

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: October 10, 2008

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 beginning a national recession or recovery before others. [JEL: C11; C32; E32]

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

*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.

cycles in the U.S., for the most part, bear a reasonable resemblance to the national cycle identified by the 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 (forthcoming), the unit of analysis is taken to be individual states and cities, respectively. Regional similarities were noted but not modeled explicitly. A common 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 as either the basic unit of analysis (e.g., Kouparitsis, 1999), or as 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 U.S. Similarly, van Dijk, Franses, Paap, and van Dijk (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 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. Kouparitsis (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,

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

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 (forthcoming) 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 a quarter before the rest of the nation. Third, a cluster of states, characterized by a high oil and agriculture share of their economy, 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 is important for determining the regional clusterings, 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 $\mathbf{y}_t = (y_{t1}, \dots, y_{tN})'$, 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

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

where the n th element of the $(N \times 1)$ vector $\boldsymbol{\mu}_0 + \boldsymbol{\mu}_1$ is the average employment growth in state n during recession, the n th 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}_τ 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:

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

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 $(\eta \times \eta)$ 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.

Our approach as in Frühwirth-Schnatter and Kaufmann (2008) is to assume that recession dynamics can be characterized in terms of a small number $K \ll 2^N$ of different clusters and by an aggregate indicator $z_t \in \{1, 2, \dots, K\}$ signifying which cluster is in recession at date t . We associate with cluster 1 an $(N \times 1)$ vector $\mathbf{h}_1 = (h_{11}, \dots, h_{N1})'$ whose n th 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 $\mathbf{h}_1, \dots, \mathbf{h}_K$, this is a standard Markov-switching framework for which inference methods are well known. The new question is how to infer the configurations of $\mathbf{h}_1, \dots, \mathbf{h}_K$ 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 idiosyncratic clusters. Thus, when $z_t = 1, 2, \dots, \kappa$, some states are in recession and others are not. The values of $\mathbf{h}_1, \dots, \mathbf{h}_\kappa$ 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

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

for $n = 1, \dots, N$; $k = 1, \dots, \kappa$. Note that state n could be affiliated with more than one idiosyncratic cluster.² Alternatively, state n would participate only in national recessions if $h_{n1} = \dots = h_{n\kappa} = 0$.

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

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

Andrews and Mallows show that if $\psi_{nk} \sim KS$ and $e_{nk} \sim N(0, 1)$, 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

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

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

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

In other words, if we thought of nature as having generated ξ_{nk} from a $N(\mathbf{x}'_{nk}\boldsymbol{\beta}_k, \lambda_{nk})$ 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).

²This approach stands in contrast with the typical notion of a "region". Government agency (Bureau of Economic Analysis, 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.

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

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

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 $\mathbf{z} = (z_1, \dots, z_T)'$ summarizes which clusters are in recession at each date, while $h = \{\mathbf{h}_1, \dots, \mathbf{h}_\kappa\}$ summarizes the cluster affiliation of each state where $\mathbf{h}_k = (h_{1k}, \dots, h_{Nk})'$ denotes the $(N \times 1)$ vector characterizing which states participate in cluster k . There are also 3 other sets of variables and parameters associated with that realization of h . Let $\boldsymbol{\xi}_k = (\xi_{1k}, \dots, \xi_{Nk})'$ and $\boldsymbol{\lambda}_k = (\lambda_{1k}, \dots, \lambda_{Nk})'$ denote the associated auxiliary variables (see Tanner and Wong 1987) that are viewed as having determined \mathbf{h}_k according to:

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

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

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

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

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

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

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

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

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

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

With independent priors across states, we then have

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

We model transition probabilities using a Dirichlet prior. Recall that for $\mathbf{w} = (w_1, \dots, w_m)'$ 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 $\{w_1, \dots, w_{m-1}\}$ is given by

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

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

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

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

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

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

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

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

Numerical values for the prior parameters are summarized in Table 1.

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

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

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 ψ and λ to obtain

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

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

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 $\mathbf{Y}_n (y_{1n}, \dots, y_{Tn})'$ and let $\boldsymbol{\theta}_n = (\mu_{n0}, \mu_{n1}, \sigma_n^{-2})'$. Then,

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

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

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

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

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 initial regime is set to expansion a priori:

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

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); Cassella 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

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

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

(13), so that

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

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

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

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

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

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

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

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

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

for

$$\begin{aligned} \mathbf{M}_n^* &= (\mathbf{M}^{-1} + \mathbf{C}_n)^{-1} \\ \mathbf{m}_n^* &= \mathbf{M}_n^* (\mathbf{M}^{-1} \mathbf{m} + \mathbf{c}_n) \\ \mathbf{C}_n &= \left[\sum_{t=1}^T \mathbf{w}(z_t, h) \mathbf{w}(z_t, h)' \right] \\ \mathbf{c}_n &= \left[\sum_{t=1}^T \mathbf{w}(z_t, h) y_{tn} \right]. \end{aligned}$$

3.5 Drawing \mathbf{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),

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

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

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

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

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

Here

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

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

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

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

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

One can calculate $p(z_t|\mathcal{Y}_t, \theta, \mathbf{P}, h)$ by iterating on equation [22.4.5] in Hamilton (1994)³, the terminal value of which ($t = T$) gives us $p(z_T|\mathcal{Y}_T, \theta, \mathbf{P}, h)$, the first term in (16). Furthermore,

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

³Here $\boldsymbol{\xi}_t$ is a $(K \times 1)$ vector whose k th element is unity when $z_t = k$ and zero otherwise, while $\boldsymbol{\eta}_t$ is a $(K \times 1)$ vector whose k th element is $\prod_{n=1}^N p(y_{tn}|\theta, z_t = k, h)$, while $\hat{\boldsymbol{\xi}}_{0|0} = (0, 0, \dots, 1)'$.

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

3.7 Generating H .

We now define $H_k = \{\mathbf{h}_k, \boldsymbol{\xi}_k, \boldsymbol{\lambda}_k\}$ and $H^{[k]} = \{\mathbf{h}_j, \boldsymbol{\xi}_j, \boldsymbol{\lambda}_j : j = 1, \dots, \kappa; j \neq k\}$. 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),

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

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

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

where

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

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

Here, we have

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

Note that if we'd 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) argue that generating ξ_{nk} from the unconditional distribution and then λ_{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 $\xi = A - \log(u^{-1} - 1)$ has a logistic distribution with mean $E(\xi) = A$.⁴ 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 define

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

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

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

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

⁴This claim may be verified directly as follows:

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

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

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,

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

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

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

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

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

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

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

which is just a standard Normal regression model for each β_k of the form

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

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{x}'_{1k} \\ \vdots \\ \mathbf{x}'_{Nk} \end{bmatrix}$$

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

$$\mathbf{W}_k = \text{diag} \lambda_{1k}, \dots, \lambda_{Nk}.$$

Thus

$$\beta_k | \mathbf{Y}, \theta, \mathbf{P}, \mathbf{z}, Q \sim N(\mathbf{b}_k^*, \mathbf{B}_k^*)$$

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

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

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 the first two columns of \mathbf{H} along with 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 $\boldsymbol{\mu}_n$ not satisfying 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 found that once we imposed this condition, the clusters are sharply differentiated by the data, so that for a typical run with 20,000 burned draws and the next 20,000 retained, all of the retained draws tended to be consistent with a particular distinct set of characterizations of the different clusters.

4 Empirical results.

The data used to measure state-level business cycles are the seasonally adjusted, annualized growth rates of quarterly payroll employment.⁵ 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). Even

⁵The 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.

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 ensure that our algorithm identifies business cycles rather than outliers, we filter the employment growth data by smoothing periods with growth rates more than 4 standard deviations away from the mean.

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 $\kappa = 4$ idiosyncratic clusters and $P_k = 6$ 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 $k = 1, \dots, \kappa$). The vector \mathbf{x}_n includes barrels of oil produced per 100 dollars of state GDP, agricultural employment as a share of total employment, manufacturing employment share, financial activities employment share, a measure of workers' compensation by state; and the share of total state employment accounted for by small firms.⁶ Values for these explanatory variables are displayed in Figure 1.

We report results for some of the parameters and unobserved latent variables of interest based on a run of 20,000 Gibbs sampler iterations having discarded an initial burn-in of 20,000 iterations. Table 2 shows the posterior medians and means for the model parameters μ_0 , μ_1 , and σ^2 for each state. Table 3 gives the posterior means of the logistic coefficients β_k associated with each of the idiosyncratic clusters ($k = 1, \dots, 4$), with a bold entry signifying that 90 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 δ_k). The i th element of δ_k has the following interpretation. Let $\bar{x}_i = N^{-1} \sum_{n=1}^N x_{in}$ denote the average value for the i th explanatory variable. Suppose we compare two states, each of which has $x_{jn} = \bar{x}_j$ for all $i \neq j$, but in the first state, characteristic i is one standard deviation below the average \bar{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 ,

⁶The 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_mbbbl_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 industry (NAICS) shares of total payroll employment from 1990-2006, also from the BLS. For agriculture's share of employment, we used the percentage of employment in 2002 that was farm jobs or farm-related jobs, which we obtained from the USDA's Economic Research Service. Workers compensation was computed as the average of the index of workers' compensation insurance costs from Table 1 in Krueger and Burton (1990). We took the average of the 1972, 1975, 1978, and 1983 data for our final series. 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 dataset.

as calculated from (2), differ between the two states? The value for this magnitude implied by the posterior mean for β_k is reported as the i th element of the vector δ_k in Table 3. Taking cluster 1 as an example, a state that was average in all respects but one standard deviation below average in the share of agricultural employment 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. A state in which financial services comprise a smaller share of employment are also more likely to be one of those that is in recession when $z_t = 1$. The second aggregate regime affects oil-producing and agricultural states, but is negatively related to the share of manufacturing in total employment. For cluster 3, states in which a higher fraction of total employment is due to large firms are more likely to be included in the group that is in recession when $z_t = 3$. Cluster 4 tends to include states in which agriculture is less important. Although one might have thought that state regulations as proxied by workers' compensation might be related to the duration of aggregate unemployment spells, we find no connection between this measure and any cluster affiliations, as reflected in small and insignificant values for β_{5k} .

Table 4 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 $p_{12} = p_{13} = p_{14} = 0$). Although we did not impose $p_{16} = 0$, the posterior mean of p_{16} in fact turns out to be quite close to zero, meaning that if the states in cluster 1 go into recession, eventually the entire nation will follow, usually with a lag of about 2 quarters. Likewise if the states in cluster 4 are in recession, we'll also see a national recession eventually arrive, usually within 3 quarters ($p_{46} = 0$, $p_{44} = 0.63$). By contrast, $z_t = 2$ corresponds to a purely idiosyncratic recession in which the subset of states in cluster 2 experience a recession which on average lasts about $1/(1 - p_{22}) = 4.3$ quarters, and transitions to a regime of all states being in expansion. The regime $z_t = 3$ would be characterized as a late recovery for the subset of states in cluster 3; we could only observe $z_t = 3$ if the previous period had seen a national recession ($z_{t-1} = 5$). If there is a national recession, these states invariably require one more quarter to recover from recession compared with the rest of the nation.

Figure 2 plots the posterior means for the regime probabilities given the data. The top panel

is calculated as the fraction out of the 20,000 simulations for which z_t for the indicated quarter is equal to 5, 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 downturn in 1956 based on state employment data that is not characterized by 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.

The shaded regions in the bottom 4 panels of Figure 1 are based on the $z_t = 5$ dates rather than the NBER dates, to clarify the nature of the estimated dynamics. The national recessions of 1980, 1981, and 2001 all began with recessions in the cluster 1 states. A new recession also seems to have begun in these states in 2007, which according to the estimated model parameters would imply a national recession should soon follow. By contrast, the 1974 and 1990 recessions began in the cluster 4 states. Every recession is characterized by a slow recovery by the cluster 4 states. The cluster 2 states experienced a uniquely idiosyncratic recession during the oil price collapse in the middle 1980s, as well as several briefer episodes in the 1950s and 1960s.

Figure 3 indicates which states are affected by the respective idiosyncratic regimes based on the posterior probabilities for each h_{nk} given the observed employment data \mathbf{Y} . Regime 1 is in fact close to being a national recession. Thirty-one states would be in recession when $z_t = 1$, and the rest historically have always followed if that happens. The main states left out of this group are the northeast, the most populous states, and some of the key oil-producing states. Cluster 2 is clearly confined to the oil-producing states and their agricultural neighbors in the central-west part of the United States. The late-recovery states of cluster 3 seem to be limited to Illinois and a few of the Great Plains states. Cluster 4, the states in which the 1974 and 1990 downturns seemed to begin, are concentrated in the northeast and southeast U.S.

5 Conclusion

Certainly there are important differences in cyclical behavior across states. But one of the striking features of our findings is the extent to which U.S. recessions appear typically to have been national events in which at some point every state participated. Historically, different recessions likely began

in different regions with different characteristics, suggesting that there is some heterogeneity in the initial shocks. But there also appears to be some common national propagating dynamic when a downturn becomes sufficiently widespread.

References

- [1] Andrews, D.F., and C.L. Mallows. "Scale Mixtures of Normal Distributions." *J. R. Statist. Soc. B*, 1974, 36, pp. 99-102.
- [2] Carlino, Gerald and DeFina, Robert. "The Differential Regional Effects of Monetary Policy." *Review of Economics and Statistics*, November 1998, 80(4), pp. 572-587.
- [3] 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.
- [4] 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.
- [5] Carter, C., and Kohn, R., 1994, "On Gibbs Sampling for State Space Models," *Biometrika* 81, 541-553.
- [6] Casella, G., and George, E., 1992, "Explaining the Gibbs Sampler," *The American Statistician* 46, 167-174.
- [7] Chib, Siddhartha. "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models." *Journal of Econometrics* 75(1), pp. 79-97.
- [8] 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.
- [9] Del Negro, Marco. "Asymmetric Shocks among U.S. States." *Journal of International Economics*, March 2002, 56(2), pp. 273-297.
- [10] Del Negro, Marco and Otrok, Christopher. "99 Luftballons: Monetary Policy and the House Price Boom across U.S. States." *Journal of Monetary Economics*, October 2007, 54(7), pp. 1962-1985.
- [11] Devroye, Luc. *Non-Uniform Random Variate Generation*. New York: Springer-Verlag, 1986.
- [12] Forni, Mario and Reichlin, Lucrezia. "Federal Policies and Local Economies: Europe and the US." *European Economic Review*, January 2001, 45(1), pp. 109-134.
- [13] Fratantoni, Michael and Schuh, Scott. "Monetary Policy, Housing, and Heterogeneous Regional Markets." *Journal of Money, Credit and Banking*, August 2003, 35(4), pp. 557-589.
- [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, A. and Smith, A., 1990, "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85,398-409.
- [16] 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.

- [17] Hamilton, James D. "What's Real About the Business Cycle?" *Federal Reserve Bank of St. Louis Review*, July/August 2005 (87, no. 4), pp. 435-452.
- [18] Holmes, Chris C., and Leonard Held. "Bayesian Auxiliary Variable Models for Binary and Multinomial Regression." *Bayesian Analysis*, 2006, 1(1), pp. 145-168.
- [19] Kim, Chang-Jin and Nelson, Charles R. *State-Space Models with Regime Switching*. Cambridge, MA: The MIT Press, 1999.
- [20] Kouparitsas, Michael A. "Is the United States an Optimal Currency Area?" *Chicago Fed Letter*, October 1999, 146, pp. 1-3.
- [21] Krueger, Alan B., and John F. Burton, Jr. "The Employers' Costs of Workers' Compensation Insurance: Magnitudes, Determinants, and Public Policy," *Review of Economics and Statistics*, May 1990, 72(2), pp. 228-240.
- [22] 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.
- [23] Owyang, Michael T.; Piger, Jeremy; Wall, Howard J.; and Wheeler, Christopher H. "The Economic Performance of Cities: A Markov-Switching Approach." *Journal of Urban Economics*, forthcoming.
- [24] 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.
- [25] 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.
- [26] van Dijk, Bram; Franses, Philip Hans; Paap, Richard; and van Dijk, Dick. "Modeling Regional House Prices." mimeo.

Table 1: Priors for Estimation			
Parameter	Prior Distribution	Hyperparameters	
$[\mu_{0n}, \mu_{1n}]'$	$N(\mathbf{m}, \sigma^2 \mathbf{M})$	$\mathbf{m} = [2, -1]'$; $\mathbf{M} = \mathbf{I}_2$	$\forall n$
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: 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	2.85	-3.93	5.00	2.85	-3.93	5.04	2.85	-2.68	3.81	2.85	-2.68	3.83
Arizona	5.56	-3.96	11.29	5.56	-3.95	11.37	5.98	-4.62	11.39	5.98	-4.62	11.46
Arkansas	3.34	-3.73	6.55	3.34	-3.73	6.60	3.54	-4.94	7.43	3.53	-4.94	7.48
California	3.24	-4.06	4.40	3.24	-4.06	4.43	2.28	-3.54	3.28	2.28	-3.54	3.30
Colorado	4.12	-3.53	6.15	4.12	-3.53	6.19	3.59	-2.33	5.78	3.59	-2.33	5.81
Connecticut	2.31	-4.03	4.76	2.31	-4.03	4.79	1.40	-3.12	2.60	1.40	-3.12	2.62
Delaware	2.75	-3.52	10.85	2.76	-3.52	10.93	3.73	-4.35	4.39	3.73	-4.35	4.43
Florida	4.85	-3.85	7.39	4.85	-3.85	7.44	2.66	-1.60	7.43	2.66	-1.60	7.47
Georgia	3.97	-4.45	4.49	3.96	-4.44	4.53	2.12	-5.22	6.19	2.12	-5.22	6.23
Idaho	3.95	-3.46	8.44	3.95	-3.46	8.49	3.07	-3.51	5.89	3.07	-3.51	5.92
Illinois	1.95	-4.02	3.83	1.95	-4.02	3.85	3.59	-5.06	6.10	3.59	-5.06	6.14
Indiana	2.63	-5.24	9.01	2.63	-5.24	9.07	1.58	-4.00	4.52	1.58	-4.00	4.56
Iowa	2.51	-3.54	5.00	2.52	-3.54	5.03	2.03	-4.40	6.70	2.03	-4.40	6.75
Kansas	2.90	-3.18	5.79	2.90	-3.18	5.82	3.53	-4.62	5.53	3.53	-4.61	5.57
Kentucky	2.97	-4.18	7.66	2.97	-4.18	7.70	3.04	-2.81	5.44	3.04	-2.80	5.47
Louisiana	2.74	-3.04	10.22	2.74	-3.04	10.27	3.31	-4.19	5.35	3.31	-4.19	5.39
Maine	2.44	-3.93	5.63	2.44	-3.93	5.67	3.88	-3.68	3.83	3.88	-3.68	3.86
Maryland	2.98	-3.65	5.20	2.98	-3.65	5.23	4.35	-3.53	6.03	4.35	-3.52	6.07
Massachusetts	2.13	-4.47	3.64	2.13	-4.47	3.67	3.04	-4.24	5.75	3.04	-4.24	5.78
Michigan	2.34	-5.66	13.03	2.34	-5.66	13.13	3.52	-3.67	3.56	3.52	-3.67	3.58
Minnesota	3.17	-3.93	4.21	3.17	-3.93	4.24	3.58	-3.82	8.31	3.58	-3.82	8.36
Mississippi	3.15	-3.63	7.07	3.16	-3.64	7.12	1.57	-3.29	11.43	1.57	-3.30	11.50
Missouri	2.34	-3.82	3.73	2.35	-3.82	3.75	2.79	-4.33	3.82	2.78	-4.33	3.85
Montana	2.87	-3.03	9.01	2.87	-3.03	9.06	3.04	-2.83	17.68	3.04	-2.82	17.80

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	β_1	δ_1	β_2	δ_2	β_3	δ_3	β_4	δ_4
constant	0.04	-	-0.17	-	-0.10	-	-0.03	-
oil production	-3.3	-0.39	25.5	1.00	2.7	0.15	-0.6	-0.07
manufacturing	-0.13	-0.24	-0.70	-0.92	-0.08	-0.07	0.12	0.21
agriculture	0.66	0.65	0.54	0.61	0.23	0.11	-0.43	-0.42
finance	-0.47	-0.25	-0.67	-0.40	0.31	0.07	0.06	0.03
workers comp	-0.08	-0.01	-0.52	-0.09	-0.17	-0.01	-0.01	0.00
small firms	-0.11	-0.25	0.06	0.17	-0.15	-0.16	0.09	0.18

	from 1	from 2	from 3	from 4	from 5	from 6
to 1	0.56	0	0	0	0.00	0.03
to 2	0	0.77	0	0	0.00	0.03
to 3	0	0	0.00	0	0.24	0.00
to 4	0	0	0	0.63	0.00	0.02
to 5	0.44	0.00	0.00	0.37	0.76	0.03
to 6	0.00	0.23	1.00	0.00	0.00	0.90

NOTES: p_{ij} for $i = 1, \dots, 4$ 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.

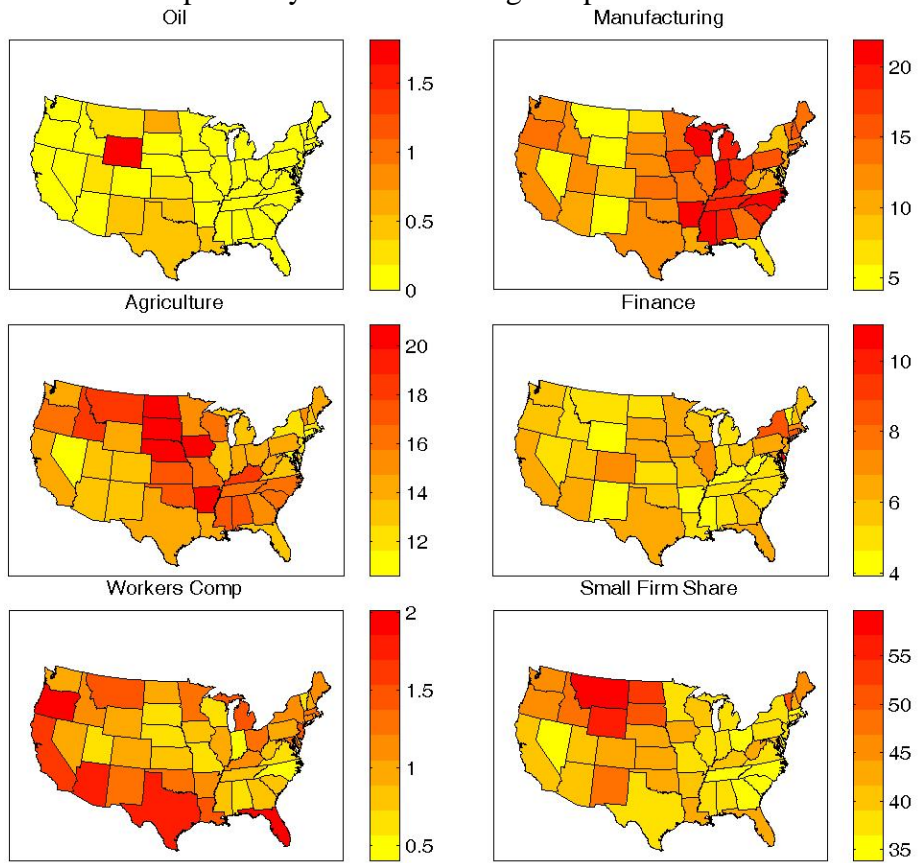
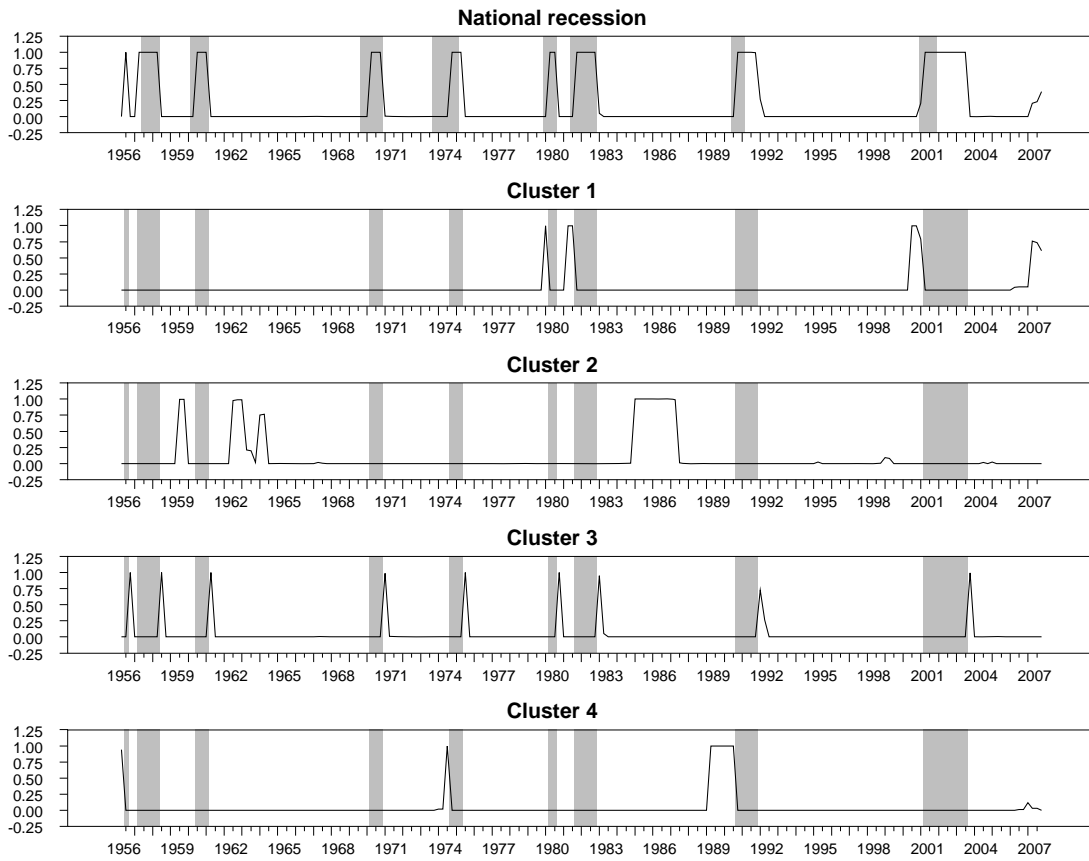
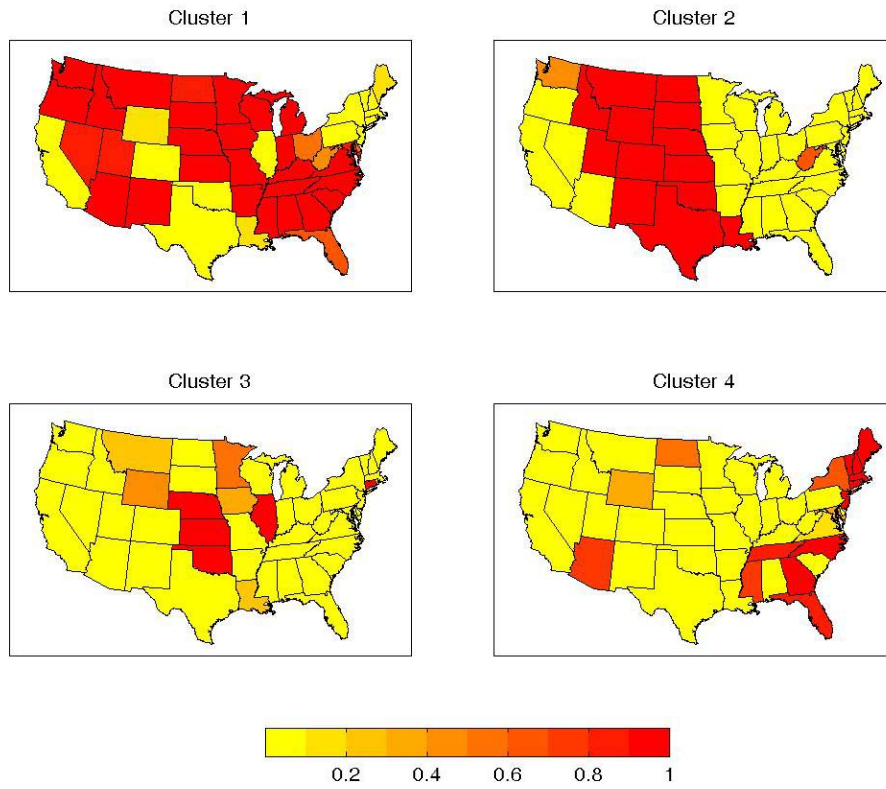


Figure 2. Posterior probabilities of aggregate regimes.



Notes to Figure 2. Top panel: posterior probability that $z_t = 5$, with shaded regions corresponding to dates of NBER recessions. Bottom panels: posterior probability that $z_t = 1, \dots, 4$, with shaded regions corresponding to dates for which posterior probability that $z_t = 5$ is greater than 0.99.

Figure 3. Posterior probabilities of cluster affiliations.



Notes to Figure 3. The color for state n in panel k indicates the fraction of 20,000 simulations for which the value of $h_{nk} = 1$.