CHAPTER 3

Macroeconomic Regimes and Regime Shifts

J.D. Hamilton
University of California, San Diego, La Jolla, CA, United States

Contents

Handbook of Macroeconomics, Volume 2A
ISSN 1574-0048, http://dx.doi.org/10.1016/bs.hesmac.2016.03.004

Abstract

Many economic time series exhibit dramatic breaks associated with events such as economic recessions, financial panics, and currency crises. Such changes in regime may arise from tipping points or other nonlinear dynamics and are core to some of the most important questions in macroeconomics. This chapter surveys the literature for studying regime changes and summarizes available methods. Section 1 introduces some of the basic tools for analyzing such phenomena, using for illustration the move of an economy into and out of recession. Section 2 focuses on empirical methods, providing a detailed overview of econometric analysis of time series that are subject to changes in regime. Section 3 discusses theoretical treatment of macroeconomic models with changes in regime and reviews applications in a number of areas of macroeconomics. Some brief concluding recommendations for applied researchers are offered in Section 4.

Keywords

Changes in regime, Markov switching, EM algorithm, Nonlinear macroeconomic dynamics, Vector autoregressions

JEL Codes

C32, E32, E37

1. INTRODUCTION: ECONOMIC RECESSIONS AS CHANGES IN REGIME

Fig. 1 plots the US unemployment rate since World War II. Shaded regions highlight a feature of the data that is very familiar to macroeconomists—periodically the US economy enters an episode in which the unemployment rate rises quite rapidly. These shaded regions correspond to periods that the Dating Committee of the National Bureau of Economic Research chose to designate as economic recessions. But what exactly does such a designation signify?

One view is that the statement that the economy has entered a recession does not have any intrinsic objective meaning. According to this view, the economy is always subject to unanticipated shocks, some favorable, others unfavorable. A recession is then held to be nothing more than a string of unusually bad shocks, with the bifurcation of the observed sample into periods of "recession" and "expansion," an essentially arbitrary way of summarizing the data.

Such a view is implicit in many theoretical models used in economics today insofar as it is a necessary implication of the linearity we often assume in order to make our models more tractable. But the convenience of linear models is not a good enough reason to assume that no fundamental changes in economic dynamics occur when the economy goes into a recession. For example, we understand reasonably well that in an expansion, GDP will rise more quickly at some times than others, depending on the pace of new technological innovations. But what exactly would we mean by a negative technology shock? The assumption that such events are just like technological improvements but

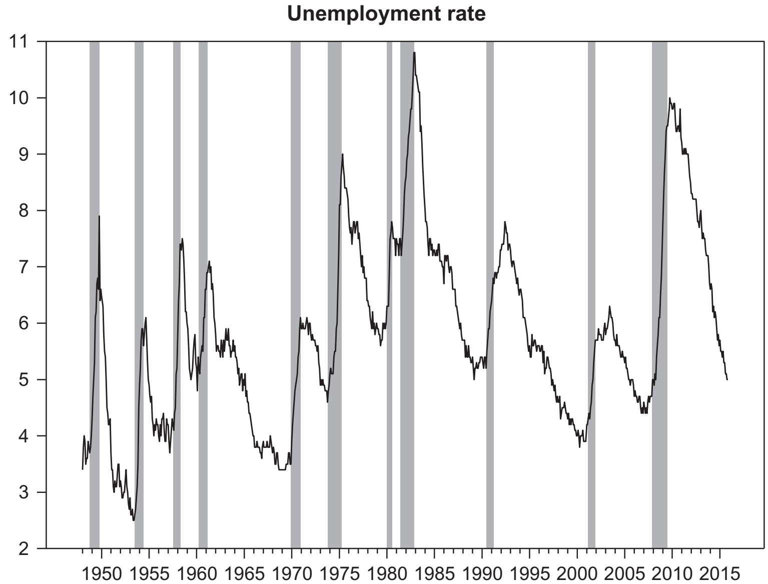

Fig. 1 US civilian unemployment rate, seasonally adjusted, 1948:M1–2015:M11. *Shaded regions* correspond to NBER recession dates.

with a negative sign does not seem like the place we should start if we are trying to understand what really happens during an economic downturn.

An alternative view is that on occasion some forces that are very different from the usual technological growth take over to determine employment and output, resulting for example when a simultaneous drop in product demand across different sectors and a rapid increase in unemployed workers introduce new feedbacks of their own. The idea that there might be a tipping point at which different economic dynamics begin to take over will be a recurrent theme in this chapter.

Let us begin with a very simple model with which we can explore some of the issues. We could represent the possibility that there are two distinct phases for the economy using the random variable s_t. When $s_t = 1$, the economy is in expansion in period t and when $s_t = 2$, the economy is in recession. Suppose that an observed variable y_t such as GDP growth has an average value of $m_1 > 0$ when $s_t = 1$ and average value $m_2 < 0$ when $s_t = 2$, as in

$$y_t = m_{s_t} + \varepsilon_t \tag{1}$$

where $\varepsilon_t \sim$ i.i.d. $N(0,\sigma^2)$. Suppose that the transition between regimes is governed by a Markov chain that is independent of ε_t,

$$\text{Prob}(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, y_{t-1}, y_{t-2}, \ldots) = p_{ij} \quad i, j = 1, 2. \tag{2}$$

Note that if both s_t and ε_t were observed directly, (1)–(2) in fact could still be described as a linear process. We can verify directly from (2) that[a]

$$E(m_{s_t}|m_{s_{t-1}}, m_{s_{t-2}}, \ldots) = a + \phi m_{s_{t-1}} \tag{3}$$

where $a = p_{21}m_1 + p_{12}m_2$ and $\phi = p_{11} - p_{21}$. In other words, m_{s_t} follows an AR(1) process,

$$m_{s_t} = a + \phi m_{s_{t-1}} + v_t. \tag{4}$$

The innovation v_t can take on only one of four possible values (depending on the realization of s_t and s_{t-1}) but by virtue of (3), v_t can be characterized as a martingale difference sequence.

Suppose however that we do not observe s_t and ε_t directly, but only have observations of GDP up through date $t-1$ (denoted $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots\}$) and want to forecast the value of y_t. Notice from Eq. (1) that y_t is the sum of an AR(1) process (namely (4)) and a white noise process ε_t. Recall (eg, Hamilton, 1994, p. 108) that the result could be described as an ARMA(1,1) process. Thus the linear projection of GDP on its own lagged values is given by

$$\hat{E}(y_t|\Omega_{t-1}) = a + \phi y_{t-1} + \theta[y_{t-1} - \hat{E}(y_{t-1}|\Omega_{t-2})] \tag{5}$$

where θ is a known function of ϕ, σ^2, and the variance of v_t (Hamilton, 1994, eq. [4.7.12]). Note that we are using the notation $\hat{E}(y_t|\Omega_{t-1})$ to denote a linear projection (the forecast that produces the smallest mean squared error among the class of all linear functions of Ω_{t-1}) to distinguish it from the conditional expectation $E(y_t|\Omega_{t-1})$ (the forecast that produces the smallest mean squared error among the class of all functions of Ω_{t-1}).

Because of the discrete nature of s_t, the linear projection (5) would not yield the optimal forecast of GDP. We can demonstrate this using the law of iterated expectations (White, 1984, p. 54):

$$\begin{aligned} E(y_t|\Omega_{t-1}) &= \sum_{i=1}^{2} E(y_t|s_{t-1} = i, \Omega_{t-1}) \text{Prob}(s_{t-1} = i|\Omega_{t-1}) \\ &= \sum_{i=1}^{2} (a + \phi m_i) \text{Prob}(s_{t-1} = i|\Omega_{t-1}). \end{aligned} \tag{6}$$

Because a probability is necessarily between 0 and 1, the optimal inference $\text{Prob}(s_{t-1} = i|\Omega_{t-1})$ is necessarily a nonlinear function of Ω_{t-1}. If data through $t-1$ have persuaded us that the economy was in expansion at that point, the optimal forecast is going to be close

[a] That is,

$$\begin{aligned} E(m_{s_{t+1}}|m_{s_t} = m_1) &= p_{11}m_1 + p_{12}m_2 = p_{11}m_1 + a - p_{21}m_1 = a + \phi m_1 \\ E(m_{s_{t+1}}|m_{s_t} = m_2) &= p_{21}m_1 + p_{22}m_2 = a - p_{12}m_2 + p_{22}m_2 = a - (1-p_{11})m_2 + (1-p_{21})m_2 = a + \phi m_2. \end{aligned}$$

to $a + \phi m_1$, whereas if we become convinced the economy was in recession, the optimal forecast approaches $a + \phi m_2$. It is in this sense that we could characterize (1) as a nonlinear process in terms of its observable implications for GDP.

Calculation of the nonlinear inference $\text{Prob}(s_{t-1} = i|\Omega_{t-1})$ is quite simple for this process. We could start for $t = 0$ for example with the ergodic probabilities of the Markov chain:

$$\text{Prob}(s_0 = 1|\Omega_0) = \frac{p_{21}}{p_{21} + p_{12}}$$
$$\text{Prob}(s_0 = 2|\Omega_0) = \frac{p_{12}}{p_{21} + p_{12}}.$$

Given a value for $\text{Prob}(s_{t-1} = i|\Omega_{t-1})$, we can arrive at the value for $\text{Prob}(s_t = j|\Omega_t)$ using Bayes's law:

$$\text{Prob}(s_t = j|\Omega_t) = \frac{\text{Prob}(s_t = j|\Omega_{t-1}) f(y_t|s_t = j, \Omega_{t-1})}{f(y_t|\Omega_{t-1})}. \tag{7}$$

Here $f(y_t|s_t = j, \Omega_{t-1})$ is the $N(m_j, \sigma^2)$ density,

$$f(y_t|s_t = j, \Omega_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(y_t - m_j)^2}{2\sigma^2}\right], \tag{8}$$

$\text{Prob}(s_t = j|\Omega_{t-1})$ is the predicted regime given past observations,

$$\text{Prob}(s_t = j|\Omega_{t-1}) = p_{1j}\text{Prob}(s_{t-1} = 1|\Omega_{t-1}) + p_{2j}\text{Prob}(s_{t-1} = 2|\Omega_{t-1}), \tag{9}$$

and $f(y_t|\Omega_{t-1})$ is the predictive density for GDP:

$$f(y_t|\Omega_{t-1}) = \sum_{i=1}^{2} \text{Prob}(s_t = i|\Omega_{t-1}) f(y_t|s_t = i, \Omega_{t-1}). \tag{10}$$

Given a value for $\text{Prob}(s_{t-1} = i|\Omega_{t-1})$, we can thus use (7) to calculate $\text{Prob}(s_t = j|\Omega_t)$, and proceed iteratively in this fashion through the data for $t = 1, 2, \ldots, T$ to calculate the necessary magnitude for forming the optimal nonlinear forecast given in (6).

Note that another by-product of this recursion is calculation in (10) of the predictive density for the observed data. Thus one could estimate the vector of unknown population parameters $\boldsymbol{\lambda} = (m_1, m_2, \sigma, p_{11}, p_{22})'$ by maximizing the log-likelihood function of the observed sample of GDP growth rates,

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{t=1}^{T} \log f(y_t|\Omega_{t-1}; \boldsymbol{\lambda}). \tag{11}$$

If the objective is to form an optimal inference about when the economy was in a recession, one can use the same principles to obtain an even better inference as more data accumulate. For example, an inference using data observed through date $t + k$ about the regime at date t is known as a k-period-ahead smoothed inference,

$$\text{Prob}(s_t = i | \Omega_{t+k}),$$

calculation of which will be explained in Eq. (22).

Though this is a trivially simple model, it seems to do a pretty good job at capturing what is being described by the NBER's business cycle chronology. If we select only those quarters for which the NBER declared the US economy to be in expansion, we calculate an average annual growth rate of 4.5%, suggesting a value for the parameter $m_1 = 4.5$. And we observe that if the NBER determined the economy to be in expansion in quarter t, 95% of the time it said the same thing in quarter $t + 1$, consistent with a value of $p_{11} = 0.95$. These values implied by the NBER chronology are summarized in column 3 of Table 1. On the other hand, if we ignore the NBER dates altogether, but simply maximize the log likelihood (11) of the observed GDP data alone, we end up with very similar estimates, as seen in column 4.

Moreover, even given the challenges of data revision, the one-quarter-ahead smoothed probabilities have an excellent out-of-sample record at tracking the NBER dates. Fig. 2 plots historical values for $\text{Prob}(s_t = 2 | \Omega_{t+1}, \hat{\boldsymbol{\lambda}}_{t+1})$ where only GDP data as they were actually released as of date $t + 1$ were used to estimate parameters and form the inference plotted for date t. Values before the vertical line are "simulated real-time" inferences from Chauvet and Hamilton (2006), that is, values calculated in 2005 using a separate historical real-time data vintage for each date t shown. Values after the vertical line are true real-time out-of-sample inferences as they have been published individually on www.econbrowser.com each quarter since 2005 without revision.

One attractive feature of this approach is that the linearity of the model conditional on s_t makes it almost as tractable as a fully linear model. For example, an optimal

Table 1 Parameter values for describing U.S. recessions

Parameter (1)	Interpretation (2)	Value from NBER classifications (3)	Value from GDP alone (4)
m_1	Average growth in expansion	4.5	4.62
m_2	Average growth in recession	−1.2	−0.48
σ	Standard deviation of growth	3.5	3.34
p_{11}	Prob. expansion continues	0.95	0.92
p_{22}	Prob. recession continues	0.78	0.74

Parameter estimates based on characteristics of expansions and recessions as classified by NBER (column 3), and values that maximize the observed sample log likelihood of postwar GDP growth rates (column 4), 1947:Q2–2004:Q2.
Source: Chauvet, M., Hamilton, J.D., 2006. Dating business cycle turning points. In: Costas Milas, P.R., van Dijk, D. (Eds.), Nonlinear Analysis of Business Cycles. Elsevier, Amsterdam, pp. 1–54.

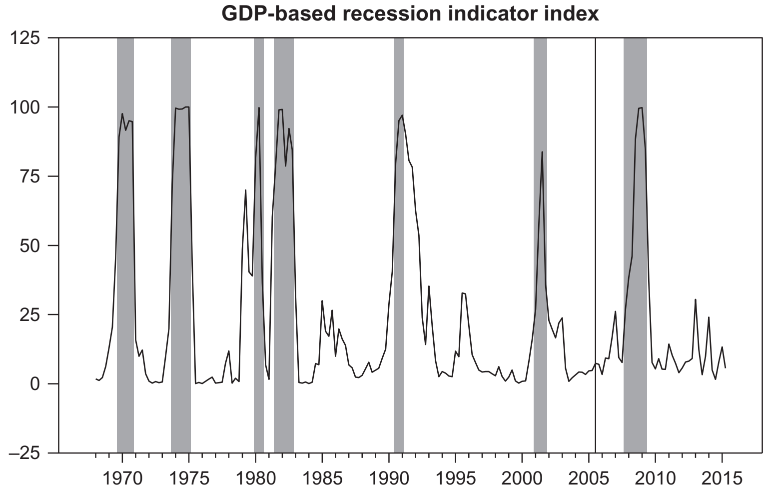

Fig. 2 One-quarter-ahead smoothed probabilities $\text{Prob}(s_t = 2|\Omega_{t+1}, \hat{\lambda}_{t+1})$, 1967:Q4–2014:Q2, as inferred using solely GDP data as reported as of date $t + 1$. *Shaded regions* correspond to NBER recession dates which were not used in any way in constructing the probabilities. Prior to 2005, each point on the graph corresponds to a simulated real-time inference that was constructed from a data set as it would have been available 4 months after the indicated date, as reported in Chauvet and Hamilton (2006). After 2005, points on the graph correspond to actual announcements that were publicly released 4 months after the indicated date. *Source: Updated from Hamilton, J.D., 2011. Calling recessions in real time. Int. J. Forecast. 27, 1006–1026 and www.econbrowser.com.*

k-period-ahead forecast of GDP growth based only on observed growth through date t can be calculated immediately using (4),

$$E(y_{t+k}|\Omega_t) = \mu + \phi^k \sum_{i=1}^{2} (m_i - \mu)\,\text{Prob}(s_t = i|\Omega_t) \tag{12}$$

for $\mu = a/(1-\phi)$. Results like this make this model of changes in regime very convenient to work with.

2. ECONOMETRIC TREATMENT OF CHANGES IN REGIME

This section discusses econometric inference for data that may be subject to changes in regime, while Section 3 examines methods to incorporate changes in regime into theoretical economic models.

2.1 Multivariate or Non-Gaussian Processes

Although the model in Section 1 was quite stylized, the same basic principles can be used to investigate changes in regime in much richer settings. Suppose we have a vector of variables $\mathbf{y}_t$ observed at date t and hypothesize that the density of $\mathbf{y}_t$ conditioned on its past history $\Omega_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots\}$ depends on parameters $\boldsymbol{\theta}$, some or all of which are different depending on the regime s_t:

$$f(\mathbf{y}_t | s_t = i, \Omega_{t-1}) = f(\mathbf{y}_t | \Omega_{t-1}; \boldsymbol{\theta}_i) \quad \text{for } i = 1, \ldots, N. \tag{13}$$

In the example in Section 1, there were $N = 2$ possible regimes with $\boldsymbol{\theta}_1 = (m_1, \sigma)'$, $\boldsymbol{\theta}_2 = (m_2, \sigma)'$, and $f(\mathbf{y}_t | \Omega_{t-1}; \boldsymbol{\theta}_i)$ the $N(m_i, \sigma^2)$ density. But the same basic approach would work for an n-dimensional vector autoregression in which some or all of the parameters change with the regime,

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\Phi}_{s_t,1}\mathbf{y}_{t-1} + \boldsymbol{\Phi}_{s_t,2}\mathbf{y}_{t-2} + \cdots + \boldsymbol{\Phi}_{s_t,r}\mathbf{y}_{t-r} + \mathbf{c}_{s_t} + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\Phi}_{s_t}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t \end{aligned} \tag{14}$$

$$\boldsymbol{\varepsilon}_t | s_t, \Omega_{t-1} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{s_t}), \tag{15}$$

a class of models discussed in detail in Krolzig (1997). Here $\mathbf{x}_{t-1}$ is an $(nr+1) \times 1$ vector consisting of a constant term and r lags of $\mathbf{y}$:

$$\mathbf{x}_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-r}, 1)'.$$

In this case the density of $\mathbf{y}_t$ conditional on its own past values and the regime s_t taking the value i would be

$$f(\mathbf{y}_t | s_t = i, \Omega_{t-1}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left[-(1/2)(\mathbf{y}_t - \boldsymbol{\Phi}_i \mathbf{x}_{t-1})' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_t - \boldsymbol{\Phi}_i \mathbf{x}_{t-1})\right]. \tag{16}$$

There is also no reason that a Gaussian density has to be used. For example, Dueker (1997) proposed a model of stock returns in which the innovation comes from a Student t distribution whose degrees of freedom parameter η changes with the regime.

2.2 Multiple Regimes

A convenient representation for a model with $N > 2$ regimes is obtained by collecting the transition probabilities in a matrix $\mathbf{P}$ whose row j, column i element corresponds to p_{ij} (so that columns of $\mathbf{P}$ sum to unity). We likewise summarize the regime at date t by an $(N \times 1)$ vector $\boldsymbol{\xi}_t$ whose ith element is unity when $s_t = i$ and is zero otherwise—in other words, $\boldsymbol{\xi}_t$ corresponds to column s_t of $\mathbf{I}_N$. Notice that $E(\boldsymbol{\xi}_t | s_{t-1} = i)$ has the interpretation

$$E(\boldsymbol{\xi}_t | s_{t-1} = i) = \begin{bmatrix} \text{Prob}(s_t = 1 | s_{t-1} = i) \\ \vdots \\ \text{Prob}(s_t = N | s_{t-1} = i) \end{bmatrix} = \begin{bmatrix} p_{i1} \\ \vdots \\ p_{iN} \end{bmatrix}$$

meaning

$$E(\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1}) = \mathbf{P}\boldsymbol{\xi}_{t-1}$$

and

$$\boldsymbol{\xi}_t = \mathbf{P}\boldsymbol{\xi}_{t-1} + \mathbf{v}_t$$

for $\mathbf{v}_t$ a discrete-valued martingale difference sequence whose elements always sum to zero. Thus the Markov chain admits a VAR(1) representation, with k-period-ahead regime probabilities conditional on the observed data $\boldsymbol{\Omega}_t$ given by

$$\begin{bmatrix} \mathrm{Prob}(s_{t+k} = 1|\boldsymbol{\Omega}_t) \\ \vdots \\ \mathrm{Prob}(s_{t+k} = N|\boldsymbol{\Omega}_t) \end{bmatrix} = \mathbf{P}^k \begin{bmatrix} \mathrm{Prob}(s_t = 1|\boldsymbol{\Omega}_t) \\ \vdots \\ \mathrm{Prob}(s_t = N|\boldsymbol{\Omega}_t) \end{bmatrix}. \tag{17}$$

Calculation of the moments and discussion of stationarity conditions for general processes subject to changes in regime can be found in Tjøstheim (1986), Yang (2000), Timmermann (2000), and Francq and Zakoïan (2001).

Although most applications assume a relatively small number of regimes, Sims and Zha (2006) used Bayesian prior information in a model with N as large as 10, while Calvet and Fisher (2004) estimated a model with thousands of regimes by imposing a functional restriction on the ways parameters vary across regimes.

2.3 Processes That Depend on Current and Past Regimes

In the original model proposed by Hamilton (1989) for describing economic recessions, the conditional density of GDP growth y_t was presumed to depend not just on the current regime but also on the r previous regimes:

$$y_t = m_{s_t} + \phi_1(y_{t-1} - m_{s_{t-1}}) + \phi_2(y_{t-2} - m_{s_{t-2}}) + \cdots + \phi_r(y_{t-r} - m_{s_{t-r}}) + \varepsilon_t. \tag{18}$$

While at first glance this might not appear to be a special case of the general formulation given in (13), this in fact is just a matter of representing (18) using the right notation. Taking $r = 1$ for illustration, define

$$s_t^* = \begin{cases} 1 & \text{when } s_t = 1 \text{ and } s_{t-1} = 1 \\ 2 & \text{when } s_t = 2 \text{ and } s_{t-1} = 1 \\ 3 & \text{when } s_t = 1 \text{ and } s_{t-1} = 2 \\ 4 & \text{when } s_t = 2 \text{ and } s_{t-1} = 2 \end{cases}.$$

Then s_t^* itself follows a four-state Markov chain with transition matrix

$$\mathbf{P}^* = \begin{bmatrix} p_{11} & 0 & p_{11} & 0 \\ p_{12} & 0 & p_{12} & 0 \\ 0 & p_{21} & 0 & p_{21} \\ 0 & p_{22} & 0 & p_{22} \end{bmatrix}$$

and the model (18) can indeed be viewed as a special case of (13), with for example

$$f(y_t|s_t^* = 2, \Omega_{t-1}) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{[y_t - m_2 - \phi_1(y_{t-1} - m_1)]^2}{2\sigma^2}\right\}.$$

2.4 Inference About Regimes and Evaluating the Likelihood for the General Case

For any of the examples above we could collect the set of possible densities conditional on one of N different possible regimes in an $(N \times 1)$ vector $\boldsymbol{\eta}_t$ whose ith element is $f(\mathbf{y}_t|s_t = i, \Omega_{t-1}; \boldsymbol{\lambda})$ for $\Omega_{t-1} = \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{y}_1\}$ and $\boldsymbol{\lambda}$ a vector consisting of all the unknown population parameters. For example, for the Markov-switching vector autoregression the ith element of $\boldsymbol{\eta}_t$ is given by (16) and $\boldsymbol{\lambda}$ is a vector collecting the unknown elements of $\{\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_N, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_N, \mathbf{P}\}$ for $\mathbf{P}$ the $(N \times N)$ matrix whose row j column i element is $\text{Prob}(s_{t+1} = j|s_t = i)$ (so columns of $\mathbf{P}$ sum to unity). We likewise can define the $(N \times 1)$ vector $\hat{\boldsymbol{\xi}}_{t|t}$ whose ith element is the probability $\text{Prob}(s_t = i|\Omega_t; \boldsymbol{\lambda})$. One goal is to take the inference $\hat{\boldsymbol{\xi}}_{t-1|t-1}$ and update it to calculate $\hat{\boldsymbol{\xi}}_{t|t}$ using the observation on $\mathbf{y}_t$. Hamilton (1994, p. 692) showed that this can be accomplished by calculating

$$\hat{\boldsymbol{\xi}}_{t|t-1} = \mathbf{P}\hat{\boldsymbol{\xi}}_{t-1|t-1}$$

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)}{\mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)} \tag{19}$$

where $\mathbf{1}$ denotes an $(N \times 1)$ vector of ones and $\odot$ denotes element-by-element vector multiplication.

If the Markov chain is known to be ergodic, we could begin the recursion for $t = 1$ by setting $\hat{\boldsymbol{\xi}}_{1|0}$ to the vector of unconditional probabilities, which as in Hamilton (1994, p. 684) can be found from the $(N+1)$th column of the matrix $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ for

$$\underset{(N+1)\times N}{\mathbf{A}} = \begin{bmatrix} \mathbf{I}_N - \mathbf{P} \\ \mathbf{1}' \end{bmatrix}. \tag{20}$$

Alternative options are to treat the initial probabilities as separate parameters,

$$\begin{bmatrix} \text{Prob}(s_1 = 1|\Omega_0) \\ \vdots \\ \text{Prob}(s_1 = N|\Omega_0) \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_N \end{bmatrix}$$

where ρ_i could reflect prior beliefs (eg, $\rho_1 = 1$ if the analyst knows the sample begins in regime 1), complete ignorance ($\rho_i = 1/N$ for $i = 1, \ldots, N$), or $\boldsymbol{\rho}$ could be a separate vector of parameters also to be chosen by maximum likelihood. Any of the last three options is particularly attractive if the EM algorithm described in Section 2.5 or Gibbs sampler in

Section 2.8 are used, or if one wants to allow the possibility of a permanent regime shift which would mean that the Markov chain is not ergodic.

In a generalization of (10) and (11), the log-likelihood function for the observed data is naturally calculated as a by-product of the above recursion:

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{t=1}^{T} \log f(\mathbf{y}_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\lambda}) = \sum_{t=1}^{T} \log[\mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)]. \tag{21}$$

From (17) the k-period-ahead forecast of the regime, $\text{Prob}(s_{t+k} = j|\boldsymbol{\Omega}_t;\boldsymbol{\lambda})$ is found from the jth element of $\mathbf{P}^k\hat{\boldsymbol{\xi}}_{t|t}$.

It is also often of interest to calculate an inference about the regime at date t conditional on the full set of all observations through the end of the sample T, known as the "smoothed probability." The smoothed $\text{Prob}(s_t = i|\boldsymbol{\Omega}_T;\boldsymbol{\lambda})$ is obtained from the ith element of $\hat{\boldsymbol{\xi}}_{t|T}$ which can be calculated as in Hamilton (1994, p. 694) by iterating backward for $t = T-1, T-2, \ldots, 1$ on

$$\hat{\boldsymbol{\xi}}_{t|T} = \hat{\boldsymbol{\xi}}_{t|t} \odot \{\mathbf{P}'[\hat{\boldsymbol{\xi}}_{t+1|T}(\div)\hat{\boldsymbol{\xi}}_{t+1|t}]\} \tag{22}$$

where $(\div)$ denotes element-by-element division.

2.5 EM Algorithm

The unknown parameters $\boldsymbol{\lambda}$ could be estimated by maximizing the likelihood function (21) using numerical search methods. Alternatively, Hamilton (1990) noted that the EM algorithm is often a convenient method for finding the maximum of the likelihood function. This algorithm is simplest if we treat the initial probabilities $\hat{\boldsymbol{\xi}}_{1|0}$ as a vector of free parameters $\boldsymbol{\rho}$ rather than using ergodic probabilities from (20). This section describes how the EM algorithm would be implemented for the case of an unrestricted Markov-switching VAR (14), in which case $\boldsymbol{\lambda}$ includes $\boldsymbol{\rho}$ along with the elements of $\{\boldsymbol{\Phi}_1,\ldots,\boldsymbol{\Phi}_N,\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_N,\mathbf{P}\}$. The EM algorithm is an iterative procedure for generating a sequence of estimates $\{\hat{\boldsymbol{\lambda}}^{(\ell)}\}$ where the algorithm guarantees that the log likelihood (21) evaluated at $\hat{\boldsymbol{\lambda}}^{(\ell+1)}$ is greater than or equal to that at $\hat{\boldsymbol{\lambda}}^{(\ell)}$. Iterating until convergence leads to a local maximum of the likelihood function.

To calculate the value of $\hat{\boldsymbol{\lambda}}^{(\ell+1)}$ we first use $\hat{\boldsymbol{\lambda}}^{(\ell)}$ in Eq. (22) to evaluate the smoothed probabilities $\text{Prob}(s_t = i|\boldsymbol{\Omega}_T;\hat{\boldsymbol{\lambda}}^{(\ell)})$ and also smoothed joint probabilities Prob $(s_t = i, s_{t+1} = j|\boldsymbol{\Omega}_T;\hat{\boldsymbol{\lambda}}^{(\ell)})$. The latter are obtained from the row i column j element of the $(N \times N)$ matrix[b]

[b] The row i column j element of this matrix corresponds to

$$\text{Prob}(s_t = i|\boldsymbol{\Omega}_t)\frac{\text{Prob}(s_{t+1} = j|\boldsymbol{\Omega}_T)}{\text{Prob}(s_{t+1} = j|\boldsymbol{\Omega}_t)}p_{ij}$$

which from equation [22.A.21] in Hamilton (1994) equals $\text{Prob}(s_t = i, s_{t+1} = j|\boldsymbol{\Omega}_T)$.

$$\{\hat{\boldsymbol{\xi}}_{t|t}(\hat{\boldsymbol{\lambda}}^{(\ell)})[\hat{\boldsymbol{\xi}}_{t+1|T}(\hat{\boldsymbol{\lambda}}^{(\ell)})(\div)(\hat{\mathbf{P}}^{(\ell)}\hat{\boldsymbol{\xi}}_{t|t}(\hat{\boldsymbol{\lambda}}^{(\ell)}))]'\}\odot\hat{\mathbf{P}}^{(\ell)'}. \tag{23}$$

We then use these smoothed probabilities to generate a new estimate $\hat{\boldsymbol{\rho}}^{(\ell+1)}$, whose ith element is obtained from $\text{Prob}(s_1 = i|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})$, along with a new estimate $\hat{\mathbf{P}}^{(\ell+1)}$ whose row j column i element is given by

$$\hat{p}_{ij}^{(\ell+1)} = \frac{\sum_{t=1}^{T-1}\text{Prob}(s_t = i, s_{t+1} = j|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})}{\sum_{t=1}^{T-1}\text{Prob}(s_t = i|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})}.$$

Updated estimates of the VAR parameters for $i = 1,\ldots,N$ are given by

$$\hat{\boldsymbol{\Phi}}_i^{(\ell+1)} = \left(\sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}'_{t-1}\text{Prob}(s_t = i|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})\right)\left(\sum_{t=1}^{T}\mathbf{x}_{t-1}\mathbf{x}'_{t-1}\text{Prob}(s_t = i|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})\right)^{-1} \tag{24}$$

$$\hat{\boldsymbol{\Sigma}}_i^{(\ell+1)} = \frac{\sum_{t=1}^{T}(\mathbf{y}_t - \hat{\boldsymbol{\Phi}}_i^{(\ell+1)}\mathbf{x}_{t-1})(\mathbf{y}_t - \hat{\boldsymbol{\Phi}}_i^{(\ell+1)}\mathbf{x}_{t-1})'\text{Prob}(s_t = i|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})}{\sum_{t=1}^{T}\text{Prob}(s_t = i|\Omega_T;\hat{\boldsymbol{\lambda}}^{(\ell)})}.$$

We thus simply iterate between calculating smoothed probabilities and OLS regressions of $\mathbf{y}_t$ on its lags weighted by those smoothed probabilities. The algorithm will converge to a point that is at least a local maximum of the log likelihood (21) with respect to $\boldsymbol{\lambda}$ subject to the constraints that $\boldsymbol{\rho}'\mathbf{1} = 1$, $\mathbf{1}'\mathbf{P} = \mathbf{1}'$, all elements of $\boldsymbol{\rho}$ and $\mathbf{P}$ are nonnegative, and $\boldsymbol{\Sigma}_j$ is positive semidefinite for $j = 1,\ldots,N$.

2.6 EM Algorithm for Restricted Models

Often we might want to use a more parsimonious representation to which the EM algorithm is easily adapted. For example, suppose that we assume that there are no changes in regime for the equations describing the first n_1 variables in the system:

$$\mathbf{y}_{1t} = \mathbf{A}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_{1t} \tag{25}$$

$$\mathbf{y}_{2t} = \mathbf{B}_{s_t}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_{2t} \tag{26}$$

$$E\left\{\begin{bmatrix}\boldsymbol{\varepsilon}_{1t}\\ \boldsymbol{\varepsilon}_{2t}\end{bmatrix}\begin{bmatrix}\boldsymbol{\varepsilon}'_{1t} & \boldsymbol{\varepsilon}'_{2t}\end{bmatrix}\bigg| s_t\right\} = \begin{bmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12,s_t}\\ \boldsymbol{\Sigma}_{21,s_t} & \boldsymbol{\Sigma}_{22,s_t}\end{bmatrix}.$$

As in Hamilton (1994, p. 310) it is convenient to reparameterize the system by premultiplying (25) by $\boldsymbol{\Sigma}_{21,s_t}\boldsymbol{\Sigma}_{11}^{-1}$ and subtracting the result from (26) to obtain

$$\mathbf{y}_{2t} = \mathbf{C}_{s_t}\mathbf{y}_{1t} + \mathbf{D}_{s_t}\mathbf{x}_{t-1} + \mathbf{v}_{2t} \tag{27}$$

where $\mathbf{C}_{s_t} = \boldsymbol{\Sigma}_{21,s_t}\boldsymbol{\Sigma}_{11}^{-1}$, $\mathbf{D}_{st} = \mathbf{B}_{s_t} - \boldsymbol{\Sigma}_{21,s_t}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{A}$, $\mathbf{v}_{2t} = \boldsymbol{\varepsilon}_{2t} - \boldsymbol{\Sigma}_{21,s_t}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\varepsilon}_{1t}$, and $E(\mathbf{v}_{2t}\mathbf{v}'_{2t}|s_t) = \mathbf{H}_{s_t} = \boldsymbol{\Sigma}_{22,s_t} - \boldsymbol{\Sigma}_{21,s_t}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12,s_t}$. Then the likelihood associated with the system (25) and (27) factors into a regime-switching component and a regime-independent component parameterized by $\mathbf{A}, \mathbf{\Sigma}_{11}$. In the absence of restrictions on $\mathbf{B}_{s_t}$, $\mathbf{\Sigma}_{21,s_t}$, and $\mathbf{\Sigma}_{22,s_t}$, the values for $\mathbf{A}$ and $\mathbf{\Sigma}_{11}$ do not restrict the likelihood for the regime-switching block, meaning full-information maximum likelihood for the complete system can be implemented by maximizing the likelihood separately for the two blocks. For the regime-independent block, the MLE is obtained by simple OLS:

$$\hat{\mathbf{A}} = \left(\sum_{t=1}^{T} \mathbf{y}_{1t}\mathbf{x}'_{t-1}\right)\left(\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}'_{t-1}\right)^{-1}$$

$$\hat{\Sigma}_{11} = T^{-1}\sum_{t=1}^{T}(\mathbf{y}_{1t} - \hat{\mathbf{A}}\boldsymbol{x}_{t-1})(\mathbf{y}_{1t} - \hat{\mathbf{A}}\boldsymbol{x}_{t-1})'.$$

The MLE for the regime-switching block can be found using the EM algorithm,

$$\hat{\mathbf{G}}_i^{(\ell+1)} = \left(\sum_{t=1}^{T} \mathbf{y}_{2t}\mathbf{z}'_t p_{it}^{(\ell)}\right)\left(\sum_{t=1}^{T} \mathbf{z}_t\mathbf{z}'_t p_{it}^{(\ell)}\right)^{-1}$$

$$\hat{\mathbf{H}}_i^{(\ell+1)} = \frac{\left(\sum_{t=1}^{T}(\mathbf{y}_{2t} - \hat{\mathbf{G}}_i^{(\ell+1)}\mathbf{z}_t)(\mathbf{y}_{2t} - \hat{\mathbf{G}}_i^{(\ell+1)}\mathbf{z}_t)' p_{it}^{(\ell)}\right)}{\left(\sum_{t=1}^{T} p_{it}^{(\ell)}\right)}$$

$$p_{it}^{(\ell)} = \mathrm{Prob}(s_t = i|\mathbf{\Omega}_T;\hat{\boldsymbol{\lambda}}^{(\ell)})$$

with $\mathbf{z}_t = (\mathbf{y}'_{1t},\mathbf{x}'_{t-1})'$ and $\mathbf{G}_j = \begin{bmatrix}\mathbf{C}_j & \mathbf{D}_j\end{bmatrix}$. The MLE for the original parameterization is then found simply by reversing the transformation that led to (27), for example, $\hat{\mathbf{\Sigma}}_{21,j} = \hat{\mathbf{C}}_j\hat{\mathbf{\Sigma}}_{11}$ and $\hat{\mathbf{B}}_j = \hat{\mathbf{D}}_j + \hat{\mathbf{\Sigma}}_{21,j}\hat{\mathbf{\Sigma}}_{11}^{-1}\hat{\mathbf{A}}$.

Alternatively, suppose we want to restrict the switching coefficients to apply only to a subset $\mathbf{x}_{2,t-1}$ of the original regressors, as for example in a VAR in which only the intercept (the last element of $\mathbf{x}_{t-1}$) is changing with regime,

$$\mathbf{y}_t = \mathbf{A}\mathbf{x}_{1,t-1} + \mathbf{B}_{s_t}\mathbf{x}_{2,t-1} + \boldsymbol{\varepsilon}_t \tag{28}$$

with $E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t) = \mathbf{\Sigma}$. In this case the EM equations take the form[c]

$$\begin{bmatrix}\hat{\mathbf{A}}^{(\ell+1)} & \hat{\mathbf{B}}_1^{(\ell+1)} & \hat{\mathbf{B}}_2^{(\ell+1)} & \cdots & \hat{\mathbf{B}}_N^{(\ell+1)}\end{bmatrix} = \mathbf{S}_{yx}(\hat{\boldsymbol{\lambda}}^{(\ell)})\mathbf{S}_{xx}^{-1}(\hat{\boldsymbol{\lambda}}^{(\ell)}) \tag{29}$$

$$\mathbf{S}_{yx}(\hat{\boldsymbol{\lambda}}^{(\ell)}) = \sum_{t=1}^{T}\mathbf{y}_t\begin{bmatrix}\mathbf{x}'_{1,t-1} & \mathbf{x}'_{2,t-1}p_{1t}^{(\ell)} & \mathbf{x}'_{2,t-1}p_{2t}^{(\ell)} & \cdots & \mathbf{x}'_{2,t-1}p_{Nt}^{(\ell)}\end{bmatrix}$$

[c] See the Appendix for more details.

$$\mathbf{S}_{xx}(\hat{\boldsymbol{\lambda}}^{(\ell)}) = \sum_{t=1}^{T} \begin{bmatrix} \mathbf{x}_{1,t-1}\mathbf{x}'_{1,t-1} & \mathbf{x}_{1,t-1}\mathbf{x}'_{2,t-1}p_{1t}^{(\ell)} & \mathbf{x}_{1,t-1}\mathbf{x}'_{2,t-1}p_{2t}^{(\ell)} & \cdots & \mathbf{x}_{1,t-1}\mathbf{x}'_{2,t-1}p_{Nt}^{(\ell)} \\ \mathbf{x}_{2,t-1}\mathbf{x}'_{1,t-1}p_{1t}^{(\ell)} & \mathbf{x}_{2,t-1}\mathbf{x}'_{2,t-1}p_{1t}^{(\ell)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{x}_{2,t-1}\mathbf{x}'_{1,t-1}p_{2t}^{(\ell)} & \mathbf{0} & \mathbf{x}_{2,t-1}\mathbf{x}'_{2,t-1}p_{2t}^{(\ell)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{x}_{2,t-1}\mathbf{x}'_{1,t-1}p_{Nt}^{(\ell)} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{2,t-1}\mathbf{x}'_{2,t-1}p_{Nt}^{(\ell)} \end{bmatrix}$$

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}^{(\ell+1)} = {} & T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{y}_t - \hat{\mathbf{A}}^{(\ell+1)}\mathbf{x}_{1,t-1} - \hat{\mathbf{B}}_i^{(\ell+1)}\mathbf{x}_{2,t-1}\right] \\ & \times\left[\mathbf{y}_t - \hat{\mathbf{A}}^{(\ell+1)}\mathbf{x}_{1,t-1} - \hat{\mathbf{B}}_i^{(\ell+1)}\mathbf{x}_{2,t-1}\right]' p_{it}^{(\ell)}. \end{aligned} \tag{30}$$

2.7 Structural Vector Autoregressions and Impulse-Response Functions

A Gaussian structural vector autoregression takes the form

$$\mathbf{A}_{s_t}\mathbf{y}_t = \mathbf{B}_{s_t}\mathbf{x}_{t-1} + \mathbf{u}_t$$

where $\mathbf{x}_{t-1} = (\mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \ldots, \mathbf{y}'_{t-r}, 1)'$ and $\mathbf{u}_t|s_t, \boldsymbol{\Omega}_{t-1} \sim N(\mathbf{0}, \mathbf{D}_{s_t})$. Here the elements of $\mathbf{u}_t$ are interpreted as different structural shocks which are identified by imposing certain restrictions on $\mathbf{A}_i$, $\mathbf{B}_i$, and $\mathbf{D}_i$. For example, the common Cholesky identification assumes that the structural equations are recursive, with $\mathbf{D}_i$ diagonal and $\mathbf{A}_i$ lower triangular with ones along the diagonal. For an identified structure we could estimate parameters by setting the ith element of $\boldsymbol{\eta}_t$ in (19) to

$$\eta_{it} = \frac{1}{(2\pi)^{n/2}}\frac{\sqrt{|\mathbf{A}_i|^2}}{\sqrt{|\mathbf{D}_i|}}\exp\left[-(1/2)(\mathbf{A}_i\mathbf{y}_t - \mathbf{B}_i\mathbf{x}_{t-1})'\mathbf{D}_i^{-1}(\mathbf{A}_i\mathbf{y}_t - \mathbf{B}_i\mathbf{x}_{t-1})\right]$$

and then choosing $\{\mathbf{A}_1,\ldots,\mathbf{A}_N,\mathbf{B}_1,\ldots,\mathbf{B}_N,\mathbf{D}_1,\ldots,\mathbf{D}_N,\mathbf{P}\}$ to maximize the likelihood (21).

A faster algorithm is likely to be obtained by first finding the MLEs $\{\hat{\boldsymbol{\Phi}}_1,\ldots,\hat{\boldsymbol{\Phi}}_N,\hat{\boldsymbol{\Sigma}}_1,\ldots,\hat{\boldsymbol{\Sigma}}_N,\hat{\mathbf{P}},\hat{\boldsymbol{\rho}}\}$ for the reduced form (14)–(15) using the EM algorithm in Section 2.5. If the model is just identified, we can just translate these into the implied structural parameters $\{\hat{\mathbf{A}}_1,\ldots,\hat{\mathbf{A}}_N,\hat{\mathbf{B}}_1,\ldots,\hat{\mathbf{B}}_N,\hat{\mathbf{D}}_1,\ldots,\hat{\mathbf{D}}_N,\mathbf{P},\hat{\rho}\}$, while for an overidentified model we could find the values for the structural parameters that are closest to the reduced form using minimum chi-square estimation (eg, Hamilton and Wu, 2012). For example, for the Cholesky formulation we would just find the Cholesky factorization $\hat{\mathbf{P}}_i\hat{\mathbf{P}}'_i = \hat{\Sigma}_i$ for each i. The row j column j element of $\hat{\mathbf{D}}_i$ is then the square of the row j column j element of $\hat{\mathbf{P}}_i$. Then $\hat{\mathbf{A}}_i = \hat{\mathbf{D}}_i^{1/2}\hat{\mathbf{P}}_i^{-1}$ and $\hat{\mathbf{B}}_i = \hat{\mathbf{A}}_i\hat{\boldsymbol{\Phi}}_i$.

Users of structural vector autoregressions are often interested in structural impulse-response functions, which in this case are functions of the regime at date t:

$$\mathbf{H}_{mj} = \frac{\partial E(\mathbf{y}_{t+m}|s_t = j, \mathbf{\Omega}_t)}{\partial \mathbf{u}'_t} = \frac{\partial E(\mathbf{y}_{t+m}|s_t = j, \mathbf{\Omega}_t)}{\partial \boldsymbol{\varepsilon}'_t} \frac{\partial \boldsymbol{\varepsilon}_t}{\partial \mathbf{u}'_t} = \mathbf{\Psi}_{mj}\mathbf{A}_j^{-1}.$$

The nonorthogonalized or reduced-form IRF, $\mathbf{\Psi}_{mj}$, can be found as follows. Suppose we first condition not just on the regime j at date t but also on a particular regime j_1 for date $t + 1, j_2$ for $t + 2$, and j_m for $t + m$, and consider the value of

$$\tilde{\mathbf{\Psi}}_{m,j,j_1,\ldots,j_m} = \frac{\partial E(\mathbf{y}_{t+m}|s_t = j, s_{t+1} = j_1, \ldots, s_{t+m} = j_m, \mathbf{\Omega}_t)}{\partial \boldsymbol{\varepsilon}'_t}.$$

Karamé (2010) noted that this $(n \times n)$ matrix can be calculated from the recursion

$$\tilde{\mathbf{\Psi}}_{m,j,j_1,\ldots,j_m} = \mathbf{\Phi}_{1,j_m}\tilde{\mathbf{\Psi}}_{m-1,j,j_1,\ldots,j_{m-1}} + \mathbf{\Phi}_{2,j_m}\tilde{\mathbf{\Psi}}_{m-2,j,j_1,\ldots,j_{m-2}} + \cdots + \mathbf{\Phi}_{r,j_m}\tilde{\mathbf{\Psi}}_{m-r,j,j_1,\ldots,j_{m-r}}$$

for $m = 1,2,\ldots$ where $\tilde{\mathbf{\Psi}}_{0j} = \mathbf{I}_n$ and $\mathbf{0} = \tilde{\mathbf{\Psi}}_{-1,.} = \tilde{\mathbf{\Psi}}_{-2,.} = \cdots$. The object of interest is found by integrating out the conditioning variables,

$$\begin{aligned}\mathbf{\Psi}_{mj} &= \sum_{j_1=1}^{N}\cdots\sum_{j_m=1}^{N}\tilde{\mathbf{\Psi}}_{m,j,j_1,\ldots,j_m}\text{Prob}(s_{t+1} = j_1, \ldots, s_{t+m} = j_m|s_t = j)\\ &= \sum_{j_1=1}^{N}\cdots\sum_{j_m=1}^{N}\tilde{\mathbf{\Psi}}_{m,j,j_1,\ldots,j_m}p_{j,j_1}p_{j_1,j_2}\cdots p_{j_{m-1},j_m}.\end{aligned}$$

These magnitudes can either be calculated analytically for modest m and N or by simulation.

Such regime-specific impulse-response functions are of interest for questions such as whether monetary policy (Lo and Piger, 2005) or fiscal policy (Auerbach and Gorodnichenko, 2012) has different effects on the economy during an expansion or recession.

2.8 Bayesian Inference and the Gibbs Sampler

Bayesian methods offer another popular approach for econometric inference. The Bayesian begins with prior beliefs about the unknown parameters $\boldsymbol{\lambda}$ which are represented using a probability density $f(\boldsymbol{\lambda})$ that associates a higher probability with values of $\boldsymbol{\lambda}$ that are judged to be more plausible. The goal of inference is to revise these beliefs in the form of a posterior density $f(\boldsymbol{\lambda}|\mathbf{\Omega}_T)$ based on the observed data $\mathbf{\Omega}_T = \{\mathbf{y}_1,\ldots,\mathbf{y}_T\}$. Often the prior distribution $f(\boldsymbol{\lambda})$ is assumed to be taken from a particular parametric family known as a natural conjugate distribution. These have the property that the prior and posterior are from the same family, as would be the case for example if the prior beliefs were based on an earlier sample of data. Natural conjugates are helpful because they allow many of the results to be obtained using known analytic solutions.

Again we will illustrate some of the main ideas using a Markov-switching vector autoregression:

$$\mathbf{y}_t = \mathbf{\Phi}_{s_t}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\varepsilon}_t | s_t = i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$$
$$\text{Prob}(s_1 = i) = \rho_i$$
$$\text{Prob}(s_t = j | s_{t-1} = i) = p_{ij}.$$

2.8.1 Prior Distributions

The Dirichlet distribution is the natural conjugate for the parameters that determine the Markov transition probabilities. Suppose $\mathbf{z} = (z_1, \ldots, z_N)'$ is an $(N \times 1)$ vector of nonnegative random variables that sum to unity. The Dirichlet density with parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)'$ is given by

$$f(\mathbf{z}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_N)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_N)} z_1^{\alpha_1 - 1} \cdots z_N^{\alpha_N - 1}$$

for $\Gamma(.)$ the gamma function, with the constant ensuring that the density integrates to unity over the set of vectors $\mathbf{z}$ satisfying the specified conditions. The beta distribution is a special case when $N = 2$, usually expressed as a function of the scalar $z_1 \in (0,1)$,

$$f(z_1) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} z_1^{\alpha_1 - 1} (1 - z_1)^{\alpha_2 - 1}.$$

We then represent prior beliefs over the $(N \times 1)$ vector of initial probabilities as $(\rho_1, \ldots, \rho_N) \sim D(\alpha_1, \ldots, \alpha_N)$ and those for transition probabilities as $(p_{i1}, \ldots, p_{iN}) \sim D(\alpha_{i1}, \ldots, \alpha_{iN})$ for $i = 1, \ldots, N$. The natural conjugate for $\boldsymbol{\Sigma}_j$, the innovation variance matrix for regime j, is provided by the Wishart distribution. Let $\mathbf{z}_i$ be independent $(n \times 1)$ $N(\mathbf{0}, \boldsymbol{\Lambda}^{-1})$ vectors and consider the matrix $\mathbf{W} = \mathbf{z}_1\mathbf{z}_1' + \cdots + \mathbf{z}_\eta\mathbf{z}_\eta'$ for $\eta > n - 1$. This matrix is said to have an n-dimensional Wishart distribution with η degrees of freedom and scale matrix $\boldsymbol{\Lambda}^{-1}$, whose density is

$$f(\mathbf{W}) = c|\boldsymbol{\Lambda}|^{\eta/2} |\mathbf{W}|^{(\eta - n - 1)/2} \exp[-(1/2)\text{tr}(\mathbf{W}\boldsymbol{\Lambda})]$$

where tr(.) denotes the trace (sum of diagonal elements). For a univariate regression ($n = 1$) this becomes Λ^{-1} times a $\chi^2(\eta)$ variable, or equivalently a gamma distribution with mean η/Λ and variance $2\eta/\Lambda^2$. The constant c is chosen so that the density integrates to unity over the set of all positive definite symmetric matrices $\mathbf{W}$ (eg, DeGroot, 1970, p. 57):

$$c = \left[2^{\eta n/2} \pi^{n(n-1)/4} \prod_{j=1}^{n} \Gamma\left(\frac{\eta + 1 - j}{2} \right) \right]^{-1}.$$

The natural conjugate prior for $\boldsymbol{\Sigma}_j^{-1}$, the inverse of the innovation variance matrix in regime j, takes the form of a Wishart distribution with η_j degrees of freedom and scale $\boldsymbol{\Lambda}_j^{-1}$:

$$f(\boldsymbol{\Sigma}_j^{-1}) = c|\boldsymbol{\Lambda}_j|^{\eta_j/2}|\boldsymbol{\Sigma}_j|^{-(\eta_j-n-1)/2}\exp\left[-(1/2)\mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Lambda}_j\right)\right].$$

Prior information about the regression coefficients $\boldsymbol{\varphi}_j = \mathrm{vec}(\boldsymbol{\Phi}_j')$ for regime j can be represented with an $N(\mathbf{m}_j,\mathbf{M}_j)$ distribution. The formulas are much simpler in the case of no useful prior information about these coefficients (which can be viewed as the limit of the inference as $\mathbf{M}_j^{-1} \to \mathbf{0}$), and this limiting case will be used for the results presented here.

2.8.2 Likelihood Function and Conditional Posterior Distributions

Collect the parameters that characterize Markov probabilities in a set $p = \{\rho_j, p_{1j}, \ldots, p_{Nj}\}_{j=1}^N$, those for variances in a set $\sigma = \{\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_N\}$, and VAR coefficients $\varphi = \{\boldsymbol{\Phi}_1,\ldots,\boldsymbol{\Phi}_N\}$. If we were to condition on all of these parameters along with a particular numerical value for the realization of the regime for every date $\mathcal{S} = \{s_1, \ldots, s_T\}$ the likelihood function of the observed data $\boldsymbol{\Omega}_T = \{\mathbf{y}_1,\ldots,\mathbf{y}_T\}$ would be

$$\begin{aligned}
f(\boldsymbol{\Omega}_T|p,\sigma,\varphi,\mathcal{S}) &= \prod_{t=1}^T \frac{1}{(2\pi)^{n/2}}|\boldsymbol{\Sigma}_{s_t}|^{-1/2}\exp\left[-(1/2)(\mathbf{y}_t-\boldsymbol{\Phi}_{s_t}\mathbf{x}_{t-1})'\boldsymbol{\Sigma}_{s_t}^{-1}(\mathbf{y}_t-\boldsymbol{\Phi}_{s_t}\mathbf{x}_{t-1})\right] \\
&= \prod_{t=1}^T \frac{1}{(2\pi)^{n/2}}\sum_{j=1}^N \delta_{jt}|\boldsymbol{\Sigma}_j|^{-1/2}\exp\left[-(1/2)(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})'\boldsymbol{\Sigma}_j^{-1}(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})\right]
\end{aligned}$$

where $\delta_{jt} = 1$ if $s_t = j$ and is zero otherwise. With independent priors the joint density of the data, parameters, and regimes is then

$$f(\boldsymbol{\Omega}_T,p,\sigma,\varphi,\mathcal{S}) = f(\boldsymbol{\Omega}_T|p,\sigma,\varphi,\mathcal{S})f(p)f(\sigma)f(\varphi)f(\mathcal{S}|p) \tag{31}$$

$$f(p) \propto \prod_{j=1}^N \rho_j^{\alpha_j-1}p_{1j}^{\alpha_{1j}-1}\cdots p_{Nj}^{\alpha_{Nj}-1}$$

$$f(\sigma) \propto \prod_{j=1}^N |\boldsymbol{\Sigma}_j|^{-(\eta_j-n-1)/2}\exp\left[-(1/2)\mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Lambda}_j\right)\right]$$

$$f(\varphi) \propto 1$$

$$f(\mathcal{S}|p) = \rho_{s_1}p_{s_1,s_2}p_{s_2,s_3}\cdots p_{s_{T-1},s_T}$$

where p_{s_{t-1},s_t} denotes the parameter p_{ij} when $s_{t-1} = i$ and $s_t = j$.

Let $\Delta(j) = \{t \in 1,\ldots,T : \delta_{jt} = 1\}$ denote the set of dates for which the regime is j. From (31) the posterior distribution of $\boldsymbol{\Sigma}_j$ conditional on $\boldsymbol{\Omega}_T, p, \varphi, \mathcal{S}$ is given by

$$f(\boldsymbol{\Sigma}_j^{-1}|\Omega_T,p,\varphi,\mathcal{S}) \propto |\boldsymbol{\Sigma}_j|^{-(\eta_j-n-1)/2}\exp\left[-(1/2)\mathrm{tr}\left(\boldsymbol{\Sigma}_j^{-1}\boldsymbol{\Lambda}_j\right)\right]\times$$
$$\prod_{t\in\Delta(j)}|\boldsymbol{\Sigma}_j|^{-1/2}\exp\left[-(1/2)(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})'\boldsymbol{\Sigma}_j^{-1}(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})\right]$$
$$=|\boldsymbol{\Sigma}_j|^{-(T_j+\eta_j-n-1)/2}\exp\left[-(1/2)\mathrm{tr}[\boldsymbol{\Sigma}_j^{-1}(\boldsymbol{\Lambda}_j+\mathbf{H}_j)\right] \tag{32}$$

for $T_j=\sum_{t=1}^T\delta_{jt}$ the number of dates characterized by regime j and $\mathbf{H}_j=\sum_{t=1}^T\delta_{jt}(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})'$ the sum of outer products of the residual vectors for those observations. In other words, $\boldsymbol{\Sigma}_j^{-1}|\Omega_T,p,\varphi,\mathcal{S}$ has a Wishart distribution with $T_j+\eta_j$ degrees of freedom and scale matrix $(\boldsymbol{\Lambda}_j+\mathbf{H}_j)^{-1}$.

Likewise for $\hat{\boldsymbol{\Phi}}_j=\left(\sum_{t=1}^T\delta_{jt}\mathbf{y}_t\mathbf{x}_{t-1}'\right)\left(\sum_{t=1}^T\delta_{jt}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}$, that is for $\hat{\boldsymbol{\Phi}}_j$ the OLS regression coefficients using only observations for regime j, the posterior distribution of $\boldsymbol{\varphi}_j=\mathrm{vec}(\boldsymbol{\Phi}_j')$ conditional on $\Omega_T,p,\sigma,\mathcal{S}$ is

$$f(\boldsymbol{\varphi}_j|\Omega_T,p,\sigma,\mathcal{S}) \propto \prod_{t\in\Delta(j)}\exp\left[-(1/2)(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})'\boldsymbol{\Sigma}_j^{-1}(\mathbf{y}_t-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})\right]$$
$$=\prod_{t\in\Delta(j)}\exp\left[-(1/2)(\mathbf{y}_t-\hat{\boldsymbol{\Phi}}_j\mathbf{x}_{t-1}+\hat{\boldsymbol{\Phi}}_j\mathbf{x}_{t-1}-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})'\boldsymbol{\Sigma}_j^{-1}\times\right.$$
$$\left.(\mathbf{y}_t-\hat{\boldsymbol{\Phi}}_j\mathbf{x}_{t-1}+\hat{\boldsymbol{\Phi}}_j\mathbf{x}_{t-1}-\boldsymbol{\Phi}_j\mathbf{x}_{t-1})\right]$$
$$\propto\prod_{t\in\Delta(j)}\exp\left[-(1/2)\mathbf{x}_{t-1}'(\hat{\boldsymbol{\Phi}}_j-\boldsymbol{\Phi}_j)'\boldsymbol{\Sigma}_j^{-1}(\hat{\boldsymbol{\Phi}}_j-\boldsymbol{\Phi}_j)\mathbf{x}_{t-1}\right]$$
$$=\prod_{t\in\Delta(j)}\exp\left\{-(1/2)(\hat{\boldsymbol{\varphi}}_j-\boldsymbol{\varphi}_j)'[\boldsymbol{\Sigma}_j^{-1}\otimes\mathbf{x}_{t-1}\mathbf{x}_{t-1}'](\hat{\boldsymbol{\varphi}}_j-\boldsymbol{\varphi}_j)\right\}$$
$$=\exp\left\{-(1/2)(\boldsymbol{\varphi}_j-\hat{\boldsymbol{\varphi}}_j)'\left[\boldsymbol{\Sigma}_j\otimes\left(\sum_{t=1}^T\delta_{jt}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}\right]^{-1}(\boldsymbol{\varphi}_j-\hat{\boldsymbol{\varphi}}_j)\right\}$$

establishing that $\boldsymbol{\varphi}_j|\Omega_T,p,\sigma,\mathcal{S}\sim N\left(\hat{\boldsymbol{\varphi}}_j,\boldsymbol{\Sigma}_j\otimes\left(\sum_{t=1}^T\delta_{jt}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\right)^{-1}\right)$ for $\hat{\boldsymbol{\varphi}}_j=\mathrm{vec}(\hat{\boldsymbol{\Phi}}_j')$.

The conditional posterior distribution of $\boldsymbol{\rho}$ is found from

$$f(\boldsymbol{\rho}|\Omega_T,\varphi,\sigma,\mathcal{S})\propto\rho_1^{a_1-1}\cdots\rho_N^{\alpha_N-1}\rho_{s_t}$$

which will be recognized as $D(\alpha_1+\delta_{11},\ldots,\alpha_N+\delta_{N1})$, in other words, a Dirichlet distribution in which we have increased the parameter associated with the realized regime for observation 1 by unity and kept all other parameters the same. We similarly have $(p_{i1},\ldots,p_{iN})|\Omega_T,\varphi,\sigma,\mathcal{S}\sim D(\alpha_{i1}+T_{i1},\ldots,\alpha_{iN}+T_{iN})$ for $T_{ij}=\sum_{t=2}^T\delta_{i,t-1}\delta_{jt}$ the number of times that regime i is followed by j in the given sequence $\mathcal{S}$.

2.8.3 Gibbs Sampler

The idea behind the Gibbs sampler is to take advantage of the above known conditional distributions to generate a sequence of random variables whose unconditional distribution will turn out to be the object we're interested in. Suppose that as the result of a previous iteration ℓ we had generated particular numerical values for $\varphi, \sigma, p, \mathcal{S}$. We could for example begin iteration $\ell = 1$ with arbitrary initial guesses for the parameters along with a possible realization of the regime for each date. Given the numbers from iteration ℓ, we could generate $\boldsymbol{\Sigma}_j^{(\ell+1)}$ from expression (32), namely, $\boldsymbol{\Sigma}_j^{(\ell+1)}$ is the inverse of a draw from a Wishart distribution with $T_j^{(\ell)} + \eta_j$ degrees of freedom and scale matrix $\left(\boldsymbol{\Lambda}_j + \mathbf{H}_j^{(\ell)}\right)^{-1}$, where $T_j^{(\ell)} = \sum_{t=1}^{T} \delta_{jt}^{(\ell)}$ is a simple count of the number of elements in $\{s_1^{(\ell)}, \ldots, s_T^{(\ell)}\}$ that take the value j and $\mathbf{H}_j^{(\ell)}$ is the sum of the residual outer products $\sum_{t=1}^{T} \delta_{jt}^{(\ell)} (\mathbf{y}_t - \boldsymbol{\Phi}_j^{(\ell)} \mathbf{x}_{t-1})$ $(\mathbf{y}_t - \boldsymbol{\Phi}_j^{(\ell)} \mathbf{x}_{t-1})'$ for those $T_j^{(\ell)}$ observations. Doing so for each $j = 1, \ldots, N$ gives us the new $\sigma^{(\ell+1)}$. We get a new value for the VAR coefficients by generating $\boldsymbol{\varphi}_j^{(\ell+1)} \sim N\left(\hat{\boldsymbol{\varphi}}_j^{(\ell+1)}, \boldsymbol{\Sigma}_j^{(\ell+1)} \otimes \left(\sum_{t=1}^{T} \delta_{jt}^{(\ell)} \mathbf{x}_{t-1} \mathbf{x}_{t-1}'\right)^{-1}\right)$ where $\hat{\boldsymbol{\varphi}}_j^{(\ell+1)} = \operatorname{vec}\left[\hat{\boldsymbol{\Phi}}_j^{(\ell+1)'}\right]$ is obtained from OLS regression on these $T_j^{(\ell)}$ observations: $\hat{\boldsymbol{\Phi}}_j^{(\ell+1)} = \left(\sum_{t=1}^{T} \delta_{jt}^{(\ell)} \mathbf{y}_t \mathbf{x}_{t-1}'\right) \left(\sum_{t=1}^{T} \delta_{jt}^{(\ell)} \mathbf{x}_{t-1} \mathbf{x}_{t-1}'\right)^{-1}$. New initial probabilities $(\rho_1^{(\ell+1)}, \ldots, \rho_N^{(\ell+1)})$ are generated from $D(\alpha_1 + \delta_{11}^{(\ell)}, \ldots, \alpha_N + \delta_{N1}^{(\ell)})$ and new Markov probabilities $(p_{i1}^{(\ell+1)}, \ldots, p_{iN}^{(\ell+1)})$ from $D(\alpha_{i1} + T_{i1}^{(\ell)}, \ldots, \alpha_{iN} + T_{iN}^{(\ell)})$ for $T_{ij}^{(\ell)}$ the number of times $s_t^{(\ell)} = i$ is followed by $s_{t+1}^{(\ell)} = j$ within the particular realization $(s_1^{(\ell)}, \ldots, s_T^{(\ell)})$.

Finally, we can get a new realization $(s_1^{(\ell+1)}, \ldots, s_T^{(\ell+1)})$ as a draw from the conditional posterior $f(\mathcal{S} | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$ by iterating backward on a variant of the smoothing algorithm in Section 2.4. Specifically, given the values $(p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$ we can iterate on (19) for $t = 1, \ldots, T$ to calculate the $(N \times 1)$ vector $\{\hat{\boldsymbol{\xi}}_{t|t}^{(\ell+1)}\}_{t=1}^{T}$ whose jth element is $\text{Prob}(s_t = j | \Omega_t, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$. To generate $s_T^{(\ell+1)}$, we first generate a $U(0,1)$ variate. If this is smaller than the calculated $\text{Prob}(s_T = 1 | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$, we set $s_T^{(\ell+1)} = 1$. If the uniform variable turns out to be between $\text{Prob}(s_T = 1 | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$ and the sum $\text{Prob}(s_T = 1 | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$ + $\text{Prob}(s_T = 2 | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})$, we set $s_T^{(\ell+1)} = 2$, and so on. After we have generated a particular value for $s_T^{(\ell+1)}$ we can use (23) to calculate the probability

$$\begin{aligned}
\text{Prob}(s_{T-1} &= i | s_T = s_T^{(\ell+1)}, \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)}) = \\
&\frac{\text{Prob}(s_{T-1} = i, s_T = s_T^{(\ell+1)} | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})}{\text{Prob}(s_T = s_T^{(\ell+1)} | \Omega_T, p^{(\ell+1)}, \sigma^{(\ell+1)}, \varphi^{(\ell+1)})}
\end{aligned}$$

with which we generate a draw $s_{T-1}^{(\ell+1)}$. Iterating backward in this manner gives us the full sequence $\mathcal{S}^{(\ell+1)}$.

We now have a complete new set $p^{(\ell+1)}, \boldsymbol{\sigma}^{(\ell+1)}, \boldsymbol{\varphi}^{(\ell+1)}, \mathcal{S}^{(\ell+1)}$, from which we can then generate values for $\ell+2, \ell+3$, and so on. The idea behind the Gibbs sampler (eg, Albert and Chib, 1993) is that the sequence corresponds to a Markov chain whose ergodic distribution under general conditions is the true posterior distribution $f(p, \boldsymbol{\sigma}, \boldsymbol{\varphi}, \mathcal{S}|\boldsymbol{\Omega}_T)$. The proposal is then to discard the first say 10^6 draws and retain the next 10^6 draws as a sample from the posterior distribution.

One can also adapt approaches like those in Section 2.6 to apply the Gibbs sampler to restricted models. For example, if regime switching is confined to a subset of the equations, we can use the parameterization (27) and perform inference on the regime-switching subset independently from the rest of the system.

Although very convenient for many applications, one caution to be aware of in applying the Gibbs sampler is the role of label switching. Strategies for dealing with this are discussed by Celeux et al. (2000), Frühwirth-Schnatter (2001), and Geweke (2007).

2.9 Time-Varying Transition Probabilities

While the calculations above assumed that regimes are characterized by an exogenous Markov chain, this is easily generalized. We could replace (2) with

$$\text{Prob}(s_t = j | s_{t-1} = i, s_{t-2} = k, \ldots, \boldsymbol{\Omega}_{t-1}) = p_{ij}(\mathbf{x}_{t-1}; \boldsymbol{\lambda}) \quad i, j = 1, \ldots, N \tag{33}$$

where $\mathbf{x}_{t-1}$ is a subset of $\boldsymbol{\Omega}_{t-1}$ or other observed variables on which one is willing to condition and $p_{ij}(\mathbf{x}_{t-1}; \boldsymbol{\lambda})$ is a specified parametric function. The generalization of (9) then becomes

$$\text{Prob}(s_t = j | \boldsymbol{\Omega}_{t-1}) = \sum_{i=1}^{N} p_{ij}(\mathbf{x}_{t-1}; \boldsymbol{\lambda}) \text{Prob}(s_{t-1} = i | \boldsymbol{\Omega}_{t-1}),$$

where the sequence $\text{Prob}(s_t = i | \boldsymbol{\Omega}_t)$ can still be calculated iteratively as in (7),

$$\text{Prob}(s_t = j | \boldsymbol{\Omega}_t) = \frac{\text{Prob}(s_t = j | \boldsymbol{\Omega}_{t-1}) f(\mathbf{y}_t | \boldsymbol{\Omega}_{t-1}; \boldsymbol{\theta}_j)}{f(\mathbf{y}_t | \boldsymbol{\Omega}_{t-1})} \tag{34}$$

with the predictive density in the denominator now

$$f(\mathbf{y}_t | \boldsymbol{\Omega}_{t-1}) = \sum_{i=1}^{N} \text{Prob}(s_t = i | \boldsymbol{\Omega}_{t-1}) f(\mathbf{y}_t | \boldsymbol{\Omega}_{t-1}; \boldsymbol{\theta}_i). \tag{35}$$

Diebold et al. (1994) showed how the EM algorithm works in such a setting, while Filardo and Gordon (1998) developed a Gibbs sampler. Other interesting applications with time-varying transition probabilities include Filardo (1994) and Peria (2002).

2.10 Latent-Variable Models with Changes in Regime

A more involved case that cannot be handled using the above devices is when the conditional density of $\mathbf{y}_t$ depends on the full history of regimes $(s_t, s_{t-1}, \ldots, s_1)$ through date t. One important case in which this arises is when a process moving in and out of recession phase is proposed as an unobserved latent variable influencing an $(n \times 1)$ vector of observed variables $\mathbf{y}_t$. For example, Chauvet (1998) specified a process for an unobserved scalar business-cycle factor F_t characterized by

$$F_t = \alpha_{s_t} + \phi F_{t-1} + \eta_t$$

which influences the observed $\mathbf{y}_t$ according to

$$\mathbf{y}_t = \boldsymbol{\psi} F_t + \mathbf{q}_t$$

for $\boldsymbol{\psi}$ an $(n \times 1)$ vector of factor loadings and elements of $\mathbf{q}_t$ presumed to follow separate autoregressions. This can be viewed as a state-space model with regime-dependent parameters in which the conditional density (13) turns out to depend on the complete history $(s_t, s_{t-1}, \ldots, s_1)$.

One approach for handling such models is an approximation to the log likelihood and optimal inference developed by Kim (1994). Chauvet and Hamilton (2006) and Chauvet and Piger (2008) demonstrated the real-time usefulness of this approach for recognizing US recessions with $\mathbf{y}_t$ a (4×1) vector of monthly indicators of sales, income, employment, and industrial production, while Camacho et al. (2014) have had success using a more detailed model for the Euro area.

The Gibbs sampler offers a particularly convenient approach for this class of models. We simply add the unobserved sequence of factors $\{F_1, \ldots, F_T\}$ as another random block to be sampled from along with p, σ, φ, and $\mathcal{S}$. Conditional on $\{F_1, \ldots, F_T\}$, draws for those other blocks can be performed exactly as in Section 2.8, while draws for $\{F_1, \ldots, F_T\}$ conditional on the regimes and other parameters can be calculated using well-known algorithms associated with the Kalman filter; see Kim and Nelson (1999a) for details.

2.11 Selecting the Number of Regimes

Often one would want to test the null hypothesis that there are N regimes against the alternative of $N+1$, and in particular to test the null hypothesis that there are no changes in regime at all $(H_0 : N = 1)$. A natural idea would be to compare the values achieved for the log likelihood (21) for N and $N+1$. Unfortunately, the likelihood ratio does not have the usual asymptotic χ^2 distribution because under the null hypothesis, some of the parameters of the model become unidentified. For example, if one thought of the null hypothesis in (1) as $m_1 = m_2$, when the null is true the maximum likelihood estimates $\hat{p}_{11}$ and $\hat{p}_{22}$ do not converge to any population values. Hansen (1992) and Garcia (1998) examined the distribution of the likelihood ratio statistic in this setting, though implementing their procedures can be quite involved if the model is at all complicated. Cho and White (2007) and Carter and Steigerwald (2012, 2013) suggested quasi-likelihood ratio tests that ignore the Markov property of s_t. For discussion of some of the subtleties and possible solutions for the case of i.i.d. regime changes, see Hall and Stewart (2005) and Chen and Li (2009).

An alternative is to calculate instead general measures that trade off the fit of the likelihood against the number of parameters estimated. Popular methods such as Schwarz's (1978) Bayesian criterion rely for their asymptotic justification on the same regularity conditions whose failure causes the likelihood ratio statistic to have a nonstandard distribution. But Smith et al. (2006) developed a simple test that can be used to select the number of regimes for a Markov-switching regression,

$$y_t = \mathbf{x}_t'\boldsymbol{\beta}_{s_t} + \sigma_{s_t}\varepsilon_t \tag{36}$$

where $\varepsilon_t \sim N(0,1)$ and s_t follows an N-state Markov chain. The authors proposed to estimate the parameter vector $\boldsymbol{\lambda} = \left(\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_N', \sigma_1, \ldots, \sigma_N, p_{ij,i=1,\ldots,N;j=1,\ldots,N-1}\right)^{-1}$ by maximum likelihood for each possible choice of N and calculate

$$\hat{T}_i = \sum_{t=1}^{T} \text{Prob}(s_t = i|\boldsymbol{\Omega}_T;\hat{\boldsymbol{\lambda}}_{\text{MLE}}) \quad \text{for } i = 1, \ldots, N$$

using the full-sample smoothed probabilities. They suggested choosing the value of N for which

$$\text{MSC} = -2\mathcal{L}(\hat{\boldsymbol{\lambda}}_{\text{MLE}}) + \sum_{i=1}^{N} \frac{\hat{T}_i(\hat{T}_i + Nk)}{\hat{T}_i - Nk - 2}$$

is smallest, where k is the number of elements in the regression vector $\boldsymbol{\beta}$. Other alternatives are to use Bayesian methods to find the value of N that leads to the largest value for the marginal likelihood (Chib, 1998) or the highest Bayes factor (Koop and Potter, 1999).

Another promising test of the null hypothesis of no change in regime was developed by Carrasco et al. (2014). Let $\ell_t = \log f(y_t|\boldsymbol{\Omega}_{t-1};\boldsymbol{\lambda})$ be the log of the predictive density of the tth observation under the null hypothesis of no switching. For the Markov-switching regression (36), $\boldsymbol{\lambda}$ would correspond to the fixed-regime regression coefficients and variance $(\boldsymbol{\beta}',\sigma^2)'$:

$$\ell_t = -(1/2)\log(2\pi\sigma^2) - \frac{(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2}{2\sigma^2}.$$

Define $\mathbf{h}_t$ to be the derivative of the log density with respect to the parameter vector,

$$\mathbf{h}_t = \left.\frac{\partial \ell_t}{\partial \boldsymbol{\lambda}}\right|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}_0}$$

where $\hat{\boldsymbol{\lambda}}_0$ denotes the MLE under the null hypothesis of no change in regime. For example,

$$\mathbf{h}_t = \begin{bmatrix} \dfrac{(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}})\mathbf{x}_t}{\hat{\sigma}^2} \\ -\dfrac{1}{2\hat{\sigma}^2} + \dfrac{(y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}})^2}{2\hat{\sigma}^4} \end{bmatrix}$$

where $\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\left(\sum_{t=1}^{T} \mathbf{x}_t y_t\right)$ and $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}})^2$. To implement the Carrasco et al. (2014) test of the null hypothesis of no change in regime against the alternative that the first element of $\boldsymbol{\beta}$ switches according to a Markov chain, let $\ell_t^{(1)}$ denote the first element of $\mathbf{h}_t$ and calculate

$$\ell_t^{(2)} = \left.\frac{\partial^2 \ell_t}{\partial \lambda_1^2}\right|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}_0}$$

$$\gamma_t(\rho) = \ell_t^{(2)} + \left[\ell_t^{(1)}\right]^2 + 2\sum_{s<t} \rho^{t-s} \ell_t^{(1)} \ell_s^{(1)}$$

where ρ is an unknown parameter characterizing the persistence of the Markov chain. We then regress $(1/2)\gamma_t(\rho)$ on $\mathbf{h}_t$, save the residuals $\hat{\varepsilon}_t(\rho)$, and calculate

$$C(\rho) = \frac{1}{2}\left[\max\left\{0, \frac{\sum_{t=1}^{T}\gamma_t(\rho)}{2\sqrt{\sum_{t=1}^{T}[\hat{\varepsilon}_t(\rho)]^2}}\right\}\right]^2.$$

We then find the value ρ^* that maximizes $C(\rho)$ over some range (eg, $\rho \in [0.2, 0.8]$) and bootstrap to see if $C(\rho^*)$ is statistically significant. This is done by generating data with no changes in regime using the MLE $\lambda = \hat{\lambda}_0$ and calculating $C(\rho^*)$ on each generated sample.

Another option is to conduct generic tests developed by Hamilton (1996) of the hypothesis that an N-regime model accurately describes the data. For example, if the model is correctly specified, the derivative of the log of the predictive density with respect to any element of the parameter vector,

$$\left.\frac{\partial \log p(\mathbf{y}_t | \boldsymbol{\Omega}_{t-1}; \boldsymbol{\lambda})}{\partial \lambda_i}\right|_{\boldsymbol{\lambda}=\hat{\boldsymbol{\lambda}}_{\mathrm{MLE}}},$$

should be impossible to predict from its own lagged values, a hypothesis that can be tested using simple regressions.

2.12 Deterministic Breaks

Another common approach is to treat the changes in regime as deterministic rather than random. If we wanted to test the null hypothesis of constant coefficients against the alternative that a certain subset of the coefficients of a regression switched at fixed known dates $t_1, t_2, \ldots, t_N$, we could do this easily enough using a standard F test (see, for example, Fisher, 1970). If we do not know the dates, we could calculate the value of the F statistic for every set of allowable N partitions, efficient algorithms for which have been described by Bai and Perron (2003) and Doan (2012), with critical values for interpreting the supremum of the F statistics provided by Bai and Perron (1998). Bai and Perron (1998) also described a sequential procedure with which one could first test the null hypothesis of no breaks against the alternative of $N = 1$ break, and then test $N = 1$ against $N = 2$, and so on.

Although simpler to deal with econometrically, deterministic structural breaks have the drawback that they are difficult to incorporate in a sensible way into models based on rational decision makers. Neither the assumption that people knew perfectly that the change was coming years in advance, nor the assumption that they were certain that nothing would ever change (when in the event the change did indeed appear) is very appealing. There is further the practical issue of how users of such econometric models are supposed to form their own future forecasts. Pesaran and Timmermann (2007) suggested estimating models over windows of limited subsamples, watching the data for an indication that it is time to switch to using a new model. Another drawback of interpreting structural breaks as deterministic events is that such approaches make no use of the fact that regimes such as business downturns may be a recurrent event.

2.13 Chib's Multiple Change-Point Model

Chib (1998) offered a way to interpret multiple change-point models that gets around some of the awkward features of deterministic structural breaks. Chib's model assumes that when the process is in regime i, the conditional density of the data is governed by parameter vector $\boldsymbol{\theta}_i$ as in (13). Chib assumed that the process begins at date 1 in regime $s_t = 1$ and parameter vector $\boldsymbol{\theta}_1$, and will stay there the next period with probability p_{11}. With probability $1 - p_{11}$ we get a new value $\boldsymbol{\theta}_2$, drawn perhaps from an $N(\boldsymbol{\theta}_1, \boldsymbol{\Sigma})$ distribution. Conditional on knowing that there were N such breaks, this could be viewed as a special case of an N-state Markov-switching model with transition probability matrix taking the form

$$\mathbf{P} = \begin{bmatrix} p_{11} & 0 & 0 & \cdots & 0 & 0 \\ 1-p_{11} & p_{22} & 0 & \cdots & 0 & 0 \\ 0 & 1-p_{22} & p_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & p_{N-1,N-1} & 0 \\ 0 & 0 & 0 & \cdots & 1-p_{N-1,N-1} & 1 \end{bmatrix}.$$

The total number of regime changes N could then be selected using one of the methods discussed above.

Again it is not clear how to form out-of-sample forecasts with this specification. Pesaran et al. (2006) proposed embedding Chib's model within a hierarchical prior with which one could forecast future changes in regime based on the size and duration of past breaks.

2.14 Smooth Transition Models

Another econometric approach to changes in regime is the smooth transition regression model (Teräsvirta, 2004):

$$y_t = \frac{\exp[-\gamma(z_{t-1}-c)]}{1+\exp[-\gamma(z_{t-1}-c)]}\mathbf{x}'_{t-1}\boldsymbol{\beta}_1 + \frac{1}{1+\exp[-\gamma(z_{t-1}-c)]}\mathbf{x}'_{t-1}\boldsymbol{\beta}_2 + u_t. \tag{37}$$

Here the scalar z_{t-1} could be one of the elements of $\mathbf{x}_{t-1}$ or some known function of $\mathbf{x}_{t-1}$. For $\gamma > 0$, as $z_{t-1} \to -\infty$, the regression coefficients go to $\boldsymbol{\beta}_1$, while when $z_{t-1} \to \infty$, the regression coefficients approach $\boldsymbol{\beta}_2$. The parameter γ governs how quickly the coefficients transition as z_{t-1} crosses the threshold c.

If $\mathbf{x}_{t-1} = (y_{t-1}, y_{t-2}, \ldots, y_{t-r})'$, this is Teräsvirta's (1994) smooth-transition autoregression, for which typically $z_{t-1} = y_{t-d}$ for some lag d. More generally, given a data-generating process for $\mathbf{x}_t$, (37) is a fully specified time-series process for which forecasts at any horizon can be calculated by simulation. One important challenge is how to choose the lag d or more generally the switching variable z_{t-1}. Although in some settings the forecast might be similar to that coming from (6), the weights $\text{Prob}(s_{t-1} = i|\Omega_{t-1})$ in the latter would be a function of the entire history $\{y_{t-1}, y_{t-2}, \ldots, y_1\}$ rather than any single value.

3. ECONOMIC THEORY AND CHANGES IN REGIME

The previous section discussed econometric issues associated with analyzing series subject to changes in regime. This section reviews how these features can appear in theoretical models of the economy.

3.1 Closed-Form Solution of DSGEs and Asset-Pricing Implications

In some settings it is possible to find exact analytical solutions for a full dynamic stochastic general equilibrium model subject to changes in regime. A standard first-order condition in many macro models holds that

$$U'(C_t) = \beta E_t[U'(C_{t+1})(1 + r_{j,t+1})] \tag{38}$$

where C_t denotes consumption of a representative consumer, β a time-discount rate, and $r_{j,t+1}$ the real return on asset j between t and $t+1$.Lucas (1978) proposed a particularly simple setting in which aggregate output comes solely from nonreproducible assets (sometimes thought of as fruit coming from trees) for which equilibrium turns out to require that C_t equals the aggregate real dividend D_t paid on equities (or the annual crop of fruit). If the utility function exhibits constant relative risk aversion ($U(C) = (1+\gamma)^{-1} C^{(1+\gamma)}$), the aggregate equilibrium real stock price must satisfy

$$P_t = D_t^{-\gamma} \sum_{k=1}^{\infty} \beta^k E_t D_{t+k}^{(1+\gamma)}.$$

Since the dividend process $\{D_{t+k}\}$ is exogenous in this model, one could simply assume that the change in the log of D_t is characterized by a process such as (1). Cecchetti et al. (1990) used calculations related to those in (12) to find the closed-form solution for the general-equilibrium stock price,

$$P_t = \rho_{s_t} D_t,$$

where the values of ρ_1 and ρ_2 are given in equations (11) and (12) in their paper.

Lucas's assumption of an exogenous consumption and dividend process is obviously quite restrictive. Nevertheless, asset-pricing relations such as (38) have to hold regardless of how we close the rest of the model. We can always use (38) or other basic asset-pricing conditions along with an assumed process for returns to find the implications of changes in regime for financial variables in more general settings. There is a very large literature investigating these issues, covering topics such as portfolio allocation (Ang and Bekaert, 2002a; Guidolin and Timmermann, 2008), financial implications of rare-event risk (Barro, 2006; Evans, 1996), option pricing (Elliott et al., 2005), and the term structure of interest rates (Ang and Bekaert, 2002b; Bansal and Zhou, 2002; Bekaert et al., 2001). For a survey of this literature, see Ang and Timmermann (2012).

3.2 Approximating the Solution to DSGEs Using Perturbation Methods

First-order conditions for a much broader class of dynamic stochastic general equilibrium models with Markov regime-switching take the form

$$E_t \mathbf{a}(\mathbf{y}_{t+1}, \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \varepsilon_{t+1}, \varepsilon_t, \boldsymbol{\theta}_{s_{t+1}}, \boldsymbol{\theta}_{s_t}) = \mathbf{0}. \quad (39)$$

Here $\mathbf{a}(.)$ is an $[(n_y + n_x) \times 1]$ vector-valued function, $\mathbf{y}_t$ an $(n_y \times 1)$ vector of control variables (also sometimes referred to as endogenous jump variables), $\mathbf{x}_t$ an $(n_x \times 1)$ vector of predetermined endogenous or exogenous variables, ε_t an $(n_\varepsilon \times 1)$ vector of innovations to those elements of $\mathbf{x}_t$ that are exogenous to the model, and s_t follows an N-state Markov chain. The example considered in the previous subsection is a special case of such a system with $n_y = n_x = 1, y_t = P_t/D_t$, $x_t = \ln(D_t/D_{t-1})$, $\theta_{s_t} = m_{s_t}$, and[d]

[d] Notice (38) can be written

$$D_t^{\gamma} = \beta E_t \left[D_{t+1}^{\gamma} \frac{P_{t+1} + D_{t+1}}{P_t} \right]$$
$$1 = \beta E_t \left[\left(\frac{D_{t+1}}{D_t} \right)^{\gamma} \left(\frac{(P_{t+1}/D_{t+1}) + 1}{P_t/D_t} \right) \left(\frac{D_{t+1}}{D_t} \right) \right].$$

$$\mathbf{a}(y_{t+1}, y_t, x_t, x_{t-1}, \varepsilon_{t+1}, \varepsilon_t, m_{s_{t+1}}, m_{s_t})$$
$$= \begin{bmatrix} \beta \exp[(1+\gamma)(m_{s_{t+1}} + \varepsilon_{t+1})][(y_{t+1}+1)/y_t] - 1 \\ x_t - m_{s_t} - \varepsilon_t \end{bmatrix}.$$

For that example we were able to find closed-form solutions of the form

$$\mathbf{y}_t = \boldsymbol{\rho}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t)$$
$$\mathbf{x}_t = \mathbf{h}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t),$$

namely $y_t = \rho_{s_t}$ and $x_t = m_{s_t} + \varepsilon_t$.

For more complicated models, solutions cannot be found analytically but can be approximated using the partition perturbation method developed by Foerster et al. (forthcoming). Their method generalizes the now-standard perturbation methods of Schmitt-Grohe and Uribe (2004) for finding linear and higher-order approximations to the solutions to DSGEs with no regime switching. Foerster et al.'s idea is to approximate the solutions $\boldsymbol{\rho}_j(.)$ and $\mathbf{h}_j(.)$ in a neighborhood around the deterministic steady-state values satisfying $\mathbf{a}(\mathbf{y}^*, \mathbf{y}^*, \mathbf{x}^*, \mathbf{x}^*, \mathbf{0}, \mathbf{0}, \boldsymbol{\theta}^*, \boldsymbol{\theta}^*) = \mathbf{0}$ where $\boldsymbol{\theta}^*$ is the unconditional expectation of $\boldsymbol{\theta}_{s_t}$ calculated from the ergodic probabilities of the Markov chain,

$$\boldsymbol{\theta}^* = \sum_{j=1}^{N} \boldsymbol{\theta}_j \mathrm{Prob}(s_t = j).$$

For the Lucas tree example from the previous subsection, $m^* = (m_1 p_{21} + m_2 p_{12})/(p_{12} + p_{21})$. We then think of a sequence of economies indexed by a continuous scalar χ such that their behavior as $\chi \to 0$ approaches the steady state, while the value at $\chi = 1$ is exactly that implied by (39):

$$\mathbf{y}_t = \boldsymbol{\rho}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) \tag{40}$$
$$\mathbf{x}_t = \mathbf{h}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi). \tag{41}$$

As $\chi \to 0$, the randomness coming from $\boldsymbol{\varepsilon}_t$ is suppressed, and it turns out to be necessary to do the same thing for any elements of $\boldsymbol{\theta}$ that influence the steady state in order to have some fixed point around which to calculate the approximation. For elements in $\boldsymbol{\theta}_{s_t}$ that may change with regime but do not matter for the steady state, Foerster et al. (forthcoming) showed that it is not necessary to shrink by χ in order to approximate the dynamic solution. The authors thus specified

$$\boldsymbol{\theta}(s_t, \chi) = \begin{bmatrix} \boldsymbol{\theta}^{A*} + \chi(\boldsymbol{\theta}^A_{s_t} - \boldsymbol{\theta}^{A*}) \\ \boldsymbol{\theta}^B_{s_t} \end{bmatrix}$$

where $\boldsymbol{\theta}^A_{s_t}$ denotes the subset of elements of $\boldsymbol{\theta}_{s_t}$ that influence the steady state. The economy characterized by a particular value of χ thus needs to satisfy

$$\mathbf{0} = \int \sum_{j=1}^{N} p_{s_t,j} \mathbf{a}[\boldsymbol{\rho}_j(\mathbf{x}_t, \chi\boldsymbol{\varepsilon}_{t+1}, \chi), \mathbf{y}_t, \mathbf{x}_t, \mathbf{x}_{t-1}, \chi\varepsilon_{t+1}, \varepsilon_t, \boldsymbol{\theta}(j,\chi), \boldsymbol{\theta}(s_t,\chi)] dF(\boldsymbol{\varepsilon}_{t+1}) \tag{42}$$

where $F(\boldsymbol{\varepsilon}_{t+1})$ denotes the cumulative distribution function for ε_{t+1}. Note (42) is satisfied by construction when evaluated at $\mathbf{y}_t = \mathbf{y}^*$, $\mathbf{x}_t = \mathbf{x}_{t-1} = \mathbf{x}^*$, $\boldsymbol{\varepsilon}_t = \mathbf{0}$, and $\chi = 0$.

We next substitute (40) and (41) into (42) to arrive at a system of $N(n_y + n_x)$ equations of the form

$$\mathbf{Q}_{s_t}(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi) = \mathbf{0} \quad s_t = 1, \ldots, N$$

which have to hold for all $\mathbf{x}_{t-1}$, $\boldsymbol{\varepsilon}_t$, and χ. Taking derivatives with respect to $\mathbf{x}_{t-1}$ and evaluating at $\mathbf{x}_{t-1} = \mathbf{x}^*$, $\boldsymbol{\varepsilon}_t = \mathbf{0}$, and $\chi = 0$ (that is, using a first-order Taylor approximation around the steady state) yields a system of $N(n_y + n_x)n_x$ quadratic polynomial equations in the $N(n_y + n_x)n_x$ unknowns corresponding to elements of the matrices

$$\underset{(n_y \times n_x)}{\mathbf{R}_j^x} = \left. \frac{\partial \boldsymbol{\rho}_j(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi)}{\partial \mathbf{x}'_{t-1}} \right|_{\mathbf{x}_{t-1} = \mathbf{x}^*, \boldsymbol{\varepsilon}_t = \mathbf{0}, \chi = 0} \quad j = 1, \ldots, N$$

$$\underset{(n_x \times n_x)}{\mathbf{H}_j^x} = \left. \frac{\partial \mathbf{h}_j(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi)}{\partial \mathbf{x}'_{t-1}} \right|_{\mathbf{x}_{t-1} = \mathbf{x}^*, \boldsymbol{\varepsilon}_t = \mathbf{0}, \chi = 0} \quad j = 1, \ldots, N.$$

The authors proposed an algorithm for finding the solution to this system of equations, that is, values for the above sets of matrices. Given these, other terms in the first-order Taylor approximation to (42) produce a system of $N(n_y + n_x)n_\varepsilon$ equations that are linear in known parameters and the unknown elements of

$$\underset{(n_y \times n_\varepsilon)}{\mathbf{R}_j^\varepsilon} = \left. \frac{\partial \boldsymbol{\rho}_j(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi)}{\partial \varepsilon'_t} \right|_{\mathbf{x}_{t-1} = \mathbf{x}^*, \boldsymbol{\varepsilon}_t = \mathbf{0}, \chi = 0} \quad j = 1, \ldots, N$$

$$\underset{(n_x \times n_\varepsilon)}{\mathbf{H}_j^\varepsilon} = \left. \frac{\partial \mathbf{h}_j(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi)}{\partial \varepsilon'_t} \right|_{\mathbf{x}_{t-1} = \mathbf{x}^*, \boldsymbol{\varepsilon}_t = \mathbf{0}, \chi = 0} \quad j = 1, \ldots, N,$$

from which $\mathbf{R}_j^\varepsilon$ and $\mathbf{H}_j^\varepsilon$ are readily calculated. Another system of $N(n_y + n_x)$ linear equations yields

$$\underset{(n_y \times 1)}{\mathbf{R}_j^\chi} = \left. \frac{\partial \boldsymbol{\rho}_j(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi)}{\partial \chi} \right|_{\mathbf{x}_{t-1} = \mathbf{x}^*, \boldsymbol{\varepsilon}_t = \mathbf{0}, \chi = 0} \quad j = 1, \ldots, N$$

$$\underset{(n_x \times 1)}{\mathbf{H}_j^\chi} = \left. \frac{\partial \mathbf{h}_j(\mathbf{x}_{t-1}, \boldsymbol{\varepsilon}_t, \chi)}{\partial \chi} \right|_{\mathbf{x}_{t-1} = \mathbf{x}^*, \boldsymbol{\varepsilon}_t = \mathbf{0}, \chi = 0} \quad j = 1, \ldots, N.$$

The approximation to the solution to the regime-switching DSGE is then

$$\begin{aligned}\mathbf{y}_t &= \mathbf{y}^* + \mathbf{R}^x_{s_t}(\mathbf{x}_{t-1} - \mathbf{x}^*) + \mathbf{R}^\varepsilon_{s_t}\boldsymbol{\varepsilon}_t + \mathbf{R}^\chi_{s_t}\\ \mathbf{x}_t &= \mathbf{x}^* + \mathbf{H}^x_{s_t}(\mathbf{x}_{t-1} - \mathbf{x}^*) + \mathbf{H}^\varepsilon_{s_t}\boldsymbol{\varepsilon}_t + \mathbf{H}^\chi_{s_t}.\end{aligned}$$

One could then go a step further if desired, taking a second-order Taylor approximation to (42). Once the first step (the linear approximation) has been completed, the second step (quadratic approximation) is actually easier to calculate numerically than the first step was, because all the second-step equations turn out to be linear in the remaining unknown magnitudes.

Lind (2014) developed an extension of this approach that could be used to form approximations to any model characterized by dramatic nonlinearities, even if regime-switching in the form of (39) is not part of the maintained structure. For example, the economic relations may change significantly when interest rates are at the zero lower bound. Lind's idea is to approximate the behavior of a nonlinear model over a set of discrete regions using relations that are linear (or possibly higher-order polynomials) over individual regions, from which one can then use many of the tools discussed above for economic and econometric analysis.

3.3 Linear Rational Expectations Models with Changes in Regime

Economic researchers often use a linear special case of (39) which in the absence of regime shifts takes the form

$$\mathbf{A}E(\mathbf{y}_{t+1}|\boldsymbol{\Omega}_t) = \mathbf{d} + \mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{x}_t \tag{43}$$

$$\mathbf{x}_t = \mathbf{c} + \boldsymbol{\Phi}\mathbf{x}_{t-1} + \mathbf{v}_t$$

for $\mathbf{y}_t$ an $(n_y \times 1)$ vector of endogenous variables, $\boldsymbol{\Omega}_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}, \mathbf{x}_t$ an $(n_x \times 1)$ vector of exogenous variables, and $\mathbf{v}_t$ a martingale difference sequence. Such a system might have been obtained as an approximation to the first-order conditions for a nonlinear DSGE using the standard perturbation algorithm, or often is instead simply postulated as the primitive conditions of the model of interest. If $\mathbf{A}^{-1}$ exists and the number of eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$ whose modulus is less than or equal to unity is equal to the number of predetermined endogenous variables, then a unique stable solution can be found of the form[e]

$$\mathbf{k}_{t+1} = \mathbf{h}_{k0} + \mathbf{H}_{kk}\mathbf{k}_t + \mathbf{H}_{kx}\mathbf{x}_t$$

$$\mathbf{d}_t = \mathbf{h}_{d0} + \mathbf{H}_{dk}\mathbf{k}_t + \mathbf{H}_{dx}\mathbf{x}_t$$

[e] Klein (2000) generalized to the case when $\mathbf{A}$ may not be invertible.

where $\mathbf{k}_t$ denotes the elements of $\mathbf{y}_t$ that correspond to predetermined variables, while $\mathbf{d}_t$ collects the control or jump variables. Algorithms for finding the values of the parameters $\mathbf{h}_{i0}$ and $\mathbf{H}_{ij}$ have been developed by Blanchard and Kahn (1980), Klein (2000), and Sims (2001).

We could also generalize (43) to allow for changes in regime,

$$\mathbf{A}_{s_t} E(\mathbf{y}_{t+1}|\Omega_t, s_t, s_{t-1}, \ldots, s_1) = \mathbf{d}_{s_t} + \mathbf{B}_{s_t}\mathbf{y}_t + \mathbf{C}_{s_t}\mathbf{x}_t \tag{44}$$

where s_t follows an exogenous N-state Markov chain and $\mathbf{A}_j$ denotes an $(n_y \times n_y)$ matrix of parameters when the regime for date t is given by $s_t = j$. To solve such a model, Davig and Leeper (2007) suggested exploiting the feature that conditional on $\mathcal{S} = \{s_t\}_{t=1}^{\infty}$ the model is linear. Let $\mathbf{y}_{jt}$ correspond to the value of $\mathbf{y}_t$ when $s_t = j$ and collect the set of such vectors for all the possible regimes in a larger vector $\mathbf{Y}_t$:

$$\underset{(Nn_y \times 1)}{\mathbf{Y}_t} = \begin{bmatrix} \underset{(n_y \times 1)}{\mathbf{y}_{1t}} \\ \vdots \\ \underset{(n_y \times 1)}{\mathbf{y}_{Nt}} \end{bmatrix}.$$

If we restrict our consideration to solutions that satisfy the minimal-state Markov property, then

$$E(\mathbf{y}_{t+1}|\mathcal{S}, \Omega_t) = E(\mathbf{y}_{t+1}|s_{t+1}, s_t, \Omega_t)$$

and

$$E(\mathbf{y}_{t+1}|s_t = i, \Omega_t) = \sum_{j=1}^{N} E(\mathbf{y}_{t+1}|s_{t+1} = j, s_t = i, \Omega_t) p_{ij}.$$

Hence when $s_t = i$,

$$\mathbf{A}_{s_t} E(\mathbf{y}_{t+1}|s_t, \Omega_t) = (\mathbf{p}_i' \otimes \mathbf{A}_i) E(\mathbf{Y}_{t+1}|\mathbf{Y}_t) \tag{45}$$

where

$$\mathbf{p}_i = \begin{bmatrix} p_{i1} \\ \vdots \\ p_{iN} \end{bmatrix}$$

denotes column i of the Markov transition probabilities, with elements of $\mathbf{p}_i$ summing to unity. Consider then the stacked structural system,

$$\mathbf{A} E(\mathbf{Y}_{t+1}|\mathbf{Y}_t) = \mathbf{d} + \mathbf{B}\mathbf{Y}_t + \mathbf{C}\mathbf{x}_t \tag{46}$$

$$\underset{(Nn_y \times Nn_y)}{\mathbf{A}} = \begin{bmatrix} \underset{(1\times N)}{\mathbf{p}_1'} \otimes \underset{(n_y\times n_y)}{\mathbf{A}_1} \\ \vdots \\ \underset{(1\times N)}{\mathbf{p}_N'} \otimes \underset{(n_y\times n_y)}{\mathbf{A}_N} \end{bmatrix} \quad \underset{(Nn_y\times 1)}{\mathbf{d}} = \begin{bmatrix} \underset{(n_y\times 1)}{\mathbf{d}_1} \\ \vdots \\ \underset{(n_y\times 1)}{\mathbf{d}_N} \end{bmatrix} \tag{47}$$

$$\underset{(Nn_y \times Nn_y)}{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_N \end{bmatrix} \quad \underset{(Nn_y\times n_x)}{\mathbf{C}} = \begin{bmatrix} \underset{(n_y\times n_x)}{\mathbf{C}_1} \\ \vdots \\ \underset{(n_y\times n_x)}{\mathbf{C}_N} \end{bmatrix}.$$

This is a simple regime-independent system for which a solution can be found using the traditional method. For example, with no predetermined variables, if all eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$ are outside the unit circle, then we can find a unique stable solution of the form

$$\underset{(Nn_y\times 1)}{\mathbf{Y}_t} = \underset{(Nn_y\times 1)}{\mathbf{h}} + \underset{(Nn_y\times n_x)}{\mathbf{H}} \underset{(n_x\times 1)}{\mathbf{x}_t} \tag{48}$$

which implies that

$$\underset{(n_y\times 1)}{\mathbf{y}_t} = \underset{(n_y\times 1)}{\mathbf{h}_{s_t}} + \underset{(n_y\times n_x)}{\mathbf{H}_{s_t}} \underset{(n_x\times 1)}{\mathbf{x}_t} \tag{49}$$

for $\mathbf{h}_i$ and $\mathbf{H}_i$ the ith blocks of $\mathbf{h}$ and $\mathbf{H}$, respectively. If (48) is a solution to (46), then (49) is a solution to (44).[f]

[f] If (49) holds, then

$$E(\mathbf{y}_{t+1}|\Omega_t, s_t = i) = \sum_{j=1}^{N} p_{ij}[\mathbf{h}_j + \mathbf{H}_j(\mathbf{c} + \mathbf{\Phi}\mathbf{x}_t)].$$

Thus for (44) to hold it must be the case that for each $i = 1,\ldots,N$,

$$\mathbf{A}_i \sum_{j=1}^{N} p_{ij}\mathbf{H}_j\mathbf{\Phi} = \mathbf{B}_i\mathbf{H}_i + \mathbf{C}_i$$
$$\mathbf{A}_i \sum_{j=1}^{N} p_{ij}(\mathbf{h}_j + \mathbf{H}_j\mathbf{c}) = \mathbf{d}_i + \mathbf{B}_i\mathbf{h}_i.$$

But if (48) is a solution to (46), then

$$\mathbf{A}[\mathbf{h} + \mathbf{H}(\mathbf{c} + \mathbf{\Phi}\mathbf{x}_t)] = \mathbf{d} + \mathbf{B}(\mathbf{h} + \mathbf{H}\mathbf{x}_t) + \mathbf{C}\mathbf{x}_t$$

block i of which requires from (47) that

$$(\mathbf{p}_i' \otimes \mathbf{A}_i)\mathbf{H}\mathbf{\Phi} = \mathbf{B}_i\mathbf{H}_i + \mathbf{C}_i$$
$$(\mathbf{p}_i' \otimes \mathbf{A}_i)(\mathbf{h} + \mathbf{H}\mathbf{c}) = \mathbf{d}_i + \mathbf{B}_i\mathbf{h}_i$$

as were claimed to hold.

However, Farmer et al. (2010) demonstrated that while (48) yields one stable solution to (44), it need not be the only stable solution. For further discussion, see Farmer et al. (2009).

3.4 Multiple Equilibria

Other economists have argued that models in which there are multiple possible solutions—for example, system (43) with no predetermined variables and an eigenvalue of $\mathbf{A}^{-1}\mathbf{B}$ inside the unit circle—are precisely those we should be most interested in, given the perception that sometimes consumers or firms seem to become highly pessimistic for no discernible reason, bringing the economy into a self-fulfilling downturn; see Benhabib and Farmer (1999) for a survey of this literature. One factor that could produce multiple equilibria is coordination externalities. The rewards to me of participating in a market may be greatest when I expect large numbers of others to do the same (Cooper, 1994; Cooper and John, 1988). Multiple equilibria could also arise when expectations themselves are a factor determining the equilibrium (Kurz and Motolese, 2001). Kirman (1993) and Chamley (1999) discussed mechanisms by which the economy might tend to oscillate periodically between the possible regimes in multiple-equilibria settings.

A widely studied example is financial market bubbles. In the special case of risk-neutral investors (that is, when $U'(C)$ is some constant independent of consumption C), Eq. (38) relating the price of the stock P_t to its future dividend D_{t+1} becomes

$$P_t = \beta E_t(P_{t+1} + D_{t+1}). \tag{50}$$

One solution to (50) is the market-fundamentals solution given by

$$P_t^* = \sum_{j=1}^{\infty} \beta^j E_t(D_{t+j}).$$

But $P_t = P_t^* + B_t$ also satisfies(50) for B_t any bubble process satisfying $B_t = \beta E_t B_{t+1}$. Hall et al. (1999) proposed an empirical test of whether an observed financial price is occasionally subject to such a bubble regime. This test has been applied in dozens of different empirical studies. However, Hamilton (1985), Driffill and Sola (1998), and Gürkaynak (2008) noted the inherent difficulties in distinguishing financial bubbles from unobserved fundamentals.

3.5 Tipping Points and Financial Crises

In other models, there may be a unique equilibrium but under the right historical conditions, a small change in fundamentals can produce a huge change in observed outcomes. Such dynamics might be well described as locally linear processes that periodically experience changes in regime. Investment dynamics constitute one possible transmission mechanism. The right sequence of events can end up triggering a big investment decline that in turn contributes to a dramatic drop in output and an effective change in regime. Acemoglu and Scott (1997) presented a model where this happens as a result of intertemporal increasing returns, for example, if an investment that leads to a significant new discovery makes additional investments more profitable for a short time. Moore and Schaller (2002), Guo et al. (2005), and Veldkamp (2005) examined different settings in which investment dynamics contribute to tipping points, often through a process of learning about current opportunities. Startz (1998) demonstrated how an accumulation of small shocks could under certain circumstances trigger a dramatic shift between alternative production technologies. Learning by market participants introduces another possible source of tipping-point or regime-shift dynamics (Hong et al., 2007; Branch and Evans, 2010). Gârleanu et al. (2015) demonstrated how tipping points could emerge from the interaction of limited market integration, leveraging, and contagion.

Brunnermeier and Sannikov (2014) developed an intriguing description of tipping points in the context of financial crises. They posited two types of agents, designated "experts" and "households." Experts can invest capital more productively than households, but they are constrained to borrow using only risk-free debt. In normal times, 100% of the economy's equity ends up being held by experts. But as negative shocks cause their net worth to decline, they can end up selling off capital to less productive households, lowering both output and investment. This results in a bimodal stationary distribution in which the economy spends most of its time around the steady state in which experts hold all the capital. But a sequence of negative shocks can lead the economy to become stuck in an inefficient equilibrium from which it can take a long time to recover.

A large number of researchers have used regime-switching models to study financial crises empirically. These include Hamilton's (2005) description of banking crises in the 19th century, Asea and Blomberg's (1998) study of lending cycles in the late 20th century, and an investigation of more recent financial stress by Hubrich and Tetlow (2015).

3.6 Currency Crises and Sovereign Debt Crises

A sudden loss of confidence in a country can lead to a flight from the currency which in turn produces a shock to credit and spending that greatly exacerbates the country's problems. A sudden wave of pessimism could be self-fulfilling, giving rise to multiple equilibria that could exhibit Markov switching (Jeanne and Masson, 2000), or could be characterized by tipping point dynamics where under the right circumstances a small change in fundamentals pushes a country into crisis. Empirical investigations of currency crises using regime-switching models include Peria (2002) and Cerra and Saxena (2005).

Similar dynamics can characterize yields on sovereign debt. If investors lose confidence in a country's ability to service its debt, they will demand a higher interest rate as compensation. The higher interest costs could produce a tipping point that indeed forces a country into default or to make drastic fiscal adjustments (Greenlaw et al., 2013). Analyses of changes in regime in this context include Davig et al. (2011) and Bi (2012).

3.7 Changes in Policy as the Source of Changes in Regime

Another source of a change in regime is a discrete shift in policy itself. One commonly studied possibility is that control of monetary policy may periodically shift between hawks and doves, the latter being characterized by either a higher inflation target or more willingness to tolerate deviations of inflation from target. Analyses using this approach include Owyang and Ramey (2004), Schorfheide (2005), Liu et al. (2011), Bianchi (2013), and Baele et al. (2015).

An alternative possibility is that changes in fiscal regime can be a destabilizing factor. Ruge-Murcia (1995) showed how a lack of credibility of the fiscal stabilization in 1984 contributed to the changes in inflation Israel experienced, while Ruge-Murcia (1999) documented the close connection between changes in fiscal regimes and inflation regimes for Brazil.

4. CONCLUSIONS AND RECOMMENDATIONS FOR RESEARCHERS

We have seen that researchers have a rich set of tools and specifications on which to draw for interpreting data and building economic models for environments in which there may be changes in regime. The chapter closes with some practical recommendations for researchers as to which options are most promising.

Although a researcher might be tempted to use the most general specification possible, with all the parameters changing across a large number of regimes and time-varying transition probabilities, in practice this is usually asking more than the data can deliver. For example, for postwar US data we have only 11 recessions, which economic theory says should be difficult or impossible to predict (Hamilton, 2011). Building a richly parameterized description of the transition into and out of recession could easily result in an overfitted and misspecified model. By contrast, using a simple time-invariant Markov chain is likely to give a reasonable and robust approximation to the key features of the data. Similarly, we know from the analytic characterization of the maximum likelihood estimates (eg, Eq. (24)) that inference about parameters that only show up in regime i can only come from observations within that regime. With postwar quarterly data that would mean about 50 observations from which to estimate all the parameters operating during recessions. One or two parameters could be estimated fairly well, but overfitting is again a potential concern in models with many parameters. For this reason researchers may want to limit the focus to a few of the most important parameters that are likely to change, such as the intercept and the variance.

Where more than two regimes are required, there again are benefits to keeping the model parsimonious. For example, a common finding is that the variance of US output growth permanently decreased in 1984 (McConnell and Perez-Quiros, 2000), while the intercept periodically shifts to negative during recessions. This requires four different regimes—the economy could be in expansion or recession and the date could be before or after the Great Moderation. A useful simplification treats the variance regime as independent of the recession regime, requiring estimation of only 4 transition probabilities rather than 12, as in Kim and Nelson (1999b).

Another feature of which researchers should be aware is that there can be multiple local maxima to the likelihood function. It is therefore good practice to begin the EM iterations from a large number of different starting points to make sure we are always ending up with the same answer, and also as a practical test of whether the algorithm has indeed converged to a fixed point. Likewise with Bayesian methods we want to make sure numerical algorithms converge to the same posterior distribution under alternative starting points and chain dynamics, and the procedure should take into account the label-switching problem.

Provided researchers make note of these issues, these approaches offer a flexible way of modeling some of the key nonlinearities in macroeconomic dynamics without sacrificing the simplicity and tractability of linear models.

APPENDIX

Derivation of EM Equations for Restricted VAR

As noted by Hamilton (1990, p. 47), the M or maximization step of the EM algorithm can be implemented by finding the first-order conditions associated with maximizing the likelihood conditional on a particular set of realizations for the regimes $\mathcal{S} = \{s_1, \ldots, s_T\}$ and then weighting these by the smoothed probability of $\mathcal{S}$ and summing over all the possible realizations of $\mathcal{S}$. For a VAR restricted as in (28), the conditional likelihood is

$$\frac{1}{(2\pi)^{nT/2}|\mathbf{\Sigma}|^{T/2}} \exp\left[-(1/2)\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})'\mathbf{\Sigma}^{-1}(\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})\right]$$

with first-order conditions

$$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})\mathbf{x}'_{1,t-1} = \mathbf{0} \tag{A.1}$$

$$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})\mathbf{x}'_{2,t-1}\delta(s_t = i) = \mathbf{0} \quad \text{for } i = 1, \ldots, N \tag{A.2}$$

$$\sum_{t=1}^{T}(1/2)\left[\mathbf{\Sigma} - (\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})(\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})'\right] = \mathbf{0} \tag{A.3}$$

where $\delta(s_t = i)$ denotes unity if $s_t = i$ and zero otherwise. Stacking (A.1)–(A.2) horizontally gives

$$\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}\mathbf{x}_{1,t-1} - \mathbf{B}_{s_t}\mathbf{x}_{2,t-1})\mathbf{z}'_{t-1} = \sum_{t=1}^{T}(\mathbf{y}_t - [\mathbf{A} \;\; \mathbf{B}_1 \;\; \mathbf{B}_2 \;\; \cdots \;\; \mathbf{B}_N]\mathbf{z}_{t-1})\mathbf{z}'_{t-1} = \mathbf{0} \tag{A.4}$$

for

$$\underset{[1\times(k_1+Nk_2)]}{\mathbf{z}'_{t-1}} = \left[\mathbf{x}'_{1,t-1} \;\; \mathbf{x}'_{2,t-1}\delta(s_t=1) \;\; \mathbf{x}'_{2,t-1}\delta(s_t=2) \;\; \cdots \;\; \mathbf{x}'_{2,t-1}\delta(s_t=N)\right].$$

Multiplying the tth term within the sums in (A.4) and (A.3) by $\text{Prob}(s_t = i|\boldsymbol{\Omega}_T;\hat{\boldsymbol{\lambda}}^{(\ell)})$, summing over $i = 1,\ldots,N$, and rearranging gives Eqs. (29) and (30).

ACKNOWLEDGMENTS

I thank Marine Carrasco, Steve Davis, Liang Hu, Òscar Jordà, Douglas Steigerwald, John Taylor, Allan Timmermann, and Harald Uhlig for helpful comments on earlier drafts.

REFERENCES

Acemoglu, D., Scott, A., 1997. Asymmetric business cycles: theory and time-series evidence. J. Monet. Econ. 40, 501–533.
Albert, J., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. J. Bus. Econ. Stat. 11, 1–15.
Ang, A.A., Bekaert, G., 2002a. International asset allocation with regime shifts. Rev. Financ. Stud. 15, 1137–1187.
Ang, A.A., Bekaert, G., 2002b. Regime switches in interest rates. J. Bus. Econ. Stat. 20, 163–182.
Ang, A.A., Timmermann, A., 2012. Regime changes and financial markets. Ann. Rev. Financ. Econ. 4, 313–337.
Asea, P.K., Blomberg, B., 1998. Lending cycles. J. Econom. 83, 89–128.
Auerbach, A., Gorodnichenko, Y., 2012. Measuring the output responses to fiscal policy. Am. Econ. J. Macroecon. 4, 1–27.
Baele, L., Bekaert, G., Cho, S., Inghelbrecht, K., Moreno, A., 2015. Macroeconomic regimes. J. Monet. Econ. 70, 51–71.
Bai, J., Perron, P., 1998. Testing for and estimation of multiple structural changes. Econometrica 66, 47–78.
Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. J. Appl. Econ. 18, 1–22.
Bansal, R., Zhou, H., 2002. Term structure of interest rates with regime shifts. J. Financ. 57, 1997–2042.
Barro, R.J., 2006. Rare disasters and asset markets in the twentieth century. Q. J. Econ. 121, 823–866.
Bekaert, G., Hodrick, R.J., Marshall, D., 2001. Peso problem explanations for term structure anomalies. J. Monet. Econ. 48, 241–270.
Benhabib, J., Farmer, R.E.A., 1999. Indeterminacy and sunspots in macroeconomics. In: Taylor, J., Woodford, M. (Eds.), Handbook of Macroeconomics. In: vol. 1. North Holland, Amsterdam.
Bi, H., 2012. Sovereign default, risk premia, fiscal limits, and fiscal policy. Eur. Econ. Rev. 56, 389–410.
Bianchi, F., 2013. Regime switches, agents' beliefs and post-World War II U.S. macroeconomic dynamics. Rev. Econ. Stud. 80, 463–490.

Blanchard, O.J., Kahn, C.M., 1980. The solution of linear difference models under rational expectations. Econometrica 48, 1305–1317.
Branch, W.A., Evans, G.W., 2010. Asset return dynamics and learning. Rev. Financ. Stud. 23, 1651–1680.
Brunnermeier, M.K., Sannikov, Y., 2014. A macroeconomic model with a financial sector. Am. Econ. Rev. 104, 379–421.
Calvet, L., Fisher, A., 2004. How to forecast long-run volatility: regime-switching and the estimation of multifractal processes. J. Financ. Econ. 2, 49–83.
Camacho, M., Perez-Quiros, G., Poncela, P., 2014. Green shoots and double dips in the Euro are: a real time measure. Int. J. Forecast. 30, 520–535.
Carrasco, M., Hu, L., Ploberger, W., 2014. Optimal test for Markov switching. Econometrica 82, 765–784.
Carter, A.V., Steigerwald, D.G., 2012. Testing for regime switching: a comment. Econometrica 80, 1809–1812.
Carter, A.V., Steigerwald, D.G., 2013. Markov regime-switching tests: asymptotic critical values. J. Econ. Methods 2, 25–34.
Cecchetti, S.G., Lam, P.S., Mark, N.C., 1990. Mean reversion in equilibrium asset prices. Am. Econ. Rev. 80, 398–418.
Celeux, G., Hurn, M., Robert, C.P., 2000. Computational and inferential difficulties with mixture posterior distributions. J. Am. Stat. Assoc. 95, 957–970.
Cerra, V., Saxena, S.C., 2005. Did output recover from the Asian crisis? IMF Staff Pap. 52, 1–23.
Chamley, C., 1999. Coordinating regime switches. Q. J. Econ. 114, 869–905.
Chauvet, M., 1998. An economic characterization of business cycle dynamics with factor structure and regime switches. Int. Econ. Rev. 39, 969–996.
Chauvet, M., Hamilton, J.D., 2006. Dating business cycle turning points. In: Costas Milas, P.R., van Dijk, D. (Eds.), Nonlinear Analysis of Business Cycles. Elsevier, Amsterdam, pp. 1–54.
Chauvet, M., Piger, J., 2008. A comparison of the real-time performance of business cycle dating methods. J. Bus. Econ. Stat. 26, 42–49.
Chen, J., Li, P., 2009. Hypothesis test for normal mixture models: the EM approach. Ann. Stat. 37, 2523–2542.
Chib, S., 1998. Estimation and comparison of multiple change-point models. J. Econom. 86, 221–241.
Cho, J.S., White, H., 2007. Testing for regime switching. Econometrica 75, 1671–1720.
Cooper, R., 1994. Equilibrium selection in imperfectly competitive economics with multiple equilibria. Econ. J. 104, 1106–1122.
Cooper, R., John, A., 1988. Coordinating coordination failures in Keynesian models. Q. J. Econ. 103, 441–463.
Davig, T., Leeper, E.M., 2007. Generalizing the Taylor principle. Am. Econ. Rev. 97, 607–635.
Davig, T., Leeper, E.M., Walker, T.B., 2011. Inflation and the fiscal limit. Eur. Econ. Rev. 55, 31–47.
DeGroot, M.H., 1970. Optimal Statistical Decisions. McGraw-Hill, New York.
Diebold, F.X., Lee, J.H., Weinbach, G.C., 1994. Regime switching with time-varying transition probabilities. In: Hargreaves, C.P. (Ed.), Nonstationary Time Series Analysis and Cointegration. Oxford University Press, Oxford.
Doan, T., 2012. RATS User's Guide, Version 8.2. http://www.estima.com.
Driffill, J., Sola, M., 1998. Intrinsic bubbles and regime-switching. J. Monet. Econ. 42, 357–374.
Dueker, M., 1997. Markov switching in GARCH processes and mean-reverting stock-market volatility. J. Bus. Econ. Stat. 15, 26–34.
Elliott, R.J., Chan, L., Siu, T.K., 2005. Option pricing and Esscher transform under regime switching. Ann. Finance 1, 423–432.
Evans, M.D.D., 1996. Peso problems: their theoretical and empirical implications. In: Maddala, G.S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 14. Elsevier, Amsterdam.
Farmer, R.E.A., Waggoner, D.F., Zha, T., 2009. Understanding Markov-switching rational expectations models. J. Econ. Theory 144, 1849–1867.
Farmer, R.E.A., Waggoner, D.F., Zha, T., 2010. Generalizing the Taylor principle: comment. Am. Econ. Rev. 100, 608–617.
Filardo, A.J., 1994. Business cycle phases and their transitional dynamics. J. Bus. Econ. Stat. 12, 299–308.

Filardo, A.J., Gordon, S.F., 1998. Business cycle durations. J. Econom. 85, 99–123.
Fisher, F.M., 1970. Tests of equality between sets of coefficients in two linear regressions: an expository note. Econometrica 38, 361–366.
Foerster, A., Rubio-Ramirez, J., Waggoner, D.F., Zha, T., forthcoming. Perturbation methods for Markov-switching DSGE models. Quant. Econom.
Francq, C., Zakoïan, J.M., 2001. Stationarity of multivariate Markov-switching ARMA models. J. Econom. 102, 339–364.
Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. J. Am. Stat. Assoc. 96, 194–209.
Garcia, R., 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. Int. Econ. Rev. 39, 763–788.
Gârleanu, N., Panageas, S., Yu, J., 2015. Financial entanglement: a theory of incomplete integration, leverage, crashes, and contagion. Am. Econ. Rev. 105, 1979–2010.
Geweke, J., 2007. Interpretation and inference in mixture models: simple MCMC works. Comput. Stat. Data Anal. 51, 3529–3550.
Greenlaw, D., Hamilton, J.D., Hooper, P., Mishkin, F., 2013. Crunch time: fiscal crises and the role of monetary policy. In: Inproceedings of the U.S. Monetary Policy Forum 2013. Chicago Booth School of Business: Initiative on Global Markets, pp. 3–58.
Guidolin, M., Timmermann, A., 2008. International asset allocation under regime switching, skew, and kurtosis preferences. Rev. Financ. Stud. 21, 889–935.
Guo, X., Miao, J., Morelle, E., 2005. Irreversible investment with regime shifts. J. Econ. Theory 122, 37–59.
Gürkaynak, R.S., 2008. Econometric tests of asset price bubbles: taking stock. J. Econ. Surv. 22, 166–186.
Hall, P., Stewart, M., 2005. Theoretical analysis of power in a two-component normal mixture model. J. Stat. Plan. Inference 134, 158–179.
Hall, S.G., Psaradakis, Z., Sola, M., 1999. Detecting periodically collapsing bubbles: a Markov-switching unit root test. J. Appl. Econ. 14, 43–154.
Hamilton, J.D., 1985. On testing for self-fulfilling speculative price bubbles. Int. Econ. Rev. 27, 545–552.
Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.
Hamilton, J.D., 1990. Analysis of time series subject to changes in regime. J. Econom. 45, 39–70.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.
Hamilton, J.D., 1996. Specification testing in Markov-switching time-series models. J. Econom. 70, 127–157.
Hamilton, J.D., 2005. What's real about the business cycle? Fed. Reserve Bank St. Louis Rev. 87, 435–452.
Hamilton, J.D., 2011. Calling recessions in real time. Int. J. Forecast. 27, 1006–1026.
Hamilton, J.D., Wu, J.C., 2012. Identification and estimation of Gaussian affine term structure models. J. Econom. 168, 315–331.
Hansen, B.E., 1992. The likelihood ratio test under non-standard conditions. J. Appl. Econ. 7, S61–S82. Erratum, 1996, 11, 195–198.
Hong, H., Stein, J.C., Yu, J., 2007. Simple forecasts and paradigm shifts. J. Financ. 62, 1207–1242.
Hubrich, K., Tetlow, R.J., 2015. Financial stress and economic dynamics: the transmission of crises. J. Monet. Econ. 70, 100–115.
Jeanne, O., Masson, P., 2000. Currency crises, sunspots, and Markov-switching regimes. J. Int. Econ. 50, 327–350.
Karamé, F., 2010. Impulse-response functions in Markov-switching structural vector autoregressions: a step further. Econ. Lett. 106, 162–165.
Kim, C.J., 1994. Dynamic linear models with Markov-switching. J. Econom. 60, 1–22.
Kim, C.J., Nelson, C.R., 1999a. State-Space Models with Regime Switching. MIT Press, Cambridge, MA.
Kim, C.J., Nelson, C.R., 1999b. Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. Rev. Econ. Stat. 81, 608–616.
Kirman, A., 1993. Ants, rationality, and recruitment. Q. J. Econ. 108, 137–156.
Klein, P., 2000. Using the generalized Schur form to solve a multivariate linear rational expectations model. J. Econ. Dyn. Control 24, 1405–1423.

Koop, G., Potter, S.N., 1999. Bayes factors and nonlinearity: evidence from economic time series. J. Econom. 88, 251–281.
Krolzig, H.M., 1997. Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis. Springer, Berlin.
Kurz, M., Motolese, M., 2001. Endogenous uncertainty and market volatility. Econ. Theory 17, 497–544.
Lind, N., 2014. Regime-switching perturbation for non-linear equilibrium models. Working paper, UCSD.
Liu, Z., Waggoner, D.F., Zha, T., 2011. Sources of macroeconomic fluctuations: a regime-switching DSGE approach. Quant. Econ. 2, 251–301.
Lo, M.C., Piger, J., 2005. Is the response of output to monetary policy asymmetric? Evidence from a regime-switching coefficients model. J. Money Credit Bank. 37, 865–886.
Lucas Jr., R.E., 1978. Asset prices in an exchange economy. Econometrica 66, 1429–1445.
McConnell, M.M., Perez-Quiros, G., 2000. Output fluctuations in the United States: what has changed since the early 1980's? Am. Econ. Rev. 90, 1464–1476.
Moore, B., Schaller, H., 2002. Persistent and transitory shocks, learning, and investment dynamics. J. Money Credit Bank. 34, 650–677.
Owyang, M., Ramey, G., 2004. Regime switching and monetary policy measurement. J. Monet. Econ. 51, 1577–1597.
Peria, M.S.M., 2002. A regime-switching approach to the study of speculative attacks: a focus on EMS crises. In: Hamilton, J.D., Raj, B. (Eds.), Advances in Markov-Switching Models. Physica-Verlag, Heidelberg.
Pesaran, M.H., Pettenuzzo, D., Timmermann, A., 2006. Forecasting time series subject to multiple structural breaks. Rev. Econ. Stud. 73, 1057–1084.
Pesaran, M.H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. J. Econom. 137, 134–161.
Ruge-Murcia, F.J., 1995. Credibility and changes in policy regime. J. Polit. Econ. 103, 176–208.
Ruge-Murcia, F.J., 1999. Government expenditure and the dynamics of high inflation. J. Dev. Econ. 58, 333–358.
Schmitt-Grohe, S., Uribe, M., 2004. Solving dynamic general equilibrium models using a second-order approximation. J. Econ. Dyn. Control 28, 755–775.
Schorfheide, F., 2005. Learning and monetary policy shifts. Rev. Econ. Dyn. 8, 392–419.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Sims, C., 2001. Solving linear rational expectations models. J. Comput. Econ. 20, 1–20.
Sims, C., Zha, T., 2006. Were there regime switches in U.S. monetary policy? Am. Econ. Rev. 96, 54–81.
Smith, A., Naik, P.A., Tsai, C.L., 2006. Markov-switching model selection using Kullback-Leibler divergence. J. Econom. 134, 553–577.
Startz, R., 1998. Growth states and shocks. J. Econ. Growth 3, 203–215.
Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. J. Am. Stat. Assoc. 89, 208–218.
Teräsvirta, T., 2004. Smooth transition regression modelling. In: Lütkepohl, H., Krätzig, M. (Eds.), Applied Time Series Econometrics. Cambridge University Press, Cambridge, UK, pp. 222–242.
Timmermann, A., 2000. Moments of Markov switching models. J. Econom. 96, 75–111.
Tjøstheim, D., 1986. Some doubly stochastic time series models. J. Time Ser. Anal. 7, 51–72.
Veldkamp, L.L., 2005. Slow boom, sudden crash. J. Econ. Theory 124, 230–257.
White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, Orlando.
Yang, M.X., 2000. Some properties of vector autoregressive processes with Markov-switching coefficients. Econ. Theory 16, 23–43.