The Propagation of Regional Recessions*

James D. Hamilton†
Department of Economics
University of California, San Diego

Michael T. Owyang‡
Research Division
Federal Reserve Bank of St. Louis

keywords: recession, regional business cycles, Markov-switching

first draft: September 13, 2008
revised: September 30, 2010

Abstract

This paper develops a framework for inferring common Markov-switching components in a panel data set with large cross-section and time-series dimensions. We apply the framework to studying similarities and differences across U.S. states in the timing of business cycles. We hypothesize that there exists a small number of cluster designations, with individual states in a given cluster sharing certain business cycle characteristics. We find that although oil-producing and agricultural states can sometimes experience a separate recession from the rest of the United States, for the most part, differences across states appear to be a matter of timing, with some states entering recession or recovering before others. [JEL: C11; C32; E32]

*The authors benefited from conversations with Michael Dueker, Jeremy Piger, and Justin Tobias. Kristie M. Engemann provided research assistance. The views expressed herein do not reflect the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

†jhamilton@ucsd.edu.

‡owyang@stls.frb.org.

1 Introduction

The formation of the European Monetary Union has sparked a resurgence of interest in regional business cycles, both in Europe and in the United States, where longer time series are available. A number of these recent studies have characterized the U.S. national economy as an agglomeration of distinct but interrelated regional economies. While some idiosyncrasies exist, regional business cycles in the United States, for the most part, bear a reasonable resemblance to the national cycle identified by the National Bureau of Economic Research (NBER) using aggregate data. Disparities in regional business cycles have often been attributed either to idiosyncratic shocks or to differences in characteristics such as the industrial composition of the regions. Conversely, commonality can be attributed to responses to common aggregate shocks for which the state responses vary but the timing is identical.[1]

Characterizing regional business cycles using a panel data set with large cross-section and time-series dimensions raises two separate questions. The first is how to model the comovements that are common across geographic divisions. In Owyang, Piger, and Wall (2005) and Owyang, Piger, Wall, and Wheeler (2008), the unit of analysis is taken to be individual states and cities, respectively. Regional similarities were noted but not modeled explicitly. One alternative for characterizing common elements across geographic divisions is to rely on factor analysis, as in Forni and Reichlin (2001) and Del Negro (2002). Another approach is to use exogenously defined regions such as those adopted by the Bureau of Economic Analysis (BEA) as either the basic unit of analysis (e.g., Kouparitsas, 1999) or an additional observable restriction on the state-level factor structure (Del Negro, 2002). A few studies define regions endogenously. Crone (2005) used $k-$means cluster analysis of state business cycle movements to define regions. While his regional definitions are similar to those used by the BEA, Crone found some discrepancies (in particular, Arizona, which may be taken as a region unto itself). Partridge and Rickman (2005) used cyclical indices to uncover common currency areas in the United States. Similarly, van Dijk et al. (2007) constructed clusters for regional housing markets in the Netherlands.

A second question concerns the manner in which the business cycle itself is defined. What exactly are we claiming to have measured when we compare the timing of a recession in one state

[1] Monetary shocks, for example, are aggregate shocks that have common timing but varying effect [see Carlino and DeFina (1998)].

with that observed in another? In a standard factor model, the cyclical component is viewed as a continuous-valued random variable, defined in terms of its ability to capture certain comovements across states. Kouparitsas (1999) and Carlino and DeFina (2004) used band-pass filters to extract the business cycle frequency from disaggregate data. Carlino and Sill (2001) and Partridge and Rickman (2005) relied on trend-cycle decompositions.

Hamilton (2005) argued that the defining characteristic of the business cycle as understood, for example, by Burns and Mitchell (1946) is a transition between distinct, discrete phases of expansion and contraction. Owyang, Piger, and Wall (2005) and Owyang, Piger, Wall, and Wheeler (2008) adopted this perspective in their application of the Markov-switching model of Hamilton (1989) to data for individual states and cities, respectively. The contribution of the present paper is to extend that effort to characterize the interactions across states in these shifts. Our paper could alternatively be viewed as an extension of factor or cluster analysis to this kind of nonlinear framework.

We account for the correlation across states by modeling both national and regional recessions. In our setup, following Frühwirth-Schnatter and Kaufmann (2008), we allow the data to define regional groupings (which we designate as "clusters") on the basis of comovement in state employment growth rates and other observable, fixed state characteristics. In particular, we model the probability of a state's inclusion in any region as a logistic variable, in which state-level characteristics affect the prior probability of state membership in a region-cluster and observed employment growth comovements inform the posterior inference about those probabilities.

The model is estimated using Bayesian methods, and we report five main findings. First, most state-level business cycle experiences are similar to those of the nation. Second, most idiosyncratic recession experiences amount to differentials in timing around the national recessions. For example, some states enter some recessions before the rest of the nation. Third, a cluster of states, characterized by an important role for oil production in their economies, does enter and exit recessions independently from the nation. Fourth, the regional clusters we find are not exclusive, i.e., a state can belong to more than one region. However, the overlapping of states in multiple regions is infrequent. Finally, while industrial composition matters for cluster determination, other factors such as the share of employment coming from small firms may also be important.

The remainder of this paper is organized as follows. Section 2 presents our characterization

of regional business cycles with particular focus on endogenous region determination. Section 3 details the estimation technique. Section 4 presents the empirical results. Section 5 concludes.

2 Characterizing regional business cycles.

Let y_{tn} denote the employment growth rate for state n observed at date t. We group observations for all states at date t in an $(N \times 1)$ vector $\mathbf{y}_t = (y_{t1}, ..., y_{tN})'$, where N denotes the number of states. Let $\mathbf{s}_t$ be an $(N \times 1)$ vector of date t recession indicators (so $s_{tn} = 1$ when state n is in recession and $s_{tn} = 0$ when state n is in expansion). Suppose that

$$\mathbf{y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 \odot \mathbf{s}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where the nth element of the $(N \times 1)$ vector $\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1$ is the average employment growth in state n during recession, the nth element of the $(N \times 1)$ vector $\boldsymbol{\mu}_0$ is the average employment growth in state n during expansion, and $\odot$ represents the Hadamard product. We assume that $\boldsymbol{\varepsilon}_t \sim$ i.i.d. $N(\mathbf{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\varepsilon}_t$ independent of $\mathbf{s}_\tau$ for all dates and that $\mathbf{s}_t$ follows a Markov chain.

Equation (1) postulates that recessions are the sole source of dynamics in state employment growth. There is no conceptual problem with adding lagged values of $\mathbf{y}_{t-j}$ or $\mathbf{s}_{t-j}$ to this equation, though that would greatly increase the number of parameters and regimes for which one needs to draw an inference. We regard the parsimonious formulation (1) as more robust than more richly parameterized models for purposes of characterizing the broad features of business cycles across states. We also adopt the simplifying assumption that $\boldsymbol{\Omega}$ is diagonal:

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}.$$

This reduces the number of variance parameters from $N(N+1)/2$ down to N, and, unfortunately, is necessary for the particular algorithms we employ to be valid. Our model thus assumes that coincident recessions, or the tendency of a recession in one state to lead to a recession in another,

are the only reason that employment growth would be correlated across states. Again, this is a stronger formulation than one might like, though we think nevertheless an interesting one for getting a broad summary of some of the ways that the business cycle may be propagated across regions.

Despite these assumptions, the model (1) is numerically intractable without further simplification. If state 1 can be in recession while 2 and 3 are not, or 1 and 2 in recession while 3 is not, there are $\eta = 2^N$ different possibilities, or 2.8×10^{14} different configurations in the case of the 48 contiguous states. Implementing the algorithm for inference and likelihood evaluation in Hamilton (1994, p. 692) would require calculation of an $(\eta \times 1)$ vector $\boldsymbol{\xi}_t$ and an $(\eta \times \eta)$ matrix $\mathbf{P}$, which is not remotely feasible. Even if it somehow could be implemented, such a formulation is trying to infer much more information from a $(T \times N)$ data set than can be reasonably justified.[2]

Our approach, as in Frühwirth-Schnatter and Kaufmann (2008), is to assume that recession dynamics can be characterized in terms of a small number $K << 2^N$ of different clusters and by an aggregate indicator $z_t \in \{1, 2, ..., K\}$ signifying which cluster is in recession at date t. We associate with cluster 1 an $(N \times 1)$ vector $\mathbf{h}_1 = (h_{11}, ..., h_{N1})'$ whose nth element is unity when state n is associated with cluster 1 and 0 if state n is not associated with the cluster. When $z_t = 1$, all the states associated with cluster 1 would be in recession. In general,

$$\mathbf{y}_t | z_t = k \sim N(\mathbf{m}_k, \boldsymbol{\Omega}),$$

where

$$\mathbf{m}_k = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 \odot \mathbf{h}_k.$$

Conditional on knowing the values of $\mathbf{h}_1, ..., \mathbf{h}_K$, this is a standard Markov-switching framework for which inference methods are well known. The new question is how to infer the configurations of $\mathbf{h}_1, ..., \mathbf{h}_K$ from the data. We impose two of these configurations a priori, stipulating that $\mathbf{h}_K$ is a column of all zeros (so that every state is in expansion when $z_t = K$), and $\mathbf{h}_{K-1}$ is a column of all ones (every state is in recession when $z_t = K - 1$). We will refer to clusters other than those characterized by $\mathbf{h}_{K-1}$ and $\mathbf{h}_K$ as "idiosyncratic" clusters and let $\kappa = K - 2$ denote the number of

[2]Others have posited alternative methods for estimating large panel Markov-switching models. See, for example, Sims, Waggoner, and Zha (2008) and Kaufmann (2010).

idiosyncratic clusters. Thus, when $z_t = 1, 2, ..., \kappa$, some states are in recession and others are not. The values of $\mathbf{h}_1, ..., \mathbf{h}_\kappa$ are unobserved variables that influence the probability distribution of the observed data $\{\mathbf{y}_t\}_{t=1}^T$.

We postulate that there is a $(P_k \times 1)$ vector $\mathbf{x}_{nk}$ that influences whether state n experiences a recession when $z_t = k$ according to

$$p(h_{nk}) = \begin{cases} 1/\left[1 + \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)\right] & \text{if } h_{nk} = 0 \\ \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)/\left[1 + \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)\right] & \text{if } h_{nk} = 1 \end{cases} \tag{2}$$

for $n = 1, ..., N$; $k = 1, ..., \kappa$. Note that state n could be affiliated with more than one idiosyncratic cluster.[3] Alternatively, state n would participate only in national recessions if $h_{n1} = \cdots = h_{n\kappa} = 0$.

We think of $\boldsymbol{\beta}_k$ as a population parameter – prior to the generation of any data, nature generated a value of h_{nk} according to (2). We will then draw a Bayesian posterior inference about the population parameter $\boldsymbol{\beta}_k$. Following Holmes and Held (2006), it is convenient for purposes of the estimation algorithm to represent this generation of h_{nk} given $\boldsymbol{\beta}_k$ as the outcome of another unobserved pair of latent variables, denoted ξ_{nk} and ψ_{nk}. The ability to do so comes from the following observation by Andrews and Mallows (1974). Let ψ_{nk} have the limiting distribution of the Kolmogorov-Smirnov test statistic, whose density Devroye (1986, p. 161) writes as

$$p\left(\psi_{nk}\right) = 8\sum_{j=1}^{\infty}\left(-1\right)^{j+1} j^2\psi_{nk}\exp\left(-2j^2\psi_{nk}^2\right). \tag{3}$$

Andrews and Mallows showed that if $\psi_{nk} \sim KS$ and $e_{nk} \sim N\left(0,1\right)$, then $\xi_{nk} = \mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k + 2\psi_{nk}e_{nk}$ has a logistic distribution with mean $\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k$ and unit scale parameter, for which the cdf is

$$\Pr\left(\xi_{nk} \leq z\right) = \frac{1}{1 + \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k - z\right)}.$$

Thus, as in Holmes and Held (2006), we have that

$$\Pr\left(\xi_{nk} > 0\right) = \frac{\exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)}{1 + \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)}.$$

[3]This approach stands in contrast with the typical notion of a "region". Government agencies (BEA, Bureau of Labor Statistics, Census, etc.) define their regions such that any state can be a member of only one region. Empirical studies (e.g., Crone, 2005) make a similar exclusivity restriction.

In other words, if we thought of nature as having generated ξ_{nk} from a $N\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k, \lambda_{nk}\right)$ distribution where $\lambda_{nk} = 4\psi_{nk}^2$ for $\psi_{nk} \sim KS$, and then selected h_{nk} to be unity if $\xi_{nk} > 0$, that is equivalent to claiming that the value of h_{nk} was generated according to the probability specified in (2).

3 Bayesian posterior inference.

The task of data analysis is to draw a Bayesian posterior inference about the values of both population parameters and the unobserved latent variables. We divide these unknown objects into several categories. The set $\theta = \{\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \boldsymbol{\Omega}\}$ characterizes the growth rates for each state in recession and expansion and the standard deviation σ_n of employment growth rates for state n around those means. The $(K \times K)$ matrix $\mathbf{P}$ contains the transition probabilities for regimes, with row i, column j element

$$p_{ji} = p\left(z_t = j | z_{t-1} = i\right),$$

where as in Hamilton (1994, p. 679) each column of $\mathbf{P}$ sums to unity.

There are also two groups of unobserved latent variables. The $(T \times 1)$ vector $\mathbf{z} = (z_1, ..., z_T)'$ summarizes which clusters are in recession at each date, while $h = \{\mathbf{h}_1, ..., \mathbf{h}_\kappa\}$ summarizes the cluster affiliation of each state where $\mathbf{h}_k = (h_{1k}, ..., h_{Nk})'$ denotes the $(N \times 1)$ vector characterizing which states participate in cluster k. There are also three other sets of variables and parameters associated with that realization of h. Let $\boldsymbol{\xi}_k = (\xi_{1k}, ..., \xi_{Nk})'$ and $\boldsymbol{\lambda}_k = (\lambda_{1k}, ..., \lambda_{Nk})'$ denote the associated auxiliary variables [see Tanner and Wong (1987)] that are viewed as having determined $\mathbf{h}_k$ according to:

$$h_{nk} = \begin{cases} 1 & \text{if } \xi_{nk} > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

$$\xi_{nk} | \boldsymbol{\beta}_k, \lambda_{nk} \sim N\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k, \lambda_{nk}\right), \tag{5}$$

$$\lambda_{nk} = 4\psi_{nk}^2,$$

$$\psi_{nk} \sim KS.$$

Collect all the latent variables associated with the cluster affiliations in a set $H = \{h, \xi, \lambda\}$, where $\xi = \{\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_\kappa\}$ and $\lambda = \{\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_\kappa\}$, while $\beta = \{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_\kappa\}$ denotes the set of all the logistic

coefficient vectors.

3.1 Priors.

Recall that a positive scalar x is said to have a $\Gamma(\alpha, \beta)$ distribution if its density is

$$p(x) = \begin{cases} \left[\beta^{\alpha}/\Gamma(\alpha)\right] x^{\alpha-1} e^{-\beta x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

We adopt a $\Gamma(\nu/2, \delta/2)$ prior for σ_n^{-2}:

$$p\left(\sigma_n^{-2}\right) \propto \sigma_n^{-\nu+2} \exp\left(-\delta\sigma_n^{-2}/2\right). \tag{7}$$

We use a $N(\mathbf{m}, \sigma^2\mathbf{M})$ prior for $\boldsymbol{\mu}_n = (\mu_{n0}, \mu_{n1})'$:

$$p\left(\boldsymbol{\mu}_n|\sigma_n\right) \propto \left|\sigma_n^2\mathbf{M}\right|^{-1/2} \exp\left\{-\left(\boldsymbol{\mu}_n - \mathbf{m}\right)' \left[\sigma_n^2\mathbf{M}\right]^{-1} \left(\boldsymbol{\mu}_n - \mathbf{m}\right)/2\right\}. \tag{8}$$

With independent priors across states, we then have

$$p(\theta) = \prod_{n=1}^{N} p\left(\boldsymbol{\mu}_n|\sigma_n\right) p\left(\sigma_n^{-2}\right).$$

We model transition probabilities using a Dirichlet prior. Recall that for $\mathbf{w} = (w_1, ..., w_m)'$ with $w_i \in [0, 1]$ and $\sum_{i=1}^{m} w_i = 1$, we say that $\mathbf{w}$ has a Dirichlet distribution with parameter vector $\boldsymbol{\alpha}$, denoted $\mathbf{w} \sim D(\boldsymbol{\alpha})$, if the joint density of $\{w_1, ..., w_{m-1}\}$ is given by

$$p(w_1, ..., w_{m-1}) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_m)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_m)} w_1^{\alpha_1 - 1} \cdots w_m^{\alpha_m - 1}.$$

We adopt the diffuse Dirichlet prior ($D(\mathbf{0})$) for each column of $\mathbf{P}$:

$$p(\mathbf{P}) \propto p_{11}^{-1} \cdots p_{KK}^{-1}.$$

Our prior distribution for $\boldsymbol{\beta}_k$ is characterized by independent Normal distributions,

$$\boldsymbol{\beta}_k \sim N\left(\mathbf{b}_k, \mathbf{B}_k\right) \quad \text{for } k = 1, ..., \kappa, \tag{9}$$

with $p\left(\beta\right)$ the product of (9) over $k = 1, ..., \kappa$. Then,

$$p\left(H, \beta\right) = p\left(H|\beta\right) p\left(\beta\right),$$

where $p\left(H|\beta\right)$ is the product of (3) through (5) over $k = 1, ..., \kappa$ and $n = 1, ..., N$.

Numerical values for the prior parameters are summarized in Table 1. Our prior expectation is that the average employment growth rate in an expansion (reported at an annual rate) would be +1%, and likely between -1% and +3%, while average employment growth in a recession would be between -3% and +1%. The prior mean for β_{kj} implies that variable x_{nj} has no effect on whether state n is included in cluster k, and the prior distribution regards the variable as equally likely to increase or decrease the probability of state n's inclusion in the cluster. The explanatory variables $\mathbf{x}_n$ are normalized to have unit mean, so that if the first element of $\boldsymbol{\beta}_k$ is unity (a high value for its prior range) and others are at zero, a state for which $\mathbf{x}_n$ is at the average value for all states would be included in cluster k with probability $e/(1+e) = 0.73$; a low value (-1) would imply an unconditional probability of $e^{-1}/(1+e^{-1}) = 0.27$. Diffuse priors were used for σ_n^{-2} and $\mathbf{P}$.

3.2 Joint distribution.

Let $\mathbf{Y}$ denote the $(T \times N)$ matrix consisting of the observed growth rates for all states at all dates, where T is the length of the time series. The joint density-distribution for data, parameters, and latent variables for the logistic clustering formulation is given by

$$\begin{aligned} p\left(\mathbf{Y}, \theta, \mathbf{P}, \mathbf{z}, H, \beta\right) &= p\left(\mathbf{Y}|\theta, \mathbf{P}, \mathbf{z}, H, \beta\right) p\left(\mathbf{z}|\theta, \mathbf{P}, H, \beta\right) p\left(\theta|\mathbf{P}, H, \beta\right) p\left(\mathbf{P}|H, \beta\right) p\left(H, \beta\right) \\ &= p\left(\mathbf{Y}|\theta, \mathbf{z}, h\right) p\left(\mathbf{z}|\mathbf{P}\right) p\left(\theta\right) p\left(\mathbf{P}\right) p\left(H, \beta\right). \end{aligned} \tag{10}$$

Note that ξ and λ affect the likelihood only through the value of h and are only relevant as auxiliary parameters to facilitate generation of posterior values of β. Specifically, one can integrate (10)

over all possible values of ξ and λ to obtain

$$\begin{aligned} p(\mathbf{Y},\theta,\mathbf{P},\mathbf{z},h,\beta) &= \int p(\mathbf{Y}|\theta,\mathbf{z},h)\,p(\mathbf{z}|\mathbf{P})\,p(\theta)\,p(\mathbf{P})\,p(H,\beta)\;\,d\xi\,d\lambda \\ &= p(\mathbf{Y}|\theta,\mathbf{z},h)\,p(\mathbf{z}|\mathbf{P})\,p(\theta)\,p(\mathbf{P})\int p(H,\beta)\;\,d\xi\,d\lambda \\ &= p(\mathbf{Y}|\theta,\mathbf{z},h)\,p(\mathbf{z}|\mathbf{P})\,p(\theta)\,p(\mathbf{P})\,p(h|\beta)\,p(\beta)\,, \end{aligned} \tag{11}$$

where $p(h|\beta)$ is the product of (2) over $k = 1, ..., \kappa$ and $n = 1, ..., N$.

The conditional likelihood $p(\mathbf{Y}|\theta,\mathbf{z},h)$ can be written as follows. Collect the state n observations for all dates in a $(T \times 1)$ vector $\mathbf{Y}_n = (y_{1n}, ..., y_{Tn})'$ and let $\boldsymbol{\theta}_n = \left(\mu_{n0}, \mu_{n1}, \sigma_n^{-2}\right)'$. Then,

$$p(\mathbf{Y}|\theta,\mathbf{z},h) = \prod_{n=1}^{N} p(\mathbf{Y}_n|\boldsymbol{\theta}_n,\mathbf{z},h) \tag{12}$$

$$p(\mathbf{Y}_n|\boldsymbol{\theta}_n,\mathbf{z},h) = \prod_{t=1}^{T} p(y_{tn}|\boldsymbol{\theta}_n,z_t,h)$$

$$p(y_{tn}|\boldsymbol{\theta}_n,z_t,h) \propto \sigma_n^{-1} \exp\left[\frac{-\left[y_{tn} - \boldsymbol{\mu}_n'\mathbf{w}(z_t,h)\right]^2}{2\sigma_n^2}\right]$$

$$\mathbf{w}(z_t,h) = (1, h_{n,z_t})'.$$

The unconditional probabilities for $\mathbf{z}$ are given by

$$p(\mathbf{z}|\mathbf{P}) = p(z_1)\prod_{t=2}^{T} p_{z_{t-1},z_t}$$

for p_{z_{t-1},z_t} the row z_t, column z_{t-1} element of $\mathbf{P}$. The $t = 1$ aggregate regime is set to expansion a priori:

$$p(z_1) = \begin{cases} 1 & \text{for } z_1 = K \\ 0 & \text{otherwise} \end{cases}.$$

The algorithm is initialized with empty clusters and aggregate recessions ($z_t = K - 1$) set to match the NBER recession dates. Regimes for all other time periods are randomized.[4]

[4]Results randomizing the regimes for all time periods were similar but converged more slowly.

3.3 Drawing $\boldsymbol{\Omega}$ given $\mathbf{Y}$,$\mu, \mathbf{P}, \mathbf{z}, H, \beta$.

Our general Bayesian inference is via the Gibbs sampler [see Gelfand and Smith (1990); Casella and George (1992); Carter and Kohn (1994)], in which we will generate a draw for one block of parameters or latent variables conditional on the others. This subsection discusses generation of $\boldsymbol{\Omega}$ conditional on the data $\mathbf{Y}$ and on the values for $\mu, \mathbf{P}, \mathbf{z}$, H, and β that were, in turn, generated by the previous step of the iteration. In the next subsection, we will discuss how to draw μ given $\mathbf{Y}, \boldsymbol{\Omega}, \mathbf{P}, \mathbf{z}, H, \beta$. Both distributions can be derived from

$$p\left(\theta|\mathbf{Y}, \mathbf{P}, \mathbf{z}, H, \beta\right) = \frac{p\left(\theta, \mathbf{Y}, \mathbf{P}, \mathbf{z}, H, \beta\right)}{\int p\left(\theta, \mathbf{Y}, \mathbf{P}, \mathbf{z}, H, \beta\right)\, d\theta}, \tag{13}$$

where the numerator is given by (10) and $\int [.]\, d\theta$ denotes the definite integral over all the possible values for θ. But multiplicative terms not involving θ cancel from the numerator and the denominator of (13), so that

$$\begin{aligned} p\left(\theta|\mathbf{Y}, \mathbf{P}, \mathbf{z}, H, \beta\right) &\propto p\left(\mathbf{Y}|\theta, \mathbf{z}, h\right) p\left(\theta\right) \\ &= \prod_{n=1}^{N} p\left(\mathbf{Y}_n|\boldsymbol{\theta}_n, \mathbf{z}, h\right) p\left(\boldsymbol{\theta}_n\right). \end{aligned}$$

Hence, the $\boldsymbol{\theta}_n$ given $\mathbf{Y}, \mathbf{P}, \mathbf{z}, H, \beta$ are independent across n with

$$\begin{aligned} p\left(\boldsymbol{\theta}_n|\mathbf{Y}, \mathbf{P}, \mathbf{z}, H, \beta\right) &\propto p\left(\mathbf{Y}_n|\boldsymbol{\theta}_n, \mathbf{z}, h\right) p\left(\boldsymbol{\theta}_n\right) \\ &\propto p\left(\boldsymbol{\theta}_n\right) \sigma_n^{-T} \exp\left[-\sum_{t=1}^{T} [y_{tn} - \boldsymbol{\mu}_n^{'} \mathbf{w}\left(z_t, h\right)]^2 / \left(2\sigma_n^2\right)\right]. \end{aligned} \tag{14}$$

Substituting (7) into (14) and dividing by the integral over $\boldsymbol{\mu}_n$, we have

$$p\left(\sigma_n^{-2}|\mathbf{Y}, \mu, \mathbf{P}, \mathbf{z}, H\right) \propto \sigma_n^{-T-\nu+2} \exp\left[-\left(\delta + \hat{\delta}\right) \sigma_n^{-2}/2\right]$$

for $\hat{\delta} = \sum_{t=1}^{T} \left[y_{tn} - \boldsymbol{\mu}_n^{'} \mathbf{w}\left(z_t, h\right)\right]^2$. Recalling (6), we thus generate σ_n^{-2} from a $\Gamma\left(\left(\nu + T\right)/2, \left(\delta + \hat{\delta}\right)/2\right)$ distribution, a standard result as in Kim and Nelson (1999, p. 181).

3.4 Drawing μ given $\mathbf{Y}, \mathbf{\Omega}, \mathbf{P}, \mathbf{z}, H, \beta$.

Using (8) in (14) and this time dividing by the integral over σ_n, we again see, as in Kim and Nelson (1999, p. 181), that

$$\boldsymbol{\mu}_n|\mathbf{Y}, \mathbf{\Omega}, \mathbf{P}, \mathbf{z}, H, \beta \sim N\left(\mathbf{m}_n^*, \sigma_n^2 \mathbf{M}_n^*\right) \tag{15}$$

for

$$\mathbf{M}_n^* = \left(\mathbf{M}^{-1} + \mathbf{C}_n\right)^{-1}$$

$$\mathbf{m}_n^* = \mathbf{M}_n^* \left(\mathbf{M}^{-1}\mathbf{m} + \mathbf{c}_n\right)$$

$$\mathbf{C}_n = \left[\sum_{t=1}^{T} \mathbf{w}\left(z_t, h\right) \mathbf{w}\left(z_t, h\right)'\right]$$

$$\mathbf{c}_n = \left[\sum_{t=1}^{T} \mathbf{w}\left(z_t, h\right) y_{tn}\right].$$

3.5 Drawing P given $\mathbf{Y}$,$\theta, \mathbf{z}, H, \beta$.

Conditional on H and $\mathbf{z}$, this is again a standard inference problem for a K-state Markov switching process, as in Chib (1996, p. 84). From (10),

$$p\left(\mathbf{P}|\mathbf{Y}, \theta, \mathbf{z}, H\right) \propto p\left(\mathbf{z}|\mathbf{P}\right) p\left(\mathbf{P}\right),$$

column i of which will be recognized as $D(\boldsymbol{\alpha}_i^*)$ distribution, where the jth element of the vector $\boldsymbol{\alpha}_i^*$ is given by

$$\alpha_{ij}^* = \frac{\sum_{t=2}^{T} \delta\left(z_{t-1} = i, z_t = j\right)}{\sum_{t=2}^{T} \delta\left(z_{t-1} = i\right)},$$

which is just the fraction of times that regime i is observed to be followed by regime j among the sequence $\{z_1, ..., z_T\}$.

3.6 Drawing z given $\mathbf{Y}$,$\theta, \mathbf{P}$,H, β.

Here,

$$p\left(\mathbf{z}|\mathbf{Y}, \theta, \mathbf{P}, H, \beta\right) \propto p\left(\mathbf{Y}|\theta, \mathbf{z}, h\right) p\left(\mathbf{z}|\mathbf{P}\right).$$

Again as in Chib (1996, p. 83),

$$p\left(\mathbf{z}|\mathbf{Y},\theta,\mathbf{P},H,\beta\right)=p\left(z_T|\mathbf{Y},\theta,\mathbf{P},h\right)\prod_{t=1}^{T-1}p\left(z_t|z_{t+1},...,z_T,\mathbf{Y},\theta,\mathbf{P},h\right).$$

But z_{t+1} conveys all the information about z_t embodied by future z or y. Thus if $\mathcal{Y}_t=\{y_{\tau n}:\tau\leq t;n=1,...,N\}$ collects observations from all states for all dates through t,

$$p\left(\mathbf{z}|\mathbf{Y},\theta,\mathbf{P},H,\beta\right)=p\left(z_T|\mathcal{Y}_T,\theta,\mathbf{P},h\right)\prod_{t=1}^{T-1}p\left(z_t|z_{t+1},\mathcal{Y}_t,\theta,\mathbf{P},h\right). \tag{16}$$

One can calculate $p\left(z_t|\mathcal{Y}_t,\theta,\mathbf{P},h\right)$ by iterating on equation [22.4.5] in Hamilton (1994)[5], the terminal value of which $(t=T)$ gives us $p\left(z_T|\mathcal{Y}_T,\theta,\mathbf{P},h\right)$, the first term in (16). Furthermore,

$$p\left(z_t|z_{t+1},\mathcal{Y}_t,\theta,\mathbf{P},h\right)=\frac{p_{z_t,z_{t+1}}p\left(z_t|\mathcal{Y}_t,\theta,\mathbf{P},h\right)}{\sum_{j=1}^{K}p_{j,z_{t+1}}p\left(z_t=j|\mathcal{Y}_t,\theta,\mathbf{P},h\right)},$$

allowing us to generate $z_T,z_{T-1},...,z_1$ sequentially.

3.7 Generating H.

We now define $H_k=\{\mathbf{h}_k,\boldsymbol{\xi}_k,\boldsymbol{\lambda}_k\}$ and $H^{[k]}=\left\{\mathbf{h}_j,\boldsymbol{\xi}_j,\boldsymbol{\lambda}_j:j=1,...,\kappa;j\neq k\right\}$. Our strategy will be to generate the elements associated with cluster k (denoted H_k) conditional on all the elements of all the other clusters (denoted $H^{[k]}$). We will, in turn, break down the generation of H_k given $\mathbf{Y},H^{[k]},\theta,\mathbf{P},\mathbf{z},\beta$ into a series of steps, first generating $\mathbf{h}_k$, then $\boldsymbol{\xi}_k$ conditional on $\mathbf{h}_k$, and finally $\boldsymbol{\lambda}_k$ conditional on $\mathbf{h}_k$ and $\boldsymbol{\lambda}_k$, all conditioning on $H^{[k]}$.

3.7.1 Drawing $\mathbf{h}_k$ given $\mathbf{Y},H^{[k]},\theta,\mathbf{P},\mathbf{z},\beta$.

From (11),

$$\begin{aligned}p\left(\mathbf{h}_k|\mathbf{Y},H^{[k]},\theta,\mathbf{P},\mathbf{z},\beta\right) &\propto p\left(\mathbf{Y}|\theta,\mathbf{z},h\right)p\left(\mathbf{h}_k|\boldsymbol{\beta}_k\right)\\ &= \prod_{n=1}^{N}p\left(\mathbf{Y}_n|h_{nk},h^{[k]},\theta,\mathbf{z}\right)p\left(h_{nk}|\boldsymbol{\beta}_k\right).\end{aligned}$$

[5]Here, $\boldsymbol{\xi}_t$ is a $(K\times 1)$ vector whose kth element is unity when $z_t=k$ and zero otherwise, while $\boldsymbol{\eta}_t$ is a $(K\times 1)$ vector whose kth element is $\prod_{n=1}^{N}p(y_{tn}|\theta,z_t=k,h)$, while $\hat{\boldsymbol{\xi}}_{0|0}=(0,0,...,1)'$.

In other words, we can generate h_{nk} for $n = 1, ..., N$ independently across states from

$$\Pr\left(h_{nk} = 1|\mathbf{Y}, h^{[k]}, \theta, \mathbf{P}, \mathbf{z}, \beta\right) = \frac{p\left(\mathbf{Y}_n|h_{nk} = 1, h^{[k]}, \theta, \mathbf{z}\right)\Pr\left(h_{nk} = 1|\boldsymbol{\beta}_k\right)}{\sum_{j=0}^{1} p\left(\mathbf{Y}_n|h_{nk} = j, h^{[k]}, \theta, \mathbf{z}\right)\Pr\left(h_{nk} = j|\boldsymbol{\beta}_k\right)},$$

where

$$\Pr(h_{nk} = j|\boldsymbol{\beta}_k) = \begin{cases} 1/\left[1 + \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)\right] & \text{for } j = 0 \\ \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)/\left[1 + \exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)\right] & \text{for } j = 1 \end{cases}.$$

3.7.2 Drawing $\boldsymbol{\xi}_k$ given $\mathbf{Y}, \mathbf{h}_k, H^{[k]}, \theta, \mathbf{P}, \mathbf{z}, \beta$.

Here, we have

$$\begin{aligned} p\left(\boldsymbol{\xi}_k|\mathbf{Y}, \mathbf{h}_k, H^{[k]}, \theta, \mathbf{P}, \mathbf{z}, \beta\right) &= p\left(\boldsymbol{\xi}_k|\mathbf{h}_k, \boldsymbol{\beta}\right) \\ &= \prod_{n=1}^{N} p\left(\xi_{nk}|h_{nk}, \boldsymbol{\beta}_k\right). \end{aligned}$$

Note that if we had conditioned on λ_{nk}, then ξ_{nk} would have a Normal distribution. However, without that conditioning, we are back to the logistic distribution that motivates the parameterization in terms of (λ_{nk}, ξ_{nk}). Holmes and Held (2006) argued that generating ξ_{nk} from the unconditional distribution and then generating λ_{nk} conditional on ξ_{nk} will give the algorithm better convergence properties. For the posterior distribution given h_{nk}, we know that ξ_{nk} is logistic with mean $\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k$ and truncated by $\xi_{nk} \geq 0$ if $h_{nk} = 1$ and $\xi_{nk} < 0$ if $h_{nk} = 0$. Recall that if $u \sim U[0, 1]$, then $\xi = A - \log\left(u^{-1} - 1\right)$ has a logistic distribution with mean $E(\xi) = A$.[6] Furthermore, $\xi \geq 0$ iff $u \geq 1/(1 + \exp(A))$. In other words, we want to generate u from a uniform distribution over the interval $[0, 1/(1 + \exp(A))]$ when $h_{nk} = 0$ and $u \sim U[1/(1 + \exp(A)), 1]$ when $h_{nk} = 1$. Note finally that if $u^* \sim U[0, 1]$, then $a + (b - a)u^* \sim U[a, b]$. Thus, we generate $u_{nk}^* \sim U[0, 1]$ and

[6] This claim may be verified directly as follows:

$$\begin{aligned} \Pr(\xi \leq z) &= \Pr\left[A - \log(u^{-1} - 1) \leq z\right] \\ &= \Pr\left[\log(u^{-1} - 1) \geq A - z\right] \\ &= \Pr\left[u^{-1} \geq 1 + \exp(A - z)\right] \\ &= \Pr\left[u \leq \frac{1}{1 + \exp(A - z)}\right] \\ &= \frac{1}{1 + \exp(A - z)}, \end{aligned}$$

which will be recognized as the cdf of a logistic variable with mean A.

define

$$u_{nk} = \begin{cases} \frac{1}{1+\exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)} u_{nk}^{*} & \text{if } h_{nk} = 0 \\ \frac{1}{1+\exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)} + \frac{\exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)}{1+\exp\left(\mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)} u_{nk}^{*} & \text{if } h_{nk} = 1 \end{cases}.$$

Then, $\xi_{nk} = \mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k - \log\left(u_{nk}^{-1} - 1\right)$.

3.7.3 Drawing $\boldsymbol{\lambda}_k$ given $\mathbf{Y}, \boldsymbol{\xi}_k, \mathbf{h}_k, H^{[k]}, \theta, \mathbf{P}, \mathbf{z}, \beta$.

Now,

$$\begin{aligned} p\left(\boldsymbol{\lambda}_k | \mathbf{Y}, \boldsymbol{\xi}_k, \mathbf{h}_k, H^{[k]}, \theta, \mathbf{P}, \mathbf{z}, \beta\right) &= p\left(\boldsymbol{\lambda}_k | \boldsymbol{\xi}_k, \boldsymbol{\beta}_k\right) \\ &\propto p\left(\boldsymbol{\xi}_k | \boldsymbol{\lambda}_k, \boldsymbol{\beta}_k\right) p\left(\boldsymbol{\lambda}_k\right) \\ &= \prod_{n=1}^{N} p\left(\xi_{nk} | \lambda_{nk}, \boldsymbol{\beta}_k\right) p\left(\lambda_{nk}\right). \end{aligned}$$

Again, as in Holmes and Held (2006), we set $r_{nk}^2 = \left(\xi_{nk} - \mathbf{x}_{nk}^{'}\boldsymbol{\beta}_k\right)^2$ and use as a proposal density a Generalized Inverse Gaussian density,

$$\hat{\lambda}_{nk} \sim GIG\left(1/2, 1, r^2\right),$$

for which a draw can be generated as follows. Generate w_{nk} the square of a standard Normal and set

$$v_{nk} = 1 + \frac{w_{nk} - \sqrt{w_{nk}\left(4r + Y\right)}}{2r}.$$

Generate a separate $\hat{u}_{nk} \sim U\left[0, 1\right]$, and set

$$\hat{\lambda}_{nk} = \begin{cases} r/v_{nk} & \text{if } \hat{u}_{nk} \leq 1/\left(1 + v_{nk}\right) \\ r v_{nk} & \text{otherwise} \end{cases}.$$

We then decide to accept $\hat{\lambda}_{nk}$ (or else repeat the above steps) using the algorithm described by Holmes and Held (2006, p. 165).

3.8 Drawing β given $\mathbf{Y}, \theta, \mathbf{P}, \mathbf{z}, H$.

Notice

$$p\left(\beta|\mathbf{Y}, \theta, \mathbf{P}, \mathbf{z}, H\right) = \prod_{k=1}^{\kappa} p\left(\boldsymbol{\beta}_k|\boldsymbol{\xi}_k, \boldsymbol{\lambda}_k\right),$$

which is just a standard Normal regression model for each $\boldsymbol{\beta}_k$ of the form

$$\boldsymbol{\xi}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \boldsymbol{\varepsilon}_k,$$

$$\underset{(N\times P_k)}{\mathbf{X}_k} = \begin{bmatrix} \mathbf{x}_{1k}^{'} \\ \vdots \\ \mathbf{x}_{Nk}^{'} \end{bmatrix},$$

$$\boldsymbol{\varepsilon}_k \sim N\left(\mathbf{0}, \mathbf{W}_k\right),$$

$$\underset{(N\times N)}{\mathbf{W}_k} = diag\left[\lambda_{1k}, ..., \lambda_{Nk}\right].$$

Thus,

$$\boldsymbol{\beta}_k|\mathbf{Y}, \theta, \mathbf{P}, \mathbf{z}, H \sim N\left(\mathbf{b}_k^*, \mathbf{B}_k^*\right),$$

where

$$\mathbf{b}_k^* = \left(\mathbf{B}_k^{-1} + \mathbf{X}_k^{'}\mathbf{W}_k^{-1}\mathbf{X}_k\right)^{-1}\left(\mathbf{B}_k^{-1}\mathbf{b}_k + \mathbf{X}_k^{\prime}\mathbf{W}_k^{-1}\boldsymbol{\xi}_k\right)$$

and

$$\mathbf{B}_k^* = \left(\mathbf{B}_k^{-1} + \mathbf{X}_k^{\prime}\mathbf{W}_k^{-1}\mathbf{X}_k\right)^{-1}.$$

3.9 Label switching.

The model described above is unidentified in two respects. First, if we were to switch the values of $\boldsymbol{\mu}_0$ with $\boldsymbol{\mu}_1$, and correspondingly switch the last two columns and then the last two rows of $\mathbf{P}$, the likelihood function would be unchanged. Likewise, switching the definition of clusters (e.g., switching $\mathbf{h}_1$ with $\mathbf{h}_2$ and switching the first two columns and first two rows of $\mathbf{P}$), the likelihood function would be unchanged.

The first is a familiar issue in the literature, and we deal with it in a typical way, by normalizing $\mu_{n1} \leq 0$. We implement this by rejecting any generated $\boldsymbol{\mu}_n$ that does not satisfy the restriction

and redrawing from (15) until obtaining a draw that satisfies the normalization restriction.

The second issue is unique to our clustering approach. We mitigate this in part by imposing the restriction that the process cannot transition from one idiosyncratic regime to another, that is, imposing $p_{ij} = 0$ if i and j are both less than $K - 1$ and if $i \neq j$. We are thus ruling out transitions in which recession for a subset of states is followed by those states going out of recession and a different set of states going into recession. We find that once these restrictions are imposed, for this data set, the posterior distribution is sufficiently tightly concentrated in the vicinity of a given representation that a given Monte Carlo Markov chain does not jump across to an alternative representation. However, different starting values can converge to different representations of the same system.

3.10 Cross-validation.

Our next objective is to choose the number of clusters. For this, we utilize cross validation [see Picard and Cook (1984); Gelfand, Dey, and Chang (1992); Shao (1993); and Bernardo and Smith (1994)], which computes a quasi-out-of-sample score by estimating the model with a subset of data and validating with the omitted data, and has been adapted for similar econometric models [e.g., Geweke and Keane (2007)]. We described above an algorithm to generate a draw for $\{\mathbf{Z}_T, \theta, \mathbf{P}, h\}$ conditional on the full data set $\mathbf{Y}_T$. We now partition the data $\mathbf{Y}_T$ into R blocks,[7]

$$\mathbf{Y}_T = \left[\begin{array}{cccc} \mathcal{Y}_1 & \mathcal{Y}_2 & \cdots & \mathcal{Y}_R \end{array} \right]$$

for

$$\mathcal{Y}_r = \left[\begin{array}{cccc} \mathbf{y}_{t_r} & \mathbf{y}_{t_r+1} & \cdots & \mathbf{y}_{t_{r+1}-1} \end{array} \right],$$

and let $\mathcal{Y}^{(r)}$ denote the full set of observations with block $\mathcal{Y}_r$ deleted:

$$\mathcal{Y}^{(r)} = \left[\begin{array}{cccccc} \mathcal{Y}_1 & \cdots & \mathcal{Y}_{r-1} & \mathcal{Y}_{r+1} & \cdots & \mathcal{Y}_R \end{array} \right].$$

[7]The method used here is sometimes described as R-fold cross validation where R is the number of subsamples over which validation is computed.

Likewise define $\mathcal{Z}^{(r)}$ to be a matrix of realizations for $\mathbf{Z}_T$ with block r deleted:

$$\mathcal{Z}^{(r)} = \left[\begin{array}{cccccc} \mathcal{Z}_1 & \cdots & \mathcal{Z}_{r-1} & \mathcal{Z}_{r+1} & \cdots & \mathcal{Z}_R \end{array} \right].$$

We propose to use the principle of cross-validation see how well a particular model that was based on data $\mathcal{Y}^{(r)}$ predicts the observed value of $\mathcal{Y}_r$ using an entropy-based loss function. Specifically, we will first generate a series of draws for $\{\mathcal{Z}^{(r)}, \theta, \mathbf{P}, h\}$ from the posterior distribution conditional on only $\mathcal{Y}^{(r)}$, implemented by running our basic procedure on the subset of data that omits block r as if the only data available were that contained in $\mathcal{Y}^{(r)}$. Let $\{\mathcal{Z}^{[r,m]}, \theta^{[r,m]}, \mathbf{P}^{[r,m]}, h^{[r,m]}\}$ denote a particular draw from this posterior distribution. We will then generate a draw from the distribution of $\{\mathcal{Z}_r\}$ conditional on $\{\mathcal{Z}^{[r,m]}, \theta^{[r,m]}, \mathbf{P}^{[r,m]}, h^{[r,m]}, \mathcal{Y}^{(r)}\}$. Let $z_t^{[r,m]}$ denote the value so generated for observation t, which implies a forecast $\mathbf{m}_{z_t^{[r,m]}}$ for the value of $\mathbf{y}_t$. We will judge a model to be superior when

$$M^{-1} \sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{t=t_r}^{t_{r+1}-1} \left\{ \log \left| \mathbf{\Omega}^{[r,m]} \right| + \left(\mathbf{y}_t - \mathbf{m}_{z_t^{[r,m]}} \right)' \left(\mathbf{\Omega}^{[r,m]} \right)^{-1} \left(\mathbf{y}_t - \mathbf{m}_{z_t^{[r,m]}} \right) \right\} \tag{17}$$

is smaller.

The necessary step for this process is to generate values for $z_t^{[r,m]}$. Note first that the values for $\mathcal{Z}^{[r,m]}$ and $\mathbf{P}^{[r,m]}$ are the only relevant conditioning information:

$$p(\mathcal{Z}_r | \mathcal{Z}^{[r,m]}, \theta^{[r,m]}, \mathbf{P}^{[r,m]}, h^{[r,m]}, \mathcal{Y}^{(r)}) = p(\mathcal{Z}_r | \mathcal{Z}^{[r,m]}, \mathbf{P}^{[r,m]}).$$

Consider first generation for the last block ($r = R$). From the Markov chain property of $\{z_t\}$, this is trivially accomplished by setting $z_t^{[r,m]} = k$ with probability $p^{[r,m]}_{z_{t-1}^{[r,m]},k}$ sequentially for $t = t_R, t_{R+1}, ..., T$, where $p_{ij}^{[r,m]}$ denotes the transition probability for a transition from $z_{t-1} = i$ to $z_t = j$ associated with the draw m from the posterior distribution given $\mathcal{Y}^{(r)}$ and $z_{t-1}^{[r,m]}$ denotes the previously generated value. This iteration begins for $t = t_R$ by setting $z_{t_R-1}^{[r,m]}$ to the draw r, m value for z_{t_R-1}.

Consider next generation for the first block ($r = 1$). Here, we make use of the reverse transition probabilities

$$q_{ij}^{[r,m]} = \Pr(z_t = i | z_{t+1} = j, \mathbf{P}^{[r,m]}),$$

which can be calculated from the following considerations:

$$\begin{aligned}\Pr(z_t = i|z_{t+1} = j) &= \frac{\Pr(z_t = i, z_{t+1} = j)}{\Pr(z_{t+1} = j)} \\ &= \frac{\Pr(z_t = i)\Pr(z_{t+1} = j|z_t = i)}{\Pr(z_{t+1} = j)} \\ &= \frac{\pi_i p_{ij}}{\pi_j}\end{aligned}$$

for p_{ij} the forward transition probabilities and π_i the ergodic probabilities. The latter can be calculated as in Hamilton (1994, eq. [22.2.26]):

$$\begin{bmatrix} \pi_1 \\ \vdots \\ \pi_K \end{bmatrix} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{e}_{K+1}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{I}_K - \mathbf{P} \\ \mathbf{1}' \end{bmatrix}$$

for $\mathbf{1}$ a $(K \times 1)$ vector of ones and $\mathbf{e}_{K+1}$ a $(K+1) \times 1$ vector whose $K+1$ element is unity and others are all zero. Thus,

$$q_{ij}^{[r,m]} = \frac{\pi_i^{[r,m]} p_{ij}^{[r,m]}}{\pi_j^{[r,m]}}.$$

We thus generate $z_t^{[1,m]} = k$ with probability $q_{k,z_{t+1}^{[r,1]}}^{[1,m]}$ sequentially backwards for $t = t_2 - 1, t_2 - 2, ..., 1$.

We can generate z's for the middle blocks $r = 2, ..., R-1$ adapting the approach in Hamilton (1994, p. 701). For $t = t_r, t_r + 1, ..., t_{r+1}$, let

$$\tilde{p}_{k,t}^{[r,m]} = \Pr(z_t = k|\mathcal{Z}_1^{[r,m]}, ..., \mathcal{Z}_{r-1}^{[r,m]}, \mathbf{P}^{[r,m]})$$

and collect these scalars in the vector $\tilde{\boldsymbol{\xi}}_t^{[r,m]} = \left(\tilde{p}_{1,t}^{[r,m]}, \tilde{p}_{2,t}^{[r,m]}, ..., \tilde{p}_{K,t}^{[r,m]}\right)'$. This can be calculated recursively from

$$\tilde{\boldsymbol{\xi}}_t^{[r,m]} = \mathbf{P}^{[r,m]}\tilde{\boldsymbol{\xi}}_{t-1}^{[r,m]}$$

for $t = t_r, t_r + 1, ..., t_{r+1}$ starting from $\tilde{\boldsymbol{\xi}}_{t_r-1}^{[r,m]}$ the $(K \times 1)$ vector whose element in position $z_{t_r-1}^{[r,m]}$ is unity and all other elements are zero, and where $\mathbf{P}^{[r,m]}$ is the matrix of transition probabilities

(arranged so that columns sum to unity). Let $\mathcal{S}^{[r,m]} = \left\{ \mathcal{Z}_1^{[r,m]}, ..., \mathcal{Z}_{r-1}^{[r,m]}, \mathbf{P}^{[r,m]} \right\}$ and notice next that for any $t \in \{t_r, ..., t_{r+1} - 1\}$,

$$\begin{aligned} \Pr(z_t = j | z_{t+1} = i, \mathcal{S}^{[r,m]}) &= \frac{\Pr(z_t = j, z_{t+1} = i | \mathcal{S}^{[r,m]})}{\Pr(z_{t+1} = i | \mathcal{S}^{[r,m]})} \\ &= \frac{\tilde{p}_{j,t}^{[r,m]} p_{ji}^{[r,m]}}{\tilde{p}_{i,t+1}} \end{aligned}$$

from which

$$\Pr(z_t = j | \mathcal{Z}_1^{[r,m]}, ..., \mathcal{Z}_{r-1}^{[r,m]}, \mathcal{Z}_{r+1}^{[r,m]}, ..., \mathcal{Z}_R^{[r,m]}, \mathbf{P}^{[r,m]}\} = \frac{\tilde{p}_{j,t}^{[r,m]} p_{ji}^{[r,m]}}{\tilde{p}_{z_{t+1}^{[r,m]},t+1}}.$$

We can thus generate a value for $z_{t_{r+1}-1}^{[r,m]}$ from the above equation for $t = t_{r+1} - 1$ given the known value $z_{t_{r+1}}^{[r,m]}$ and can generate the values for $t = t_{r+1} - 2, t_{r+1} - 3, ..., t_r$ recursively by iterating on the equation backwards.

4 Empirical results.

The data used to measure state-level business cycles are the seasonally adjusted, annualized quarter-to-quarter growth rates of payroll employment.[8,9] The sample period is 1956:Q2 to 2007:Q4; Alaska and Hawaii are excluded. These data were obtained from the Bureau of Labor Statistics (BLS).

In addition to the time series data, the model in the preceding section requires a set of state-level covariates characterizing the *ex ante* likelihood of membership in a given cluster. We report results for a specification with $P_k = 4$ covariates used to explain the cluster affiliations of each state, with the same vector of explanatory variables used for each cluster ($\mathbf{x}_{nk} = \mathbf{x}_n$ for $k = 1, ..., \kappa$). The vector $\mathbf{x}_n$ includes barrels of oil produced per 100 dollars of state GDP, manufacturing employment share, financial activities employment share, and the share of total state employment accounted for by small firms.[10] We normalize each variable by dividing by the sample mean. Values for these

[8]The measure most synonymous with GDP at the state level is Gross State Product (GSP). Unfortunately, GSP is available only at an annual frequency and at a two-year lag, making it nonviable for a study of business cycles.

[9]Even at the quarterly frequency, the growth rate in state-level employment can experience large swings caused by idiosyncratic state experiences (for example, mining strikes in West Virginia). To focus on the estimation of the business cycle, we check for outliers defined as observations more than three standard deviations from each series' mean. We then set these values at two standard deviations from the series mean.

[10]The oil share was calculated as 100 times the number of barrels of crude oil produced in the state in 1984 (from the Energy Information Administration, http://tonto.eia.doe.gov/dnav/pet/pet_crd_crpdn_adc_mbbl_m.htm) divided by 1984 state personal income (from the Census Bureau, http://www.census.gov/compendia/statab/tables/08s0658.xls). The manufacturing and financial activities shares of employment by state were calculated as the average of the annual

explanatory variables are displayed in Figure 1.

We report results for some of the parameters and unobserved latent variables of interest based on the five pooled runs of 25,000 Gibbs sampler iterations, having discarded an initial burn-in of 250,000 iterations each. Table 2 displays the cross validation results using $R = 10$ subsamples for various values of k, the number of idiosyncratic clusters. Cross validation chooses $\kappa = 3$ idiosyncratic clusters with $\kappa = 4$ being the second closest alternative. We also calculated the comparable cross-validation measure when a single-equation Markov-switching model is estimated separately for each state.[11] Note that the latter specification estimates 96 separate regime transition probabilities ($p_{n,11}$ and $p_{n,22}$ for $n = 1, 2, ..., 48$), whereas the cluster specification requires only $2(\kappa + 1) + \kappa(\kappa - 1)$ transition probabilities – for example, 14 parameters for our favored case of $\kappa = 3$. Although the cluster specification is much less richly parameterized, its substantially better fit reflects the feature in the data that knowing whether state n is in recession at date t is extremely helpful for predicting whether state ℓ will be in recession at date $t + 1$.

Table 3 shows the posterior medians and means for the model parameters μ_0, μ_1, and σ^2 for each state. Table 4 gives the posterior means of the logistic coefficients $\boldsymbol{\beta}_k$ associated with each of the idiosyncratic clusters ($k = 1, ..., 3$), with a bold entry signifying that 68 percent of the posterior draws were on the same side of zero as the reported posterior mean. We also translate these coefficients into discrete derivatives (denoted $\boldsymbol{\delta}_k$). The ith element of $\boldsymbol{\delta}_k$ has the following interpretation. Let $\overline{x}_i = N^{-1} \sum_{n=1}^{N} x_{in}$ denote the average value for the ith explanatory variable. Suppose we compare two states, each of which has $x_{jn} = \overline{x}_j$ for all $j \neq i$, but in the first state, characteristic i is one standard deviation below the average $\overline{x}_i$, and in the other state, characteristic i is one standard deviation above the average. How would the probability of inclusion in cluster k, as calculated from (2), differ between the two states? The value for this magnitude implied by the posterior mean for $\boldsymbol{\beta}_k$ is reported as the ith element of the vector $\boldsymbol{\delta}_k$ in Table 4. For example, a state that was average in all respects but one standard deviation below average in the importance of oil production would be rather unlikely to be included in cluster 1, whereas a state one standard deviation above the average would be quite likely to be included. An important role

industry (NAICS) shares of total payroll employment from 1990-2006, also from the BLS. The share of small firms was computed as an average of the share of total employment in firms with fewer than 100 employees and was taken from the Statistics of U.S. Businesses data set.

[11] For the standard Markov-switching model, the cross validation is taken as the sum of (17) over all 48 states.

for manufacturing or finance makes a state less likely to be part of this cluster. States with an important role for finance were less likely to be part of cluster 2 and more likely to be included in cluster 3.

Table 5 reports posterior means of the regime transition probabilities p_{ij}. Starting with the first column, suppose that $z_t = 1$ in quarter t, which would mean that only those states that are included in cluster 1 would be in recession. We have ruled out a priori the possibility that these states go out of recession and a new different subset of states begins a recession at $t + 1$ (that is, we imposed $p_{12} = p_{13} = 0$). Although we did not impose $p_{14} = 0$, the posterior mean of p_{14}, in fact, turns out to be quite close to zero. Thus, if the states in cluster 1 go into recession, a national recession is unlikely. Moreover, the cluster 1 recession is relatively persistent, lasting an average of 3.2 quarters. By contrast, if the states in cluster 2 are in recession (i.e., $z_t = 2$), we see a national recession eventually arrive, usually within two and a half quarters ($p_{24} = 0.40$, $p_{25} = 0$). Similarly, the regime $z_t = 3$ would be characterized as the subset of states that have extended recessions. The regime $z_t = 3$ can be entered either through expansion or recession and typically signals a forthcoming national recession.

Figure 2 plots the posterior means for the regime probabilities given the data. The top panel is calculated as the fraction out of the 125,000 simulations for which z_t for the indicated quarter is equal to 4 – that is, it shows the posterior probability of a national recession. These correspond fairly closely to the traditional NBER dates, which are indicated by shaded regions in the top panel, with the exception of a few short downturns (no longer than one quarter) based on state employment data that are not characterized by the NBER as a national recession. Also, our framework would date both the 1990-91 and 2001 recessions as substantially longer based on state employment growth than the traditional NBER dates specify.[12] Interestingly, the recession that the NBER dates as 2007:Q4-2009:Q2 was recognized by this algorithm as beginning in 2007:Q2, and this assessment was made using data only through 2007:Q4. The approach thus provided an earlier signal of the most recent downturn than was provided by most other approaches at the time.

The shaded regions in the bottom three panels of Figure 2 are based on the $z_t = 4$ dates (when $\Pr(z_t = 4) > 0.99$) rather than the NBER dates, to clarify the nature of the estimated dynamics.

[12]This result is consistent with the so-called jobless recovery periods [see Koenders and Rogerson (2005) for a survey].

The cluster 1 states experienced a uniquely idiosyncratic recession during the oil price collapse in the mid-1980s, as well as several briefer episodes in the 1950s and 1960s. The recession of 1957-58 was preceded by a downturn in the cluster 2 states, and there is also some possibility that the recessions of 1980 and 2001 began in these states. By contrast, a downturn in cluster 3 states preceded the national downturns in 1973-75, 1990-91, and 2007-09, suggesting a possible role of financial factors in precipitating those recessions. Cluster 3 was also slow to recover from the 1990-91 recession.

Figure 3 indicates which states are affected by the respective idiosyncratic regimes and conveys some idea of the role played by the exogenous state characteristics $\mathbf{x}_n$ and observed employment growth rates $\mathbf{Y}$ in associating states with particular clusters. The first column of Figure 3 summarizes the inference we would draw if we knew nothing about the state other than the state characteristics $\mathbf{x}_n$ and the likely values for $\boldsymbol{\beta}_k$ as inferred from the employment data; that is, it reports for each state n and cluster k the value of

$$\int \frac{\exp(\mathbf{x}_n^{'}\boldsymbol{\beta}_k)}{1+\exp(\mathbf{x}_n^{'}\boldsymbol{\beta}_k)} f(\boldsymbol{\beta}_k|\mathbf{Y}).$$

The second column of Figure 3 reports the posterior probability of cluster designations given all the observed data:

$$p(h_{nk}=1|\mathbf{Y}).$$

The information based on state characteristics alone – specifically, the importance of oil for the state – gives a fairly sharp designation for the states included in cluster 1 (see row 1, column 1 of Figure 3). The first and second columns of the first row of Figure 3 have much in common. However, this appears to be because the particular pattern for the employment behavior of states in this cluster is so closely aligned with the importance of oil production for the state. A cluster designation similar to what we see in the first row of Figure 3 has emerged from virtually all of the specifications we have studied. Specifically, we also estimated a version of the model with no explanatory variables at all and found a similar grouping of states that experienced their own separate recession in the mid-1980s. For that matter, we found the same pattern when we

estimated models separately for each state in isolation. The conclusion that the oil-producing states experienced their own recession at the time of the oil price collapse appears to be fairly robust.

For cluster 2, the information content in the prior based on the state characteristics is not as sharp. A priori, states belonging to cluster 2 appear to be those not belonging to the oil-producing cluster. The observed employment growth rates refine and sharpen these designations considerably (row 2, column 2). Based on both information from the prior and the state-level employment data, the cluster appears to include primarily the manufacturing states in the southeast, Indiana, Michigan, and a few scattered states around the Pacific Northwest and Maine.

Based on state-level characteristics, cluster 3 appears a priori likely to consist of states with a higher employment share in financial industries but without a high component of oil production (row 3, column 1). Again, the business cycle data refine these designations. The posterior cluster probabilities for cluster 3 place high likelihood on including a number of states on the East Coast with the addition of California and Arizona.

Results for $\kappa = 4$ clusters have a similar character, with one cluster capturing the mid-1980s recession in the oil-producing states, and the other three clusters characterized by groups of states that go into recession a little earlier or come out of recession a little later than other states. It is interesting that this feature – i.e., that recessions tend to be a national phenomenon, with idiosyncrasies manifest in the timing of when they start or stop in each state – is a broad finding from different specifications of our approach. This suggests that although different recessions may have different original causes, the key defining characteristic may be their breadth– what makes an episode a recession is the fact that everybody is experiencing problems at about the same time.

5 Conclusion

Two broad conclusions emerge from our results. First, we have found substantial heterogeneity across recessions. Different recessions seemed to begin in different ways. In distinct episodes, different parts of the country could have manifested the first signs of a downturn, and the oil and agricultural states have on occasion experienced a recession while the rest of the country appears to be doing fine. Based on the geographic patterns, recessions are not all alike, but appear to differ

in their causes and propagation.

On the other hand, we were surprised that, despite this clear heterogeneity, there nevertheless appears to be a strong national component to most recessions. Although our framework allowed for the possibility of groups of states at times moving in complete isolation of the rest of the nation, we find such behavior to be the exception rather than the rule. The primary differences we find across states come down to timing – when did the recession begin and end for that state – and not whether the state was able to avoid a national downturn altogether. This suggests to us that although recessions are different in terms of their causes, there is something similar about the event itself. We would propose that a salient characteristic of a recession is the comovement across states and the eventual tendency for the entire nation or at least a very large region to experience contraction at the same time.

References

[1] Andrews, D.F. and Mallows, C.L. "Scale Mixtures of Normal Distributions." *Journal of the Royal Statistical Society: Series B,* 1974, 36(1), pp. 99-102.

[2] Bernardo, José M. and Smith, Adrian F. M. *Bayesian Theory.* Chichester: Wiley, 1994.

[3] Burns, Arthur F. and Mitchell, Wesley C. *Measuring Business Cycles.* New York: National Bureau of Economic Research, 1946.

[4] Carlino, Gerald and DeFina, Robert. "The Differential Regional Effects of Monetary Policy." *Review of Economics and Statistics*, November 1998, 80(4), pp. 572-587.

[5] Carlino, Gerald A. and DeFina, Robert H. "How Strong Is Co-movement in Employment over the Business Cycle? Evidence from State/Sector Data." *Journal of Urban Economics*, March 2004, 55(2), pp. 298-315.

[6] Carlino, Gerald and Sill, Keith. "Regional Income Fluctuations: Common Trends and Common Cycles." *Review of Economics and Statistics*, August 2001, 83(3), pp. 446-456.

[7] Carter, C.K. and Kohn, R. "On Gibbs Sampling for State Space Models." *Biometrika,* August 1994, 81(3), pp. 541-553.

[8] Casella, George and George, Edward I. "Explaining the Gibbs Sampler." *American Statistician,* August 1992, 46(3), pp. 167-174.

[9] Chib, Siddhartha. "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models." *Journal of Econometrics,* November 1996, 75(1), pp. 79-97.

[10] Crone, Theodore M. "An Alternative Definition of Economic Regions in the United States Based on Similarities in State Business Cycles." *Review of Economics and Statistics*, November 2005, 87(4), pp. 617-626.

[11] Del Negro, Marco. "Asymmetric Shocks among U.S. States." *Journal of International Economics*, March 2002, 56(2), pp. 273-297.

[12] Devroye, Luc. *Non-Uniform Random Variate Generation.* New York: Springer-Verlag, 1986.

[13] Forni, Mario and Reichlin, Lucrezia. "Federal Policies and Local Economies: Europe and the US." *European Economic Review*, January 2001, 45(1), pp. 109-134.

[14] Frühwirth-Schnatter, Sylvia and Kaufmann, Sylvia. "Model-Based Clustering of Multiple Times Series." *Journal of Business and Economic Statistics*, January 2008, 26(1), pp. 78-89.

[15] Gelfand, Alan E.; Dey, Dipak K.; and Chang, Hong. "Model Determination Using Predictive Distributions with Implementation via Sampling-Based Methods." J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), *Bayesian Statistics 4.* Oxford: Oxford University Press, 1992.

[16] Gelfand, Alan E. and Smith, Adrian F.M. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association,* June 1990, 85(410), pp. 398-409.

[17] Geweke, John and Keane, Michael. "Smoothly Mixing Regressions" *Journal of Econometrics*, May 2007, 138(1), pp. 252-291

[18] Hamilton, James D. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, March 1989, 57(2), pp. 357-384.

[19] Hamilton, James D. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.

[20] Hamilton, James D. "What's Real About the Business Cycle?" *Federal Reserve Bank of St. Louis Review,* July/August 2005, 87(4), pp. 435-452.

[21] Holmes, Chris C. and Held, Leonhard. "Bayesian Auxiliary Variable Models for Binary and Multinomial Regression." *Bayesian Analysis,* 2006, 1(1), pp. 145-168.

[22] Kaufmann, Sylvia. "Dating and Forecasting Turning Points by Bayesian Clustering with Dynamic Structure: A Suggestion with an Application to Austrian Data." *Journal of Applied Econometrics*, March 2010, 25(2), pp. 309-344.

[23] Kim, Chang-Jin and Nelson, Charles R. *State-Space Models with Regime Switching*. Cambridge, MA: The MIT Press, 1999.

[24] Koenders, Kathryn and Rogerson, Richard. "Organizational Dynamics over the Business Cycle: A View on Jobless Recoveries." *Federal Reserve Bank of St. Louis Review,* July/August 2005, 87(4), pp. 555-579.

[25] Kouparitsas, Michael A. "Is the United States an Optimal Currency Area?" *Chicago Fed Letter*, October 1999, No. 146, pp. 1-3.

[26] Owyang, Michael T.; Piger, Jeremy; and Wall, Howard J. "Business Cycle Phases in U.S. States." *Review of Economics and Statistics*, November 2005, 87(4), pp. 604-616.

[27] Owyang, Michael T.; Piger, Jeremy M.; Wall, Howard J.; and Wheeler, Christopher H. "The Economic Performance of Cities: A Markov-Switching Approach." *Journal of Urban Economics*, November 2008, 64(3), pp. 538-550.

[28] Partridge, Mark D. and Rickman, Dan S. "Regional Cyclical Asymmetries in an Optimal Currency Area: An Analysis Using US State Data." *Oxford Economic Papers,* July 2005, 57(3), pp. 373-397.

[29] Picard, Richard R. and Cook, R. Dennis. "Cross-Validation of Regression Models." *Journal of the American Statistical Association*, September 1984, 79(387), 575-583.

[30] Shao, Jun. "Linear Model Selection by Cross-Validation." *Journal of the American Statistical Association*, June 1993, 88(422), pp. 486-494.

[31] Sims, Christopher A.; Waggoner, Daniel F.; and Zha, Tao. "Methods for Inference in Large Multiple-Equation Markov-Switching Models." *Journal of Econometrics*, October 2008, 146(2), pp. 255-274.

[32] Tanner, Martin A. and Wong, Wing Hung. "The Calculation of Posterior Distributions by Data Augmentation." *Journal of the American Statistical Association*, June 1987, 82(398), pp. 528-540.

[33] van Dijk, Bram; Franses, Philip Hans; Paap, Richard; and van Dijk, Dick. "Modeling Regional House Prices." Econometric Institute Report EI 2007-55, 2007.

Table 1: Priors for Estimation			
Parameter	Prior Distribution	Hyperparameters	
$[\mu_{0n}, \mu_{1n}]'$	$N\left(\mathbf{m}, \sigma^2\mathbf{M}\right)$	$\mathbf{m} = [1, -2]'$; $\mathbf{M} = \mathbf{I}_2$	$\forall n$
σ_n^{-2}	$\Gamma\left(\frac{\nu}{2}, \frac{\delta}{2}\right)$	$\nu = 0$; $\delta = 0$	$\forall n$
$\mathbf{P}$	$\mathbf{D}(\boldsymbol{\alpha})$	$\alpha_i = 0$	$\forall i$
$\boldsymbol{\beta}_k$	$N(\mathbf{b}, \mathbf{B})$	$\mathbf{b} = \mathbf{0}_p$; $\mathbf{B} = \frac{1}{2}\mathbf{I}_p$	$\forall k$

Table 2: Cross-Validation Results[1]							
	Number of Clusters						
	2	**3**	**4**	**5**	**6**	**7**	Markov[3]
Score[2]	2122.4	2056.3	2071.5	2160.2	2146.3	2167.6	2179.1
[1] Cross validation uses 10 subsample splits.							
[2] $Score = M^{-1}\sum_{m=1}^{M}\sum_{r=1}^{R}\sum_{t=t_r}^{t_{r+1}-1}\left\{\log\left\|\boldsymbol{\Omega}^{[r,m]}\right\| + \left(\mathbf{y}_t - \mathbf{m}_{z_t^{[r,m]}}\right)'\left(\boldsymbol{\Omega}^{[r,m]}\right)^{-1}\left(\mathbf{y}_t - \mathbf{m}_{z_t^{[r,m]}}\right)\right\}$							
[3] Markov is the model with independent Markov-switching states [e.g., Owyang, Piger, and Wall (2005)].							

Table 3: Estimated model coefficients (posterior medians and means)													
	Median			Mean				Median			Mean		
	μ_0	μ_1	σ^2	μ_0	μ_1	σ^2		μ_0	μ_1	σ^2	μ_0	μ_1	σ^2
Alabama	2.92	−3.96	4.82	2.92	−3.96	4.86	Nebraska	2.68	−2.77	3.91	2.68	−2.78	3.94
Arizona	5.88	−4.20	10.21	5.88	−4.20	10.28	Nevada	6.17	−4.27	11.49	6.16	−4.27	11.57
Arkansas	3.40	−3.58	6.59	3.40	−3.58	6.64	New Hampshire	3.82	−4.45	7.56	3.82	−4.45	7.61
California	3.54	−3.76	4.09	3.54	−3.76	4.12	New Jersey	2.49	−3.27	3.21	2.49	−3.27	3.23
Colorado	4.20	−2.88	6.84	4.20	−2.89	6.89	New Mexico	3.55	−2.26	5.89	3.55	−2.26	5.93
Connecticut	2.41	−3.78	4.92	2.41	−3.78	4.95	New York	1.56	−2.68	2.74	1.56	−2.68	2.75
Delaware	3.05	−3.32	10.56	3.05	−3.32	10.63	North Carolina	3.68	−4.23	4.66	3.68	−4.23	4.69
Florida	5.06	−3.71	7.18	5.05	−3.71	7.23	North Dakota	2.83	−2.08	6.92	2.83	−2.07	6.97
Georgia	4.12	−4.09	4.71	4.12	−4.09	4.74	Ohio	2.21	−4.93	6.31	2.21	−4.94	6.36
Idaho	3.82	−3.07	9.16	3.82	−3.08	9.20	Oklahoma	3.03	−3.06	6.54	3.01	−3.06	6.59
Illinois	1.93	−3.78	4.11	1.93	−3.79	4.14	Oregon	3.71	−4.92	6.03	3.70	−4.92	6.07
Indiana	2.70	−5.07	9.08	2.70	−5.08	9.15	Pennsylvania	1.74	−3.69	4.48	1.74	−3.69	4.51
Iowa	2.52	−3.55	4.96	2.52	−3.56	5.00	Rhode Island	2.20	−3.75	7.30	2.20	−3.74	7.35
Kansas	2.68	−3.33	5.93	2.68	−3.33	5.96	South Carolina	3.59	−4.37	5.71	3.59	−4.38	5.75
Kentucky	3.03	−4.13	7.59	3.03	−4.14	7.64	South Dakota	2.90	−2.70	5.72	2.92	−2.69	5.74
Louisiana	2.86	−2.99	10.09	2.85	−2.99	10.17	Tennessee	3.35	−4.15	5.32	3.36	−4.16	5.35
Maine	2.67	−3.48	5.78	2.67	−3.47	5.82	Texas	3.90	−3.11	4.45	3.90	−3.12	4.48
Maryland	3.24	−3.56	4.87	3.24	−3.56	4.90	Utah	4.17	−3.38	6.50	4.18	−3.37	6.54
Massachusetts	2.28	−3.72	4.26	2.28	−3.72	4.29	Vermont	3.28	−3.86	5.79	3.28	−3.86	5.83
Michigan	2.35	−5.03	14.24	2.36	−5.07	14.30	Virginia	3.74	−3.43	3.45	3.74	−3.43	3.48
Minnesota	3.19	−4.04	4.06	3.19	−4.04	4.09	Washington	3.55	−3.96	8.16	3.56	−3.96	8.21
Mississippi	3.17	−3.75	6.89	3.17	−3.75	6.93	West Virginia	1.53	−3.19	11.59	1.54	−3.19	11.67
Missouri	2.38	−3.75	3.72	2.38	−3.76	3.75	Wisconsin	2.83	−4.26	3.79	2.83	−4.27	3.82
Montana	2.92	−2.97	9.04	2.91	−2.98	9.10	Wyoming	3.06	−2.65	17.92	3.06	−2.64	18.03

Table 4: Estimated logistic coefficients and derivatives (posterior means)						
	Cluster 1		Cluster 2		Cluster 3	
	β_1	δ_1	β_2	δ_2	β_3	δ_3
Constant	**-0.62**	-	-0.24	-	-0.02	-
Oil Production	**1.22**	**0.84**	-0.16	-0.17	**-1.02**	**-0.76**
Manufacturing	-1.03	-0.11	0.35	0.05	**-0.98**	**-0.11**
Finance	**-0.61**	**-0.04**	**-0.64**	**-0.06**	**0.75**	**0.06**
Small Firms	-0.34	-0.02	-0.11	-0.01	-0.09	0.00
Notes: Bold indicates zero is outside the 68 percent coverage interval.						
Oil production is measured as the share of income.						
Manufacturing and Finance are measured as the industry share of employment.						
Small firms are measured as share of employment in firms with < 100 employees.						

Table 5: Estimated regime transition probabilities (posterior means)					
	from Cluster 1	from Cluster 2	from Cluster 3	from Recession	from Expansion
to Cluster 1	0.69	**0**	**0**	0.03	0.02
to Cluster 2	**0**	0.60	**0**	0.00	0.00
to Cluster 3	**0**	**0**	0.73	0.03	0.02
to Recession	0.01	0.40	0.24	0.76	0.06
to Expansion	0.29	0.00	0.03	0.18	0.89
NOTES: p_{ij} for $i = 1, ..., 3$ and $i \neq j$ were restricted a priori to be zero (indicated by boldface).					

Figure 1. Values of explanatory variables for logistic probabilities across states.

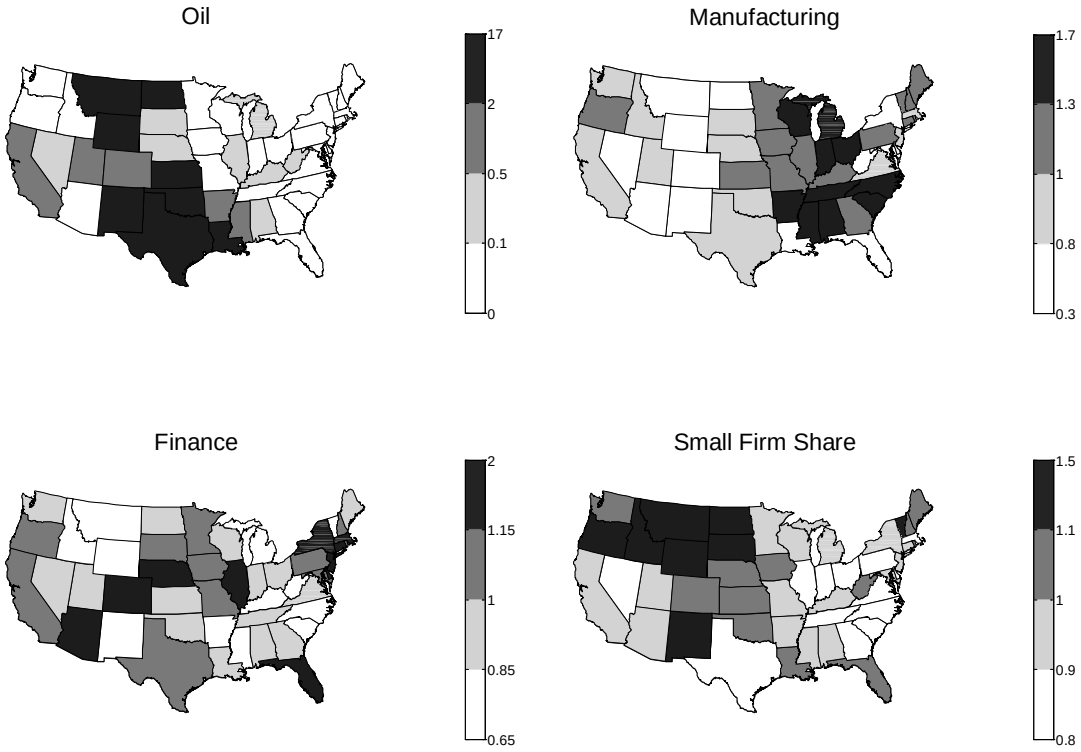
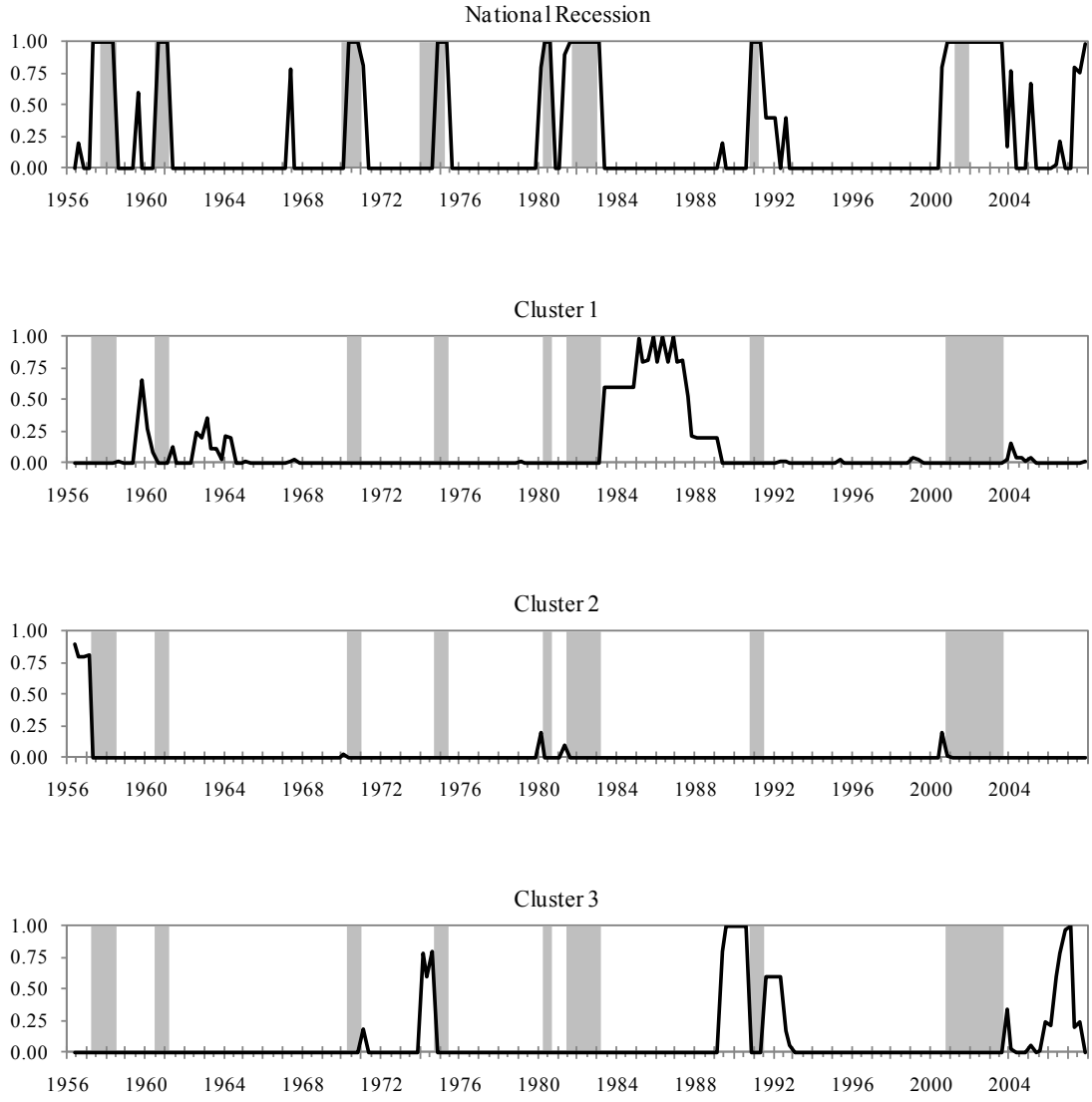
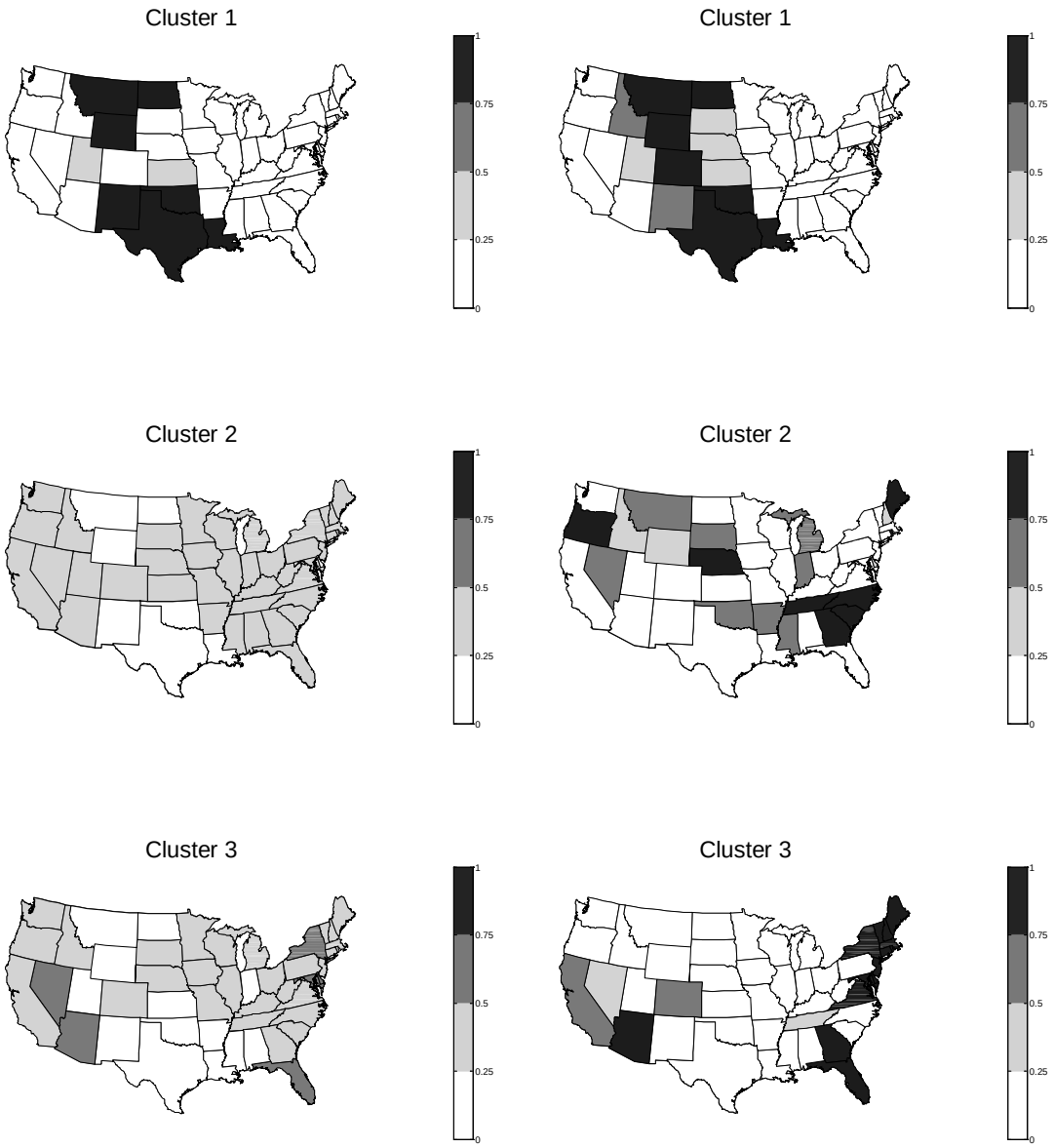


Figure 2. Posterior Probabilities of aggregate regimes.



Notes to Figure 2. Top panel: posterior probability that $z_t = 4$, with shaded regions corresponding to dates of NBER recessions. Bottom three panels: posterior probability that $z_t = 1, 2, 3$, with shaded regions corresponding to dates for which posterior probability that $z_t = 4$ is greater than 0.99.

Figure 3. Probabilities of cluster affiliations based on exogenous explanatory variables alone (first column) and based on exogenous explanatory variables plus observed employment growth patterns (second column).



Notes to Figure 3. *First column*: the color for state *n* for cluster *k* indicates the average value of $\exp(\mathbf{x}_n'\beta_k)/\left[1+\exp(\mathbf{x}_n'\beta_k)\right]$ across 125,000 simulated draws for β_k. *Second column*: the color for state *n* for cluster *k* indicates the average value of h_{nk} across 125,000 simulated draws for h_{nk}.