. 7.7, pp. 387) is reproduced[32] in Table 12 under the heading "Tong". A TAR model is replicated by imposing constraints in the transition equation. One of the two local maxima found is reported in the table under the heading "TAR". (For reasons unknown, however, I was unable to replicate Tong's coefficients exactly. Only the underlined digits are the same.)

The data is from Tong (1990, 470), and the same log transformation is used as Tong.

Two STAR-like[33] estimates (among 4 local maxima found) that improved the likelihood significantly compared to the TAR model are also reported in the table under the headings "STAR-Like1" and "STAR-Like2". The first estimate

---

[32] For an explanation of how to read the table, see footnote 18 on page 42.
Additionally, an entry with "<>" around it means that it is fixed to the value during estimation. Furthermore,
An entry with a "[]" around it indicates that the entry is constrained, except for the sign, to be the same as a corresponding parameter entry, or a group of parameter entries.
"FP(τ)" is the time in τ-periods it takes for the series to reach the long run equilibrium point, starting from the T-th (i.e., the most current) observation during forecast simulation. A value of ∞ implies that it is a limit cycle.

[33] It is STAR-like in the sense that the transition is smoother than a TAR model. However, since a STAR model has no filtering and no Markov-switching dynamics, SSM with constraints in the transition equation is only an approximation to it.

displays limit cycle behavior when forecasting out 500 periods. The second estimate oscillates for a long time (approximately 200 periods) before it reaches its limit point. Another noteworthy feature of the two estimates is that although the first estimate has a set of coefficients very close to Tong's TAR model (within one standard error), the second has a much higher likelihood value.

Two estimates out of 5 local maxima found in a 2-state self-swithing model are reported in the table under the headings "SSM1" and "SSM2". Similar to the STAR model estimates, the first estimate displays limit cycle when forecasting out 500 periods, and the second estimate will take a fairly long time (approximately 80 periods) to reach its limit point.

An attempt to model SSM using the semi-normal mapping was not as successful as the logistic mapping in terms of the AIC. The results are reported in the table under the heading "SSM3". It is a 2-state model. In terms of (3-9), $I(S_t^*, h) = S_t^*$ for all h. The low bootstrap likelihood indicates that it is not a robust model.

Since the SSM2 model nests a TAR model as a constrained model, the likelihood ratio test (LRT) is a valid test. It yields a value of 11.17 (2*[0.3241-0.2719]*107), and is greater than the 5% $\chi^2_{(3)}$ of 7.81. The LRT between the STAR-like model and the TAR model is 4.109 (=2*[0.2911-0.2719]) — greater than the 5% $\chi^2_{(1)}$ of 3.84. A LRT between an SSM model and a STAR-like model is invalid, since each has a different set of constraints. Thus one is not a nested model of the other. However, using the AIC, it appears that the best SSM model is superior to the best STAR-Like model in Table 12.

Table 12 Models of the Lynx series

	Tong	TAR	STAR-Like1	STAR-Like2	SSM1	SSM2	SSM3
μ_0	3.81	3.950	3.952	NS	NS	3.251	NS
μ_1	NS	NS	NS	4.109	4.454	NS	4.473
C_0	0.546	$0.516_{1.9}$	$0.518_{2.0}$	$1.182_{3.3}$	$1.145_{3.5}$	$0.492_{2.1}$	$1.154_{3.6}$
$\phi_{0,1}$	1.032	$1.037_{11.}$	$1.032_{11.}$	$1.336_{11.}$	$1.349_{11.}$	$1.133_{13.}$	$1.347_{11.}$
$\phi_{0,2}$	-0.173	$-0.163_{1.1}$	$-0.153_{1.0}$	$-0.836_{3.3}$	$-0.729_{3.0}$	$-0.245_{1.9}$	$-0.774_{3.3}$
$\phi_{0,3}$	0.171	$0.169_{1.1}$	$0.157_{1.1}$	$0.463_{1.5}$	$0.288_{1.0}$	$0.001_{0.0}$	$0.348_{1.3}$
$\phi_{0,4}$	-0.431	$-0.437_{2.8}$	$-0.424_{2.8}$	$-0.645_{3.2}$	$-0.624_{3.2}$	$-0.288_{2.1}$	$-0.625_{3.3}$
$\phi_{0,5}$	0.332	$0.347_{2.1}$	$0.341_{2.1}$	$0.121_{0.6}$	$0.231_{1.2}$	$0.232_{1.7}$	$0.192_{1.1}$
$\phi_{0,6}$	-0.284	$-0.293_{1.8}$	$-0.293_{1.8}$	$-0.111_{0.6}$	$-0.119_{0.7}$	$-.0194_{1.6}$	$-0.112_{0.6}$
$\phi_{0,7}$	0.210	$0.209_{2.1}$	$0.209_{2.1}$	$0.254_{2.0}$	$0.203_{1.8}$	$0.210_{2.7}$	$0.219_{2.0}$
η_0		<3.116>	<3.116>	<3.116>	$-8.202_{3.0}$	$9.524_{2.8}$	$3.631_{3.3}$
$\beta_{0,1}$					$0.395_{0.4}$	$4.344_{3.1}$	
$\beta_{0,2}$		<-1>	<-1>	<1>	$2.586_{2.3}$	$-6.751_{3.4}$	$-0.955_{2.9}$
σ_0^2	0.0259	$0.0256_{9.5}$	$0.0247_{11.}$	$0.028_{8.9}$	$0.0313_{9.8}$	$0.0237_{12.}$	$0.0285_{9.5}$
C_1	2.632	$2.359_{3.8}$	$1.745_{2.5}$	$0.557_{4.9}$	$0.502_{4.3}$	$3.637_{10.}$	$0.495_{4.6}$
$\phi_{1,1}$	1.492	$1.514_{15.}$	$1.571_{14.}$	$1.129_{20.}$	$1.117_{19.}$	$1.251_{11.}$	$1.120_{20.}$
$\phi_{1,2}$	-1.324	$-1.264_{6.6}$	$-1.147_{5.7}$	$-0.265_{4.3}$	$-0.230_{3.9}$	$-1.433_{7.9}$	$-0.231_{4.0}$
η_1		<3.116>	<3.116>	<3.116>	[8.202]	[-9.524]	$-1.000_{0.8}$
$\beta_{1,1}$					[-0.395]	[-4.344]	
$\beta_{1,2}$		<1>	<1>	<-1>	[-2.586]	[6.751]	$0.691_{1.4}$
σ_1^2	0.0505	$0.050_{9.5}$	$0.046_{4.3}$	$0.015_{8.3}$	$0.014_{7.7}$	$0.013_{5.6}$	$0.0134_{8.0}$
δ		$<10^{-5}>$	$0.085_{5.4}$	$0.400_{2.7}$	<1>	<1>	<1>
p^{00}		1	0.927	0.359	0.609	0.927	0.475
p^{11}		0	0.073	0.641	0.391	0.073	0.365
AIC		-0.3008	-0.3012	-0.3205	-0.2869	-0.3492	-0.3445
L L		0.2719	0.2815	0.2911	0.2930	0.3241	0.3311
LR/P-v[34]			4.109/0.0426		11.17/0.0108		

[34] For both likelihood rations, the null is the TAR model.

Table 12 continued

Tong	TAR	STAR-Like1	STAR-Like2	SSM1	SSM2	SSM3
BootL	0.242	0.269	0.272	0.333	0.321	0.252
BtL V	0.109	0.075	0.162	0.099	0.113	0.195
SP^0	0.58	0.61	0.47	0.58	0.79	0.56
$\approx Cyc^0$	5.2	5.3	3.5	3.3	6.4	3.1
$\approx$Cyc	9.2	9.2	6	6.6	8.4	5.7
Est	1/2	2/4	1/4	3/5	1/5	1/5
$FP(\tau)$	∞	∞	>200	∞	>80	∞

5.2. The Sunspot Numbers

Tong's model using sunspot numbers (1990, eq. 7.15, 421) is reproduced[35] in Table 13 under the heading "Tong". A TAR model is replicated in the table under the heading "TAR", just as in the Lynx series. (This time, I was only unable to reproduce the identical coefficients for the first AR equation.)

The data is also from Tong (1990, 471), and is transformed by Tong's (1990, 420) equation: $Y_t=2\{(1+X_t)^{1/2}-1\}$, where X_t is the annual mean of daily sunspot observations.

Two STAR-like[36] estimates (among 3 local maxima found) are also reported in the table under the headings "STAR-Like1" and "STAR-Like2". The likelihood ratio between "STAR-Like1" and "TAR" is significant. The former also has a lower forecast error. An interesting feature of "STAR-Like2" is that it is in reality a TAR model, with a similar set of coefficients and the same limiting behavior as those in "TAR".

---

[35] For an explanation of the symbols used in the table, see footnote 18 on page 42, and footnote 32 on page 66. Additionally,

"$\xi_{8,1}$" is the mean post-sample prediction error of the transformed data from 1980 to 1987.

[36] See footnote 33 on page 82.

One of 3 SSM estimates is reported in the table under the heading "SSM2". The likelihood ratio between "SSM2" and "TAR" is significant. The SSM model also has the lowest forecast error reported in the table. One of the most striking features of "SSM2", however, is the transition equation coefficient estimates η_0 and $\beta_{0,8}$. The first came very close to Tong's choice of 11.93 for the threshold. The second came very close to -1, the value of the delay factor is set to implicitly. Tong used an auxiliary procedure (grid search) to find the threshold value. In "SSM2" it is estimated directly from the data, showing the advantage of the SSM model.

Table 13 Models of the Sunspot Numbers

	Tong	TAR	STAR-Like1	STAR-Like2	SSM2
μ_0	8.591	8.147	7.355	8.147	7.538
μ_1	12.24	12.25	12.95	12.25	12.857
C_0	1.89	$1.742_{3.6}$	$2.021_{4.5}$	$1.742_{3.6}$	$1.943_{4.6}$
$\phi_{0,1}$	0.86	$0.870_{11.}$	$0.856_{12.}$	$0.870_{11.}$	$0.859_{12.}$
$\phi_{0,2}$	0.08	$0.075_{0.7}$	$0.014_{0.1}$	$0.075_{0.7}$	$0.025_{0.3}$
$\phi_{0,3}$	-0.32	$-0.329_{3.2}$	$-0.265_{2.7}$	$-0.329_{3.2}$	$-0.276_{3.0}$
$\phi_{0,4}$	0.16	$0.164_{1.7}$	$0.200_{2.2}$	$0.164_{1.7}$	$0.196_{2.3}$
$\phi_{0,5}$	-0.21	$-0.189_{2.1}$	$-0.247_{2.8}$	$-0.189_{2.1}$	$-0.230_{2.7}$
$\phi_{0,6}$	-0.00	$-0.024_{0.3}$	$-0.018_{0.2}$	$-0.024_{0.3}$	$-0.035_{0.4}$
$\phi_{0,7}$	0.19	$0.212_{2.3}$	$0.218_{2.5}$	$0.212_{2.3}$	$0.231_{2.8}$
$\phi_{0,8}$	-0.28	$-0.291_{3.1}$	$-0.340_{4.0}$	$-0.291_{3.1}$	$-0..327_{4.0}$
$\phi_{0,9}$	0.20	$0.191_{2.0}$	$0.162_{1.8}$	$0.191_{2.0}$	$0.168_{2.0}$
$\phi_{0,10}$	0.10	$0.106_{1.5}$	$0.144_{2.3}$	$0.106_{1.5}$	$0.132_{2.2}$
η_0		<11.93>	<11.93>	<11.93>	$11.156_{2.9}$
$\beta_{0,8}$		<-1>	<-1>	<-1>	$-0.849_{2.8}$
σ_0^2	1.946	$1.99_{18.}$	$1.546_{14.}$	$1.988_{18.}$	$1.591_{17.}$
C_1	4.53	$4.532_{9.6}$	$4.787_{9.1}$	$4.532_{9.7}$	$5.280_{8.4}$
$\phi_{1,1}$	1.41	$1.411_{22.}$	$1.410_{21.}$	$1.411_{22.}$	$1.402_{20.}$
$\phi_{1,2}$	-0.78	$-0.781_{10.}$	$-0.779_{9.6}$	$-0.781_{10.}$	$-0.813_{8.8}$
η_1		<-11.93>	<-11.93>	<-11.93>	[-11.156]
$\beta_{1,8}$		<1>	<1>	<1>	[0.849]
σ_1^2	6.302	$6.278_{15.}$	$6.045_{14.}$	$6.278_{15.}$	$6.086_{13.}$
δ		<10^{-4}>	$1.13_{0.3}$	$0.0003_{0.1}$	<1>
p^{∞}		1	0.73	1	0.88
AIC		4.112	4.065	4.119	4.055
L L		-1.9968	-1.9697	-1.9967	-1.9606
LR/P-v[37]			14.63/0.0000		19.54/0.0000
BootL		-2.01	-1.978	-1.991	-1.971
Std BtL		0.053	0.047	0.048	0.048

[37] The null hypotheses for both likelihood ratios are the nested TAR model.

Table 13: Continued

	Tong	TAR	STAR-Like1	STAR-Like2	SSM2
$\xi_{8.1}$		6.033	5.49	6.034	4.68
SP^o		0.59	0.59	0.59	0.65
$\approx Cyc^0$		6.9	6.4	6.9	7.7
$\approx Cyc$		11.7	10.7	11.7	11.7
Est		1/2	1/3	2/3	1/3
FP(τ)		∞	>200	∞	>100

6. References

Adelman, I. and F. L. Adelman 1959. The dynamic properties of the Klein-Goldberger model. *Econometrica* 596-625.

Bak, Per, Kan Chen, José Scheinkman, and Michael Woodford 1993. Aggregate fluctuations from independent sectorial shocks: Self-organized criticality in a model of production and inventory dynamics, *NBER* #4241 & *Ricerche Econ.* 47(1):3-30.

Balducci, Renato, G. Candela, and G. Ricci 1984. A generalization of R. Goodwin's model with rational behavior of economic agents, pp. 47-66. In R. M. Goodwin et al. (ed.), *Nonlinear Models of Fluctuating Growth*, New York: Springer-Verlag.

Beaudry, Paul and Gary Koop 1993. Do recessions permanently change output? *Journal of Monetary Economics* 31:149-63.

Benhabib, Jess, ed. 1992. *Cycles and Chaos in Economic Equilibrium*, Princeton: Princeton University Press.

Blatt, John M. 1978. On the econometric approach to business-cycles. *Oxford Econ. Pap.* 30: 292-300.

Blatt, John M. 1980. On the Frisch model of business cycles. *Oxford Econ. Pap.* 32: 467-479.

Blatt, John M. 1983. *Dynamic economic systems: A post-Keynesian approach.* Armonk, N.Y. M.E. Sharpe.

Boldrin, Michele 1994. Competitive markets and endogenous cycles: An evaluation. In Semmler (ed.), *Business Cycles: Theory and Empirical Methods*, Boston: Kluwer Academic Publishers.

Boldin, Michael 1996. A Check on the robustness of Hamilton's Markov switching model approach to the economic analysis of the business cycle. Studies in Nonlinear Dynamics and Economics 1(1): 35-46.

Brook, William A. and Chera L. Sayers 1992. Is the business cycle characterized by deterministic chaos? In Benhabib (ed.), *Cycles and Chaos in Economic Equilibrium*, Princeton: Princeton University Press.

Brunner, Allan D. 1992. Conditional asymmetries in real GNP: A seminonparametric approach. *Journal of Business & Economic Statistics* vol. 10, no. 1.

Caballero, Ricardo J., and Mohamad L. Hammour 1994. The cleansing effect of recessions. *American Economic Review* 84(5):1350-1368.

Cao, C. Q, and Ruey S. Tsay 1993. Nonlinear time-series analysis of stock volatilities. In Pesaran and Potter (eds.), *Nonlinear Dynamics, Chaos and Econometrics.* New York, John Wiley & Sons.

Cheng, Leonard K., and Elias Dinopoulos 1992. Endogenous growth in open economies. *American Economic Review* 82(2): 409-414.

Clarke, Simon 1994. *Marx's Theory of Crisis.* New York, St. Martin's Press.

Cooley, Thomas F., ed. 1995. *Frontiers of Business Cycle Research*, Princeton: Princeton University Press.

Danthine, Jean-Pierre and John B. Donaldson 1995. Non-Walrasian economics. In Cooley, T. F. (ed.), *Frontiers of Business Cycle Research*, Princeton: Princeton University Press.

Davidson, Paul 1991. Is probability theory relevant for uncertainty? A post Keynesian perspective. *Journal of Economic Perspectives* 5(1):129-143.

Day, Richard H. 1982. Irregular growth cycles. *American Economic Review* 72(3): 406-14.

De Gooijer, Jan G. and Kuldeep Kumar 1992. Some recent developments in non-linear time series modeling, testing, and forecasting. *International Journal of Forecasting* 8:135-156.

DeLong, J. Bradford. and Lawrence H. Summers 1986. Are business cycles symmetrical? In R. J. Gordon (Ed.), *The American Business Cycle*, National Bureau of Economic Research, Chicago: University of Chicago Press.

Deneckere, Raymond and Kenneth Judd 1992. Cyclical and chaotic behavior in a dynamic equilibrium model, with implications for fiscal policy. In Benhabib (ed.), *Cycles and Chaos in Economic Equilibrium*, Princeton: Princeton University Press.

Devereux, Michael B., Allen C. Head, and Beverly J. Lapham 1996. Aggregate fluctuations with increasing returns to specialization and scale. *Journal of Economic Dynamics and Control* 20:627-656.

Diebold, Francis X., Joon-Haeng Lee, and Gretchen C. Weinbach 1994. Regime switching with time-varying transition probabilities. In Hargreaves, C. (ed.), *Nonstationary Time Series Analysis and Cointegration*. Oxford: Oxford University Press.

Filardo, Andrew J. 1994. Business cycle phases and their transitional Dynamics. *Journal of Business and Economic Statistics* 12(3):299-308.

Flaschel, Peter, Reiner Franke, and Willi Semmler 1997. *Dynamic Macroeconomics: Instability, Fluctuations and Growth in Monetary Economies*. Cambridge, Mass.: MIT Press.

Foley, Duncan 1992. A contribution to the theory of business cycles. *The Quarterly Journal of Economics* 1070-88.

Goodwin, Richard M. 1951. The nonlinear accelerator and the persistence of business cycles. *Econometrica* 19:1-17.

Goodwin, Richard M. 1967. A growth cycle. In C. H. Feinstein (ed.), *Socialism, Capitalism, and Economic Growth*, Cambridge: Cambridge University Press. Reprinted in E. K. Hunt and J. G. Schwartz (eds.), 1972. *A Critique of Economic Theory*, Harmandsworth (UK): Penguin Books. And in Goodwin, R. M. 1982. *Essays in Economic Dynamics*. London: Macmillan.

Goodwin, Richard M. 1990. *Chaotic Economic Dynamics*, Oxford, Oxford University Press.

Grandmont, Jean-Michel 1985. On endogenous competitive cycles. *Econometrica* 53:995-1045.

Grandmont, Jean-Michel 1986. Stabilizing competitive cycles. *Journal of Economic Theory* 40:57-76. Reprint in Grandmont, J. (ed.), 1987. *Nonlinear Economic Dynamics*. New York. Academic Press.

Grandmont, Jean-Michel 1987. *Nonlinear Economic Dynamics*. New York. Academic Press.

Hamilton, James D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357-384.

Hamilton, James D. 1990. Analysis of time series subject to changes in regime. *Journal of Econometrics* 45:39-70.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton, Princeton University Press.

Hansen, Bruce E. 1997. Inference in TAR models. *Studies in Nonlinear Dynamics and Econometrics* 2(1):1-14.

Hansen, Gary D. and Prescott, Edward C. 1993. Did Technology Shocks Cause the 1990-1991 Recession? *American Economic Review* 83(2):280-86.

Harvey, Andrew C. 1990. *The econometric analysis of time series.* 2d ed. Cambridge, The MIT Press.

Hicks, John R. 1950. *A Contribution to the Theory of the Trade Cycle.* Oxford. Clarendon Press.

Hornstein, Andreas 1993. Monopolistic competition, increasing returns to scale, and the importance of productivity shocks. *Journal of Monetary Economics* 31:299-316.

Hussey, Robert 1992. Nonparametric evidence on asymmetry in business cycles using aggregate employment time series. *Journal of Econometrics* 51:217-31.

Kaldor, Nicholas 1940. A model of the trade cycle. *Economic Journal* 50:78-92. Reprinted in Kaldor, N. 1980. *Essays on Economic Stability and Growth.* New York: Holmes & Meier.

Kalecki, Michal 1935. A macro-dynamic theory of business cycles. *Econometrica* 3:327-344.

Keynes, John Maynard 1936. *The General Theory of Employment, Interest and Money,* London, Macmillan

Klein, Lawrence R. and A. S. Goldberger 1955. *An Econometric Model of the United States Economy 1922-1952.* Amsterdam: North-Holland.

Klein, Lawrence R. and R. S. Preston 1969. Stochastic nonlinear models. *Econometrica* (January):95-106.

Kydland, F. E. and Edward C. Prescott 1982. Time to build and aggregate fluctuations. *Econometrica* 50:1345-70.

Laroque, Guy and Guillaume Rabault 1995. The inventory cycle: from theory to empirical evidence. *The Economic Journal* 105:283-301.

Leijonhufvud, Axel 1993. Towards a not-too-rational macroeconomics. *Southern Economic Journal* 60(1):1-13.

Lorenz, Hans-Walter 1993. *Nonlinear Dynamical Economics and Chaotic Motion.* 2d ed. Berlin, Springer-Verlag.

McQueen, Grant and Steven Thorley 1993. Asymmetric business cycle turning points. *Journal of Monetary Economics* 31:341-62.

McCulloch, Robert and Ruey Tsay 1994. Statistical analysis of economic time series via Markov switching models. *Journal of Time Series Analysis* 15(5):523-39.

Medio, Alfredo 1992. *Chaotic Dynamics: Theory and Applications to Economics.* Cambridge, Cambridge University Press.

Meltzer, Allan H. 1982. Rational expectation, risk, uncertainty, and market responses. In Wachtel, P. L. (ed.), *Crises in the Economic and Financial Structure,* Lexington, Mass.: Lexington Books

Mitchell, Wesley 1927. *Business Cycles: The problem and its Setting.* New York: National Bureau of Economic Research.

Mullineux, Andy 1990. *Business Cycles and Financial Crises.* Ann Arbor: University of Michigan Press.

Mullineux, Andy and WenSheng Peng 1993. Nonlinear business cycle modeling. *Journal of Economic Surveys* 7(1):41-83.

Neftci, Salih, N. 1984. Are economic time series asymmetric over the business cycle? *Journal of Political Economy* 92:307-28.

Nishimura, Kazuo and Gerhard Sorger 1996. Optimal cycles and chaos: a survey. *Studies in Nonlinear Dynamics and Econometrics* 1(1):11-28.

Potter, Simon M. 1993. A nonlinear approach to U.S. GNP. Working Paper #692, Depart. of Economics, University of California, Los Angeles.

Potter, Simon M. 1994. Asymmetric economic propagation mechanisms. In Semmler (ed.), *Business Cycles: Theory and Empirical Methods,* Boston: Kluwer Academic Publishers.

Puu, Tonu 1993. *Nonlinear Economic Dynamics.* 3rd Edition. Berlin, Springer-Verlag.

Ramsey, James B., Chera L. Sayers and Philip Rothman 1992. The statistical properties of dimension calculations using small data sets: Some economic applications. In Benhabib (ed.), *Cycles and Chaos in Economic Equilibrium,* Princeton: Princeton University Press.

Ramsey, James B. and Philip Rothman 1996. Time irreversibility and business cycle asymmetry. *Journal of Money, Credit, and Banking* 28(1):1-21.

Rose, H. 1967. On the nonlinear theory of the employment cycle. *Review of Economic Studies* 34:132-52

Rothman, Philip 1991. Further evidence on the asymmetric behavior of unemployment rates over the business cycle. *Journal of Macroeconomics* 13(2):291-298

Rouwenhorst, k. Geert 1991. Time to build and aggregate fluctuations: a reconsideration, *Journal of Monetary Economics* 27(2):241-54.

Sayers, Chera L. 1994. Testing for chaos and nonlinearities in macroeconomic time series. In Semmler (ed.), *Business Cycles: Theory and Empirical Methods,* Boston: Kluwer Academic Publishers.

Scheinkman, José 1990. Nonlinearities in economic dynamics. *The Economic Journal* 100:33-48.

Schumpeter, Joseph A. 1939. *Business Cycles - A Theoretical, Historical, and Statistical Analysis of the Capitalist Process, Vol. 1,* New York: McGrow-Hill.

Semmler, Willi, ed. 1989. *Financial Dynamics and Business Cycles.* Armonk: M. E. Sharpe.

Semmler, Willi, ed. 1994. *Business Cycles: Theory and Empirical Methods,* Boston: Kluwer Academic Publishers.

Sichel, Daniel E. 1989. Are business cycles asymmetric: a correction. *Journal of Political Economy* 97:1255-60.

Sichel, Daniel E. 1991. Business cycle duration dependence: a parametric approach. *The Review of Economics and Statistics* 73:254-60.

Sims, Christopher. 1980. Macroeconomics and reality. *Econometrica* 48:1-48.

Stadler, George W. 1994. Real business cycles. *Journal of Economic Literature* 32(4):1750-83.

Stiglitz, Joseph E. 1993. Endogenous growth and cycles. *NBER* #4286.

Stiglitz, Joseph E. 1994. *Whither Socialism?,* Cambridge, MIT Press

Teräsvirta, Timo 1996. Power properties of linearity tests for time series. *Studies in Nonlinear Dynamics and Econometrics* 1(1):3-10.

Teräsvirta, Timo. and Heather M. Anderson 1993. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. In Pesaran and Potter (ed.). *Nonlinear Dynamics, Chaos and Econometrics.* New York, John Wiley & Sons.

Tong, Howell 1983. *Threshold Models in Non-linear Time Series Analysis.* New York, Springer-Verlag.

Tong, Howell 1990. *Non-linear Time Series: A Dynamical System Approach*, Oxford , Oxford University Press.

Yule, G. U. 1927. On a method of investigating periodicities in disturbed series with special reference to Walfer's sunspot numbers. *Philos. Trans. R. Soc.*, A226

Zarnowitz, Victor 1992. *Business Cycles: Theory, History, Indicators, and Forecasting*, Chicago, University of Chicago Press.

Vita

Fred Engst

1987	B.A. in Economics, Temple University, Philadelphia, Pennsylvania.
1987-88	Consultant, The Farm Bureau, Beijing, The People's Republic of China.
1988-90	Teaching Assistant, Department of Economics, New Brunswick, New Jersey.
1990	M.S. in economics, Rutgers University, New Brunswick, New Jersey.
1990-95	Staff manager, Business Operation Analysis, AT&T, Basking Ridge, New Jersey.
1997	Ph.D. in economics.