Empirical Economics (2002) 27:149–162

EMPIRICAL ECONOMICS

New directions in business cycle research and financial analysis

James D. Hamilton[1], **Baldev Raj**[2]

[1] Department of Economics, 0508, University of California, San Diego, La Jolla, CA 92093-0508 (e-mail: jhamilton@ucsd.edu)
[2] School of Business and Economics, Wilfrid Laurier University, 75 University Aveneu, Waterloo, Ontario, Canada N2L 3C5 (e-mail: braj@wlu.ca)

First Version Received: August 2001/Final Version Received: October 2001

Abstract. This paper serves as a partial introduction to and survey of the literature on Markov-switching models. We review the history of this class of models, describe their mathematical structure, and exposit the basic ideas behind estimation and inference. The paper also describes how the approach can be extended in a variety of directions, such as non-Gaussian distributions, time-varying transition probabilities, vector processes, state-space and GARCH models, and surveys recent methodological advances. The contributions of the other papers in this volume are reviewed. A final section offers conclusions and implications for policy.

Key words: Markov-switching, regime-switching, business cycles

JEL classification: C32, E32

1. Introduction: The basic Markov-switching framework

The normal behavior of economies is occasionally disrupted by dramatic events that seem to produce quite different dynamics for the variables that economists study. Chief among these is the business cycle, in which capitalist economies depart from their normal growth behavior and a variety of indicators go into decline. Other examples include currency crises, stock market bubbles, and sharp changes in the volatility of asset prices or exchange rates.

One natural way to describe such features is with an autoregressive process. Suppose that y_t represents the growth rate of real GDP in quarter t. In normal times, its dynamic behavior might be well characterized with a first-order autoregression,

$$y_t = c_1 + \phi_1 y_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \sigma^2)$. In such times, one would forecast the growth of GDP for quarter $t+1$ according to

$$\hat{y}_{t+1|t} = c_1 + \phi_1 y_t.$$

During a recession, however, an alternative forecasting rule might be better, perhaps using different values for the coefficients c and ϕ:

$$y_t = c_2 + \phi_2 y_{t-1} + \varepsilon_t.$$

One can write the above two expressions compactly by letting the regime indicator $s_t = 2$ if the economy is in a recession in quarter t and $s_t = 1$ otherwise:

$$y_t = c_{s_t} + \phi_{s_t} y_{t-1} + \varepsilon_t. \tag{1}$$

A full description of the dynamics of y_t could be obtained if we had a probabilistic description of how the economy changes from one regime to another. The simplest such model would be a Markov chain:

$$\Pr(s_t = j \mid s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) = \Pr(s_t = j \mid s_{t-1} = i) = p_{ij}. \tag{2}$$

Suppose that the econometrician observes y_t directly but can only make an inference about the value of s_t based on what we see happening with y_t. Then equations (1) and (2) constitute an example of a simple regime-switching process, more elaborate versions of which underlie all the papers of this volume. Theoretical descriptions of why the economy might behave in such a way have been proposed by Cooper and John (1988), Diamond and Fudenberg (1989), Howitt and McAfee (1992), Cooper (1994), Acemoglu and Scott (1997), Startz (1998), Chamley (1999), and Jeanne and Mason (2000). See Raj (2002) and the concluding section of this paper for a review of this and other related research.

Such a description is sometimes called a "hidden Markov model," since the realizations of the Markov chain s_t are not observed directly by the econometrician. The first treatments of such processes appear to have been by Lindgren (1978) and Baum, et. al. (1980). Hidden Markov models were an important tool in speech recognition algorithms developed in the 1980's; see Rabiner (1989) for a survey. In such applications, the "signal" y_t is often discrete-valued, so that in place of an autoregression such as (1) we would have

$$\Pr(y_t = k \mid s_t = j) = w_{jk}. \tag{3}$$

These early applications did not include the time-series components ϕ_1 or ϕ_2. Instead, the assumption was that, if one could condition on the regime $\{s_1, s_2, \ldots, s_T\}$, the sequence of observations $\{y_1, y_2, \ldots, y_T\}$ would be i.i.d. (independent and identically distributed).

The suggestion of using a switching autoregressive model such as (1) and (2) to describe the business cycle was made independently by Neftci (1982) and Sclove (1983), though neither proposed methods for calculating the likelihood function of y_t or trying to forecast such a process. The solutions were developed by Hamilton (1989), and turn out to be closely related to the algorithms for estimating Markov-switching regression functions discovered by Lindgren (1978) and Cosslett and Lee (1985). The key component in Hamilton's solution is an iterative algorithm analogous to the Kalman filter. The Kalman filter can be described as an algorithm for forming an estimate of an unobserved state variable x_t based on observations of y through date t:

$$\hat{x}_{t|t} = E(x_t \mid y_t, y_{t-1}, \ldots, y_1).$$

In the usual formulation of the Kalman filter, the unobserved state variable x_t itself follows a Gaussian autoregression and y_t is a linear function of x_t plus Gaussian measurement error, in which case the optimal inference $\hat{x}_{t|t}$ turns out to be a linear function of $(y_t, y_{t-1}, \ldots, y_1)$; see for example Hamilton (1994b). The Kalman filter is an iterative algorithm, whose input for step t is the value of $\hat{x}_{t-1\,|\,t-1}$ and y_t, and whose output is $\hat{x}_{t|t}$.

In a model such as (1) and (2), the unobserved state variable s_t follows a discrete-valued Markov chain, in contrast to the continuous-valued Gaussian autoregression usually assumed for x_t in the Kalman filter. The optimal inference about such a variable would take the form of a probability. Conditional on observing $(y_t, y_{t-1}, \ldots, y_1)$, for example, the observer might conclude that there is a probability 0.8 that the economy has entered a recession and a probability of 0.2 that the expansion is continuing.

The variable that takes the place of $\hat{x}_{t|t}$ is thus $\hat{\boldsymbol{\xi}}_{t|t}$ which represents a vector of probabilities. The first element of $\hat{\boldsymbol{\xi}}_{t|t}$ is $\Pr(s_t = 1 \mid y_t, y_{t-1}, \ldots, y_1)$, and the second element is $\Pr(s_t = 2 \mid y_t, y_{t-1}, \ldots, y_1)$. If there are K different possible regimes, $\hat{\boldsymbol{\xi}}_{t|t}$ would be a $(K \times 1)$ vector, each of whose elements is between zero and unity. If one knew the value $\hat{\boldsymbol{\xi}}_{t-1\,|\,t-1}$, it would be a simple matter to form a forecast of the regime for period t given what is known at time $t-1$. For example,

$$\Pr(s_t = 1 \mid y_{t-1}, y_{t-2}, \ldots, y_1) = \sum_{i=1}^{K} p_{i1} \Pr(s_{t-1} = i \mid y_{t-1}, y_{t-2}, \ldots, y_1). \tag{4}$$

A vector that collected the corresponding terms for the probabilities of $s_t = 2, 3, \ldots, K$ would naturally be denoted $\hat{\boldsymbol{\xi}}_{t\,|\,t-1}$.

A model such as (1) or (3) can be interpreted as specifying the probability law of the observed variable y_t conditional on s_t and its own past values, $f(y_t|s_t, y_{t-1}, y_{t-2}, \ldots, y_1)$. For example, for (1) we see

$$f(y_t \mid s_t = 1, y_{t-1}, y_{t-2}, \ldots, y_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y_t - c_1 - \phi_1 y_{t-1})^2}{2\sigma^2}\right]. \tag{5}$$

This could be collected along with the corresponding expressions for $s_t = 2, \ldots, K$ in a $(K \times 1)$ vector $\boldsymbol{\eta}_t$. The joint probability of y_t and s_t is then given by the product

$$\begin{aligned} &f(y_t, s_t = 1 \mid y_{t-1}, y_{t-2}, \ldots, y_1) \\ &\qquad = f(y_t \mid s_t = 1, y_{t-1}, y_{t-2}, \ldots, y_1) \Pr(s_t = 1 \mid y_{t-1}, y_{t-2}, \ldots, y_1). \end{aligned}$$

The conditional density of the tth observation is the sum of these terms over the different possible values for s_t, which will be recognized as simply the inner product of $\boldsymbol{\eta}_t$ with $\hat{\boldsymbol{\xi}}_{t|t-1}$:

$$f(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1) = \sum_{j=1}^{K} f(y_t, s_t = j \mid y_{t-1}, y_{t-2}, \ldots, y_1) = \boldsymbol{\eta}_t' \hat{\boldsymbol{\xi}}_{t|t-1}. \tag{6}$$

The ith element of $\hat{\boldsymbol{\xi}}_{t|t}$ is then given by

$$\frac{f(y_t, s_t = i \mid y_{t-1}, y_{t-2}, \ldots, y_1)}{f(y_t \mid y_{t-1}, y_{t-2}, \ldots, y_1)}.$$

Hence with these simple calculations one can go from the input $\hat{\boldsymbol{\xi}}_{t-1|t-1}$ to the output $\hat{\boldsymbol{\xi}}_{t|t}$, calculating the conditional density of the tth observation (6) as a by-product. The same principles can be used to form an inference about the regime the economy was in at date t, where the inference is based on information observed through the end of the sample,

$$\Pr(s_t = j \mid y_T, y_{T-1}, \ldots, y_1),$$

which, following the parallel with calculations for the Kalman filter, are called the "smoothed" probabilities. See Hamilton (1994a, Chapter 22) for further details.

2. Extensions of the basic framework

Although the above approach was described in terms of the particular examples (1) of a Gaussian autoregression or (3) a discrete observed scalar, the procedure works exactly the same way for an arbitrary collection of different densities $\boldsymbol{\eta}_t$. For example, instead of Gaussian innovations, we could have assumed a Student t distribution with v degrees of freedom:

$$f(y_t \mid s_t = 1, y_{t-1}, y_{t-2}, \ldots, y_1)$$
$$= \frac{\Gamma[(v+1)/2]}{(\pi v)^{1/2}\Gamma(v/2)\sigma}\left[1 + \frac{(y_t - c_1 - \phi_1 y_{t-1})^2}{\sigma v}\right]^{-(v+1)/2}. \tag{7}$$

The variance could be time-varying with an ARCH process whose parameters themselves are subject to change (e.g., Hamilton and Susmel, 1994), the degrees of freedom parameter v could itself change (e.g., Dueker, 1997), or the elements of $\boldsymbol{\eta}_t$ could even come from different families of densities. Whatever the form of the elements of $\boldsymbol{\eta}_t$, they are used in the recursion in the same way through equation (6).

The above example also assumed that changes in regime follow a Markov chain. However, the calculations go through in exactly the same way if $\Pr(s_t = j \mid s_{t-1} = i)$ is replaced with $\Pr(s_t = j \mid s_{t-1} = i, y_{t-1}, y_{t-2}, \ldots, y_1)$, allowing time-varying transition probabilities. For example, suppose for $k = 2$ states that the probability of a change in regime is modeled as

$$\Pr(s_t = j \mid s_{t-1} \neq j, y_{t-1}, y_{t-2}, \ldots, y_1) = \frac{\exp(\beta y_{t-1})}{1 + \exp(\beta y_{t-1})},$$

then (4) generalizes to

$$\begin{aligned}
&\Pr(s_t = 1 \mid y_{t-1}, y_{t-2}, \ldots, y_1) \\
&\quad = \frac{1}{1 + \exp(\beta y_{t-1})} \Pr(s_{t-1} = 1 \mid y_{t-1}, y_{t-2}, \ldots, y_1) \\
&\qquad + \frac{\exp(\beta y_{t-1})}{1 + \exp(\beta y_{t-1})} \Pr(s_{t-1} = 2 \mid y_{t-1}, y_{t-2}, \ldots, y_1),
\end{aligned}$$

See Diebold, Lee, and Weinbach (1994), Filardo (1994) and the papers by Peria and by Schaller and van Norden in this volume for illustrations of regime-switching models with time-varying transition probabilities.

Furthermore, no modification of the above discussion is necessary when $\mathbf{y}_t$ is a vector of observations on n different variables. In this case, expressions such as (5) or (7) might be replaced by a scalar such as

$$\begin{aligned}
&f(\mathbf{y}_t \mid s_t = 1, \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_1) \\
&\quad = \frac{1}{(2\pi)^{n/2}} |\boldsymbol{\Omega}|^{-1/2} \\
&\qquad \times \exp[-(\mathbf{y}_t - \mathbf{c}_1 - \Phi_1 \mathbf{y}_{t-1})' \boldsymbol{\Omega}^{-1} (\mathbf{y}_t - \mathbf{c}_1 - \Phi_1 \mathbf{y}_{t-1})/2].
\end{aligned}$$

One could also allow different elements of $\mathbf{y}_t$ to be governed by different regimes, or some coefficients to depend on the current regime and some on the lagged regime, with a simple redefinition of variables. For example, suppose that s_{1t} determines the coefficients of the first row of Φ and s_{2t} determines the coefficients of the second row, with s_{1t} and s_{2t} each taking on one of two possible values. One can define a summary regime s_t according to

$$s_t = \begin{cases} 1 & \text{if } s_{1t} = 1 \text{ and } s_{2t} = 1 \\ 2 & \text{if } s_{1t} = 1 \text{ and } s_{2t} = 2 \\ 3 & \text{if } s_{1t} = 2 \text{ and } s_{2t} = 1 \\ 4 & \text{if } s_{1t} = 2 \text{ and } s_{2t} = 2 \end{cases}$$

and s_t follows a 4-state Markov chain. Illustrations are provided by Hamilton and Lin (1996) and Hamilton and Perez-Quiros (1996). For a survey of these and other extensions and applications of the Markov-switching framework, see Raj (2002).

3. State-space and GARCH models

One important dimension in which a generalization is less obvious is when the dynamic behavior depends not just on a finite collection of regimes (such as $s_t, s_{t-1}, \ldots, s_{t-p}$) but instead depends on the infinite past $(s_t, s_{t-1}, s_{t-2}, \ldots)$. An example is when one wishes to apply the regime-switching approach to a GARCH as opposed to an ARCH model, such as

$$y_t = \sigma_t v_t,$$

with $v_t \sim$ i.i.d. $N(0, 1)$ and

$$\sigma_t^2 = \zeta_{s_t} + \alpha_{s_t} y_{t-1}^2 + \delta_{s_t} \sigma_{t-1}^2. \tag{8}$$

By recursive substitution in (8), we see

$$\sigma_t^2 = \zeta_{s_t} + \alpha_{s_t} y_{t-1}^2 + \delta_{s_t}(\zeta_{s_t-1} + \alpha_{s_{t-1}} y_{t-2}^2) + \delta_{s_t}\delta_{s_{t-1}}(\zeta_{s_t-2} + \alpha_{s_{t-2}} y_{t-2}^2) + \cdots.$$

The whole history of regimes is necessary to calculate the density of the tth observation. A related problem arises in a regime-switching state-space model, with state equation

$$\mathbf{x}_t = \mathbf{F}_{s_t}\mathbf{x}_{t-1} + \mathbf{v}_t \tag{9}$$

and observation equation

$$\mathbf{y}_t = \mathbf{H}'\mathbf{x}_t + \mathbf{w}_t \tag{10}$$

implying

$$\mathbf{y}_t = \mathbf{w}_t + \mathbf{H}'\mathbf{v}_t + \mathbf{H}'\mathbf{F}_{s_t}\mathbf{v}_{t-1} + \mathbf{H}'\mathbf{F}_{s_t}\mathbf{F}_{s_{t-1}}\mathbf{v}_{t-2} + \mathbf{H}'\mathbf{F}_{s_t}\mathbf{F}_{s_{t-1}}\mathbf{F}_{s_{t-2}}\mathbf{v}_{t-2} + \cdots.$$

Such regime-switching state-space models are extremely important for studying the business cycle, in which we anticipate that a large number of observed series may be influenced by common unobserved components.

One solution to this problem was proposed by Gray (1996), who suggested that we consider an alternative data-generating process for which the problem does not appear, replacing (8) with

$$\begin{aligned} h_{it} &= \zeta_i + \alpha_i y_{t-1}^2 + \delta_i h_{t-1}^2 \\ h_{t-1} &= \Pr(s_{t-1} = 1 \mid y_{t-2}, y_{t-3}, \ldots, y_1) h_{1,t-1} \\ &\quad + \Pr(s_{t-1} = 2 \mid y_{t-2}, y_{t-3}, \ldots, y_1) h_{2,t-1} \end{aligned} \tag{11}$$

and $\sigma_t^2 = h_{1t}$ when $s_t = 1$ and h_{2t} when $s_t = 2$. Thus, instead of the lagged magnitude σ_{t-1}^2, which depends on the infinite history of past regimes, the value of σ_t^2 depends only on the current regime s_t and the previous period's *inference* about s_{t-1}. The inference itself depends only on observed data through date $t-1$, so there is no multiplication of states as t increases. In this volume, Klaassen proposes that instead of (11) we use

$$\begin{aligned} h_{t-1} &= \Pr(s_{t-1} = 1 \mid y_{t-1}, y_{t-2}, \ldots, y_1) h_{1,t-1} \\ &\quad + \Pr(s_{t-1} = 2 \mid y_{t-1}, y_{t-2}, \ldots, y_1) h_{2,t-1} \end{aligned}$$

which turns out to be a substantially more convenient formulation for purposes of calculating multi-period-ahead forecasts.

A closely related method of collapsing the history using the current filtered probabilities was developed for regime-switching state-space models by Kim (1994), albeit with a slightly different motivation. Kim proposed a state-space analog of (11) (using the smoothed probabilities) not as a data-generating process but as a tool for approximating the likelihood function of data that were truly generated by a process such as (9); See Kim and Nelson (1999b) for details. This approach is adopted in the papers by Kim and Murray and Mills and Wang in this volume.

An alternative solution uses recent advances in numerical Bayesian methods to solve this problem, as developed by Albert and Chib (1993) and Shephard (1994). In the Bayesian formulation, rather than estimating parameters such as ϕ_1 and ϕ_2 by maximum likelihood, the goal is to characterize one's subjective probability distribution of these parameters, given the observed data $(\mathbf{y}_T, \mathbf{y}_{T-1}, \ldots, \mathbf{y}_1)$. The mean of the subjective distribution of ϕ_1 is often close to the maximum likelihood estimate $\hat{\phi}_1$, and if the former is much easier to calculate, even classical econometricians can be drawn to the Bayesian approach. The basic idea behind numerical Bayesian methods is that, even though one may not be able to characterize analytically the Bayesian posterior distribution, one can generate draws from it. Specifically, in a state-space model such as (9) and (10) it turns out to be straightforward to generate a draw from the distribution of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ conditional on $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$, $(s_1, s_2, \ldots, s_T)$, and the conventional parameters $\boldsymbol{\theta}$ which includes $(\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_K)$, $\mathbf{H}$, the variances of $\mathbf{v}_t$ and $\mathbf{w}_t$, and the p_{ij} parameters. When $(s_1, s_2, \ldots, s_T)$ is regarded as a known set of numbers, this distribution of $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ is known analytically from the standard Kalman filtering and smoothing recursions. Likewise one can characterize the distribution of $(s_1, s_2, \ldots, s_T)$ conditional on $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$, $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, and $\boldsymbol{\theta}$ – if one could observe $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, none of the problems of infinite expansion of states would arise in calculating the likelihood function, so the methods described in Section 1 above could be applied directly. Finally, if one could observe $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T)$, $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, and $(s_1, s_2, \ldots, s_T)$, none of the problems associated with either state-space models of regime-switching would be present, and the posterior distribution of $\boldsymbol{\theta}$ given these other magnitudes is readily obtained from known results about Bayesian GLS.

One can thus construct a chain of simulations, first generating $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$, then taking these simulated values as if observed data to simulate $(s_1, s_2, \ldots, s_T)$, taking these as data in turn to generate $\boldsymbol{\theta}$, and then generating a new set $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$. This sequence constitutes a Markov chain, and the average value of $\boldsymbol{\theta}$ across the chain should converge to the ergodic distribution of the chain, which is the unconditional distribution of $\boldsymbol{\theta}$ given the data, the object of interest. An excellent introduction to and survey of this approach are provided by Kim and Nelson (1999b). Numerical Bayesian methods are also extremely convenient for estimating regime-switching models with time-varying transition probabilities, as shown by Filardo and Gordon (1998). Papers in this volume that apply these advances in numerical Bayesian methods include Kaufmann and Chauvet, Juhn, and Potter.

This volume collects some of the most exciting and promising new applications of these methods to understanding economic phenomena. The current volume features state-of-the-art applications and development of some innovative approaches to regime-switching models, with substantive contributions in new descriptions of the business cycle, both in the U.S. and elsewhere, new descriptions of financial markets, and methodological contributions. The following sections briefly summarize these respective contributions.

4. The business cycle in the US

Three papers contribute new discoveries about the nature of the U.S. business cycle. Kim and Murray, in their paper "Permanent and Transitory Components of Recessions," propose a model that allows them to measure the permanent and transitory consequences of recessions separately. They study a monthly vector system including industrial production, personal income, sales, and employment in which there is a common permanent and common transitory component, both of which undergo potential changes in regime. They find that the transitory component of recessions accounts for more than three-fourths of the variance of the individual series, and suggest that the U.S. business cycle exhibits three separate phases: recession, rapid growth as the economy initially pulls out of the recession, followed by a normal growth phase.

Clements and Krolzig, in "Can Oil Shocks Explain Asymmetries in the US Business Cycle," study the role of oil price shocks in the business cycle. They investigate a cointegrated quarterly vector system including employment and output subject to shifts in regime, and look at measures of business cycle asymmetry in such a system based on the steepness and deepness of contractions. They find business cycle asymmetry regardless of whether one conditions on oil prices, and conclude that U.S. business cycle asymmetry can not be attributed to the role of oil shocks in propagating recessions. They also find evidence in support of a three-regime description of the business cycle: recessions are typically followed by a rapid-growth recovery phase, moving from there into either a slower growth phase or perhaps back into recession.

"Markov Switching in Disaggregate Unemployment Rates" by Chauvet, Juhn, and Potter, studies quarterly U.S. unemployment rates disaggregated by age. The disaggregation helps separate the role of secular demographic changes from business cycle effects. In contrast to the other two papers on the U.S. business cycle, this paper uses a two-state model, one state corresponding to a high degree of labor market mismatch (whose inferred probabilities adhere remarkably to conventional NBER recession dates) and the second a more normal state. The paper finds that demographic changes are very important in explaining the secular trends in unemployment, while business cycles dominate the short run movements.

5. The business cycle in other countries

Three other papers study business cycles in other countries. "A Markov-Switching Vector Equilibrium Correction Model of the UK Labour Market," by Krolzig, Marcellino, and Mizon, studies quarterly data on UK output, employment, labor force, and real wage. The system is cointegrated with the vector of constant terms in the VAR subject to shifts in regime. The paper provides abundant evidence that the data are well described by a three-regime model. In regime 1, output, employment, and the labor supply all fall, while in regime 3, they all rise. In regime 2, output grows even as employment is stagnant. The paper finds that allowing for the changes in regime can significantly change the inferences one would draw from linear impulse-response functions.

"Plucking Models of Business Cycle Fluctuations: Evidence from the G-7 Countries," by Mills and Wang, studies quarterly real GDP for each of the G-7 countries. They use Kim and Nelson's (1999) model in which recessions bring output temporarily below trend growth, with recovery then allowing reversion back to trend. They find that this offers a good description of the business cycle for the U.S., U.K., France, and Italy, with somewhat weaker performance for Canada, Germany, and Japan.

6. Financial applications

Sylvia Kaufmann's "Is There an Asymmetric Effect of Monetary Policy over Time?" studies the relation between monetary policy and the business cycle for Austria. Her model relates quarterly GDP to its own lags and lagged changes in the 3-month interest rate, with the constant term, coefficients on the interest rates, and standard error of this relation allowed to change between two regimes. The results suggest that monetary policy (as measured by the interest rate) makes a difference for the Austrian economy primarily when the economy is in a downturn.

Three other papers contribute to our understanding of regimes in asset prices. "A Regime-Switching Approach to the Study of Speculative Attacks: A Focus on EMS Crises" by Peria studies monthly data on exchange rates, central bank reserves, and interest rate differentials. She estimates trivariate vector systems separately for 7 different European countries (Belgium, Denmark, France, Ireland, Italy, Spain, and UK). Speculative currency attacks show up as sharp movements in one or more of these indicators for any country. By describing a speculative attack as a shift in the coefficients of a VAR, the paper gives new insights into the causes of speculative attacks, concluding that budget deficits are a key factor.

Schaller and van Norden, in their paper "Fads or Bubbles," investigate dramatic changes in the joint behavior of an index of monthly U.S. stock prices and dividends over 1926–1989. They allow the coefficients in a regression of stock returns on alternative measures of the deviation between the lagged stock price and the "fundamentals" value to change with different regimes. Such a structure could be consistent either with investors following fads or bubbles. The paper shows how to distinguish between these two, and reports evidence of regime-switching in stock returns that could be interpreted as reflecting investing fads.

"Improving GARCH Volatility Forecasts with Regime-Switching GARCH" by Klaassen studies the volatility of daily dollar-pound, dollar-mark, or dollar-yen exchange rates. He uses a GARCH representation for expected squared exchange rate changes, allowing the coefficients to change with different regimes. This model yields substantially improved forecasts of exchange rate volatility.

7. Methodological contributions

Although all of these papers produce new empirical results of substantive economic interest, they are also methodologically innovative and up-to-date, representing current best known methods for estimating regime-switching processes. These include how to handle vector autoregressions with changes in regime, both with cointegration (Krolzig-Marcellino-Mizon and Clements-Krolzig) and without (Peria). Several papers illustrate the approach to systems in which there is a latent unobserved variable besides the regime indicator itself. In dynamic factor models, this unobserved variable is the state of the economy or business cycle, while in GARCH models, this unobserved variable is the lagged conditional variance. In dynamic factor models, the technical problems this introduces can be solved either with Kim's (1994) approximate maximum likelihood algorithm for collapsing the infinite past into approximate indexes, as in Kim-Murray and Mills-Wang, or with numerical Bayesian methods, as in Chauvet-Juhn-Potter. In the GARCH context, Klaassen proposes an improvement on Gray's (1996) earlier idea in which the collapsing approximation is in fact regarded as the data-generating process. Numerical Bayesian methods and Frühwirth-Schnatter's (2001) permutation approach are also featured in Kaufmann's paper. Markov-switching models in which the regimes follow more complicated dynamics than a simple Markov chain are illustrated by the papers by Peria and Schaller and van Norden.

Finally, the volume includes a pure methodological contribution by Patrick Coe entitled "Power Issues when Testing the Markov Switching Model with the Sup Likelihood Ratio Test Using U.S. Output." A common problem in these applications is testing the null hypothesis that there are no changes in regime. One popular approach is a test suggested by Garcia (1998). Coe's paper uses Monte Carlo experiments to study the power properties (the probability of correctly concluding that there is regime-switching) of Garcia's test, and finds the test does reasonably well. This result may be quite useful in future applications of these methods.

8. Policy implications and concluding remarks

In earlier sections we provided the motivation for regime-switching models and gave an account of advances in this area as related to business cycle analysis, financial applications and methodological developments. Ten papers in the special issue give a representative picture of the scope and breadth of the regime-switching framework in these areas. These papers plus this introduction and overview paper can serve as a window to the regime-switching literature for not only experienced readers but also new readers. Specifically, five papers are on business cycle analysis, four on financial applications and one on methodology, although several other papers made contributions to this dimension. These provide a good account of the background and new advances on the topic. Below we provide some additional remarks plus give a brief account of some policy implications of regime switching models, especially in the context of business cycle research, for completeness.

Regime-switching models are designed to capture the asymmetry observed in the business cycle, along with some of its other known stylized features. The interest in this feature of the business cycle has a long history in economics that dates back to Keynes (1936). It is concerned with one of the key stylized facts of business cycles – expansions and contractions are quite different from each other, with the former typically long-lived and the latter more violent. Specifically, when output expands towards a peak, the economy is said to be in boom or an expansion. Conversely, when output falls towards a low point or trough, the economy is in recession. This period is characterized by negative rates of growth.

One of the main goals of theoretical macroeconomics is to understand the reasons for the asymmetry and other features of the business cycle. The applied or empirical analysis, on the other hand, is concerned with testing alternative theories and description of how well one or more theoretical models of the business cycle fit the observed features of business fluctuations found in the data. It can also be concerned with establishing stylized facts of the business cycle.

A good way to assess the business cycle is by examining the departure of the economic variable from its trend line that depicts the long-term growth or economic development of the economy. The tendency over the business cycle of variables such as labor productivity to move in the same directions as the gross domestic output in a procyclical way is often also of interest. Similarly the tendency of other economic variables to move in the opposite direction or countercyclically, such as the unemployment rate, is also part of business cycle research. The tendency of business cycles of various industrialized countries to be related to each other can also be of interest. Finally, the tendency of economic variables to depict stochastic equilibrium (or form a cointegration relation) among several variables being jointly studied for the business cycle analysis can also be of interest.

A theoretical description of why the economy might exhibit regime-switching behavior is that there can be multiple equilibria in aggregate activity, as developed by Cooper and John (1988). This theory uses a representative agent framework where the agent chooses the value of some variable, which one could call output for specificity, by taking other agents' choices as given, and is concerned about his position relative to others. Moreover, if his reaction function is nonlinear, and is bounded between zero and some positive value, then multiple equilibria are possible. In this set up, more than one equilibrium occurs since the reaction function crosses with the equilibrium line of the representative agent in relation to the output level of others an odd number of times. Moreover, since the reaction function must begin above the equlibria line and end up below it, some equilibria will be unstable while others stable under plausible assumptions about dynamics. A model with multiple equilibria, when outputs are Pareto-ranked, is sometimes also known as a coordination-failure model.

One of the implications of these models is that fundamentals do not fully determine outcomes since economic activity can also be influenced by other factors such as animal spirits, self-fulfilling prophecies, and sunspots. This occurs because agents' believe such nonfundamentals affect aggregate economic activity. Furthermore, the possibility of coordination failures suggests that the economy can get stuck in unemployment equilibrium where the output is low in relation to what it could be since everyone believes that it will be. Moreover, there would be absence of an inherent force to restore the aggregate output to a normal level. Thus, there is scope for government policies that coordinate expectations of agents to a high-output equilibrium such as a temporary fiscal and monetary stimulus that might move the economy to a better equilibrium state. However, in this framework there can be a potential for a strategic interplay between private economic agents and government or institutions or policy makers. Accordingly, the role of the reputation or commitment and time consistency becomes important for the success of the policy. These implications are obtained by applying game theory tools to macroeconomic policy analysis. Furthermore, the possibility of low level of output equilibrium underscores the importance of information and media management in regard to communicating policy tools to the citizens to remedy business cycle problems at the hands of political leaders and public officials by enlisting the cooperation of most economic agents. Also, the role of coalition building becomes important for the success of policy to avoid coordination failures. Finally, it suggests the role of leadership and governance can prove to be quite important to take the economy out of the low-output state and move towards a high-output state.

While business cycles occur in the coordination-failure models from movements among multiple equilibria, business cycles in the traditional Keynesian and real-business cycle models can arise from some type of market imperfection and/or real shocks. Business cycles can occur in a Keynesian model where markets fail to clear due to presence of imperfection in one or more markets of the nominal rigidity variety. Conversely, fluctuations where markets clear arise in the real-business models due to technology and other real shocks, and incomplete information. Finally, in new Keynesian business cycle models, the primary focus is on the presence of real rigidities or imperfections in one or more markets of the economy. In these models, the concern is about one or more factors that can produce real rigidity to nominal adjustments. The consequence of such rigidity can be that the firm's profit function becomes less sensitive to the price of its product. On the cost front, the firm's incentive to cut its price could become smaller, the smaller the fall in the marginal cost. For instance, thick-market externalities may make it more likely that a smaller downward shift in the profit function will occur when there is fall in aggregate demand. Within this type of setup, the purchase of inputs and the sale of final products by firms becomes easier when there is a high level of economic activity, reducing the incentive for a firm to lower its price relative to other firms. Similarly, agglomeration economies may lower the relative cost of some firms, another form of imperfection that can produce second-best type equilibria in the economy. Imperfections in the labor market are another important source of cost swings that could almost swamp the effect of other real rigidities. Imperfections in the labor market such as search and contracting models (e.g. see Diamond, 1982, and Howitt, 1988) and efficiency wage models (see Shapiro and Stiglitz, 1985) are two prime examples of such real rigidities. Such imperfections are important since firms' incentive to vary prices in situations of downward shifts of demand is reduced.

On the revenue side, a real rigidity can arise from imperfect information that makes existing customers more responsive to price increases than decreases relative to prospective new customers (see Stigitz, 1979). Another source of imperfection is related to capital-market imperfection that induces firms to raise prices during recessions if they are facing liquidity constraints (see Greenwald et al., 1984, and Chevalier and Schafstein, 1996).

It is important to note that there is a close link between real rigidity models and the coordination failure model of Cooper and John. One implication of this link is that as there are many potential sources of real rigidity, there are many potential sources of coordination failure, and hence many more models. Another implication is that real rigidities and/or coordination failure can make the equilibrium quite sensitive to technology and other real shocks, credit market imperfections or other nominal shocks, and uncertainty about future government policies. A complete discussion of the strengths and limitations of various approaches to business cycle theory along with applied examples is provided by Romer (1996).

In summary, there is a growing consensus among economists that regime changes might be more appropriately modeled as arising from a probability process such as the Markov process instead of deterministic structural changes. Moreover, the regime-switching framework can capture regime changes in diverse set of macroeconomics and financial time series. Also, this framework is well suited to modeling asymmetry of business cycles analysis. In addition, the framework is flexible to allow regime changes in the trend component as in Hamilton (1989) and extensions, and regime switching in the transitory component as in Kim and Nelson (1999). Another model of regime switching allows for the autoregressive coefficients to change as in Ang and Bekaert (2001), which is suitable for modeling the Fed asymmetric policy response to alternate inflationary regimes. The regime-switching framework is easily extended to deal with a variety of situations encountered in economics, as demonstrated by various papers in this special issue. Other applications are covered in the recent survey of the regime-switching models by Raj (2002). Finally, the regime-switching framework is well suited to modeling volatility in financial time series, including fads and bubbles.

References

Acemoglu D, Scott A (1997) Asymmetric business cycles: Theory and time-series evidence. Journal of Monetary Economics 40:501–33

Albert J, Chib S (1993) Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business and Economic Statistics 11:1–15

Ang A, Bekaert G (2001) Regime switches in interest rates. Journal of Business and Economic Statistics, forthcoming

Baum LE, Ted P, George S, Weiss N (1980) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics 41:164–171

Chamley C (1999) Coordinating regime switches. Quarterly Journal of Economics 114:869–905

Chevalier JA, Scharfstein DS (1996) Capital market imperfections and countercyclical markups: Theory and evidence. American Economic Review 86:703–725

Cooper R (1994) Equilibrium selection in imperfectly competitive economics with multiple equilibria. Economic Journal 104:1106–1122

Cooper R, John A (1988) Coordinating coordination failures in keynesian models. Quarterly Journal of Economics 103:441–463

Cosslett SR, Lee L-F (1985) Serial correlation in discrete variable models. Journal of Econometrics 27:79–97

Diamond PA (1982) Aggregate demand management in search equilibrium. Journal of Political Economy, 90 (October): 881–894. Reprinted in Mankiw and Romer (1991) Eds. New Keynesian Economics 2 Volumes. Cambridge: MIT Press

Diamond PA, Fudenberg D (1989) Rational expectations business cycles in search equilibrium. Journal of Political Economy 97:606–19

Diebold FX, Lee J-H, Weinbach GC (1994) Regime switching with time-varying transition probabilities. In: Hargreaves C (ed) Nonstationary Time Series Analysis and Cointegration, Oxford: Oxford University Press

Dueker M (1997) Markov switching in GARCH processes and mean-reverting stock-market volatility. Journal of Business and Economic Statistics 15:26–34

Filardo AJ (1994) Business cycle phases and their transitional dynamics. Journal of Business and Economic Statistics 12:299–308

Filardo AJ, Gordon SF (1998) Business cycle durations. Journal of Econometrics 85:99–123

Frühwirth-Schnatter S (2001) Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. Journal of the American Statistical Association 96:194–209

Garcia R (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models. International Economic Review 39:763–788

Gray SF (1996) Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42:27–62

Greenwald BC, Stiglitz JE, Andrew W (1994) Informational imperfections in capital markets and macroeconomic fluctuations. American Economics Review 74 (May): 194–199

Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57:357–384

Hamilton JD (1994a) Time series analysis. Princeton, NJ: Princeton University Press

Hamilton JD (1994b) State-space models. In: Engle R, McFadden D (eds) Handbook of Econometrics, Volume 4, pp. 3039–3080, Amsterdam: North Holland

Hamilton JD, Lin G (1996) Stock market volatility and the business cycle. Journal of Applied Econometrics 11:573–593

Hamilton JD, Perez-Quiros G (1996) What do the leading indicators lead? Journal of Business 69:27–49

Hamilton JD, Susmel R (1994) Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64:307–333

Howitt P (1988) Business cycles with costly search and recruiting. Quarterly Journal of Economics 103 (February): 147–165

Howitt P, McAfee P (1992) Animal spirits. American Economic Review 82:493–507

Jeanne O, Masson P (2000) Currency crises, sunspots, and Markov-switching regimes. Journal of International Economics 50:327–350

Keynes JM (1936) The general theory of employment, interest and money. London: Macmillan

Kim CJ (1994) Dynamic linear models with Markov-switching. Journal of Econometrics 60:1–22

Kim CJ, Nelson CR (1999a) Friedman's plucking model of business fluctuations: Tests and estimation of permanent and transitory components. Journal of Money, Credit, and Banking 31:317–334

Kim CJ, Nelson CR (1999b) State-space models with regime switching. Cambridge, Massachusetts: MIT Press

Lindgren G (1978) Markov regime models for mixed distributions and switching regressions. Scandinavian Journal of Statistics 5:81–91

Neftci SN (1982) Optimal prediction of cyclical downturns. Journal of Economic Dynamics and Control 4:225–241

Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77:257–286

Raj B (2002) Asymmetries of business cycle: The Markov-switching approach. In: Ullah A, Wan A, Chaturvedi A (eds) Handbook of Applied Econometrics and Statistical Inference, Marcel Dekker, Chapter 31, pp 687–710

Romer D (1996) Advanced macroeconomics. NY: McGraw-Hill Companies, Inc

Sclove SL (1983) Time-series segmentation: A model and a method. Information Sciences 29:7–25

Shaprio C, Stiglitz JE (1984) Equilibrium unemployment as a worker discipline device. American Economic Review 74 (June): 433–444. Reprinted in Mankiw and Romer (1991) Eds. New Keynesian Economics 2 Volumes. Cambridge: MIT Press

Shephard N (1994) Partial non-gaussian state space. Biometrika 81:115–131

Startz R (1998) Growth states and shocks. Journal of Economic Growth 3:203–15

Stiglitz JE (1979) Equilibrium in product markets with imperfect information. American Economic Review 69 (May): 339–345