Shibaev, Sergei S.

Working Paper

Recession propagation in small regional economies: Spatial spillovers and endogenous clustering

Queen's Economics Department Working Paper, No. 1369

Provided in Cooperation with:
Queen’s University, Department of Economics (QED)

Suggested Citation: Shibaev, Sergei S. (2016) : Recession propagation in small regional economies: Spatial spillovers and endogenous clustering, Queen's Economics Department Working Paper, No. 1369, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:
https://hdl.handle.net/10419/149095

Queen's Economics Department Working Paper No. 1369

Recession Propagation in Small Regional Economies: Spatial Spillovers and Endogenous Clustering

Sergei S. Shibaev
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

11-2016

Recession Propagation in Small Regional Economies: Spatial Spillovers and Endogenous Clustering*

Sergei S. Shibaev†
Queen's University

November 10, 2016

Abstract

This paper develops a statistical model for measuring spatial interactions when estimating macroeconomic regimes and regime shifts. The model is applied to study the contagion and propagation of recessions in small regional economies in the United States from 1990 to 2015. The empirical analysis identifies two geographical concentrations (or clusters) where small regional economies were affected by recessions in similar ways. These clusters are interpreted as groups of regions that are potentially at-risk to collective economic distress, which is useful for national and regional policy makers. The first identified cluster is characterized by regional economies with important roles in the financial sector, while the second cluster is characterized by the oil and gas extraction sector. The empirical findings uncover an important propagation dynamic that would be overlooked if one were to apply the model without the spatial extension developed in this paper. Specifically, the evidence shows significant spatial spillovers between small regional economies, meaning that shocks to regions are expected to be higher, when shocks to neighboring regions are high on average. The magnitude of this effect is amplified for the period spanning and following the Great Recession.

JEL codes: C11, C31, C34, E32, R10

Keywords: Bayesian statistics, business cycles, endogenous clustering, regime-switching, regional economic analysis, spatial econometrics, time series econometrics

*I am grateful to my supervisors Taylor Jaworski, James G. MacKinnon and Morten Ø. Nielsen for their guidance and support. This research is supported by the John Deutsch Institute (JDI) for the Study of Economic Policy. I thank Martin Burda (discussant), Robert Clark, Christopher Cotton, Alfonso Flores-Lagunes, Beverly Lapham, Allen Head, Hashem Pesaran, participants at the Doctoral Workshop in Applied Econometrics (DWAE) 2016 and seminar participants at Queen's University for helpful comments and suggestions. I express appreciation to James D. Hamilton and Michael T. Owyang for making their computer programs freely available to the public. All errors are my own.

†E-mail: shibaevs@econ.queensu.ca

1 Introduction

This paper delivers two contributions. The first contribution is to econometric methods. A new statistical model is developed that is capable of estimating spatial interactions in a class of regime-switching models. The second contribution is to empirical macroeconomics. The developed model is used to study the contagion and propagation of recessions in small regional economies in the United States, namely the 177 contiguous economic areas classified by the Bureau of Economic Analysis (BEA)[1]. Investigating small geographical units is motivated by the view that larger regions such as states, may not provide sufficient geographical detail to identify regional contractions in the economy[2]. The proposed model has a broad range of potential applications, including, but not limited to, studying bankruptcies, credit, trade, housing, financial markets, urbanization and political topics like electoral support.

To the best knowledge of the author, this is the first paper to analyze regional business cycles in small regional economies in the United States using a multivariate regime-switching model, and the first to allow for explicit modeling of spatial interactions when dealing with common Markov-switching components.

The empirical results characterize several geographical concentrations of regions that have been impacted by recessions in similar ways, in addition to economic downturns that were spread nationwide. The composition of these concentrations is studied to understand the influence of region-specific characteristics and observed employment growth patterns

[1]BEA economic areas are generally smaller in size than individual States and they can cross state borders. They are defined by mutually exclusive groups of counties that constitute relevant regional markets surrounding metropolitan or micropolitan statistical areas in the United States. United States metropolitan and micropolitan statistical areas are defined by the United States Office of Management and Budget (OMB). A metropolitan statistical area is defined as one or more adjacent counties or county equivalents that have at least one urban core area of at least 50,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties. A micropolitan statistical area is an urban area in the United States centered on an urban cluster (urban area) with a population at least 10,000 but less than 50,000.

[2]This is empirically supported by simply comparing observed employment growth patterns between states and smaller regions.

that determine their geography. Furthermore, the findings strongly support the need to explicitly account for spatial correlation between geographical units and the evidence shows significant degrees of spatial spillovers between regions in the United States for the period 1990–2015.

The evidence presented in this paper provides a unified analysis of two types of economic phenomena that have received substantial attention in the empirical macroeconomics literature on regional business cycles. The first phenomenon is that all regions in the aggregate economy are connected, which motivates the belief that there exist spatial spillovers (or interactions) between regions in the economy - see Artis et al. (2011), Fogli et al. (2015) and Beraja et al. (2016). Notably, the work of Beraja et al. (2016) provides evidence that particular regional patterns (e.g. regional business cycle fluctuations) contain important underlying connections with the aggregate business cycle. The second phenomenon has to do with the synchronicities (*i.e.* co-movements) in regional business cycle behavior, which can be captured through large geographical concentrations (or clusters) of regions that exhibit co-movements over the business cycle - see Crone (2005), Partridge and Rickman (2005), van Dijk et al. (2007) and Hamilton and Owyang (2012).

The methodological contribution of this paper is to explicitly model spatial interactions between regional shocks within a specific category of statistical models, namely regime-switching (or Markov-switching) models. The proposed model is an extension of the recently developed regime-switching model of Hamilton and Owyang (2012). The extended version of this model retains the many attractive properties of the original model. Specifically, the model is capable of endogenously identifying groupings (or clusters) of regions that exhibit similar business cycle characteristics. Regime-switching models were first shown to be a flexible methodology for empirical business cycle analysis in the work of Hamilton (1994, 1989). Henceforth, the proposed model will be referred to as the spatial model and the original model will be referred to as the restricted model, given that the original model will

be nested in the spatial specification.

Typically multivariate models focus on either large regional divisions such as the eight BEA regions – Kouparitsas (1999) and Crone (2005) – or the 50 US states and 48 lower US states; Forni and Reichlin (2001), Del Negro (2002), Partridge and Rickman (2005), Artis et al. (2011) and Hamilton and Owyang (2012). When it comes to regime-switching models of the type considered in Hamilton (1989, 1994), regional business cycle analysis has been constrained to univariate models, which have been used to analyze both state-level and large metropolitan statistical areas in the United States; see Owyang et al. (2005) and Owyang et al. (2008), respectively. The regime-switching model of Hamilton and Owyang (2012) has resolved the curse of dimensionality that arises when extending these types of models to panel data. The empirical application considered by Hamilton and Owyang has shown that common Markov-switching components provide valuable insights into regional co-movements in the lower 48 states using state-level employment data for the period 1956–2007. Analyzing small regional economies provides more geographical detail for capturing both the cyclical and the spatial interdependencies between regions.

Popular alternatives to regime-switching models for this type of analysis are factor models and to a lesser extent structural vector autoregressive models. Where regime-switching models capture the mechanism of dynamic change governing the transitions between business cycle phases, factor models measure co-movement of many time series (regions) and are designed to help identify business cycle turning points. Prominent examples of factor models used for regional analysis are Forni and Reichlin (2001) and Del Negro (2002). Forni and Reichlin (2001) look at the lower 48 states and counties, but assume that the local shock in a region cannot affect other regions. In contrast, the model proposed in this paper explicitly allows for this type of connection between regions. Del Negro (2002) also look at states, and uses geographic proximity to impose model restrictions. Del Negro argues that this acts as a proxy for regional productive structure and income levels in geographical units. This view

motivates the spatial weighting structure considered in this paper's empirical investigation. An example of a structural vector autoregressive approach is the work of Thorsrud (2013), which simultaneously analyzes global and regional business cycles.

Another advantage of the methodology applied in this paper is the way in which the model can endogenously group (or define) regions. Specifically, the mechanism that groups regions in the model does not restrict groups to be mutually exclusive. This is convenient for characterizing regional economic downturns at different points in time. One can easily imagine a scenario where a small regional economy may be part of a regional economic downturn concentrated in one part of a country in a particular decade, and also part of another regional contraction with a completely different geography in another decade. The model's mechanism that learns about regional synchronicities through time allows for such occurrences. This mechanism allows for a multi-sector analysis of the economy by relying on region-specific characteristics, for example, industrial composition, to supplement the characterization of similarities and differences between regions.

Typically, empirical regional business cycle studies use exogenously defined regions, e.g. Kouparitsas (1999) and Del Negro (2002), where either state borders or the eight BEA regional divisions are used to define geographical units. Other studies have focused on endogenously defining regions, e.g. Crone (2005), Partridge and Rickman (2005) and van Dijk et al. (2007). Crone (2005) uses pattern recognition methodology of k-means cluster analysis to the cyclical components of Stock-Watson-type coincident indices, estimated at the state level, to group the 48 contiguous states into eight regions with similar cycles and compare to the eight BEA regions. van Dijk et al. (2007) use a latent-class clustering to identify co-movements between regions in the Netherlands. They use likelihood-based information criteria to identify a total of two clusters in the Netherlands. Partridge and Rickman (2005) present an alternative way of endogenously defining regions by analyzing co-movements between US states with static bivariate correlations, but not their exact numerical

correlation values.

More recently, spatial interaction models have received attention for business cycle analysis. Artis et al. (2011) use a spatial ARMA model to examine business cycle convergence for 41 Euro area regions and 48 US states. They find that including spatial effects like spatial lag and spatial error components improves the fit of their model. Fogli et al. (2015) use a spatial autoregressive lag model to analyze county level unemployment and housing price data in the United States. They find that unemployment rates are spatially dispersed and spatially correlated (estimated correlation between 0.63–0.83 over the Great Recession). They refer to increasing spatial correlation as a clustering dynamic and focus on different channels through which these characteristics change during recessions. Their argument is centered on local geographic factors being very important for aggregate business cycle dynamics. These results will serve as benchmark comparisons for the estimated spatial correlations in the empirical analysis of this paper.

The remainder of the paper is structured as follows. Section 2 describes the model and its structure for characterizing regional business cycles, regional clusters and spatial dependence. Section 3 describes the statistical inference and estimation of the model, including the model selection procedure used for identifying the number of regional clusters. All aspects of the original model that are affected by the spatial dependence specification are discussed in this section. Section 4 discusses the data and spatial weighting structure considered in the empirical investigation, including alternative measures. The empirical results are presented and analyzed in Section 5 including a discussion of some policy implications using the results. Section 6 concludes. There are two appendices at the end of the paper. Appendix A contains supplementary figures and tables. Appendix B contains supplementary details on the statistical inference, estimation algorithms and derivations.

2 Model

This section discusses the model in three parts. Section 2.1 motivates the use of Markov-switching models to characterize regional business cycles and explains the model specification of Hamilton and Owyang (2012). Section 2.2 describes the endogenous clustering mechanism. Section 2.3 motivates and describes the extended model specification which introduces a spatial dependence component into the existing model.

2.1 Characterizing regional business cycles

Suppose a researcher is interested in estimating a simple statistical model to understand the cyclical behavior of a particular region's economy. If the defining characteristic of a business cycle is taken as the transition between distinct discrete phases of contraction and expansion, then an appropriate starting point is the widely known and studied model of Hamilton (1989, 1994), which is an autoregressive Markov-switching model that Hamilton first used to characterize national recessions in the United States. The simplest version of Hamilton's model for business cycles is a model with only a Markov-switching mean,

$$\begin{aligned} y_t &= \mu_0 + \mu_1 s_t + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{N}(0, \sigma^2), \\ s_t &= \begin{cases} 1, & \text{if recession,} \\ 0, & \text{if expansion.} \end{cases} \end{aligned} \tag{1}$$

Model (1) postulates that only the recession state variable, s_t, explains y_t, the variable that fluctuates over the business cycle. This model is useful when analyzing national recessions or regional recessions separately, and facilitates a regional analysis of one region at a time. As shown in Owyang et al. (2005) and Owyang et al. (2008), this albeit basic model offers a flexible framework for understanding the timing of transitions between phases of regional contraction and expansion in comparison with the aggregate economy.

A more compelling approach to understanding regional business cycles is a statistical

model that allows for both national and regional analysis, while accounting for multiple regions. Furthermore, a desired property for the model is that the framework has the ability to capture similarities between regions explicitly. A model for this purpose has recently been introduced by Hamilton and Owyang (2012), who develop a framework for inferring common Markov-switching components in a panel data model. The model they are initially interested in is

$$\boldsymbol{y_t} = \boldsymbol{\mu_0} + \boldsymbol{\mu_1} \odot \boldsymbol{s_t} + \boldsymbol{\varepsilon_t}, \quad \boldsymbol{\varepsilon_t} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{2}$$

where $\boldsymbol{y_t} = (y_{1t}, \ldots, y_{Nt})'$ is a vector of employment growth rates for N regions at date t, $\boldsymbol{s_t} = (s_{1t}, \ldots, s_{Nt})'$ is a vector of recession (contraction phase) indicators ($s_{nt} = 1$ when region n is in recession and $s_{nt} = 0$ when region n is in expansion), and $\odot$ is the Hadamard (element-by-element) product. The n^{th} element of the $N \times 1$ vector $\boldsymbol{\mu_0} + \boldsymbol{\mu_1}$ is the average employment growth in region n during recession and the n^{th} element of the $N \times 1$ vector $\boldsymbol{\mu_0}$ is the average employment growth in region n during expansion. The choice of the variable that fluctuates over the business cycle is discussed in detail in Section 4.

2.2 Endogenous clustering

For the empirical application in Hamilton and Owyang (2012) which looks at the lower 48 US states, the model defined in (2) is intractable, since it requires $\eta = 2^{48}$ regimes to be inferred from the $T \times N$ data set[3], so the standard approach of Hamilton (1994) is not feasible. To mitigate the curse of dimensionality, the model that Hamilton and Owyang develop is an augmented version of the model in (2), which follows Frühwirth-Schnatter and Kaufmann (2008) and assumes that a small number of clusters, $K << \eta$, capture business cycle dynamics. This approach greatly reduces the state space dimension (the number of regime-states to consider), and the mechanism of dynamic change governing the transitions between regimes is now a $K \times K$ dimensional transition probability matrix, $\mathbf{P}$. The model

[3]The binary recession indicator s_{nt} for region $n \in \{1, 2, \ldots, N\}$ implies $\eta = 2^N$ distinct regimes.

is written as

$$\boldsymbol{y_t}\Big|_{\{z_t=k\}} = \boldsymbol{\mu_0} + \boldsymbol{\mu_1} \odot \mathbf{h}_k + \boldsymbol{\varepsilon_t}, \quad \boldsymbol{\varepsilon_t} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}), \tag{3}$$

where z_t is an aggregate indicator, $z_t \in \{1, 2, \ldots, K\}$, indicating which cluster of regions is in recession at date t. Each cluster k, has an associated $N \times 1$ state vector $\mathbf{h}_k = (h_{1k} \ldots h_{Nk})'$, where the n^{th} element is unity when region n is associated with the cluster k, and is zero otherwise. Conditional on $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K$, the standard Markov-switching framework applies.

Following Hamilton and Owyang (2012)'s inference procedure for the configurations of the cluster affiliation vectors $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K$, two clusters are imposed a priori to capture nationwide recessions and expansions. The nationwide expansion cluster is $\mathbf{h}_K$ (a column of zeros) where every region is in expansion when the aggregate indicator $z_t = K$, and the nationwide recession cluster is $\mathbf{h}_{K-1}$ (a column of ones), where every region is in recession when the aggregate indicator is $z_t = K-1$. This formulation allows the model to account for instances where only a subset of the all regions experience economic downturns separately from the rest of the nation, determined by clusters $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{K-2}$ and economic downturns that spread nationwide, determined by clusters $\mathbf{h}_{K-1}, \mathbf{h}_K$.

Whether region n belongs to cluster k when $z_t = k$ is determined by observed employment growth patterns and a group of regional-level covariates $\boldsymbol{x}_n$, which serve to identify similarities between regions. The model first identifies the probability that region n belongs to cluster k based only on the fixed regional covariates, computed from

$$\Pr(h_{nk} = 1 | \boldsymbol{\beta}_k) = \frac{\exp(\mathbf{x}'_{nk}\boldsymbol{\beta}_k)}{1 + \exp(\mathbf{x}'_{nk}\boldsymbol{\beta}_k)}. \tag{4}$$

The model then updates this probability with observed regional employment growth patterns. If the probability in (4) is low (high) for region n, but the updated probability is high (low), then the observed employment growth patterns (regional characteristics) strongly designate

this region to cluster k. The covariates used in the empirical investigation are discussed in Section 4. As in Hamilton and Owyang (2012), it is assumed that the same covariates influence each cluster, and that any given region is not restricted to belong to only one cluster grouping. Supplementary details on the logistic clustering methodology are given in Section B.1 and Appendix B.1.

2.3 Characterizing spatial dependence

When dealing with geographical units, such as regional divisions of a country, cross section observations are unlikely to be independent of one another. This requires specific attention and is often overlooked in applied work since accounting for this type of dependency is not always straightforward. Up to this point, the described model has no explicit specification for allowing this type of behavior in the data. The only channels through which co-movements are captured by the model are through the observed employment growth rates and the information contained in the fixed regional-level covariates used in the clustering mechanism. Furthermore, to be operational, the parametric model specified in Hamilton and Owyang (2012) imposes a distributional assumption on the error vector where the variance-covariance matrix is diagonal, which under the normality assumption implies cross-sectional independence. Hamilton and Owyang (2012, p. 936) admit that unfortunately this is necessary because relaxing this assumption would greatly increase the number of parameters for which they need to draw inference and would invalidate their estimation algorithms.

The argument in this paper is that extending the original framework, to model spatial dependence explicitly, offers a more direct treatment of spatial interactions along the cross-section dimension. The proposed approach delivers a richer spatial analysis that is appropriate for analyzing regional business cycles in the economy. This is due to the extra information it is able to capture, namely, co-variation within geographical space. Extending the model in this form provides a parsimonious solution for alleviating the severity of the

required distributional assumption (diagonality of $\mathbf{\Omega}$) in the original model.

The empirical application in Section 5 will show that significant positive spatial correlation is prevalent, according to (i) testing for spatial correlation in the considered panel data set of regional employment growth rates, and (ii) the estimates obtained from the spatial dependence component in the developed model. The latter will be shown to be strongly robust to various specifications of the model.

To introduce an explicit characterization of spatial interactions, it is important to clearly define what is postulated on the nature of the spatial relationship and what are the desirable testable results. The proposed approach postulates that every region in the economy has ties to one or more other regions, stemming from specific factors or characteristics that define the connections. Given the structure of these connections, a region may be influenced by shocks that are contemporaneously related to the average magnitude of shocks to other regions in the economy. This formulation in the context of a regional business cycle analysis is conveniently nested in a more general class of spatial models common to the spatial econometrics literature. These types of models provide a flexible framework for assessing spatial relationships for specific spatial weighting structures. This paper argues that out of many potential candidate specifications, which will be subsequently discussed, the one with the most attractive properties for the regime-switching model being analyzed is the spatial autoregressive error (SAE) component. The SAE component is adapted into the model to facilitate the analysis of various spatial dependence structures beyond the regional similarities captured by the information entering the clustering mechanism. This specification allows the degree of spatial correlation along the cross-sectional dimension to be quantified and tested.

Specifically, the SAE component dictates that the unobserved errors (or shocks) in a region can be spatially correlated with the shocks of other regions. The proposed SAE version of the original regime-switching model is

$$
\begin{aligned}
\boldsymbol{y_t}\Big|_{\{z_t=k\}} &= \boldsymbol{\mu_0} + \boldsymbol{\mu_1} \odot \mathbf{h}_k + \boldsymbol{\varepsilon_t} \\
\boldsymbol{\varepsilon_t} &= \rho \mathbf{W} \boldsymbol{\varepsilon_t} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}).
\end{aligned} \tag{5}
$$

If $\rho = 0$, then there is no spatial dependence, and the model is that of Hamilton and Owyang (2012). A positive value (negative value) of ρ indicates that shocks are expected to be higher (lower), if on average, shocks to neighboring regions are high. $\mathbf{W}$ is a row-standardized matrix of spatial weights, and each row of $\mathbf{W}$, $\boldsymbol{w}_i$, dictates the spatial dependency of region i to all other regions. The spatial weights used in the empirical investigation are discussed in detail in Section 4.3.

An alternative SAE specification would be to generalize the spatial parameter to vary across the N regions, which would result in the following model

$$
\begin{aligned}
\boldsymbol{y_t}\Big|_{\{z_t=k\}} &= \boldsymbol{\mu_0} + \boldsymbol{\mu_1} \odot \mathbf{h}_k + \boldsymbol{\varepsilon_t}, \\
\boldsymbol{\varepsilon_t} &= \boldsymbol{\Psi} \mathbf{W} \boldsymbol{\varepsilon_t} + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}), \\
\boldsymbol{\Psi} &= \begin{bmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \rho_N \end{bmatrix}.
\end{aligned} \tag{6}
$$

The model in (6) quantifies region-specific degrees of spatial spillovers. This model is not chosen for the analysis because the desired objective is to quantify the overall degree of spatial spillovers in the macroeconomy. Furthermore, using any of the discussed models (including the restricted original model) to analyze 177 BEA economic areas, as opposed to the lower 48 states, substantially increases the computational burden of estimating the model. Therefore, despite there being no conceptual problem with implementing the model in (6), the more parsimonious SAE specification in (5) with a single common spatial parameter is more desirable for addressing the questions posed on the nature of spatial interactions.

A well known alternative to the SAE formulation is the spatial autoregressive lag (SAL) component, which in the context of this study specifies that the employment growth in a

region can be spatially dependent on the employment growth in other regions. The proposed SAL version of the model would be

$$\boldsymbol{y_t}\Big|_{\{z_t=k\}} = \rho \mathbf{W} \boldsymbol{y_t} + \boldsymbol{\mu_0} + \boldsymbol{\mu_1} \odot \mathbf{h}_k + \boldsymbol{\varepsilon_t}, \quad \boldsymbol{\varepsilon_t} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega}). \tag{7}$$

This paper advocates the SAE formulation over the SAL formulation due to the specific role played by the clustering mechanism in the model. The concern is that introducing the SAL component into the model would capture co-movements in the data that are net of the co-movements identified through the fixed groupings of regions into clusters. Therefore it would become difficult to interpret the magnitude and sign of the final estimates of the spatial parameter ρ. On the other hand, the SAE specification allows for spatial interactions between regional unobserved shocks captured by the error term, which is a less restrictive specification that provides a very interesting interpretation of the spatial parameter ρ as capturing the spatial regional interactions in the macroeconomy.

The proposed model (5) nests the non-spatial (restricted) model (3) as a special case $\rho = 0$. Therefore, the posterior density of the spatial parameter ρ provides a natural framework to test the hypothesis that the spatial parameter is significantly different from zero for a given spatial weighting structure embodied by $\mathbf{W}$. In the frequentist framework one would find it appropriate to motivate the need for a spatial dependence component by estimating the non-spatial model and testing for no cross-sectional dependence in the residuals using a test of the type considered in Pesaran (2004, 2015). The latent structure of the model would require generalized residuals to be computed following Gourieroux et al. (1987). Forming residuals relies on computing fitted values based on point estimates of the parameters. Bayesian posterior inference provides the entire surface of parameter densities, which provides all the relevant information for obtaining quantities of interest and making probability statements on the parameters. However, residual computation is not standard or straightforward. There are two approaches one could take, and it is not clear which is

more appropriate. The first is to form residuals for every sampled draw of the Markov chain Monte Carlo (MCMC) sequence retained for posterior inference. This would result in a chain of residual vectors corresponding to the posterior draws retained for inference. How to summarize this chain into one residual quantity to build the test statistic is not clear. Alternatively, one can proceed by using posterior means or medians as point estimates to compute residuals, which results in residuals that have a completely different interpretation from the first approach. To obtain a more clearly defined assessment of spatial dependence in the Bayesian framework, quantities of interest for testing hypotheses on the spatial parameter will be taken from the posterior. Posterior-based hypothesis testing will be discussed in Section 5.

In all three of the discussed spatial specifications, the dependence structure is based on a fixed metric embodied by the the spatial weighting matrix $\mathbf{W}$. This matrix may be specified by the analyst using physical (e.g. distance), economic (e.g. trade) or technological measures, or it can be estimated using data. This paper considers fixed specifications for $\mathbf{W}$ and favors weighting structures that give a relatively more sparse matrix, in order to reduce computational cost. A variety of connections between regions can be examined thoroughly through an explicitly defined spatial dependence structure in the model. Using an SAE specification for this purpose is a natural extension to the original model and is common to other models in the spatial econometrics and statistics literatures.

The SAE component has the advantage of being a substantially more parsimonious approach than relaxing the diagonal assumption for the variance-covariance matrix of the model's error vector. To illustrate this point it is sufficient to compare the two possible parameterizations. The data set analyzed in this paper has a time-series dimension of $T = 102$ and a cross-section dimension of $N = 177$. For the non-spatial model, excluding the auxiliary variables, this results in 1321 parameters. The SAE specification adds only a single parameter to be estimated, albeit one that plays a very important role through the fixed

spatial weighting metric embodied by the matrix $\mathbf{W}$. Conversely, relaxing the off-diagonal restrictions on the variance-covariance matrix $\mathbf{\Omega}$ would require inferences to be made on $N(N-1)/2$ additional parameters.

3 Statistical inference and estimation

The inference for the spatial dependence component of the model is integrated into the Bayesian posterior inference procedure developed in Hamilton and Owyang (2012). The notation and definitions are kept as close as possible to the original framework. This section describes the all aspects of the model that are affected by the introduction of the spatial dependence component.

Section 3.1 discusses the joint distribution of all random variables in the model. The likelihood function is discussed in Section 3.2. Section 3.3 formally states all prior distributions. The conditional distributions for the population parameters affected by the spatial parameter ρ are derived in Section 3.4. Section 3.5 describes the procedure used for model selection. Section 3.6 outlines the estimation algorithm for the model. Supplementary details regarding the inference and estimation procedures are discussed in Appendix B.

3.1 Joint distribution

The joint distribution for the data ($\mathbf{Y}$), population parameters ($\rho, \boldsymbol{\mu}, \mathbf{\Omega}, \beta$), variables for the dynamic change mechanism (regime transition probabilities, $\mathbf{P}$) and latent variables ($\boldsymbol{z}, h$)is given by (8), where, to be consistent with the compact notation in Hamilton and Owyang (2012), the cluster affiliation variables are grouped in $\mathrm{H} = \{h, \xi, \lambda\}$. The spatial weighting matrix $\mathbf{W}$ is suppressed as an argument in the notation because it is fixed.

$$p(\mathbf{Y}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta) = p(\mathbf{Y}|\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, h)p(\boldsymbol{z}|\mathbf{P})p(\rho)p(\boldsymbol{\mu}, \boldsymbol{\Omega})p(\mathbf{P})p(h|\beta)p(\beta) \tag{8}$$

The compact expression for the arguments of the joint distribution in (8) makes use of a logical grouping of the parameters. The exact composition and dimension of these arguments are now explained in the order that they appear in (8). $\mathbf{Y}$ is the $T \times N$ dimensional matrix of the regional employment growth rates. ρ is a scalar parameter measuring the degree of spatial dependence. $\boldsymbol{\mu}$ is a $N \times 2$ dimensional matrix of the average employment growth rates in each regime, where each row, $\boldsymbol{\mu}_n$, is defined as $\boldsymbol{\mu}_n = [\mu_{n0} \;\; \mu_{n1}]$. The $N \times N$ dimensional error variance-covariance matrix, $\boldsymbol{\Omega}$, is diagonal with N distinct elements: $\text{diag}(\boldsymbol{\Omega}) = (\sigma_1^2 \;\; \sigma_2^2 \;\; \ldots \;\; \sigma_N^2)$. The aggregate regime indicator, $\boldsymbol{z}$, is a $T \times 1$ vector, where each element is $z_t \in \{1, 2, \ldots, K\}$, indicating which cluster of regions is in recession at date t. $\mathbf{P}$ is the matrix of transition probabilities that governs the mechanism of dynamic change between regimes in the model. h is a set of vectors, $h = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{K-2}\}$, where each vector $\mathbf{h}_k = (h_{1k} \;\; h_{2k} \;\; \ldots \;\; h_{Nk})'$ determines which regions belong to cluster k. β is the set of all logistic coefficient vectors: $\beta = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_{K-2}\}$, where $\boldsymbol{\beta}_k = (1 \;\; \beta_{k1} \;\; \ldots \;\; \beta_{kP_k})'$ and P_k is the number of covariates explaining the cluster affiliations.

3.2 Likelihood function

This section defines the functional form of the likelihood for the model (5). Prior to deriving the likelihood function, it is convenient to note that the model (5) can be written equivalently for the n^{th} observation at time t as

$$y_{tn} = \rho \sum_{j=1}^{N} W_{nj} y_{tj} + \mu_{n0} + \mu_{n1} h_{n,z_t} - \rho \sum_{j=1}^{N} W_{nj} (\mu_{j0} + \mu_{j1} h_{j,z_t}) + u_{tn}. \tag{9}$$

The likelihood function $\mathcal{L}(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, h; \mathbf{Y})$ is derived as

$$\mathcal{L}(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, h; \mathbf{Y}) \propto |\mathbf{I}_N - \rho\mathbf{W}|^T \left(\prod_{n=1}^{N} \sigma_n^{-T} \right) \times \exp\Bigg(-\frac{1}{2} \sum_{n=1}^{N} \Big[\sigma_n^{-2} \sum_{t=1}^{T} \Big\{ y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} - \mu_{n0} - \mu_{n1} h_{n,z_t} + \rho \sum_{j=1}^{N} W_{nj} (\mu_{j0} + \mu_{j1} h_{j,z_t}) \Big\}^2 \Big] \Bigg).$$

The term $|\mathbf{I}_N - \rho\mathbf{W}|$ takes into account the endogeneity of $\sum_{j=1}^{N} W_{nj} y_{tj}$ (see Anselin (1988) pp. 61–62). To see how the term appears in the likelihood, it is convenient to re-write the model as

$$\boldsymbol{y_t} = \boldsymbol{\mu_0} + \boldsymbol{\mu_1} \odot \mathbf{h_k} + (\mathbf{I}_N - \rho\mathbf{W})^{-1} \boldsymbol{u_t}. \tag{10}$$

Since the error covariance matrix is diagonal for $\boldsymbol{u_t}$, there exists a vector of homoskedastic random errors $\boldsymbol{v}_t$

$$\boldsymbol{v}_t = \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{u_t}. \tag{11}$$

Substituting (11) into equation (10) gives

$$\boldsymbol{\Omega}^{-\frac{1}{2}} \Big((\mathbf{I}_N - \rho\mathbf{W})(\boldsymbol{y_t} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \boldsymbol{h_k}) \Big) = \boldsymbol{v}_t. \tag{12}$$

The error terms $\boldsymbol{v}_t$ have a known distribution but are unobserved. Therefore it is necessary to introduce a Jacobian term to derive the joint distribution for $\boldsymbol{y_t}$ from the joint distribution of these error terms through the relationships in (12). The Jacobian for the transformation of the random variables vector $\boldsymbol{v}_t$ into the random variables vector $\boldsymbol{y_t}$ is

$$\mathbf{J} = \det \left(\frac{\partial \boldsymbol{v}_t}{\partial \boldsymbol{y_t}} \right) = |\boldsymbol{\Omega}^{-\frac{1}{2}}| |\mathbf{I}_N - \rho\mathbf{W}|. \tag{13}$$

This section defined the functional form of the likelihood with the spatial parameter ρ. An important implication of the Jacobian term defined in (13) is that its presence will require an augmentation to the MCMC estimation procedure (sampling algorithm), which will be discussed in Section 3.6.

3.3 Prior distributions

This section provides a formal statement of the model parameters. The beliefs and information pertaining to the uncertainty around all parameters is captured by their respective prior distributions (adopted prior to observing the data). Table 1 presents all priors and hyperparameters for all variables in the model. With the exception of the parameters related to the spatial dependence component in the model, all assumptions follow from Hamilton and Owyang (2012).

The priors for the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ are standard for Markov-switching models (see Kim and Nelson, 1999). The parameters $\boldsymbol{\mu}_n$, which characterize the growth rates for region n during regimes of recession and expansion are assigned a Normal prior distribution (Equation 14). The Inverse Gamma distribution is specified for σ_n^2, the n^{th} diagonal element of the diagonal variance-covariance matrix $\boldsymbol{\Omega}$, which implies a Gamma prior for the precision, σ_n^{-2} (Equation 15).

$$\pi(\boldsymbol{\mu}_n|\sigma_n) \propto \left|\sigma_n^2\mathbf{M}\right|^{-0.5}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_n-\mathbf{m})'[\sigma_n^2\mathbf{M}]^{-1}(\boldsymbol{\mu}_n-\mathbf{m})\right] \tag{14}$$

$$\pi(\sigma_n^{-2}) \propto \sigma_n^{-\nu+2}\exp\left(-\frac{1}{2}\delta\sigma_n^{-2}\right) \tag{15}$$

The β_k coefficients for the logistic clustering procedure adopt a Normal prior distribution, $\pi(\beta_k) \sim \mathrm{N}(\boldsymbol{b}_k, \boldsymbol{B}_k)$. Each column of the transition probability matrix, $\mathbf{P}$, adopts a diffuse Dirichlet prior, $\pi(\mathbf{P}_p) \sim D(\mathbf{0})$. The priors used for the latent variables $\boldsymbol{z}, h, \xi, \lambda$ are given in the cluster affiliation category in Table 1, they follow all assumptions made in the operational clustering procedure for the original model (see Hamilton and Owyang (2012) for prior elicitation details for these parameters).

Further consideration regarding prior choice for the population parameters, transition probabilities and latent variables is not considered. The reason for this is that the devel-

oped spatial model nests the original model, and perturbing the distributional assumptions would make the comparison of results between the two models less clearly attributed to the introduction of the spatial dependence component. Therefore, the operational framework of Hamilton and Owyang (2012) is taken as given.

The discussion is hereby focused on prior elicitation for the spatial population parameter ρ. The approach taken follows the spatial econometrics literature (see LeSage and Pace (2009) for a comprehensive treatment). The first step is to determine how the spatial parameter in the model (5) enters the joint prior distribution $\pi(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega})$. Following the spatial econometrics literature – LeSage and Pace (2009, p.129) – the priors for $\boldsymbol{\mu}_n$ and σ_n^{-2} are assumed to be independent of the spatial parameter ρ, which is constant across all N regions. The joint prior is given as

$$\pi(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}) = \pi(\rho) \prod_{n=1}^{N} \pi(\boldsymbol{\mu}_n|\sigma_n)\pi(\sigma_n^{-2}). \tag{16}$$

The spatial parameter is interpreted as the spatial correlation coefficient. Therefore, an appropriate prior distribution choice for the parameter is any valid density with support on the $(-1, 1)$ interval. This reflects the notion that values of ρ less than -1 are indicative of either model or spatial weighting misspecification (see discussion in LeSage and Pace (2009)). A popular choice is the uniform prior distribution such as $\pi(\rho) \sim \mathrm{U}(\gamma_{min}^{-1}, \gamma_{max}^{-1})$ or $\pi(\rho) \sim \mathrm{U}(-1, 1)$, where γ_{min} and γ_{max} represent the minimum and maximum eigenvalues of the spatial weight matrix $\mathbf{W}$. Other than restricting ρ to lie in a bounded interval, these distributions are uninformative with respect to ρ and assign an equal probability to any realization of ρ on that interval. These two choices are appropriate to use when the magnitude and significance of ρ is of particular interest. In the context of this paper's empirical investigation, the estimation results for ρ are of particular importance. If $\hat{\rho} \neq 0$ (estimated by the posterior mean) and ρ has a high sign certainty probability[4] then there are

[4]The probability mass on the same side of zero as the posterior mean.

Table 1: Formal statement of the model parameters.

Parameter	Prior distributions	Hyperparameters
Average employment growth in region n	$\pi \begin{pmatrix} \mu_{n0} \\ \mu_{n1} \end{pmatrix} \sim \mathrm{N}(\mathbf{m}, \sigma^2 \mathbf{M})$	$\mathbf{m}= \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \quad \mathbf{M}=\mathbf{I}_2$
Variance of errors	$\pi(1/\sigma_n^2) \sim \Gamma(\nu/2, \delta/2)$	$\nu = 0, \quad \delta = 0$
Spatial dependence	$\pi(\rho) \sim U(-1, 1)$	
Spatial weight matrix	$\mathbf{W} = \begin{bmatrix} w_{11} & \dots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \dots & w_{NN} \end{bmatrix}$	w_{ij} are row standardized spatial weights s.t. $\sum_{i=1}^{N} w_{ni} = 1, \ \forall n$
Cluster affiliation	$\pi(h_{nk}) = \begin{cases} \frac{1}{1+\exp(\mathbf{x}'_{nk}\boldsymbol{\beta}_k)}, & \text{if } h_{nk} = 0 \\ \frac{\exp(\mathbf{x}'_{nk}\boldsymbol{\beta}_k)}{1+\exp(\mathbf{x}'_{nk}\boldsymbol{\beta}_k)}, & \text{if } h_{nk} = 1 \end{cases}$	
	$h_{nk} = \begin{cases} 1, & \text{if } \xi_{nk} > 0 \\ 0, & \text{otherwise} \end{cases}$	
	$\pi(\xi_{nk} \vert \boldsymbol{\beta}_k, \lambda_{nk}) \sim \mathrm{N}(\mathrm{x}'_{nk}\boldsymbol{\beta}_k, \lambda_{nk})$	
	$\pi(\boldsymbol{\beta}_k) \sim \mathrm{N}(\boldsymbol{b}_k, \boldsymbol{B}_k)$	$\mathbf{b} = \mathbf{0}_p, \quad \mathbf{B} = 0.5\boldsymbol{I}_p$
	$\pi(\lambda_{nk}) \sim \mathrm{GIG}\left(\frac{1}{2}, 1, r_{nk}^2\right)$	$r_{nk} = \xi_{nk} - \mathbf{x}'_{nk}\boldsymbol{\beta}_k$
	GIG $\equiv$ Generalized Inverse Gaussian	
Transition probabilities	$\pi(\mathbf{P}_p) \sim \mathbf{D}(\boldsymbol{\alpha})$ (Dirichlet)	$\boldsymbol{\alpha} = \mathbf{0}$

significant spatial interactions between regions. Otherwise, the model is that of Hamilton and Owyang (2012) with the restriction $\rho = 0$.

3.4 Conditional distributions

The conditional distribution of any given subset of parameters $\theta \in \{\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta\}$ is given by

$$p(\theta|\mathbf{Y}, \rho, \mathbf{P}, \boldsymbol{z}, h, \beta) = \frac{p(\mathbf{Y}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta)}{\int p(\mathbf{Y}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta) d\theta} \tag{17}$$

Conditioning on the data and all other parameters makes all factors that are not functions of the individual parameter go into the proportionality constant (normalization constant).

Conditional distribution of population parameters μ, Ω

The conditional posterior distributions of each individual parameter $\boldsymbol{\mu}_n$ and σ_n^{-2} are derived from

$$p(\theta_n|\mathbf{Y},\mathbf{P},\boldsymbol{z},h,\beta) \propto \pi(\theta_n)\sigma_n^{-T}\exp\left[-\frac{1}{2}\sigma_n^{-2}\sum_{t=1}^{T}\left(y_{tn}-\rho\sum_{j=1}^{N}W_{nj}y_{tj}-\mu_{n0}-\mu_{n1}h_{n,z_t}\right)^2\right].$$

Conditioning on the data and all other parameters makes all factors that are not functions of the individual parameter go into the proportionality constant (normalization constant).

Both conditional posterior distributions for $\boldsymbol{\mu}_n$ and σ_n^{-2} are affected by ρ, which enters through the likelihood function. Compared to the original model with no spatial lag, these distributions remain in standard known form, which merits the use of the Gibbs sampling procedure for their estimation.

The conditional posterior distribution of $\boldsymbol{\mu_n}$ is (see Appendix B.3 for proof)

$$p(\boldsymbol{\mu}_n|\mathbf{Y},\rho,\sigma_n^{-2},\mathbf{P},\boldsymbol{z},h,\beta) \propto \exp\left[-\frac{1}{2}(\boldsymbol{\mu}_n-\mathbf{m}^*)'\boldsymbol{\Sigma_n}^{-1}(\boldsymbol{\mu}_n-\mathbf{m}^*)\right],$$

$$\boldsymbol{\mu}_n|\mathbf{Y},\rho,\sigma_n^{-2},\mathbf{P},\boldsymbol{z},h,\beta \sim \mathrm{N}\Big(\mathbf{m}^*,\boldsymbol{\Sigma_n}\Big),$$

where

$$\boldsymbol{\Sigma_n} = \mathbf{A}^{-1},$$

$$\mathbf{m}^* = \mathbf{A}^{-1}\mathbf{b},$$

$$\mathbf{A} = \sigma_n^{-2}(1+\rho W_{nn})^2\sum_{t=1}^{T}\mathbf{w}(z_t,h)\mathbf{w}(z_t,h)' + [\sigma_n^2\mathbf{M}]^{-1},$$

$$\mathbf{b} = \sigma_n^{-2} \sum_{t=1}^{T} \Big(\mathbf{w}(z_t, h)(1 + \rho W_{nn}) \Big) \Big(y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \Big) + [\sigma_n^2 \mathbf{M}]^{-1} \mathbf{m}.$$

The conditional posterior distribution of σ_n^{-2} (see Appendix B.4 for proof) is (as in Kim and Nelson (1999))

$$\sigma_n^{-2} | \mathbf{Y}, \rho, \boldsymbol{\mu}_n, \mathbf{P}, \boldsymbol{z}, h, \beta \sim \Gamma \left(\frac{\nu + T}{2}, \frac{\delta + \hat{\delta}}{2} \right), \tag{18}$$

for $\hat{\delta} = \sum_{t=1}^{T} \Big(y_{tn} - \mu_{n0} - \mu_{n1} h_{n,z_t} - \rho \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \Big)^2$.

Conditional distribution of population parameter ρ

Any uniform prior distribution for ρ, that is a valid density, e.g. $\pi(\rho) \sim \mathrm{U}[\gamma_{min}, \gamma_{max}]$, $\pi(\rho) \sim \mathrm{U}[-1, 1]$, results in the following conditional posterior distribution of the spatial parameter ρ is (see Appendix B.5 for complete proof)

$$p(\rho | \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta) \propto \; |\boldsymbol{I}_N - \rho \mathbf{W}|^T \exp \left[-\frac{1}{2} \rho \Big(\rho \mathbf{B}_1 - 2 \mathbf{B}_2 \Big) \right], \tag{19}$$

where

$$\mathbf{B}_1 = \sum_{n=1}^{N} \left(\sigma_n^{-2} \sum_{t=1}^{T} \Big(\sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \Big)^2 \right),$$

$$\mathbf{B}_2 = \sum_{n=1}^{N} \left(\sigma_n^{-2} \sum_{t=1}^{T} \Big((y_{tn} - \mu_{n0} - \mu_{n1} h_{n,z_t}) \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \Big) \right).$$

The distribution in (19) has no known standard form. This has implications for the sampling procedure required to estimate the model, which is discussed in Section 3.6 that describes the MCMC sampling algorithm for ρ and for the full model.

3.5 Cluster selection

Following Hamilton and Owyang (2012) the number of clusters is determined by computing a quasi-out-of-sample score. This procedure can also be used to select which weighting structure to impose on the matrix $\mathbf{W}$. The score is built through an $R-$fold cross validation procedure that segments the full data set into $R-$equal blocks ($R = 10$ in the empirical investigation), denoted as $r = 1, \ldots, R$. Each of the r blocks is retained as a validation set, and the $R - 1$ blocks serve as a training set. The model estimates for the training set are tested on the omitted validation set. All R blocks serve exactly one time as the validation set. Cross-validation deems the model with the lowest aggregate score as superior to the other model specifications. The treatment for this procedure in a similar class of models can be found in Geweke and Keane (2007).

The aggregate quasi-out-of-sample score is defined as

$$\text{Score} = \frac{1}{M}\sum_{m=1}^{M}\sum_{r=1}^{R}\sum_{t=t_r}^{t_{r+1}-1}\Big(\log|\boldsymbol{\Omega}^{[r,m]}| + (\boldsymbol{y}_t - \boldsymbol{y}_t^f)'(\boldsymbol{\Omega}^{[r,m]})^{-1}(\boldsymbol{y}_t - \boldsymbol{y}_t^f)\Big), \tag{20}$$

where M is the number of sampling iterations, R is the number of equal blocks, $\boldsymbol{\Omega}^{[r,m]}$ is the iteration m diagonal variance-covariance matrix draw for block r, t_r is the first observation of the omitted block, $t_{r+1}-1$ is the last observation of the omitted block, $\boldsymbol{y_t} = (y_{1t}, \ldots, y_{Nt})'$ is the vector of employment growth rates for N regions at date t, $\boldsymbol{y}_t^f$ is the forecast of $\boldsymbol{y}_t$ conditional on the aggregate indicator $z_t^{[r,m]}$, and $(\boldsymbol{y}_t - \boldsymbol{y}_t^f)$ is the forecast error vector.

The score will serve as an objective criterion for selecting the number of clusters.This is supported by two arguments. The first is for the empirical investigation to be consistent with the original application of the restricted non-spatial model, which allows the results to be compared to the state-level analysis. The second is that due to the complicated latent-structure of the model, approximating the marginal likelihood function is a difficult task. Therefore, implementing a cross-validation procedure is more straightforward than relying

on approximate Bayes Factors (BF) for selection, which explains why the measure was not employed by Hamilton and Owyang (2012). For completeness, a suggested approach for obtaining BF values for the spatial model is now discussed.

Model comparison for econometric models that use Bayesian posterior inference typically rely on the Bayes Factors (BF) to discriminate between models. The BF judges which of two given models is better supported by the data through their respective marginal likelihood functions. The BF that compares model i to model j is defined as

$$\begin{aligned} BF_{ij} &= \frac{m_i(\mathbf{Y})}{m_j(\mathbf{Y})} \\ &= \frac{\int \mathcal{L}_i(\Phi;\mathbf{Y})\pi_i(\Phi)d\Phi}{\int \mathcal{L}_j(\Phi;\mathbf{Y})\pi_j(\Phi)d\Phi}, \end{aligned} \tag{21}$$

where $\Phi = \{\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, \mathrm{H}, \beta\}$. BF has the advantage of penalizing more heavily parameterized models, which is not relevant for comparing weighting matrix structures, but relevant for specifying the number of clusters. An exact or approximate evaluation of marginal likelihoods is needed to compute the Bayesian factor integrals in (21). This can be accomplished by employing the approach in Chib (1995) for computing the marginal likelihood given the parameter draws from the posterior distribution, which facilitates the computation of Bayes factors as a by-product of the simulation. The central equation for evaluating the marginal density is given by re-arranging Bayes' Rule to isolate the unconditional density of $\mathbf{Y}$ (the normalization constant of the posterior density) denoted as $m(\mathbf{Y})$ and given by

$$m(\mathbf{Y}) = \frac{\mathcal{L}(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, \mathrm{H}, \beta; \mathbf{Y})\pi(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, \mathrm{H}, \beta)}{p(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, \mathrm{H}, \beta|\mathbf{Y})} \tag{22}$$

which Chib refers to as the basic marginal likelihood identity (BMI). The proposed estimate for the marginal likelihood is obtained through the estimated log-likelihood function as follows

$$\begin{aligned}\ln \widehat{m}(\mathbf{Y}) =& \ln \mathcal{L}(\rho^*, \boldsymbol{\mu}^*, \boldsymbol{\Omega}^*, \boldsymbol{z}^*, H^*, \beta^*; \mathbf{Y}) + \ln \pi(\rho^*, \boldsymbol{\mu}^*, \boldsymbol{\Omega}^*, \boldsymbol{z}^*, \mathbf{P}^*, H^*, \beta^*) \\ & - \ln p(\rho^*, \boldsymbol{\mu}^*, \boldsymbol{\Omega}^*, \boldsymbol{z}^*, \mathbf{P}^*, H^*, \beta^* | \mathbf{Y}),\end{aligned} \tag{23}$$

which requires the evaluation of the log-likelihood function, prior and the joint posterior distribution (third term in (23)) for a given set $\{\rho^*, \boldsymbol{\mu}^*, \boldsymbol{\Omega}^*, \boldsymbol{z}^*, \mathbf{P}^*, H^*, \beta^*\}$. The computational challenge is due to the fact that the joint posterior functional form now includes all normalization constants for the individual posterior conditional distributions of the variables. For the Metropolis-within-Gibbs sampler these constants can be suppressed when they are fixed conditional on the variable for which the conditional posterior densities is obtained. Obtaining estimates of (23) would base model selection on the estimated BF given by

$$\widehat{B}_{ij} = \exp\left(\ln \widehat{m}_i(\mathbf{Y}) - \ln \widehat{m}_j(\mathbf{Y}) \right) \tag{24}$$

Recognizing the two existing competing options for model selection, the cross-validation scores will be used for choosing the number of idiosyncratic clusters in the model to be consistent with the empirical investigation in Hamilton and Owyang (2012). Furthermore, analyzing 177 regional divisions of the United States, excluding the auxiliary variables, implies a total of 1322 parameters in the model, a much greater parameter space than if one was analyzing the 48 lower US states. The cross-validation program of Hamilton and Owyang for this model provides a convenient implementation of this model selection procedure for investigating smaller regions.

3.6 MCMC estimation

The estimation procedure will employ the Metropolis-within-Gibbs (M-G) algorithm to sample from the conditional distributions of the parameters. The M-G algorithm takes its name from the fact that it uses the Gibbs algorithm to sample from conditional distributions where

Table 2: M-H algorithm with random walk tuning

Step 1: Initialize the draw, setting $\rho^{(1)}$ and $c^{(1)}$ to an arbitrary number.
Step 2: For $j = 1, \ldots, N^{burn-in}, N^{burn-in} + 1, \ldots, N^{burn-in} + N^{keep}$

(a) A candidate value ρ^* is drawn from the candidate distribution:

$$\rho^* = \rho^{(j)} + c^{(j)} \cdot \mathrm{N}(0, 1) \tag{25}$$

(b) The value is accepted as $\rho^{(j+1)} = \rho^*$ with probability:

$$\phi(\rho^{(j)}, \rho^*) = \min\left[1, \frac{p(\rho^*|\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta)}{p(\rho^{(j)}|\mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, h, \beta)}\right] \tag{26}$$

(c) Adjust the tuning parameter by monitoring the acceptance rate

$$c^{(j+1)} = \begin{cases} 1.1c^{(j)}, & \text{if } \phi(\rho^{(j)}, \rho^*) > 0.60 \\ \frac{c^{(j)}}{1.1}, & \text{if } \phi(\rho^{(j)}, \rho^*) \leq 0.40 \end{cases} \tag{27}$$

the distributional form is known, and the Metropolis-Hastings (M-H) algorithm to sample from conditional distributions where the distributional form is unknown. The estimation in Hamilton and Owyang (2012) did not require the M-H algorithm because all distributions were of known standard form. With the introduction of a spatial dependence component, inference on the spatial parameter ρ requires sampling from a distribution of unknown form (see discussion in Section 3.2). The sampling algorithm for ρ is discussed first, followed by the complete sampling algorithm for the full model.

The M-H algorithm will be implemented with a tuned random-walk procedure (see LeSage and Pace (2009) and Holloway et al. (2002)). The candidate distribution will be the normal distribution and the tuning parameter is denoted as c. The sampler proceeds according to the algorithm in Table 2. Tuning the draws from the candidate normal distribution is governed by the tuning parameter c, which ensures that the M-H algorithm moves over the entire conditional distribution. The restriction that $-1 < \rho^* < 1$ is imposed for the M-H algorithm. This ensures that candidate values lie inside the desired interval

when drawn from the candidate distribution.

The algorithm in Table 2 is nested into the Gibbs sampling algorithm for the full model, which collectively is referred to as the Metropolis-within-Gibbs (M-G) algorithm. A concise summary of the algorithm is given in Table 3, which suppresses all conditioning variables, hyperparameters and functional forms. This is followed by a complete and detailed outline of the full sampling procedure. After initializing parameters (Step 0), the samplers iteratively draw following Steps 1 through 6 for $j = 1, \ldots, N^{burn-in}, N^{burn-in} + 1, \ldots, N^{burn-in} + N^{keep}$, where $N^{burn-in}$ is the specified number of burn-in iterations that are discarded and N^{keep} is the number of iterations retained for inference as posterior draws.

Table 3: Short summary of Metropolis-within-Gibbs (M-G) algorithm

Step 0: Initialize all parameters
Step 1: Draw cluster
(a) $\beta_k^{(j+1)}$,$k = 1, \ldots, K-2$
(b) $h_{nk}^{(j+1)}$, $k = 1, \ldots, K-2$ and $n = 1, \ldots, N$
(c) $\xi_{nk}^{(j+1)}$, $k = 1, \ldots, K-2$ and $n = 1, \ldots, N$
(d) $\lambda_{nk}^{(j+1)}$, $k = 1, \ldots, K-2$ and $n = 1, \ldots, N$
Step 2: Draw $\boldsymbol{\mu}_n^{(j+1)}$, $n = 1, \ldots, N$
Step 3: Draw $\sigma_n^{-2(j+1)}$, $n = 1, \ldots, N$
Step 4: Draw $\rho^{(j+1)}$ using the M-H algorithm defined in Table 2.
Step 5: Draw aggregate regime indicator, $z_t^{(j+1)}$, $t = 1, 2, \ldots, T$
Step 6: Draw the transition probabilities, $\mathbf{P}^{(j+1)}$

Notes: All conditioning variables are suppressed in this outline of the algorithm.

The complete M-G algorithm for estimating the model is given in B.2.

4 Data and spatial weighting

This section describes the choice of aggregation level, the data sets, and the spatial weighting structure. Section 4.1 describes BEA economic areas, explains their relevance to classifying

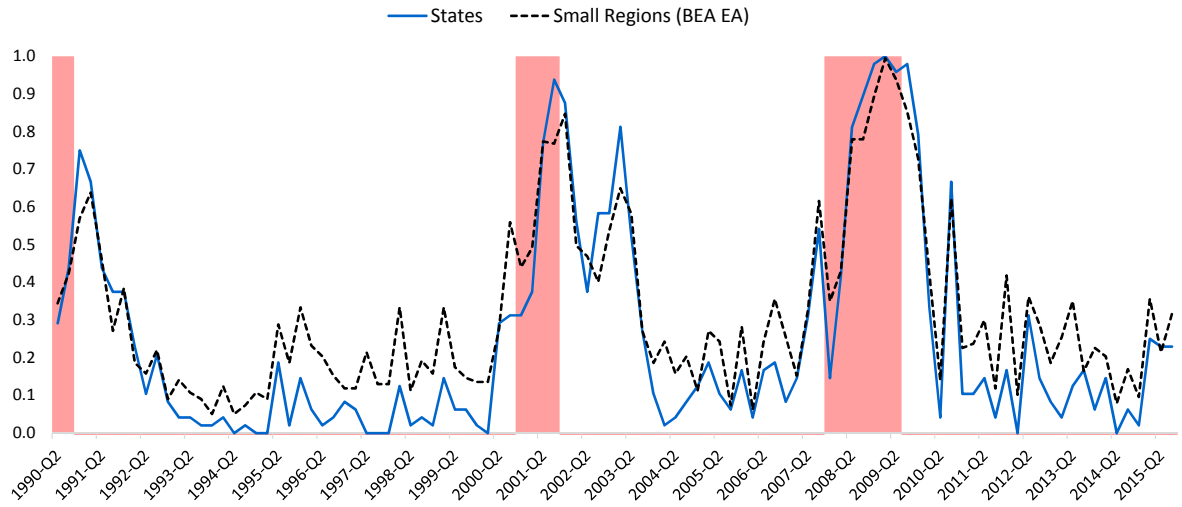
small regional economies in the US, and describes the employment data used to construct the variable that fluctuates over the business cycle. Section 4.2 describes the regional covariates to be used in the clustering mechanism. Lastly, Section 4.3 defines the main specification of the spatial weighting structure and existing alternatives, some of which will serve as robustness checks for the estimation results.

4.1 County-level employment – BEA Economic Areas

The goal of this paper is to conduct an empirical investigation of regional business cycles in small regional economies using a multivariate Markov-switching model. Because the model draws inferences on larger regional groupings of geographical units, it is appropriate to consider smaller regions as the unit of analysis than the lower 48 states. This is also motivated by the fact that smaller areas provide more geographic detail. Geographical detail is important for analyzing business cycle characteristics as it provides a more in-depth view of regional economies. For example, it allows the model to potentially identify economic downturns that affect regions covering only a certain part of a state and that cross state borders.

Looking at the state level can be misleading for identifying regional contractions. This is evident if one compares the proportion of states and small regions (as given by the 177 BEA economic areas) that experience negative employment growth rates; see Figure 1. The first observation is that the figure clearly shows periods when none of the 48 lower states exhibit contraction in observed employment, while over the same periods as much as 21% of smaller regions exhibit contractions in observed employment. The second observation is that, overall, in-between periods of nationwide negative employment growth, a higher proportion of the country is experiencing negative employment growth when looking at smaller regions than at the state level.

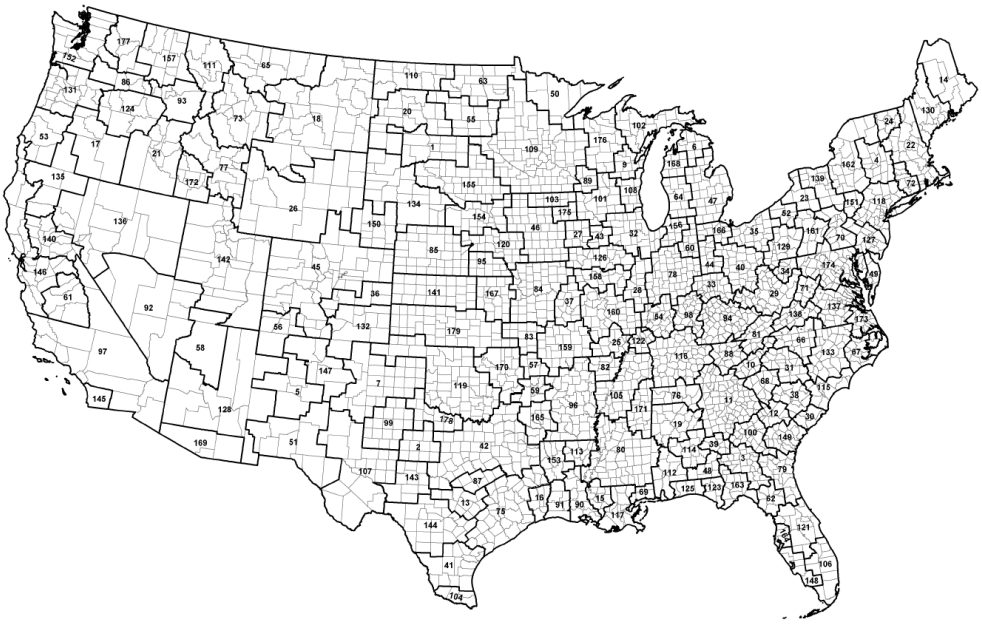
Figure 1: Proportion of regions with negative employment growth



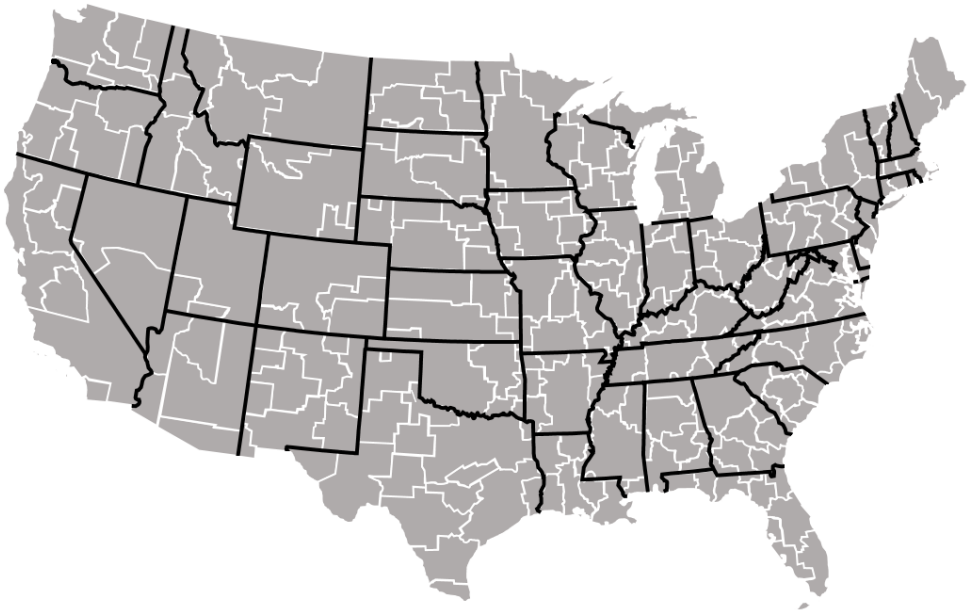
The choice of disaggregation level is motivated by two criteria. The first is that the computational cost of the model is manageable, and the second is that, ideally, the geographical units share physical borders, which preserves the contiguity property exhibited by the lower 48 US states. Both criteria would be invalidated if one were to consider core-based statistical areas (metropolitan and micropolitan statistical areas; see footnote 1 for details). A regional classification that is less commonly analyzed but is appropriate for both the computational cost of the model and the empirical context is the economic area classification by the Bureau of Economic Analysis (BEA). This classification was defined in 1995 and subsequently updated in 2004. The 2004 classification defined 177 contiguous groups of counties that represent relevant regional markets surrounding metropolitan and micropolitan statistical areas in the United States. See Figure 2 for a map of BEA Economic Areas in relation to county borders and state borders, and for a map of surrounding statistical areas see Figure 11 in Appendix A.

Figure 2: BEA Economic Areas

(a) Counties



(b) Overlay with state borders



The 2004 BEA economic area classification is based on the counties and county-equivalents enumerated in the 2000 census, which remained unchanged from the 1990 census. This enumeration does not include Broomfield County in Colorado which was established in

November, 2001 from parts of four Colorado counties in the vicinity of Denver. Statistical agencies do not typically publish data using this classification. All of the data sets used in the analysis will be constructed for this classification based on county-level data. All county-level data for the contiguous US counties are based on the 3106 contiguous counties and county-equivalents enumerated in the 1990 and 2000 census. That is the 3001 contiguous counties, 64 Louisiana Parishes, 41 independent cities in Virgina and the District of Columbia, for a total of 3106. The 5 counties of Hawaii and the 19 organized boroughs and 11 census areas of Alaska are excluded.

The business cycle characteristics in the model will be inferred from county-level employment data, which are aggregated to the 177 contiguous BEA economic areas. The data were obtained as total private payroll employment for the period 1990–2015 from the Quarterly Census of Employment and Wages (QCEW) at the Bureau of Labour Statistics (BLS). The aggregate series are seasonally adjusted by the X13 seasonal adjustment programs of the US Census Bureau. The data enter the model as annualized quarter-over-quarter growth rates.

The empirical investigation of Hamilton and Owyang (2012) used state-level employment (1956–2007) as the variable that fluctuates over the business cycle, while the data for the covariates for the most part only spanned the period 1990–2006[5]. Due to the importance these covariates exert on assessing the probability that a given region n belongs to a given cluster k, an unbalanced coverage of the data has drawbacks. This paper takes on a different approach by ensuring that the period spanned by all data used in the empirical application coincides with the main period of analysis as closely as possible.

[5]The empirical investigation in Hamilton and Owyang (2012) used manufacturing employment shares (1990-2006), oil production (1984), financial services shares (1990-2006), and small-firm employment shares (unspecified)

4.2 Regional covariates

Cluster affiliation is driven by regional-level covariates. These covariates serve to capture the influences that specific industry sectors have on designating regions to be affiliated with a cluster. This operates through the clustering mechanism and influences the probability that a given region n belongs to a given cluster k.

A total of six covariates are considered in the empirical investigation (see Table 4 for details). Five of the covariates are average industrial employment composition variables. This is partially motivated by the availability of comprehensive time series data for these variables over the examined period. More importantly, the direct representation of industry-specific characteristics using employment shares ought to provide an intuitive geographical/spatial comparison of each regional cluster grouping to the higher and lower concentrations of specific industries employment composition. To keep the model computationally feasible, only the discussed six variables are used, which are postulated to be informative for identifying similar co-movements relevant to business cycles. Four of the covariates are direct analogs to those considered originally in Hamilton and Owyang (2012). The other two are motivated as important goods-producing economic sectors relevant for the considered time period, 1990–2015, and meaningful when the model is estimated at a more disaggregated level. All covariates enter the model as averages over the full period, meaning they serve to capture the relative regional concentration of the labor force in these industries, not their fluctuations over the period.

Manufacturing employment shares represent the largest sector of the goods-producing sector of the economy. The second largest sub-sector of the goods-producing sector is the construction sector, which for small regional divisions exhibits a lot of spatial variability, see Figure 3. Natural resource extraction is given by oil and gas extraction employment shares and mining and quarrying employment shares. To capture the regional concentrations of employment in small firms, average employment shares of firms with fewer than 100

Table 4: Data summary - Employment shares

Industry		County-level data		Aggregated variable description	
Name	NAICS	Source	Period	# Regions	Enters the model as
Oil & gas extraction	211	QCEW	1990-2014	177	↑
Mining & quarrying	212	QCEW	1990-2014	177	average (%) share
Construction	23	QCEW	1990-2014	177	of total employment
Manufacturing	33-33	QCEW	1990-2014	177	↓
Financial & insurance	52	QCEW	1990-2014	177	
Small firms (less than 100 employees)	All	SUSB	2010	177	(%) share of total employment

Notes: All variables are constructed for the 177 BEA Economic Areas, and are aggregated from county-level data obtained from the Quarterly Census of Employment and Wages (QCEW) of the Bureau of Labor Statistics with the exception of small firm shares, which are aggregated from county-level data of the Statistics of U.S. Businesses of the United States Census Bureau.

employees are considered.

For the service-producing sector, financial activities employment shares for the finance and insurance sub-sector (NAICS 52) represent the financial sector. Consideration was given to accounting for the information and cultural industries employment shares to serve as a broad measure of regional employment concentration due to high-tech industries (e.g. computer software and Internet sub-sectors). This sector is not included because the relative regional concentrations of employment shares for this sector are very similar to the financial sector. A subset of industries related to high-tech stocks such as computer and electronic products, aerospace, pharmaceutical and medicine are covered by the manufacturing industry classification.

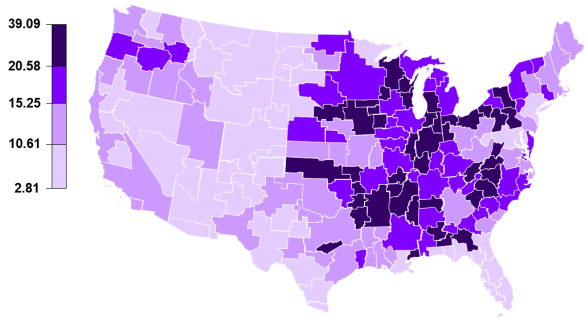
4.3 Spatial weighting matrices

This section defines the weighting structure considered in the empirical investigation and the existing alternatives.

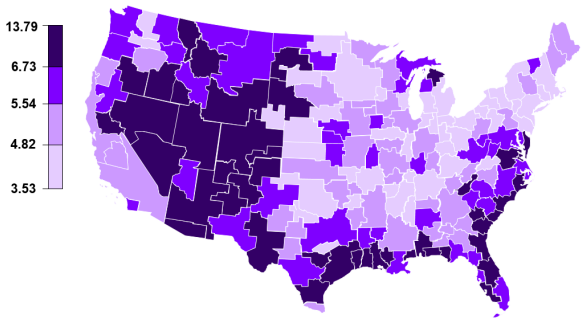
Estimating the model using a data set with a cross-section dimension of $N = 177$ (as

Figure 3: Regional covariates - employment shares (%)

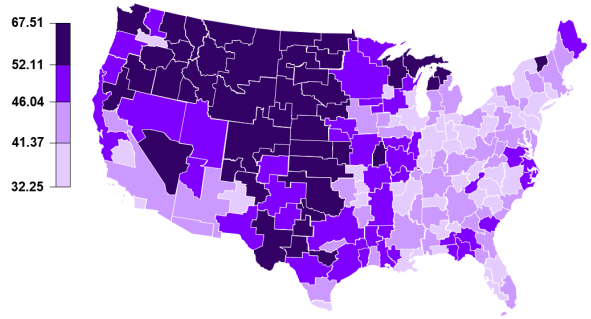
(a) Manufacturing sector



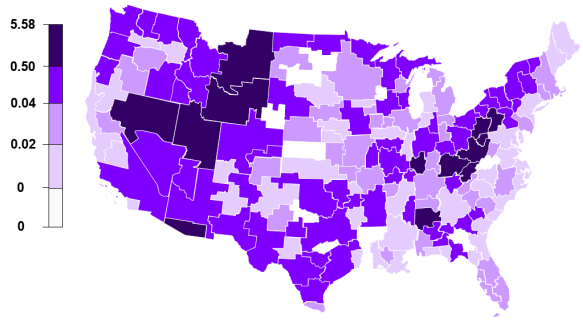
(b) Construction sector



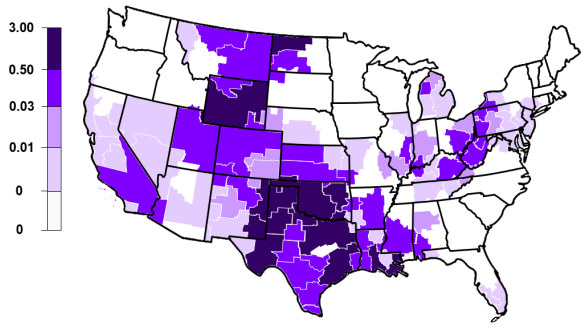
(c) Small firm (less than 100 employees)



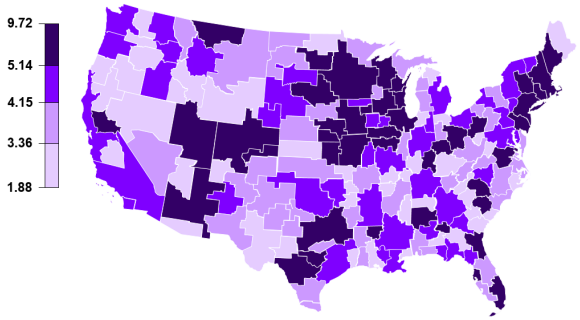
(d) Mining & quarrying



(e) Oil & gas extraction



(f) Finance and Insurance sector



compared to $N = 48$ when looking at the state level) substantially increases computational cost. The advantage of introducing spatial interactions via a spatial error lag component is that inference is required only for a single additional parameter, ρ, compared to the original model. However, the computational intensity of the MCMC estimation algorithm depends on the sparsity of the spatial weighting matrix $\mathbf{W}$.

This paper advocates the use of a spatial weighting structure based on shared physical borders between economic areas. This contiguity weighting is defined in 28 and is illustrated

with an example in Figure 4.

Contiguity-based weighting matrix

$$\mathbf{W} = \begin{bmatrix} w_{11} & \dots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \dots & w_{NN} \end{bmatrix}, \quad w_{ij} = \frac{\gamma_{ij}}{\sum_{n=1}^{N} \gamma_{in}}$$

$$\gamma_{ij} = \begin{cases} 1, & \text{if region } j \text{ shares a common border with region } i \\ 0, & \text{otherwise} \end{cases} \tag{28}$$

This weighting structure is motivated by two observations. The first is that BEA economic areas define relevant regional markets surrounding metropolitan and micropolitan areas, hence this regional classification accounts for socioeconomic similarities between counties surrounding statistical areas. Therefore, the spatial dependence structure embodied in $\mathbf{W}$ doesn't necessarily need to use economic distance attributes of these regions to establish connections, as one would expect when analyzing larger regional groupings (e.g. states or the eight BEA regions). The second observation is that a weighting matrix defined through shared physical borders between regions is one of the most sparse specifications of $\mathbf{W}$ in comparison to well-known existing alternatives. This can be seen in Table 5, which shows how many non-zero weights various $\mathbf{W}$ specifications have for the 177 BEA economic areas. A contiguity-based weighting provides a lighter computational load on the model while leveraging the inherent spatial similarities accounted for by the BEA economic area classification.

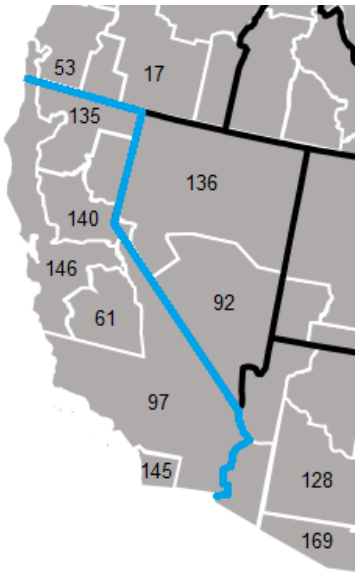
Table 5: Spatial weights summary

	$\mathbf{W}_{\text{contig}}$	$\mathbf{W}_{\text{400km}}$	$\mathbf{W}_{\text{600km}}$	$\mathbf{W}_{\text{800km}}$	$\mathbf{W}_{\text{inv-dist}}$
# of weights	912	1882	3926	6552	31152

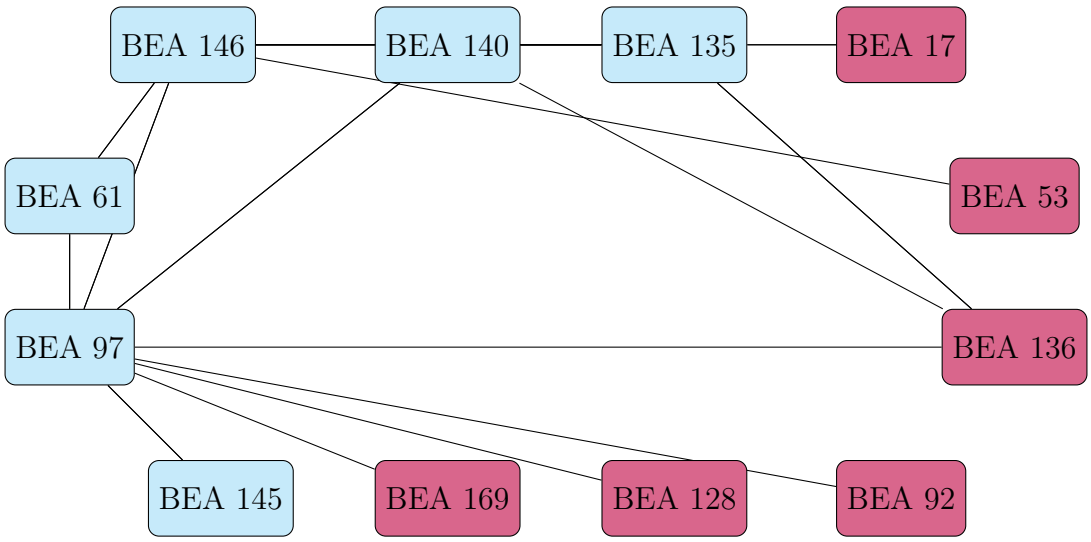
Notes: $\mathbf{W}_{\text{contig}}$ is the contiguity weighting, $\mathbf{W}_{\text{400km}}$, $\mathbf{W}_{\text{400km}}$ and $\mathbf{W}_{\text{800km}}$ are 400, 600 and 800 kilometer distance band weightings, and $\mathbf{W}_{\text{inv-dist}}$ is the symmetric inverse distance weighting. parentheses.

Figure 4: Contiguity weighting example

(a) California and surrouding economic areas



(b) Network: Blue (light) - all or majority of area contained in the state, Red (dark) - all or majority of area outside of the state [7]



Geography-based restrictions have been supported in empirical factor model and spatial econometric studies of regional business cycles. An example is the work of Del Negro (2002), where model restrictions are based on geographic proximity, which serves as a proxy

[1]Edges connecting bordering economic areas outside of the California state border are suppressed to illustrate the spatial weighting for the specific state.

for regional productive structure and income levels. This is motivated by the belief that geography plays an important role in characterizing the productive structure of a region. Fogli et al. (2015) also argue that local geographical factors play an important role for aggregate business cycle dynamics.

A contiguity-based weighting structure is a common choice for $\mathbf{W}$ in the spatial econometrics literature, and its use is substantiated by the regional similarities accounted for through the BEA economic area classification. The following are well-known alternative weighting structures that can be used for specifying $\mathbf{W}$:

1. Distance-based (K-nearest neighbors)
$$\gamma_{ij} = \begin{cases} 1, & \text{if region } j \text{ is one of the K-nearest neighbours of region } i \\ 0, & \text{otherwise} \end{cases}$$
2. Economic or technological-distance matrices (trade flows)
$$\gamma_{ij} = \begin{cases} 1, & \text{if region } j \text{ is the largest trading partner of region } i \\ 0, & \text{otherwise} \end{cases}$$
or
$$w_{ij} = \text{proportion of region } i\text{'s trade with region } j.$$
3. Distance band (symmetric):
$$\gamma_{ij} = \begin{cases} 1, & \text{if regions } i, j \text{ are at a critical distance cut-off point} \\ 0, & \text{otherwise} \end{cases}$$
4. Inverse distance weights (symmetric):
$$w_{ij} = 1/d_{ij}^2$$
5. Kernel matrices (fixed and adaptive)
6. Distance by relative frontier longitude (non-symmetric):
$$w_{ij} = (d_{ij})^{-a}(d_{ij})^{b}$$
7. Accessibility matrices: networks

Where γ_{ij} are elements of a matrix that is then row-standardized, and w_{ij} are the exact

weights used for the weighting matrix.

5 Empirical results and discussion

This section presents and discusses the results of the empirical investigation. Section 5.1 outlines the selection of idiosyncratic clusters using a cross-validation procedure. Section 5.2 analyzes the regional composition of each idiosyncratic cluster grouping. Section 5.3 analyzes how each phase of the regional (non-nationwide) and nationwide business cycles propagate in the economy. Section 5.4 discusses the assessment of spatial regional interactions, including a robustness analysis. Section 5.5 discusses some policy implications based on the findings.

5.1 Selection of idiosyncratic clusters

The cross-validation procedure guides the choice of how many idiosyncratic clusters $\kappa = K-2$ are specified in the model. All estimation results are based on MCMC runs of 250,000 burn-in iterations followed by 25,000 subsequent iterations retained for posterior inference. The long burn-in period is the same length as in Hamilton and Owyang (2012) and ensures reliable posterior inference. Convergence is discussed in detail in Appendix A.3. Cross-validation is conducted for $R = 5$ folds (refer to Section 3.5 for details on cross-validation) for equal MCMC sequence lengths and considers up to six ($\kappa = 6$) clusters.

Cross-validation results are reported in Table 6. The procedure selects two idiosyncratic clusters ($\kappa = 2$) for both the spatial and restricted non-spatial models, based on their lowest total scores. The total entropy scores for the spatial model are lower than the corresponding scores for the restricted model, indicating that cross-validation favors the spatial specification. All subsequent analysis and results will be for the spatial model specification

Table 6: Cluster selection - cross-validation entropy scores

Clusters (κ)	Spatial Model						Restricted Model ($\rho = 0$)
	Block 1	Block 2	Block 3	Block 4	Block 5	**Total**	**Total**
1	6856.0	8455.4	3679.6	4321.5	3279.0	26591.5 (4)	34780.0 (2)
2	6847.5	8434.6	3679.1	4161.7	3344.5	26467.4 (1)	34230.6 (1)
3	6845.4	8460.7	3693.0	4132.1	3406.2	26537.4 (2)	36548.0 (4)
4	6864.0	8462.6	3703.1	4221.6	3392.7	26644.0 (5)	35222.6 (3)
5	6847.4	8469.9	3707.3	4156.7	3527.7	26709.0 (6)	38621.2 (6)
6	6851.3	8457.4	3693.1	3920.0	3660.7	26582.5 (3)	37052.5 (5)

Notes: The aggregate scores for each number of clusters and weighting matrix. Numbers in parentheses rank the model specification from the highest ranking (1) to the lowest ranking (6). Results based on $R = 5$ cross-validation, up to 6 clusters considered with every MCMC run comprised of $N^{burn-in} = 250000$ burn-in iterations and $N^{keep} = 25000$ samples retained for inference.

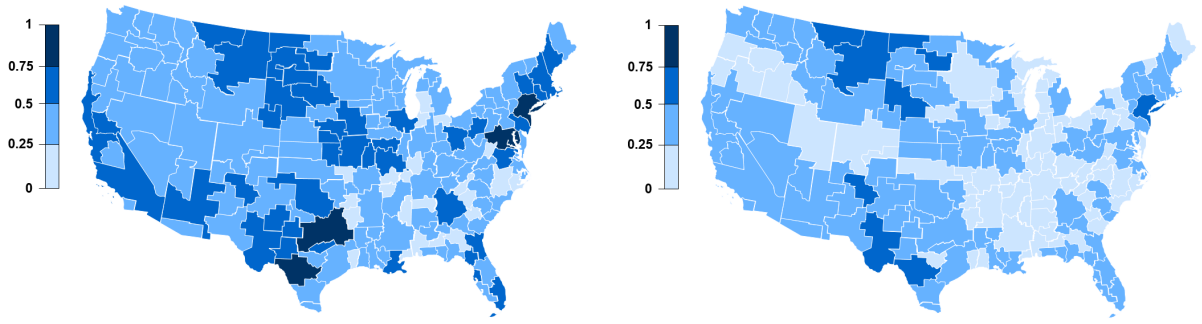
with two idiosyncratic clusters ($\kappa = 2$, four regimes in total). A formal assessment of MCMC output convergence for this specification is given in Appendix A.3. The results show that all 1322 parameters converge, indicating that the target posterior distributions based on the retained MCMC draws are reliable for inference and there is very little potential scale reduction from continuing the MCMC algorithm.

5.2 Regional spatial clusters

The mechanism of endogenous clustering is best illustrated in the following two steps. The first step draws inference based only on the exogenous industrial composition variables that capture fixed regional characteristics. This provides a geographical illustration (see Figure 5) for all regions according to their respective probabilities of belonging to each idiosyncratic cluster. The second step updates these probabilities with the observed regional employment growth rates. For every region and cluster, these probabilities are based on the posterior draws of the cluster affiliation indicator h_{nk}. Figure 6 shows the spatial illustration for the posterior probabilities of regions belonging to each cluster based on all of the data entering through the clustering mechanism and the likelihood function. Together these figures convey

Figure 6: Spatial Model
Posterior cluster affiliation probabilities updated with observed regional employment growth [posterior means of h_{nk}]

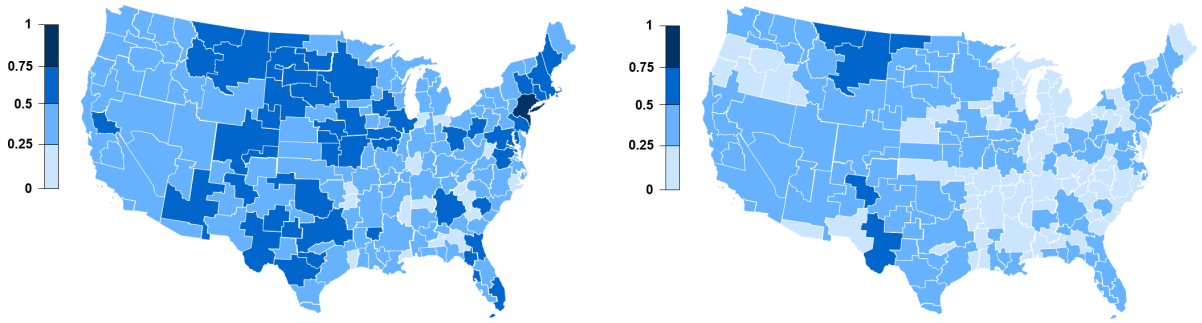
(a) Cluster 1 (b) Cluster 2



the hierarchal structure of the model to show how the informational content of the regional data shapes the cluster groupings.

Figure 5: Spatial Model
Posterior cluster affiliation probabilities based on exogenous industrial composition variables alone [posterior means of $\frac{\exp(\mathbf{x}'_{nk}\beta_k)}{1+\exp(\mathbf{x}'_{nk}\beta_k)}$]

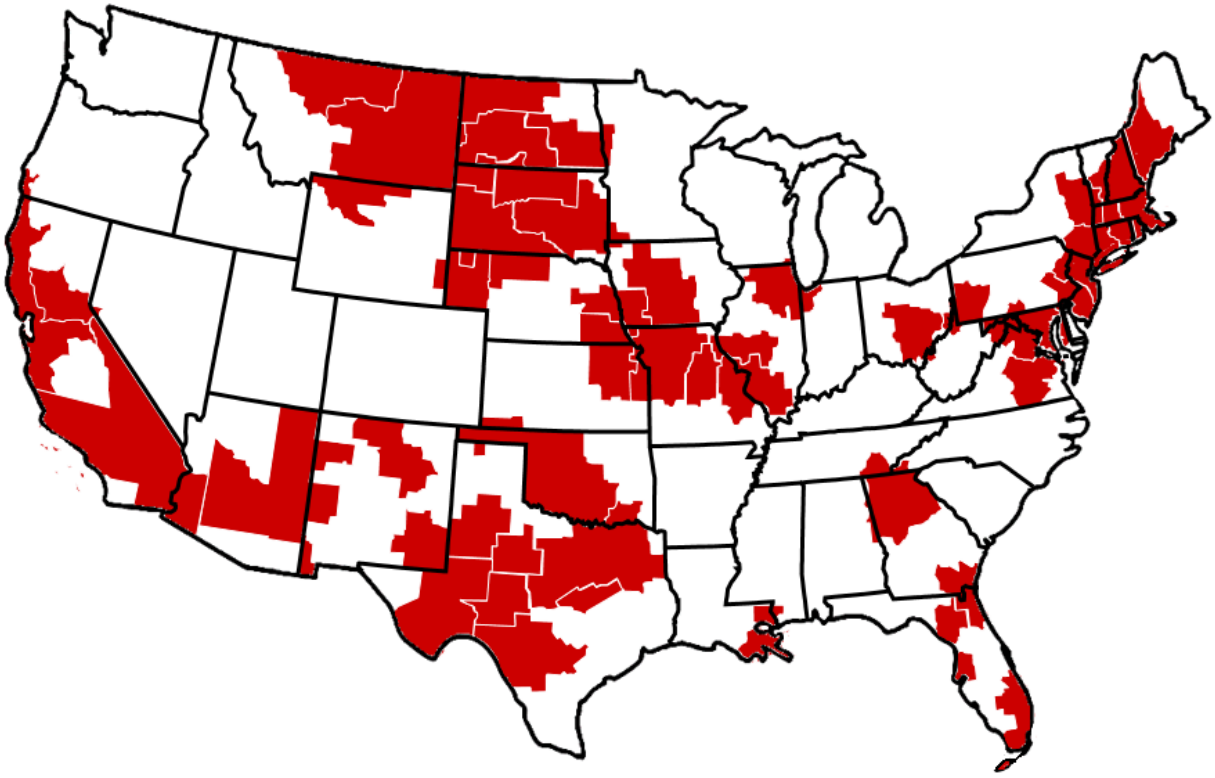
(a) Cluster 1 (b) Cluster 2



Comparing Figures 5(a) and 6(a) shows that for cluster one the geographical concentrations of regions changes noticeably after the probabilities are updated with observed regional employment growth. The same is true for the second cluster.

The probability thresholds for designating a BEA economic area to a cluster are based on the updated probabilities in Figure 6. Regions with a higher than 0.5 probability of being affiliated with a cluster are deemed to be strongly affiliated to that grouping, while

Figure 7: Geographical concentrations in Cluster 1

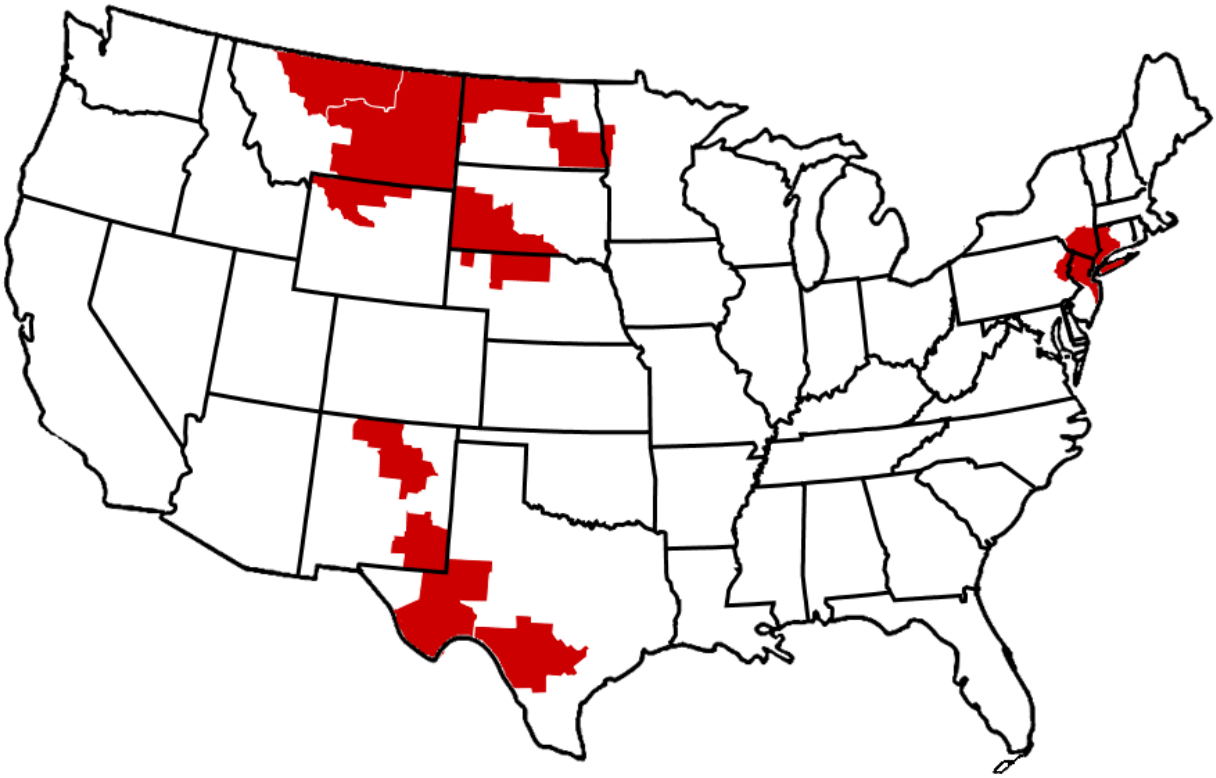


regions with probabilities between 0.25 and 0.50 are deemed to be weakly affiliated. The geographical concentrations based on the strong affiliation threshold are shown in Figures 7 and 8.

Having spatially identified the regional composition of each cluster, the analysis turns to the parameters in the clustering mechanism that influence the probability that a region belongs to a given cluster. These parameters are logistic coefficients. Their estimates provide information regarding which regional characteristics play an important role in designating any given economic area to specific cluster groupings.

The logistic coefficients for each cluster do not have a direct magnitude interpretation by themselves. Therefore, the analysis is supplemented with discrete derivatives of the cluster affiliation probabilities using the estimated logistic coefficients. These derivatives are difference quotients specifically defined to answer which regional characteristics are designating

Figure 8: Geographical concentrations in Cluster 2



small regional divisions into which clusters.

Let a discrete derivative for the industrial characteristic i in cluster k be denoted as δ_{ki}. This value will be implied by the estimated logistic coefficients, and its purpose is to quantify the magnitude by which the cluster affiliation probability differs between any regions in the economy. Specifically, δ_{ki} is calculated for two hypothetical regions that differ only with regards to a single characteristic, with region q having that characteristic one standard deviation below the national average and region s having that characteristic one standard deviation above the national average. The other characteristics of each region are equal to the national average.

Let the average value of characteristic i for all 177 regions including region j, be denoted by $\bar{x}_i$, and the standard deviation denoted as s_{x_i}. Then the vector of industrial characteristics

(including a constant term) for region q is given by

$$\mathbf{x}_q' = \begin{bmatrix} 1 & \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_i - s_{x_i} & \dots & \bar{x}_{P_k-1} & \bar{x}_{P_k} \end{bmatrix},$$

and the vector of industrial characteristics for region s is given by

$$\mathbf{x}_s' = \begin{bmatrix} 1 & \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_i + s_{x_i} & \dots & \bar{x}_{P_k-1} & \bar{x}_{P_k} \end{bmatrix}.$$

Given the posterior means of the logistic coefficients for cluster k,

$$\hat{\boldsymbol{\beta}}_k = \begin{bmatrix} 1 & \hat{\beta}_{k1} & \hat{\beta}_{k2} & \dots & \hat{\beta}_{ki} & \dots & \hat{\beta}_{kP_k-1} & \hat{\beta}_{kP_k} \end{bmatrix}',$$

δ_{ki} is calculated as the change in the cluster affiliation probability between regions s and q, which for each region n is given by $\Pr(h_{nk} = 1|\hat{\beta}_k, \mathbf{x}_n)$ computed from

$$\Pr(h_{nk} = i|\hat{\boldsymbol{\beta}}_k, \mathbf{x}_n) = \begin{cases} \frac{1}{1+\exp(\mathbf{x}_n'\hat{\boldsymbol{\beta}}_k)} & \text{if } i = 0 \\ \frac{\exp(\mathbf{x}_n'\hat{\boldsymbol{\beta}}_k)}{1+\exp(\mathbf{x}_n'\hat{\boldsymbol{\beta}}_k)} & \text{if } i = 1 \end{cases}$$

where the same industrial characteristics are specified for each cluster k. Therefore, for any given cluster and any given characteristic i the quantity of interest is the following difference quotient

$$\begin{aligned} \delta_{ki} &= \Pr(h_{sk} = 1|\hat{\boldsymbol{\beta}}_k, \mathbf{x}_s) - \Pr(h_{qk} = 1|\hat{\boldsymbol{\beta}}_k, \mathbf{x}_q) \\ &= \frac{\exp(\mathbf{x}_s'\hat{\boldsymbol{\beta}}_k)}{1 + \exp(\mathbf{x}_s'\hat{\boldsymbol{\beta}}_k)} - \frac{\exp(\mathbf{x}_q'\hat{\boldsymbol{\beta}}_k)}{1 + \exp(\mathbf{x}_q'\hat{\boldsymbol{\beta}}_k)} \end{aligned} \tag{29}$$

A positive (negative) value of δ_{ki} implies that regions with higher concentrations of employment in sector i and close to average employment shares in other sectors are likely (not likely) to be affiliated with cluster k.

Table 7: Estimated logistic coefficients (posterior means $\hat{\beta}_{ki}$) and discrete derivatives δ_{ki}

	Cluster 1		Cluster 2	
	$\hat{\beta}_{1i}$	δ_{1i}	$\hat{\beta}_{2i}$	δ_{2i}
Constant	0.055		0.019	
	(53)		(51)	
Manufacturing	−0.208*	−0.501*	−0.115*	−0.087*
	(81)		(79)	
Finance & insurance	0.443*	0.219*	0.104	0.015
	(83)		(58)	
Mining & quarrying	−0.097	−0.021	−0.067	−0.004
	(55)		(53)	
Oil & gas extraction	0.138	0.019	0.220	0.008
	(58)		(62)	
Small firms	0.031	0.089	−0.011	−0.009
	(57)		(43)	
Construction	−0.215*	−0.124*	−0.162	−0.027
	(70)		(62)	

Notes: The percentage of posterior draws on the same side of zero as the posterior mean (sign certainty probability) are given in parentheses. $*$ – indicates that at least 68 percent of the posterior draws were on the same side of zero as the reported posterior mean. Posterior means computed for MCMC sequence based on $N^{burn-in} = 250000$ burn-in iterations and $N^{keep} = 25000$ samples retained for inference.

Table 7 presents the estimated logistic coefficients and the implied discrete derivatives. The financial sector has the highest positive estimated logistic coefficient (0.443) for cluster 1 and the oil and gas extraction sector has the highest positive estimated logistic coefficient (0.220) for cluster 2, albeit with a lower sign certainty probability of 0.62. The evidence in Table 7 suggests that the manufacturing, financial and construction sectors are significant industrial characteristics for cluster 1. Based on the current definition of the discrete derivatives, the δ_{ki} magnitudes are the most pronounced for this grouping. The probability of belonging to cluster 1 is substantially higher for a region that has a higher concentration of labor in the financial sector ($\delta_{12} = 0.219$). While the probability of belonging to cluster 1 is substantially lower for regions with higher concentrations of labor in manufacturing ($\delta_{11} = -0.501$) and construction ($\delta_{16} = -0.124$). The magnitudes for cluster 2 are not as pronounced, because the current definition of these quotients is not adequately characteriz-

ing the distinctions between regions in the data. This cluster is highly resemblant of the oil and gas producing regions, which is evident when the labor concentrations in the oil and gas sector in Figure 3 are compared to the geographical concentrations in cluster 2 in Figure 8. The evidence in Table 7 confirm the importance of the oil & gas extraction covariate, as it has the highest sign certainty probability (0.62) among all positive logistic coefficients and the highest positive estimated logistic coefficient ($\hat{\beta}_{24} = 0.220$).

5.3 Timing of business cycle phases

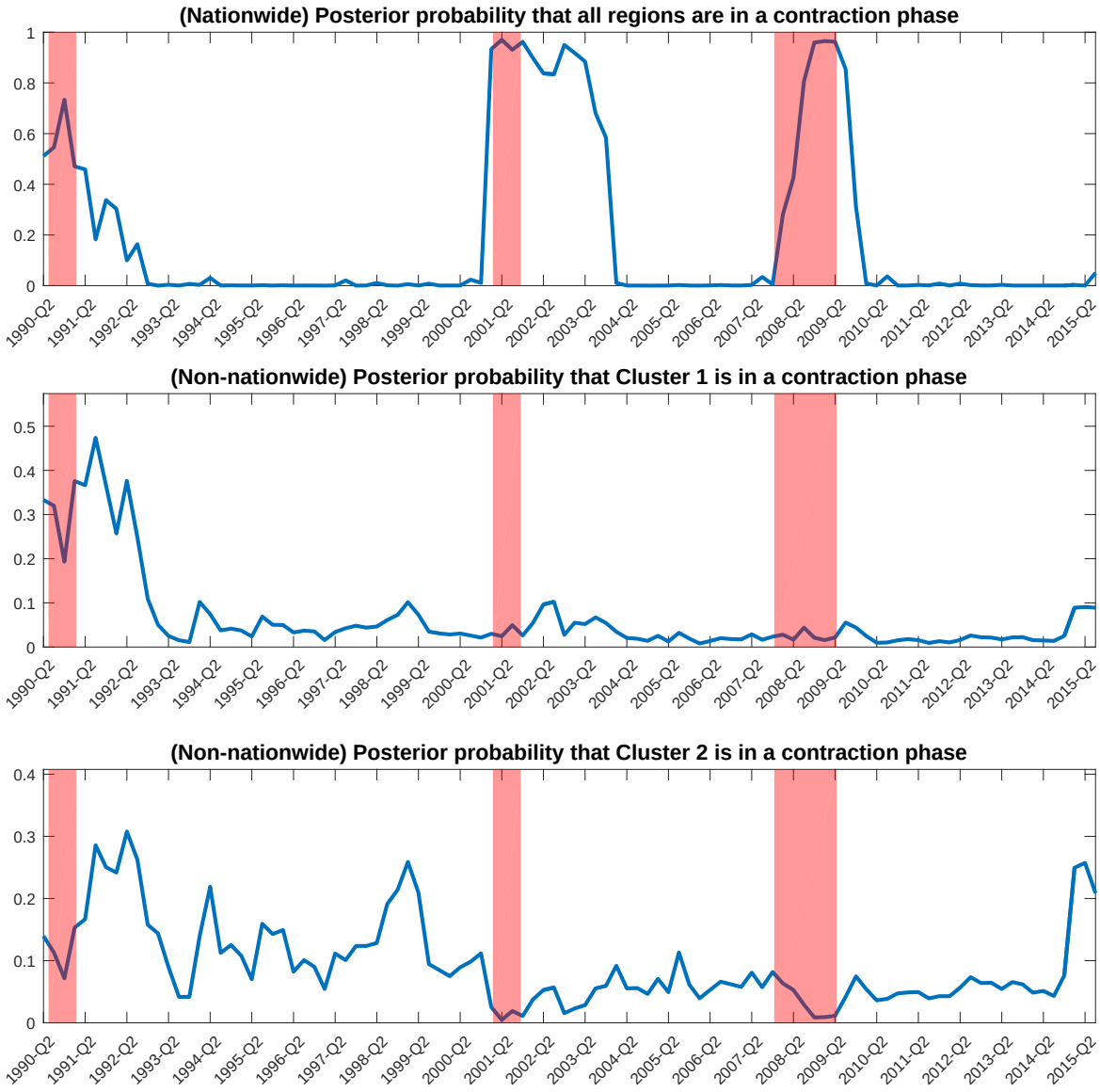
The analysis now turns to the timing of specific business cycle phases for the nationwide grouping and the idiosyncratic (non-nationwide) clusters. Figure 9 presents the time series of the aggregate regime indicator z_t, given by the posterior mean probabilities that regimes associated with contraction phases, $z_t = 1, 2, 3$, are active at time t. $z_t = 3$ is the *a priori* nationwide grouping designating that every single 177 economic area unit is in a state of contraction at time t. It is important to note that this is a much stronger definition of a negative national growth phase compared to a negative growth phase at higher levels of aggregation (e.g. state-level or national). Therefore, whenever the probability of this regime being active is high, the nationwide economic downturn is deemed to have propagated throughout most relevant markets surrounding metropolitan areas in the United States. The reported posterior probabilities of this grouping being active during the the early 2000's recession and the Great Recession following the financial crisis of 2008 are very intuitive. Based on employment data for 177 small regional divisions, the top panel of Figure 9 shows that the early 2000's recession was quite severe and persistent, lasting considerably longer than what is given by the NBER recession dates. This result is consistent with what was observed using state-level employment in Hamilton and Owyang (2012). The results for the period spanning the Great Recession show a distinct propagation dynamic. This is characterized by the fact that the probability that all regions are in an active contraction

phase at time t increases towards one monotonically from the second quarter of 2007. This observation illustrates that in the earlier stages of the crisis, around 2007-2008, the observed employment growth patterns where not sufficient to designate a nationwide contraction in all 177 areas, but as the crisis progressed it's propagation to other regions in the country drove employment growth substantially downward across the country to designate a very high probability of the recession affecting all areas by the end of 2008.

The contraction phase timings for the idiosyncratic clusters, based on posterior probabilities, are not as distinct. They provide information regarding which periods were the most likely to see either cluster active. The second panel of Figure 9 shows the first cluster having the highest probability of being active following the early 1990's recession. This is the cluster for which regions with high concentrations in the finance and insurance sector were most likely to be designated, given by a positive influence on a the cluster affiliation probability. The last panel of Figure 9 shows the regime indicator for the second cluster. The probabilities of being active for the period 1990-2015 are lower for this cluster, which has the highest probability of entering a phase of contraction in-between the early 1990's and early 2000's national recessions, and in the second and third quarters of 2015. The evolution of the regime indicator for this cluster provides only weak evidence of regions with a high probability of belonging to this cluster as being at-risk to negative co-movements in employment growth patterns.

The second instrument for understanding the business cycle phases are the estimates associated with the mechanism of dynamic change in the model. The estimated transition matrix is given in Table 8. The top-left two-by-two block shows the transition probabilities between nationwide states of contraction and expansion. These results are consistent with the results in Hamilton (1989, 1994) and Hamilton and Owyang (2012), all of which accurately characterize the observed fact that expansionary phases ($p_{11} = 0.86$) in the economy are more persistent than phases of contraction ($p_{22} = 0.72$), with contractionary (expansionary)

Figure 9: Posterior probabilities of aggregate regime indicator z_t



Notes: Shaded regions indicate NBER recessions. Posterior means computed for MCMC sequence based on $N^{burn-in} = 250000$ burn-in iterations and $N^{keep} = 25000$ samples retained for inference.

Table 8: Estimated regime transition probabilities (posterior means)

	From nationwide expansion	From nationwide contraction	From cluster 1 contraction	From cluster 2 contraction
To nationwide expansion	0.86	0.10	0.25	0.40
To nationwide contraction	0.03	0.72	0.35	0.21
To cluster 1 contraction	0.03	0.10	0.40	**0**
To cluster 2 contraction	0.08	0.08	**0**	0.39

Notes: Bold zeros are transition probabilities that are restricted to take on zero values to ensure that the MCMC sequence does not switch to a specification with a reverse order of clusters under which the likelihood would be unchanged.

phases more (less) likely to be followed by a phase of expansion (contraction). The last two columns of Table 8 show that the idiosyncratic clusters one and two are both equally persistent ($p_{33} = 0.40$ and $p_{44} = 0.39$, respectively), with cluster one being much more likely to be followed by a nation-wide phase of contraction than cluster two ($p_{23} = 0.35$ and $p_{24} = 0.21$, respectively).

The estimated transition probabilities generate a measure of duration of all regimes. The expected recession durations are reported in Table 9. Over the period 1990-2015 nationwide expansions are expected to last an average of 7.14 quarters, while nationwide phases of contraction last an average 3.57 quarters. Contraction phases in clusters one and two last an average of 1.67 and 1.64 quarters, respectively. If either of these clusters are active and transition into a nationwide phase of expansion, then these expected durations fall short of the formal recession definition of two consecutive quarters of negative growth, which in the empirical context is not based on state or national GDP growth but on employment growth in small regional divisions. However, the primary interest lies in estimating the probability that phases of contraction in the clusters are followed by a nationwide contraction, which would capture the propagation of a regional economic downturn. The probability that a cluster transitions from a phase of contraction into a nationwide contraction is 0.35 and 0.21 for cluster one and two, respectively.

Table 9: Expected duration of expansion and contraction phases (in quarters)

	Nationwide expansion	Nationwide contraction	Cluster 1 in contraction	Cluster 2 in contraction
Average number of quarters	7.14	3.57	1.67	1.64

Notes: Average length of each regime as implied by the estimated regime transition probabilities (posterior means).

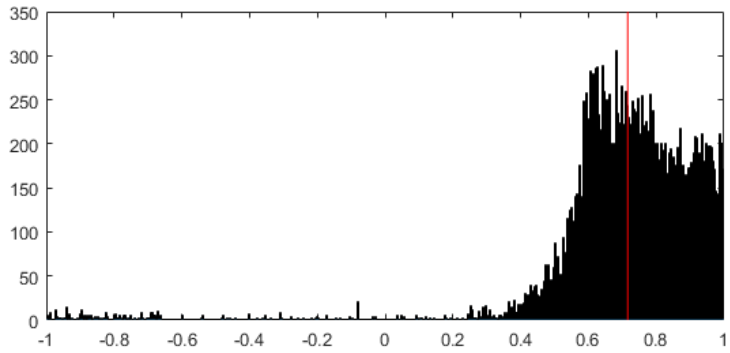
5.4 Spatial spillovers and robustness

The observed employment growth rate data is tested for evidence of spatial autocorrelation using five different spatial weighting structures; see Appendix A.2, Figures 12 and 13. For each spatial weighting, the vector of employment growth rates, $\boldsymbol{y}_t$, is tested for zero spatial autocorrelation using Moran's (1950) I Index and test. Overall, the results strongly reject the null hypothesis of no spatial autocorrelation in the data across all spatial weightings. Furthermore, testing indicates positive spatial autocorrelation. This is given by the fact that Moran's I Index is generally significantly positive, meaning that values in the dataset tend to cluster spatially (high values cluster near other high values; low values cluster near other low values). This supports the application of the proposed spatial model for this data set.

The degree of spatial interactions captured by the SAE component is measured by ρ, which alleviates the restrictive assumption of a diagonal variance-covariance matrix for the error vector. The posterior draws of the spatial parameter ρ are shown in Figure 10. Inference is drawn on the entire surface of the posterior distribution for which the draws of ρ are summarized by central tendency measures: mean, median and mode of 0.72, 0.73 and 0.68, respectively. The spatial parameter in the spatial autoregressive error (SAE) component is interpreted as capturing the overall degree of spatial correlation between the unobserved regional shocks. With a contiguity spatial weighting a positive value of ρ indicates that shocks are expected to be higher if, on average, shocks to neighboring regions are high. The strong positive degree of spatial dependence implies that for the period 1990–2015 there were substantial spatial spillovers between regions.

Figure 10: Posterior draws of ρ
$\hat{\rho} = 0.72$ (red line)

(a) Histogram



(b) Kernel density estimate

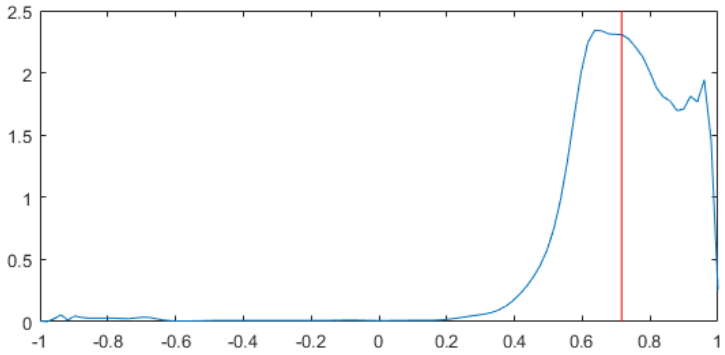


Figure 10 provides evidence that the SAE component in the model is capturing information that the restricted (*i.e.* non-spatial) model overlooks. The posterior distribution has a very high sign certainty probability of 99 percent, meaning that almost all posterior draws of ρ lie on the same side of zero as the mean. This strongly rejects the hypothesis of $\rho = 0$, ruling in favor of the spatial model over the nested restricted model. This is in agreement with Moran's test and the cross-validation scores, which also favor the spatial specification.

Recall that the spatial weighting structure, given by the matrix $\mathbf{W}$, is defined based on shared physical borders between regions. For this weighting structure the evidence identifies a strong positive degree of spatial correlation. The full model is estimated for four other weighting specifications of $\mathbf{W}$ to assess how robust the finding of a strong positive spatial

Table 10: Spatial spillovers spanning the period 1990–2015

	$\mathbf{W}_{\text{contig}}$	$\mathbf{W}_{400\text{km}}$	$\mathbf{W}_{600\text{km}}$	$\mathbf{W}_{800\text{km}}$	$\mathbf{W}_{\text{inv-dist}}$
$\hat{\rho}$	0.72 (0.47,0.97)	0.77 (0.58, 0.97)	0.85 (0.72,0.98)	0.89 (0.78,0.99)	0.94 (0.88,0.99)
Number of weights (max = 31152)	912	1882	3926	6552	31152

Notes: The posterior means of ρ are obtained by estimating the full model with two ($\kappa = 2$) idiosyncratic clusters on different spatial weighting matrices $\mathbf{W}$. $\mathbf{W}_{\text{contig}}$ is the contiguity weighting, $\mathbf{W}_{400\text{km}}$, $\mathbf{W}_{600\text{km}}$ and $\mathbf{W}_{800\text{km}}$ are 400, 600 and 800 kilometer distance band weightings, and $\mathbf{W}_{\text{inv-dist}}$ is the symmetric inverse distance weighting. The 90 percent equal-tailed credible intervals are given in parentheses.

interaction is to various spatial weights. The robustness results are given in Table 10. The estimates of ρ for each spatial weighting are given in an increasing order of the number of weights defined in each $\mathbf{W}$ matrix for the 177 geographical units. The most sparse matrix is the contiguity weighting, $\mathbf{W}_{\text{contig}}$, with 912 non-zero weights. The least sparse matrix is the symmetric inverse distance weighting, $\mathbf{W}_{\text{inv-dist}}$, with 31152 non-zero weights, which is the maximum number of weights allowed in any given specification. The results show that the finding of a strong positive degree of spatial correlation is robust to various spatial weighting measures. Furthermore, $\hat{\rho}$ increases as more connections are specified between regions.

For the main model specification, the posterior mean (0.72) of the spatial parameter for the period 1990-2015 is comparable to the findings of recent empirical work analyzing the United States economy. Fogli et al. (2015), who use a non-regime-switching spatial autoregressive lag (SAL) model to analyze county-level unemployment and housing price data in the United States, find that unemployment rates are spatially dispersed and spatially correlated. They estimate the degree of correlation to vary between 0.46–0.64 over the early 1990's recession, 0.58–0.82 over the early 2000's recession and 0.63–0.83 over the Great Recession. Where their approach concentrates on the evolution of the spatial correlation parameter through time, with national recession dates exogenous to the model, the spatial model in this paper captures a non-time varying degree of spatial correlation in a framework that endogenously identifies business cycle phases and regional clusters. To draw comparable

inferences estimating the model for different subsamples of the data provides a time varying assessment of spatial correlation. Table 11 shows the varying degrees of spatial spillovers for disjoint subsets of the data spanning each of the three observed national recessions for the period 1990-2015. Focusing on the periods spanning only the national recessions, using either NBER recession dates or the nationwide contraction periods identified in the model, provides insufficient data to obtain meaningful estimates in a regime-switching model. Therefore, the full data set for the period 1990–2015 is split into subsamples of approximately equal length. The results show that the degree of spatial correlation is substantially higher (0.82) for the period spanning and following the Great Recession, compared to 0.71 and 0.70 for 1990–1999 and 2000–2006, respectively. The stronger spatial interactions between regional shocks for the period 2007–2015 suggest that the importance of geographical proximity between regional markets was amplified during the crisis and in the period that followed. This amplification is robust to other spatial weightings.

5.5 Policy implications

This section discusses some of the policy implications based on the empirical findings. Suppose you are a regional policy maker or researcher interested in a specific United States county or small regional area (e.g. BEA Economic Area 146 San Jose-San Francisco-Oakland, CA). You need to know if your region is likely to become economically at-risk or potentially distressed separately from the national economy, and to do so you require an informative assessment of any synchronicities (*i.e.* co-movements) with other regions in the country regarding how your small region's economy has evolved over the last several decades. Furthermore, you have existing knowledge regarding several types of connections to other regions that you know are important for your local economy, and you wish to explore and compare them through time. Insights into these dynamics are provided by the spatial and clustering components in the model, which provide evidence for several policy-relevant questions.

Table 11: Spatial spillovers spanning the national recessions for the period 1990-2015

Weighting Structure	Number of weights (max = 31152)	1990–1999	2000–2006	2007–2015
$\mathbf{W}_{\text{contig}}$	912	0.70 (0.44,0.97)	0.70 (0.42,0.96)	0.82 (0.66,0.98)
$\mathbf{W}_{\text{400km}}$	1882	0.75 (0.53,0.97)	0.72 (0.47,0.97)	0.85 (0.71,0.98)
$\mathbf{W}_{\text{600km}}$	3926	0.85 (0.71,0.98)	0.81 (0.64,0.98)	0.90 (0.82,0.99)
$\mathbf{W}_{\text{800km}}$	6552	0.89 (0.79,0.98)	0.86 (0.73,0.98)	0.93 (0.87,0.99)
$\mathbf{W}_{\text{inv-dist}}$	31152	0.94 (0.88,0.99)	0.93 (0.86,0.99)	0.97 (0.93,0.99)

Notes: The posterior means of ρ are obtained by estimating the full model with two ($\kappa = 2$) idiosyncratic clusters on three subsets of the full data set: 1990q2–1999q4, 2000q1–2006q4 and 2007q1–2015q3. The same subsets are used to estimate the model with different spatial weighting matrices $\mathbf{W}$. $\mathbf{W}_{\text{contig}}$ is the contiguity weighting, $\mathbf{W}_{\text{400km}}$, $\mathbf{W}_{\text{600km}}$ and $\mathbf{W}_{\text{800km}}$ are 400, 600 and 800 kilometer distance band weightings, and $\mathbf{W}_{\text{inv-dist}}$ is the symmetric inverse distance weighting.The 90 percent equal-tailed credible intervals are given in parentheses.

For example, (Q1) "Have there been any geographical concentrations (or clusters) of small regions in the United States, over the last several decades, that have been impacted by recessions in similar ways, which have included the counties surrounding the San Jose-San Francisco-Oakland area?", (Q2) "What geographical or economic factors receive the highest degree of spatial interaction over that period? Is it shared physical borders with neighboring regions? Is it connections beyond immediate neighboring regions?", and (Q3) "Over the last several decades, when was the degree of spatial spillovers between regions the highest?" For exposition the discussion will use BEA Economic Area 146 as a working example.

The degree of spatial interactions is found to be strong and positive for neighboring regions over the period spanning the last three national recessions (see Section 5.4 for details). The degree of spatial spillovers was highest for the period 2007–2015 ($\hat{\rho} = 0.82$) than for

the periods spanning the early 1990's and early 2000's national recessions, where $\hat{\rho} = 0.70$ for both periods. This is important for policy analysis because it identifies that during the Great Recession and in its aftermath, the integration of the United States economy played a more important role. Specifically, the impact of regional shocks on neighboring regions during this period are expected to be higher than in previous decades. This speaks to the growing importance of accounting for neighboring regions in any analysis or modeling that focuses on a specific geographical unit such as a county. These results provide answers to policy questions like (Q2) and (Q3) in the previous paragraph.

BEA Economic Area 146 is comprised of 22 counties that define the relevant regional markets surrounding San Jose, San Francisco and Oakland in the state of California. This region is designated to Cluster 1 with a high probability of 0.69; see Figures 11 and 7. It is important to refer to Figures 5(a) and 5(b) to see how this region was designated to the grouping. This region has a low probability of being designated to Cluster 1 based only on the industrial characteristics of that region, see Figures 5(a). The region is designated to Cluster 1 with a high probability when the model updates with observed regional employment growth. Another way to confirm this is to look at the regions characteristics and follow the discrete derivative analysis in Section 5.2. The industrial characteristics of this region are given in Table 12, along with the national averages and values of one standard deviation above the national averages. The numbers in the table confirm that the industrial composition of this region is very comparable to the national average, which given the models estimated logistic coefficients discussed earlier, means that it is not likely that the model would assign this region to Cluster 1 based solely on these covariates. The designation to Cluster 1 is therefore strongly driven by the observed employment growth patterns, see Figures 5(b). This means that co-movements in observed employment growth for the period 1990–2015 are more important for affiliating the counties in BEA Economic Area 146 with other regions in Cluster 1. The estimated average employment growth for this region over periods of national recessions and regional recessions is $\hat{\mu}_{n0} + \hat{\mu}_{n1} = -2.20$. In the context of this region,

Table 12: Industrial Characteristics of BEA EA 146
San Jose-San Francisco-Oakland, CA

	Manufacturing	Finance & insurance	Mining & quarrying	Oil & gas extraction	Small firms	Construction
EA 146	13.86	4.62	0.02	0.00	48.06	5.53
$\bar{x}_i$	15.86	4.32	0.18	0.12	47.31	5.93
$\bar{x}_i + s_{x_i}$	22.71	5.67	0.77	0.49	55.00	7.50

Notes: $\bar{x}_i$ is the national average of covariate i. $\bar{x}_i + s_{x_i}$ is the one standard deviation bound on the national average of covariate i.

the answer for the policy-relevant question (Q1) posed earlier would be: "Yes, counties surrounding San Fransisco, CA belong to a geographical concentration in Cluster 1 that has been impacted by recessions in similar ways in the past several decades."

Identifying geographical areas that are likely to experience collective regional recessions helps determine which regions would stand to benefit from industry-specific economic stimulus to prevent chronic economic distress. The importance of specific industries for each cluster provides guidance for industry-specific analysis. Although industrial composition is represented by averages over the full period analyzed, having identified the composition of each cluster, one can turn to a more in-depth industry/sector analysis to better understand similarities and differences across regions, which is beyond the scope of this paper.

The notion of prolonged economic distress is explained by how each cluster of regions exhibiting common economic downturns transition between business cycle phases over the observed period. The mechanism of dynamic change embodied by the estimated transition probabilities provides a measure of persistence, transition and expected duration for specific business cycle phases. For the first cluster, where high labor concentration in the financial sector and low labor concentrations in construction and manufacturing play an important role, the probability of transitioning into a nationwide state of contraction is highest (0.35). When active, this cluster has an expected duration of 1.67 quarters and if followed by a nationwide contraction the joint expected duration is 5.24 quarters. For the second cluster

the expected duration of a similar occurrence is only slightly lower (5.21 quarters), but the probability of this cluster transitioning into a nationwide economic downturn is much lower (0.21). This evidence suggests that over the period 1990–2015 the regions with a higher probability of being affiliated to the first cluster were more likely group to exhibit prolonged economic distress.

6 Conclusion

This paper has developed a new framework for measuring spatial interactions when estimating macroeconomic regimes and regime shifts. The developed model was used to conduct an empirical investigation of regional business cycles characteristics of the 177 contiguous BEA economic areas in the United States for the period 1990–2015. Investigating small regional economies has provided greater geographical detail for understanding regional contagion. The proposed methodological extension delivers a parsimonious framework for capturing spatial interactions in a multivariate Markov-switching model, an approach that enables the spatial interaction aspect of regional business cycles to be measured. This extension also improves the reliability of inference when investigating a large number of geographical units by better accounting for spatial regional (cross-sectional) dependence. Significant positive spatial spillovers between regional shocks have been identified due to the importance of geographical factors. This regional propagation dynamic would be overlooked if one were to apply the model without the proposed spatial extension developed in this paper. The estimated degree of spatial dependence implies that shocks to small regions are expected to be higher, when shocks to neighboring regions are high on average. The magnitude of this effect, which speaks to the importance of geographical proximity between regional markets, is found to be amplified for the period spanning and following the Great Recession, 2007–2015.

Two groupings of regions that tend to co-move during regional economic downturns

have been endogenously identified through common business cycle and industrial characteristics. The general observation is that the first grouping is driven by regional economies with important roles in the financial sector, while the second grouping is driven by regional economies with important roles in oil and gas extraction. Economic downturns experienced by regions in the first grouping are the most likely to be followed by a nationwide economic downturn. The empirical results also provide a region-by-region assessment for explaining the region-specific characteristics that designate regions to the same grouping. Regions with a high probability of affiliation with either grouping are interpreted as potentially at-risk to collective economic distress.

References

Andrews, D. and C. Mallows (1974). Scale Mixtures of Normal Distributions. *Journal of the Royal Statistical Society: Series B 36*, 99–102.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Springer.

Artis, M., C. Dreger, and K. Kholodilin (2011). What drives regional business cycles? The role of common and spatial components. *The Manchester School 79*, 1035–1044.

Beraja, M., E. Hurst, and J. Ospina (2016). The Aggregate Implications of Regional Business Cycles. *NBER Working Paper No. 21956*.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association 90*, 1313–1321.

Chib, S. (1996). Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models. *Journal of Econometrics 75*, 79–97.

Crone, T. M. (2005). An Alternative Definition of Economic Regions in the United States Based on Similarities in State Business Cycles. *The Review of Economics and Statistics 87*, 617–626.

Del Negro, M. (2002). Asymmetric Shocks among U.S. States. *Journal of International Economics 56*, 273–297.

Fogli, A., E. Hill, and F. Perri (2015). The Geography of the Great Recession/Spatial Business Cycles. *Working Paper*.

Forni, M. and L. Reichlin (2001). Federal policies and local economies: Europe and the US. *European Economic Review 45*, 109–134.

Frühwirth-Schnatter, S. and S. Kaufmann (2008). Model-Based Clustering of Multiple Times Series. *Journal of Business and Economic Statistics 26*, 78–89.

Gelman, A. and D. B. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*, 457–472.

Geweke, J. and M. Keane (2007). Smoothly Mixing Regressions. *Journal of Econometrics 138*, 252–291.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalised Residuals. *Journal of Econometrics 34*, 5–32.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle. *Econometrica 57*, 357–384.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hamilton, J. D. and M. T. Owyang (2012). The Propagation of Regional Recessions. *The Review of Economics and Statistics 94*, 935–947.

Holloway, G., B. Shankar, and S. Rahman (2002). Bayesian spatial probit estimation: a primer and an application to high-yielding variety (HYV) rice adoption. *Agricultural Economics 27*, 383–402.

Holmes, C. C. and L. Held (2006). Bayesian Auxiliary Variable Models for Binary and Multinomial Regression. *Bayesian Analysis 1*, 145–168.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching*. MIT Press.

Kouparitsas, M. A. (1999). Is the United States an Optimal Currency Area? *Chicago Fed Letter 146*, 1–3.

LeSage, J. and K. Pace (2009). *Introduction to Spatial Econometrics*. Taylor and Francis Group, LLC.

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*.

Owyang, M. T., J. Piger, and H. J. Wall (2005). Business Cycle Phases in U.S. States. *The Review of Economics and Statistics 87*, 605–616.

Owyang, M. T., J. Piger, H. J. Wall, and C. H. Wheeler (2008). The economic performance of cities: A Markov-switching approach. *Journal of Urban Economics 64*, 538–550.

Partridge, M. D. and D. S. Rickman (2005). Regional Cyclical Asymmetries in an Optimal Currency Area: An Analysis Using US State Data. *Oxford Economic Papers 57*, 373–397.

Pesaran, M. H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels. *IZA Discussion Papers 1240, Institute for the Study of Labor (IZA)*.

Pesaran, M. H. (2015). Testing Weak Cross-Sectional Dependence in Large Panels. *Econometric Reviews 6*, 1089–1117.

Thorsrud, L. A. (2013). Global and Regional Business Cycles – Shocks and Propagations. *Norges Bank Working Paper*.

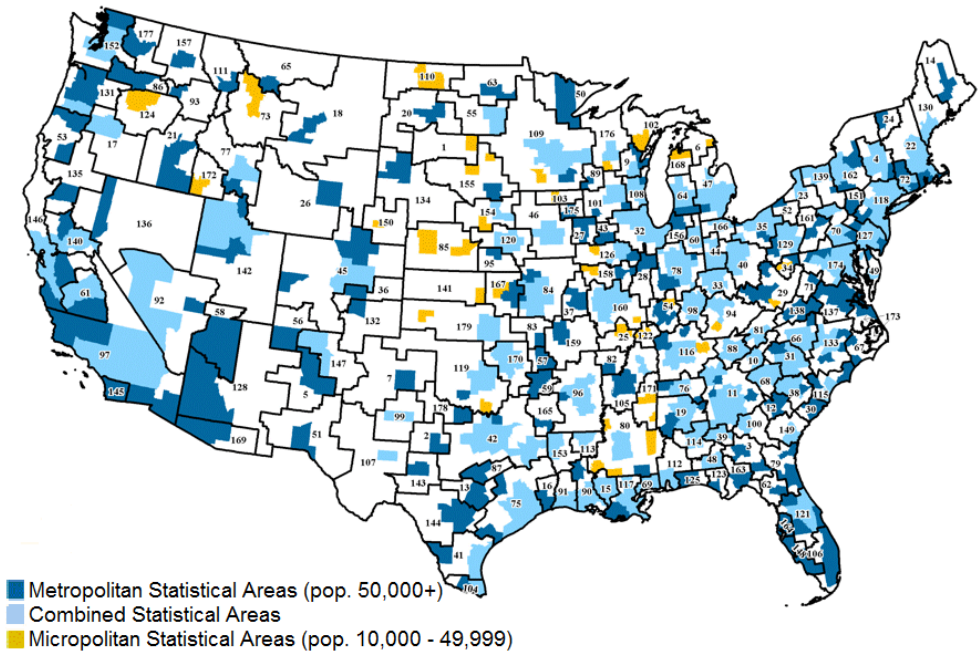
van Dijk, B., P. H. Franses, R. Paap, and D. van Dijk (2007). Modeling Regional House Prices. *Econometric Institute report no. EI 2007-55*.

Appendix A Figures, Tables and Convergence

A.1 BEA Economic Areas

Figure 11: BEA Economic Areas

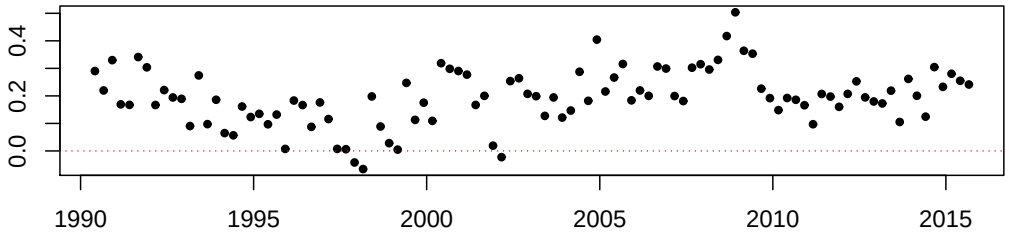
(a) Surrounding statistical areas



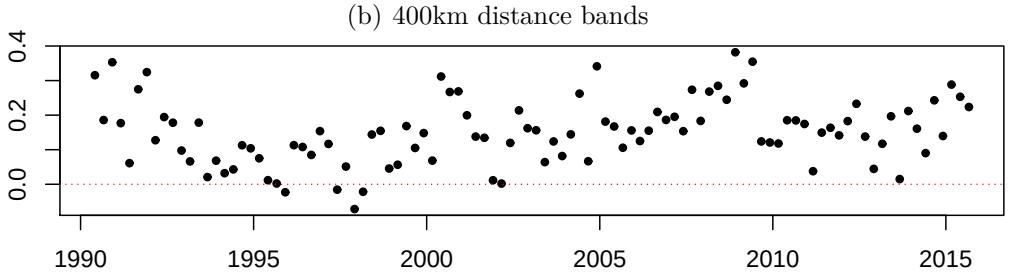
A.2 Testing for spatial correlation

Figure 12: Moran's I Statistic: Positive (negative) values indicate positive (negative) spatial autocorrelation

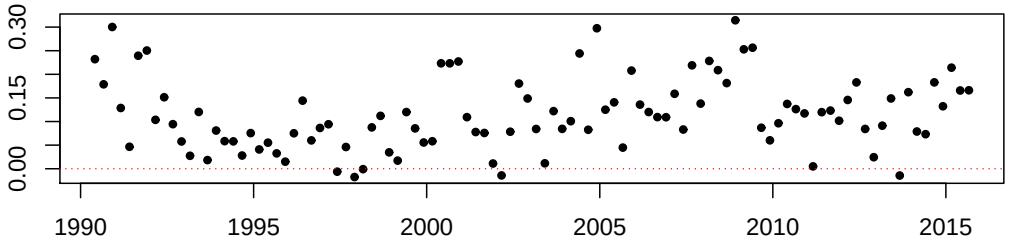
(a) Contiguity



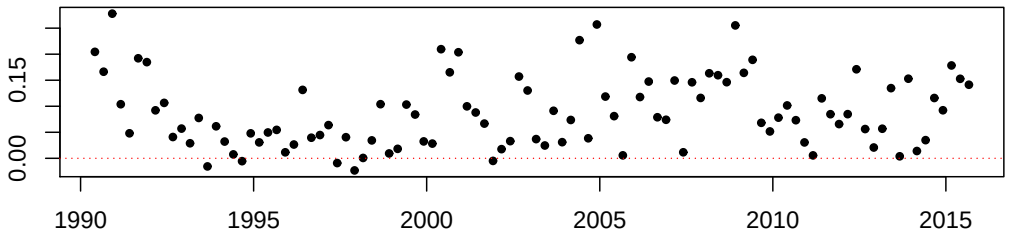
(b) 400km distance bands



(c) 600km distance bands



(d) 800km distance bands



(e) Inverse-distance bands

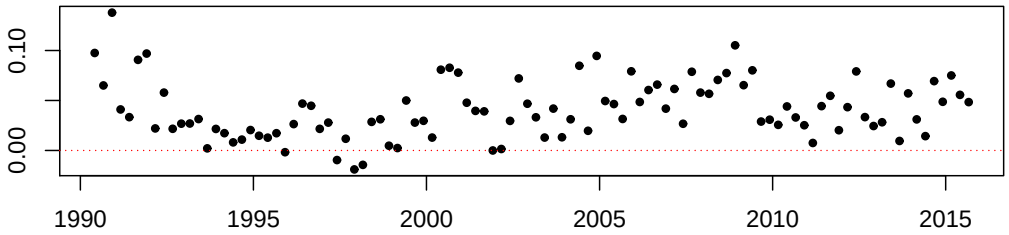
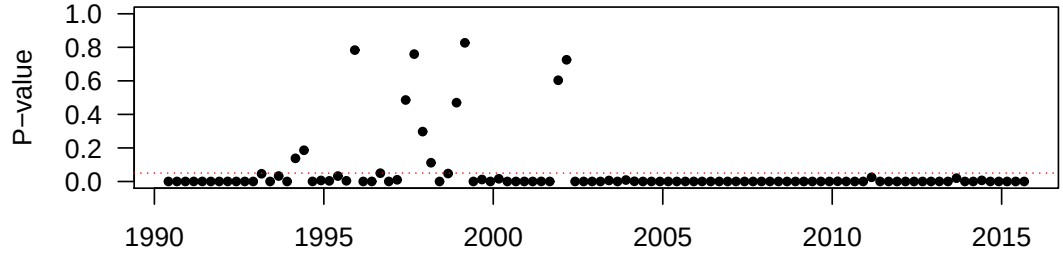
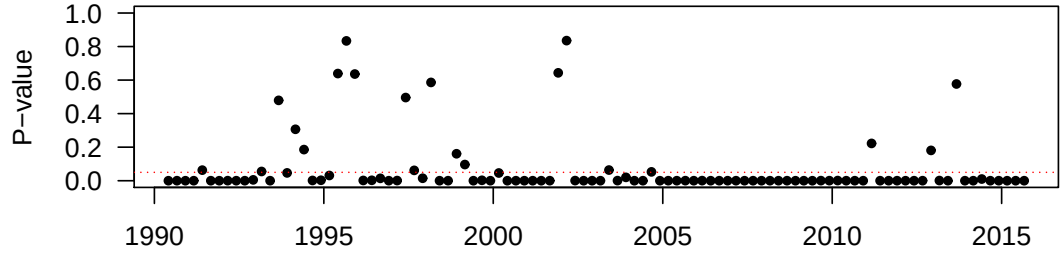


Figure 13: Moran's I Test for Spatial Correlation: P-values (H_0 : no spatial autocorrelation)

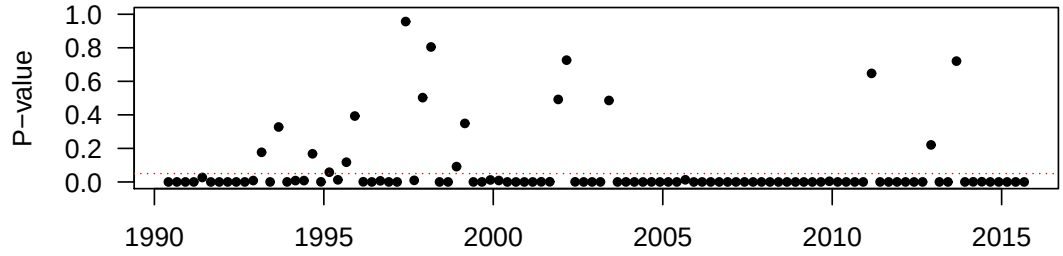
(a) Contiguity



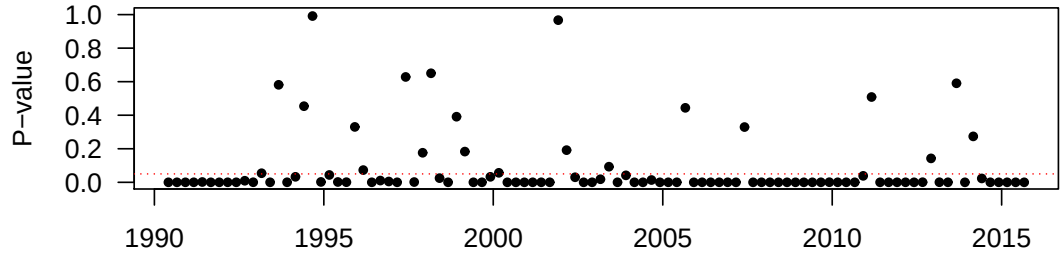
(b) 400km distance bands



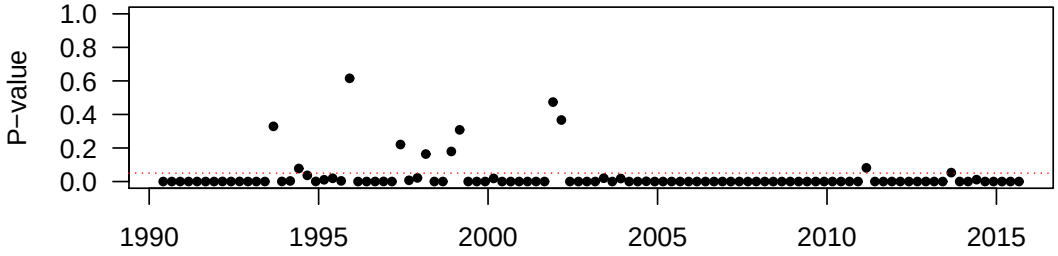
(c) 600km distance bands



(d) 800km distance bands



(e) Inverse-distance bands



A.3 Convergence

This section provides a formal assessment of MCMC output convergence. The numerical diagnostic for chain convergence will follow Gelman and Rubin (1992) (henceforth GR diagnostic). The diagnostic is formed for each individual scalar parameter (θ) and is based on four separate MCMC runs of the full model. The GR measures the potential scale reduction from continuing the MCMC run further. If the chain has converged the GR value for θ should be close to one, indicating that further MCMC draws are not needed to improve our inference about the posterior distribution of θ.

GR is constructed based on within and between chain variability of posterior parameter draws. The between-chain variance for M chains is defined as

$$\text{Var}^b = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\theta}_m - \hat{\theta})^2 \tag{30}$$

where $\hat{\theta}_m$ is the posterior mean of chain m, and $\hat{\theta}$ is the pooled posterior mean. The within-chain variance is the average of the M within-sequence variances and is defined as

$$\text{Var}^w = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{N^{\text{burn-in}}-1}\sum_{n=N^{\text{burn-in}}+1}^{N^{\text{keep}}}(\theta_{mn} - \hat{\theta}_m)^2 \tag{31}$$

where θ_{mn} is the n^{th} posterior draw for chain m. The weighted average of Var^b and Var^w gives an estimate of the target variance,

$$\hat{\sigma}_\theta^2 = \frac{N^{\text{keep}}-1}{N^{\text{keep}}}\text{Var}^w + \frac{1}{N^{\text{keep}}}\text{Var}^b \tag{32}$$

which overestimates the population variance σ_θ^2, while Var^w underestimates σ_θ^2, both converge in expectation as $n \to \infty$.

The GR diagnostic is given by

$$GR = \sqrt{\frac{\hat{V}}{\text{Var}^w} \frac{d}{d-2}}, \tag{33}$$

where $\hat{V}$ is the scale of a Student's t-distribution centered at the pooled posterior mean $\hat{\theta}$ and given by

$$\hat{V} = \hat{\sigma}_\theta^2 + \frac{\text{Var}^b}{MN^{\text{keep}}}, \tag{34}$$

and d are the degrees of freedom estimated as

$$d = \frac{2\hat{V}^2}{\widehat{Var}^{\hat{V}}} \tag{35}$$

Table 13: GR convergence diagnostic results

Parameter Group	Parameter count	All GR < 1.1
$\boldsymbol{\rho}$	1	Y
$\boldsymbol{\mu_0}$	177	Y
$\boldsymbol{\mu_1}$	177	Y
$\boldsymbol{\Omega}$	177	Y
$\mathbf{z}$	408	Y
$\mathbf{P}$	14	Y
$\mathbf{h}_1$	177	Y
$\mathbf{h}_2$	177	Y
$\boldsymbol{\beta_1}$	7	Y
$\boldsymbol{\beta_2}$	7	Y

Notes: GR diagnostic values based on four MCMC runs of $N^{burn-in} = 250000$ burn-in iterations and $N^{keep} = 25000$ posterior draws retained for inference. The GR value is calculated for each individual sequence of every single parameter separately.

The results in Table 13 show that all 1322 parameters converge, indicating that the target posterior distribution based on the retained MCMC draws are reliable for inference and there is very little potential scale reduction from continuing the MCMC algorithm.

A.4 Other parameter estimates

Table 14: Regional average employment growth during expansions
$\kappa = 2$, estimated coefficients μ_0 (posterior means)

Region	μ_0	Region	μ_0	Region	μ_0	Region	μ_0
1.	0.25	47.	0.45	93.	0.31	138.	0.40
2.	0.33	48.	0.14	94.	0.59	139.	0.28
3.	0.33	49.	1.02	95.	0.79	140.	0.86
4.	0.18	50.	0.41	96.	0.22	141.	0.18
5.	0.42	51.	0.49	97.	0.26	142.	1.49
6.	0.60	52.	0.40	98.	0.70	143.	0.19
7.	0.58	53.	0.45	99.	0.53	144.	1.05
9.	0.98	54.	0.20	100.	0.22	145.	1.02
10.	0.77	55.	1.06	101.	0.85	146.	0.91
11.	1.57	56.	0.92	102.	0.22	147.	0.58
12.	0.16	57.	2.34	103.	0.17	148.	1.26
13.	2.76	58.	0.49	104.	1.70	149.	1.14
14.	0.14	59.	0.33	105.	0.49	150.	0.30
15.	0.99	60.	0.36	106.	0.43	151.	0.27
16.	0.30	61.	1.03	107.	1.60	152.	0.90
17.	2.11	62.	0.74	108.	0.18	153.	0.30
18.	0.43	63.	0.21	109.	0.80	154.	0.43
19.	0.34	64.	0.86	110.	2.36	155.	1.26
20.	1.09	65.	0.22	111.	1.41	156.	0.58
21.	1.84	66.	0.18	112.	0.82	157.	0.90
22.	0.80	67.	0.41	113.	0.45	158.	0.15
23.	0.12	68.	0.54	114.	0.23	159.	1.02
24.	1.12	69.	1.11	115.	0.85	160.	0.27
25.	0.33	70.	0.56	116.	1.34	161.	0.26
26.	0.75	71.	0.67	117.	0.66	162.	0.06
27.	1.03	72.	0.16	118.	0.75	163.	0.92
28.	0.29	73.	1.58	119.	0.82	164.	0.46
29.	0.12	75.	1.25	120.	0.75	165.	0.19
30.	1.35	76.	0.29	121.	1.20	166.	0.34
31.	1.08	77.	0.37	122.	0.52	167.	0.30
32.	0.47	78.	0.70	123.	0.94	168.	0.98
33.	0.51	79.	1.07	124.	0.26	169.	0.59
34.	0.99	80.	0.18	125.	1.12	170.	0.51
35.	0.15	81.	0.17	126.	0.41	171.	0.25
36.	1.73	82.	0.47	127.	0.17	172.	0.87
37.	0.88	83.	0.26	128.	1.75	173.	0.27
38.	0.52	84.	0.66	129.	0.15	174.	0.82
39.	0.69	85.	0.42	130.	0.94	175.	0.42
40.	1.01	86.	1.03	131.	0.97	176.	0.33
41.	0.50	87.	0.24	132.	0.36	177.	1.17
42.	1.76	88.	0.66	133.	1.14	178.	0.19
43.	0.23	89.	0.90	134.	0.48	179.	0.21
44.	0.10	90.	0.92	135.	0.21		
45.	1.63	91.	0.91	136.	0.39		
46.	0.62	92.	2.61	137.	0.65		

Notes: Region number corresponds to the BEA code associated with that economic area. All posterior draws of μ_0 lie on the same side of zero as their respective posterior means.

Table 15: Regional average employment growth during recessions
$\kappa = 2$, estimated coefficients $\mu_0 + \mu_1$ (posterior means)

Region	μ_1	Region	μ_1	Region	μ_1	Region	μ_1
1.	-0.81	47.	-1.76	93.	-0.82	138.	-1.02
2.	-0.19	48.	-0.52	94.	-0.47	139.	-0.73
3.	-1.28	49.	0.55	95.	0.13	140.	-0.03
4.	-0.13	50.	-1.09	96.	-0.26	141.	-0.29
5.	-0.50	51.	-0.41	97.	-1.43	142.	-0.48
6.	-0.34	52.	-0.85	98.	-0.58	143.	-0.54
7.	0.17	53.	-0.58	99.	-0.35	144.	0.39
9.	0.34	54.	-0.25	100.	-0.81	145.	0.16
10.	-0.46	55.	0.58	101.	0.19	146.	-2.20
11.	-0.71	56.	-0.03	102.	-1.32	147.	0.04
12.	-0.85	57.	1.73	103.	-0.73	148.	0.13
13.	0.07	58.	-1.01	104.	0.81	149.	0.25
14.	-0.48	59.	-1.14	105.	-0.98	150.	-1.04
15.	0.10	60.	-0.87	106.	-1.07	151.	-0.53
16.	-1.02	61.	0.10	107.	-0.53	152.	-1.59
17.	0.16	62.	0.12	108.	-1.15	153.	-0.50
18.	-0.06	63.	-0.27	109.	-0.84	154.	-0.56
19.	-0.23	64.	-0.28	110.	0.25	155.	0.45
20.	0.48	65.	-1.04	111.	0.21	156.	-1.29
21.	-0.14	66.	-1.53	112.	-0.95	157.	-0.73
22.	-1.30	67.	-1.22	113.	-0.58	158.	-1.00
23.	-0.68	68.	-1.51	114.	-0.76	159.	-0.25
24.	-0.04	69.	-0.33	115.	-0.14	160.	-1.07
25.	-0.27	70.	-0.05	116.	0.18	161.	-0.33
26.	-0.10	71.	-0.43	117.	-0.73	162.	-0.72
27.	-0.22	72.	-1.24	118.	-1.03	163.	-0.61
28.	-0.50	73.	0.53	119.	-0.10	164.	-1.42
29.	-0.25	75.	-0.14	120.	-0.53	165.	-0.78
30.	-0.03	76.	-0.47	121.	-0.39	166.	-0.61
31.	-1.08	77.	-0.35	122.	-0.79	167.	-0.43
32.	-1.06	78.	-0.19	123.	-0.30	168.	-0.06
33.	-0.60	79.	-0.79	124.	-0.75	169.	-0.47
34.	0.53	80.	-0.52	125.	-0.16	170.	-1.14
35.	-1.03	81.	-0.55	126.	-0.73	171.	-1.38
36.	-0.34	82.	-0.97	127.	-0.81	172.	0.04
37.	0.29	83.	-0.75	128.	-0.47	173.	-0.07
38.	-0.75	84.	-0.68	129.	-0.55	174.	-0.58
39.	-0.44	85.	0.10	130.	-0.80	175.	-0.34
40.	0.15	86.	-0.02	131.	-1.08	176.	-0.22
41.	-0.21	87.	-0.24	132.	-0.69	177.	-0.41
42.	-0.78	88.	0.18	133.	-0.14	178.	-1.18
43.	-0.81	89.	0.28	134.	0.20	179.	-0.82
44.	-1.36	90.	-0.33	135.	-0.55		
45.	-1.22	91.	-0.50	136.	-0.77		
46.	-0.19	92.	1.20	137.	-0.55		

Notes: Region number corresponds to the BEA code associated with that economic area. All posterior draws of μ_1 lie on the same side of zero as their respective posterior means.

Table 16: $\kappa = 2$, Estimated σ^2 (posterior means)
$N^{burn-in} = 250000$ $N^{keep} = 25000$

Region	σ^2	Region	σ^2	Region	σ^2	Region	μ_1
1.	6.63	47.	17.26	93.	10.38	138.	2.79
2.	6.93	48.	5.48	94.	2.27	139.	2.28
3.	7.55	49.	4.64	95.	3.35	140.	5.01
4.	1.44	50.	4.72	96.	1.77	141.	4.34
5.	4.65	51.	5.98	97.	3.96	142.	2.76
6.	10.21	52.	2.98	98.	2.11	143.	6.45
7.	3.53	53.	4.12	99.	7.10	144.	3.73
9.	3.12	54.	2.26	100.	6.76	145.	4.40
10.	5.71	55.	3.64	101.	2.30	146.	3.78
11.	2.82	56.	10.13	102.	5.11	147.	10.36
12.	6.10	57.	5.56	103.	4.68	148.	14.92
13.	4.60	58.	22.32	104.	9.35	149.	4.17
14.	4.31	59.	6.37	105.	3.46	150.	11.83
15.	7.94	60.	4.57	106.	4.78	151.	4.14
16.	11.87	61.	19.20	107.	16.75	152.	15.47
17.	12.33	62.	9.01	108.	1.89	153.	5.54
18.	5.74	63.	5.52	109.	1.93	154.	3.25
19.	1.82	64.	3.32	110.	33.47	155.	2.05
20.	5.51	65.	11.16	111.	9.45	156.	5.32
21.	11.44	66.	1.89	112.	5.46	157.	12.05
22.	1.17	67.	4.72	113.	7.91	158.	5.26
23.	2.34	68.	3.70	114.	4.59	159.	4.35
24.	3.65	69.	25.45	115.	4.78	160.	1.88
25.	4.40	70.	1.19	116.	2.92	161.	2.86
26.	7.32	71.	4.60	117.	10.79	162.	1.20
27.	4.13	72.	1.73	118.	0.84	163.	21.71
28.	4.22	73.	5.69	119.	2.80	164.	6.46
29.	1.63	75.	2.77	120.	2.93	165.	5.82
30.	6.45	76.	4.09	121.	5.41	166.	2.83
31.	3.41	77.	5.83	122.	5.35	167.	3.41
32.	1.23	78.	1.74	123.	13.90	168.	6.72
33.	2.25	79.	5.54	124.	9.64	169.	13.46
34.	6.00	80.	2.98	125.	8.74	170.	3.80
35.	1.21	81.	5.32	126.	3.99	171.	4.93
36.	4.58	82.	5.84	127.	0.97	172.	7.97
37.	3.06	83.	3.86	128.	6.10	173.	2.87
38.	6.17	84.	1.95	129.	2.35	174.	1.70
39.	5.44	85.	3.58	130.	4.02	175.	5.43
40.	1.69	86.	34.34	131.	3.35	176.	4.21
41.	5.12	87.	6.98	132.	5.05	177.	68.12
42.	1.59	88.	3.22	133.	2.12	178.	6.73
43.	5.04	89.	4.93	134.	3.53	179.	3.17
44.	2.10	90.	8.71	135.	8.16		
45.	3.41	91.	13.50	136.	4.19		
46.	2.41	92.	13.96	137.	2.02		

Notes: Region number corresponds to the BEA code associated with that economic area.

Appendix B Statistical Inference, Estimation Algorithms and Proofs

This appendix provides supplementary information regarding the Bayesian statistical inference; see Section B.1, the MCMC estimation algorithm; see Section B.2, and the complete derivations of the posterior conditional distributions that are impacted by the introduction of the spatial parameter ρ; see Sections B.3, B.4 and B.5.

B.1 Statistical inference

Conditional distribution of transition proability matrix P

Conditional on H and $\mathbf{z}$ the inference is that of a standard K-state Markov switching process. The conditional posterior distribution does not involve ρ. Chib (1996)

$$p(\mathbf{P}|\mathbf{Y}, \rho, \boldsymbol{\mu_n}, \boldsymbol{\Omega}, H) \propto \pi(\boldsymbol{z}|\mathbf{P})\pi(\mathbf{P})$$

The transition probability matrix, $\mathbf{P}$, is obtained by independently drawing every column $\mathbf{P}_p$ from $\pi(\mathbf{P}_p) \sim \mathbf{D}(\boldsymbol{\alpha}_p^*)$ with the q^{th} hyperparameter, $\boldsymbol{\alpha}_{pq}^*$, of vector $\boldsymbol{\alpha}_p^*$ calculated as

$$\boldsymbol{\alpha}_{pq}^* = \frac{\sum_{t=2}^{T} \delta(z_{t-1} = p, z_t = q)}{\sum_{t=2}^{T} \delta(z_{t-1} = p)}, \tag{36}$$

the fraction of times regime p is followed by regime q in the drawn sequence $\{z_1, z_2, \ldots, z_T\}$.

Conditional distributions of unobserved latent variables $\boldsymbol{z}, h, \xi, \lambda$

The aggregate regime indicator, $\boldsymbol{z}$, for which $z_t \in \{1, 2, \ldots, K\}$, designates which regime is in active state at date t. The conditional posterior distribution of $\boldsymbol{z}$ is computed using the two step filter of Hamilton (1994). ρ only enters the forecast error computation (*i.e.* the non-constant portion of the likelihood) Chib (1996). In the first step, the Hamilton Filter is used to obtain the filtered transition probabilities. The second step sequentially draws $z_T, z_{T-1}, \ldots, z_1$ by recursively iterating backwards. This is accomplished by multiplying the filtered probabilities by the forward transition probability.

$$p(\boldsymbol{z}|\mathbf{Y}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, H, \beta) \propto \mathcal{L}(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, h; \mathbf{Y})\pi(\boldsymbol{z}|\mathbf{P}) \tag{37}$$

Let $\mathrm{H} = \{h, \xi, \lambda\}$. For the set $h = \{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{K-2}\}$, each vector $\mathbf{h}_k = (h_{1k} h_{2k} \ldots h_{Nk})'$ determines which regions belong to cluster k. For $\mathbf{h}_k$, introducing ρ affects the likelihood function.

$$\begin{aligned} p(\mathbf{h}_k|\mathbf{Y}, \mathrm{H}^{[k]}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathrm{P}, \beta) &\propto \mathcal{L}(\rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{z}, h; \mathbf{Y})\pi(\mathbf{h}_k|\boldsymbol{\beta}_k) \\ &= \prod_{n=1}^{N} \mathcal{L}(\rho, \mu_{n0}, \mu_{n1}, \sigma_n^{-2}, \boldsymbol{z}, h_{nk}, h^{[k]}; \mathbf{Y}_n)\pi(h_{nk}|\boldsymbol{\beta}_k), \end{aligned} \tag{38}$$

where $\mathrm{H}^{[k]}$ is the set of all elements belonging to the cluster affiliation parameters not associated with cluster k. That is, $\mathrm{H}^{[k]} = \{\mathbf{h}_j, \boldsymbol{\xi}_j, \boldsymbol{\lambda}_j : j = 1, \ldots, K-2; j \neq k\}$ and $h^{[k]} = \{h_{ni} : i = 1, \ldots, K-2; i \neq k\}$.

Each individual h_{nk} indicates the affiliation of region n to cluster k independently across regions - follows from (17) - and is drawn from

$$\begin{aligned} &\Pr(h_{nk} = 1|\mathbf{Y}_n, h^{[k]}, \rho, \boldsymbol{\mu}_n, \sigma_n^{-2}, \mathbf{P}, \boldsymbol{z}, \boldsymbol{\beta}_k) \\ &= \frac{\mathcal{L}(h_{nk} = 1, h^{[k]}, \rho, \boldsymbol{\mu}_n, \sigma_n^{-2}, \boldsymbol{z}; \mathbf{Y}_n)\Pr(h_{nk} = 1|\boldsymbol{\beta}_k)}{\sum_{i=0}^{1} \mathcal{L}(h_{nk} = i, h^{[k]}, \rho, \boldsymbol{\mu}_n, \sigma_n^{-2}, \boldsymbol{z}; \mathbf{Y}_n)\Pr(h_{nk} = i|\boldsymbol{\beta}_k)}, \end{aligned} \tag{39}$$

where $\Pr(h_{nk} = i|\boldsymbol{\beta}_k)$ is computed from

$$\Pr(h_{nk} = i|\boldsymbol{\beta}_k) = \begin{cases} \frac{1}{1+\exp(\mathbf{x}'_{nk}\beta_k)} & \text{if } i = 0 \\ \frac{\exp(\mathbf{x}'_{nk}\beta_k)}{1+\exp(\mathbf{x}'_{nk}\beta_k)} & \text{if } i = 1 \end{cases}. \tag{40}$$

Nothing is affected by the inclusion of ρ for the auxiliary parameters ξ and λ, the outcomes of which influence the generation of h_{nk}. Hamilton and Owyang (2012) introduce ξ and λ to generate h_{nk} following the auxiliary variable approaches to Bayesian binary and multinomial regression of Holmes and Held (2006)[8]. As in Hamilton and Owyang (2012), the conditional posterior distribution of $\boldsymbol{\xi}_k$ is

$$\begin{aligned} p(\boldsymbol{\xi}_k|\mathbf{Y}, \mathbf{h}_k, \mathrm{H}^{[k]}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathrm{P}, \boldsymbol{z}, \beta) &= \pi(\boldsymbol{\xi}_k|\mathbf{h}_k, \boldsymbol{\beta}_k) \\ &= \prod_{n=1}^{N} \pi(\xi_{nk}|h_{nk}, \boldsymbol{\beta}_k). \end{aligned} \tag{41}$$

Each element ξ_{nk} is computed from $\xi_{nk} = \mathbf{x}'_{nk}\boldsymbol{\beta}_k - \log(u_{nk}^{-1} - 1)$, with u^{-1} computed from

$$u^{-1} = \begin{cases} \frac{1}{1+\exp(\mathbf{x}'_{nk}\beta_k)} u^*_{nk} & \text{if } h_{nk} = 0 \\ \frac{\exp(\mathbf{x}'_{nk}\beta_k)}{1+\exp(\mathbf{x}'_{nk}\beta_k)} + \frac{1}{1+\exp(\mathbf{x}'_{nk}\beta_k)} u^*_{nk} & \text{if } h_{nk} = 1 \end{cases}, \tag{42}$$

where u^*_{nk} is drawn from $u \sim \mathrm{U}[0, 1]$.

The conditional posterior distribution of $\boldsymbol{\lambda}_k$ is

$$\begin{aligned} p(\boldsymbol{\lambda}_k|\mathbf{Y}, \boldsymbol{\xi}_k, \mathbf{h}_k, \mathrm{H}^{[k]}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathrm{P}, \boldsymbol{z}, \beta) &= \pi(\boldsymbol{\lambda}_k|\boldsymbol{\xi}_k, \boldsymbol{\beta}_k) \\ &\propto \pi(\boldsymbol{\xi}_k|\mathbf{h}_k, \boldsymbol{\beta}_k)\pi(\boldsymbol{\lambda}_k) \\ &= \prod_{n=1}^{N} \pi(\xi_{nk}|\lambda_{nk}, \boldsymbol{\beta}_k)\pi(\lambda_{nk}). \end{aligned} \tag{43}$$

Following Holmes and Held (2006), each element λ_{nk} is a draw from $\lambda_{nk} \sim \mathrm{GIG}\left(\frac{1}{2}, 1, r^2_{nk}\right)$ where r^2_{nk} is calculated from[9]

$$r_{nk} = \xi_{nk} - \mathbf{x}'_{nk}\beta_k. \tag{44}$$

[8]The feasibility of their approach for logistic regression with auxiliary variables is based on the observations of Andrews and Mallows (1974)

[9]Generalized Inverse Gaussian distribution

Conditional distribution of population parameter β

$$p(\beta|\mathbf{Y}, \rho, \boldsymbol{\mu}, \boldsymbol{\Omega}, \mathbf{P}, \mathbf{z}, \mathrm{H}) = \prod_{k=1}^{\kappa} p(\boldsymbol{\beta}_k|\boldsymbol{\xi}_k, \boldsymbol{\lambda}_k) \tag{45}$$

The conditional posterior distribution is just a standard Normal regression model for each $\boldsymbol{\beta}_k$. ρ has no impact here. $\kappa = K - 2$ is the number of clusters. The conditional posterior of $\boldsymbol{\beta}_k$ is given by

$$\boldsymbol{\beta}_k|\mathbf{Y}, \boldsymbol{\mu}_n, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, \boldsymbol{H} \sim \mathrm{N}\Big(\mathbf{b}_k^*, \mathbf{B}_k^*\Big),$$

with a mean $\mathbf{b}_k^*$ and variance $\mathbf{B}_k^*$

$$\begin{aligned} \mathbf{B}_k^* &= \Big(\mathbf{B}_k^{-1} + \mathbf{X}_k \mathbf{V}_k^{-1} \mathbf{X}_k'\Big)^{-1} \\ \mathbf{b}_k^* &= \mathbf{B}_k^* \Big(\mathbf{B}_k^{-1} \mathbf{b}_k + \mathbf{X}_k \mathbf{V}_k^{-1} \xi_k\Big) \end{aligned} \tag{46}$$

which is a standard Normal regression of the form

$$\begin{aligned} \xi_k &= \mathbf{X}_k \boldsymbol{\beta}_k + \varepsilon_k, \quad \varepsilon_k \sim \mathrm{N}\Big(\mathbf{0}, \mathbf{V}_k\Big) \\ \mathbf{V}_k &= \mathrm{diag}(\lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{Nk}) \end{aligned} \tag{47}$$

B.2 Complete MCMC algorithm

The complete Metropolis-within-Gibbs sampling algorithm, corresponding to the compact outline of the algorithm given in Table 3 is:

Step 0: Initialize all parameters

Step 1: Draw cluster

(a) Draw $\beta_k^{(j+1)}$ from $\beta_k|\mathbf{Y}, \boldsymbol{\mu}_n, \boldsymbol{\Omega}, \mathbf{P}, \boldsymbol{z}, \boldsymbol{H} \sim \mathrm{N}\Big(\mathbf{b}_k^*, \mathbf{B}_k^*\Big)$ with a mean $\mathbf{b}_k^*$ and variance $\mathbf{B}_k^*$ calculated from:

$$\begin{aligned}\mathbf{B}_k^{*(j+1)} &= \left(\mathbf{B}_k^{-1} + \mathbf{X}_k \mathbf{V}_k^{-1(j)} \mathbf{X}_k'\right)^{-1} \\ \mathbf{b}_k^{*(j+1)} &= \mathbf{B}_k^{*(j+1)} \left(\mathbf{B}_k^{-1} \mathbf{b}_k + \mathbf{X}_k \mathbf{V}_k^{-1(j)} \xi_k^{(j)}\right)\end{aligned} \tag{48}$$

which is a standard Normal regression of the form

$$\begin{aligned}\xi_k &= \mathbf{X}_k \boldsymbol{\beta}_k + \varepsilon_k, \quad \varepsilon_k \sim \mathrm{N}\left(\mathbf{0}, \mathbf{V}_k\right) \\ \mathbf{V}_k &= \mathrm{diag}(\lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{Nk})\end{aligned} \tag{49}$$

(b) Draw $h_{nk}^{(j+1)}$, the affiliation of region n to cluster k independently across regions (follows from (17)) from

$$\begin{aligned}&\Pr(h_{nk}^{(j+1)} = 1 | \mathbf{Y}_n, h^{[k]}, \rho^{(j)}, \boldsymbol{\mu}_n^{(j)}, \sigma_n^{-2(j)}, \mathbf{P}^{(j)}, \boldsymbol{z}^{(j)}, \beta_k^{(j+1)}) \\ &= \frac{\mathcal{L}(h_{nk}^{(j+1)} = 1, h^{[k]}, \rho^{(j)}, \boldsymbol{\mu}_n^{(j)}, \sigma_n^{-2(j)}, \boldsymbol{z}^{(j)}; \mathbf{Y}_n) \Pr(h_{nk}^{(j+1)} = 1 | \beta_k^{(j+1)})}{\sum_{i=0}^{1} \mathcal{L}(h_{nk}^{(j+1)} = i, h^{[k]}, \rho^{(j)}, \boldsymbol{\mu}_n^{(j)}, \sigma_n^{-2(j)}, \boldsymbol{z}^{(j)}; \mathbf{Y}_n) \Pr(h_{nk}^{(j+1)} = i | \beta_k^{(j+1)})}\end{aligned} \tag{50}$$

where $\Pr(h_{nk}^{(j+1)} = i | \beta_k^{(j+1)})$ is computed from

$$\Pr(h_{nk}^{(j+1)} = i | \beta_k^{(j+1)}) = \begin{cases} \frac{1}{1+\exp(\mathbf{x}_{nk}' \beta_k^{(j+1)})} & \text{if } i = 0 \\ \frac{\exp(\mathbf{x}_{nk}' \beta_k^{(j+1)})}{1+\exp(\mathbf{x}_{nk}' \beta_k^{(j+1)})} & \text{if } i = 1 \end{cases} \tag{51}$$

and $h^{[k]} = \{h_{ni} : i = 1, \ldots, K-2; i \neq k\}$

(c) Draw $\xi_{nk}^{(j+1)} = \mathbf{x}_{nk}' \beta_k^{(j+1)} - \log(u_{nk}^{-1(j+1)} - 1)$ where $u^{-1(j+1)}$ is computed from

$$u^{-1(j+1)} = \begin{cases} \frac{1}{1+\exp(\mathbf{x}_{nk}' \beta_k^{(j+1)})} u_{nk}^* & \text{if } h_{nk}^{(j+1)} = 0 \\ \frac{\exp(\mathbf{x}_{nk}' \beta_k^{(j+1)})}{1+\exp(\mathbf{x}_{nk}' \beta_k^{(j)})} + \frac{1}{1+\exp(\mathbf{x}_{nk}' \beta_k^{(j+1)})} u_{nk}^* & \text{if } h_{nk}^{(j+1)} = 1 \end{cases} \tag{52}$$

where u_{nk}^* is a draw from $u \sim \mathrm{U}[0,1]$

(d) Draw $\lambda_{nk}^{(j+1)}$ from $\lambda_{nk} \sim \mathrm{GIG}\left(\frac{1}{2}, 1, r_{nk}^2\right)$ where r_{nk}^2 is calculated from[10]:

$$r_{nk} = \xi_{nk}^{(j+1)} - \mathbf{x}_{nk}' \beta_k^{(j+1)} \tag{53}$$

Step 2: Draw $\boldsymbol{\mu}_n^{(j+1)}$ from $\boldsymbol{\mu}_n | \mathbf{Y}, \rho, \sigma_n^{-2}, \mathbf{P}, \boldsymbol{z}, h, \beta \sim \mathrm{N}\left(\mathbf{m}^*, \boldsymbol{\Sigma}_{\boldsymbol{n}}\right)$ with a mean $\mathbf{m}^* = \mathbf{A}^{-1}\mathbf{b}$ and variance $\boldsymbol{\Sigma}_{\boldsymbol{n}} = \mathbf{A}^{-1}$ calculated from:

[10] Generalized Inverse Gaussian distribution

$$\mathbf{A} = \sigma_n^{-2(j)}(1+\rho^{(j)}W_{nn})^2\sum_{t=1}^{T}\mathbf{w}(z_t^{(j)},h^{(j+1)})\mathbf{w}(z_t^{(j)},h^{(j+1)})' + [\sigma_n^{2(j)}\mathbf{M}]^{-1}$$
$$\mathbf{b} = \sigma_n^{-2(j)}\sum_{t=1}^{T}\Big(\mathbf{w}(z_t^{(j)},h^{(j+1)})(1+\rho^{(j)}W_{nn})\Big)\Big(y_{tn}-\rho^{(j)}\sum_{j=1}^{N}W_{nj}y_{tj}\Big) + [\sigma_n^{2(j)}\mathbf{M}]^{-1}\mathbf{m} \tag{54}$$

where $\mathbf{w}(z_t^{(j)},h^{(j+1)}) = [1, h^{(j+1)}_{n,z_t^{(j)}}]'$.

Step 3: Draw $\sigma_n^{-2(j+1)}$ from $\sigma_n^{-2}|\mathbf{Y},\rho,\boldsymbol{\mu}_n,\mathbf{P},\boldsymbol{z},h,\beta \sim \Gamma\left(\frac{\nu+T}{2},\frac{\delta+\hat{\delta}}{2}\right)$ with the hyperparameter:

$$\begin{aligned}\hat{\delta} = \sum_{t=1}^{T}\Big(&y_{tn}-\mu_{n0}^{(j+1)}-\mu_{n1}^{(j+1)}h^{(j+1)}_{n,z_t^{(j)}}-\rho^{(j)}\sum_{i\neq n}W_{nj}(y_{ti}-\mu_{i0}^{(j+1)}-\mu_{i1}^{(j+1)}h^{(j)}_{i,z_t^{(j)}})\\&-\rho^{(j)}W_{nn}(y_{tn}-\mu_{n0}^{(j+1)}-\mu_{n1}^{(j+1)}h^{(j+1)}_{n,z_t^{(j)}})\Big)^2\end{aligned} \tag{55}$$

Step 4: Draw $\rho^{(j+1)}$ using the M-H algorithm defined in Table 2.

Step 5: Draw the aggregate regime indicator, $\boldsymbol{z}^{(j+1)}$ (recall that $z_t \in \{1,2,\ldots,K\}$ signifies which cluster is in recession at date t.).

Filter Step Apply the Hamilton Filter to obtain the filtered transition probabilities

Generation Step Sequentially draw $z_T^{(j+1)}, z_{T-1}^{(j+1)},\ldots,z_1^{(j+1)}$ by recursively iterating backwards. This is accomplished by multiplying the filtered probabilities by the forward transition probability

Step 6: Draw the transition probabilities, $\mathbf{P}^{(j+1)}$, by independently drawing every column $\mathbf{P}_p^{(j+1)}$ from $\mathbf{P}_p \sim \mathbf{D}(\boldsymbol{\alpha}_p^*)$ with the q^{th} hyperparameter, $\boldsymbol{\alpha}_{pq}^*$, of vector $\boldsymbol{\alpha}_p^*$ calculated as

$$\boldsymbol{\alpha}_{pq}^{*(j+1)} = \frac{\sum_{t=2}^{T}\delta(z(j+1)_{t-1}=p, z(j+1)_t=q)}{\sum_{t=2}^{T}\delta(z(j+1)_{t-1}=p)} \tag{56}$$

the fraction of times regime p is followed by regime q in the drawn sequence $\{z_1^{(j+1)}, z_2^{(j+1)},\ldots,z_T^{(j+1)}\}$.

B.3 Conditional posterior distribution of μ

This section derives the conditional posterior distribution of μ. The key to deriving the conditional posterior distribution of $\boldsymbol{\mu_n}$ is the *multivariate completion of squares* or *ellipsoidal rectification* identity:

$$\boldsymbol{u}'A\boldsymbol{u} - 2\boldsymbol{\alpha}'\boldsymbol{u} = (\boldsymbol{u} - A^{-1}\boldsymbol{\alpha})'A(\boldsymbol{u} - A^{-1}\boldsymbol{\alpha}) - \boldsymbol{u}'A^{-1}\boldsymbol{\alpha}. \tag{57}$$

The conditional posterior distribution of μ is:

$$\begin{aligned}
p(\boldsymbol{\mu}_n|\mathbf{Y}, \rho, \sigma_n^{-2}, \mathbf{P}, \boldsymbol{z}, h, \beta) \propto\ & \pi(\boldsymbol{\mu}_n|\sigma_n^{-2})\mathcal{L}(\rho, \boldsymbol{\mu}_n, \sigma_n^{-2}\,\boldsymbol{z}, h; \boldsymbol{Y_n}) \\
=\ & \left|\sigma_n^2\mathbf{M}\right|^{-0.5}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_n - \mathbf{m})'[\sigma_n^2\mathbf{M}]^{-1}(\boldsymbol{\mu}_n - \mathbf{m})\right] \\
& \times \sigma_n^{-T}|\boldsymbol{I}_N - \rho\mathbf{W}|^T\exp\left[-\frac{1}{2}\sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\left(y_{tn} - \rho\sum_{j=1}^{N}W_{nj}y_{tj} - \boldsymbol{\mu_n'}\mathbf{w}(z_t, h) + \rho\sum_{j=1}^{N}W_{nj}(\mu_{j0} + \mu_{j1}h_{j,z_t})\right)^2\right)\right] \\
\propto\ & \exp\left[-\frac{1}{2}\Big(\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\boldsymbol{\mu}_n - 2\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m} + \mathbf{m}'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m}\Big)\right] \\
& \times \exp\Bigg[-\frac{1}{2}\sigma_n^{-2}\Bigg(\sum_{t=1}^{T}\left(y_{tn} - \rho\sum_{j=1}^{N}W_{nj}y_{tj}\right)^2 \\
& - 2\sum_{t=1}^{T}\left(\boldsymbol{\mu_n'}\mathbf{w}(z_t, h) + \rho\sum_{j=1}^{N}W_{nj}(\mu_{j0} + \mu_{j1}h_{j,z_t})\right)\left(y_{tn} - \rho\sum_{j=1}^{N}W_{nj}y_{tj}\right) \\
& + \sum_{t=1}^{T}\left(\boldsymbol{\mu_n'}\mathbf{w}(z_t, h) + \rho\sum_{j=1}^{N}W_{nj}(\mu_{j0} + \mu_{j1}h_{j,z_t})\right)^2\Bigg)\Bigg].
\end{aligned}$$

All terms constant with respect to $\boldsymbol{\mu_n}$ including all $\{\boldsymbol{\mu_j} : j \neq n\}$ drop out into the proportionality constant.

$$\propto\ \exp\left[-\frac{1}{2}\Big(\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\boldsymbol{\mu}_n - 2\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m} + \mathbf{m}'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m}\Big)\right]$$

$$\times \exp\Bigg[-\frac{1}{2}\sigma_n^{-2}\Bigg(-2\sum_{t=1}^{T}\Big(\boldsymbol{\mu_n'}\mathbf{w}(z_t,h)(1+\rho W_{nn})\Big)\Big(y_{tn}-\rho\sum_{j=1}^{N}W_{nj}y_{tj}\Big)+\sum_{t=1}^{T}\Big(\boldsymbol{\mu_n'}\mathbf{w}(z_t,h)(1+\rho W_{nn})\Big)^2\Bigg)\Bigg]$$

$$= \exp\Bigg[-\frac{1}{2}\Big(\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\boldsymbol{\mu}_n - 2\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m}+\mathbf{m}'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m}\Big)\Bigg]$$

$$\times \exp\Bigg[-\frac{1}{2}\sigma_n^{-2}\Bigg(-2\sum_{t=1}^{T}\Big(\boldsymbol{\mu_n'}\mathbf{w}(z_t,h)(1+\rho W_{nn})\Big)\Big(y_{tn}-\rho\sum_{j=1}^{N}W_{nj}y_{tj}\Big)$$

$$+\Big(\boldsymbol{\mu_n'}(1+\rho W_{nn})^2\sum_{t=1}^{T}\mathbf{w}(z_t,h)\mathbf{w}(z_t,h)'\boldsymbol{\mu_n}\Big)\Bigg)\Bigg]$$

$$\propto \exp\Bigg[-\frac{1}{2}\Bigg(\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\boldsymbol{\mu}_n+\boldsymbol{\mu_n'}\sigma_n^{-2}(1+\rho W_{nn})^2\sum_{t=1}^{T}\mathbf{w}(z_t,h)\mathbf{w}(z_t,h)'\boldsymbol{\mu_n}$$

$$-2\boldsymbol{\mu}_n'[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m}-2\boldsymbol{\mu_n'}\sigma_n^{-2}\sum_{t=1}^{T}\Big(\mathbf{w}(z_t,h)(1+\rho W_{nn})\Big)\Big(y_{tn}-\rho\sum_{j=1}^{N}W_{nj}y_{tj}\Big)\Bigg)\Bigg]$$

Collecting terms before completing the multivariate square gives

$$= \exp\Bigg[-\frac{1}{2}\Bigg(\boldsymbol{\mu}_n'\Big(\sigma_n^{-2}(1+\rho W_{nn})^2\sum_{t=1}^{T}\mathbf{w}(z_t,h)\mathbf{w}(z_t,h)'+[\sigma_n^2\mathbf{M}]^{-1}\Big)\boldsymbol{\mu}_n$$

$$-2\boldsymbol{\mu}_n'\Big(\sigma_n^{-2}\sum_{t=1}^{T}\Big(\mathbf{w}(z_t,h)(1+\rho W_{nn})\Big)\Big(y_{tn}-\rho\sum_{j=1}^{N}W_{nj}y_{tj}\Big)+[\sigma_n^2\mathbf{M}]^{-1}\mathbf{m}\Big)\Bigg)\Bigg].$$

Considering the term in the exponential

$$-\frac{1}{2}(\boldsymbol{\mu}_n'\mathbf{A}\boldsymbol{\mu}_n-2\boldsymbol{\mu}_n'\mathbf{b})$$

and introducing a term to complete the square, along with using $\mathbf{I}=\mathbf{A}\mathbf{A}^{-1}$, gives

$$-\frac{1}{2}(\boldsymbol{\mu}_n'\mathbf{A}\boldsymbol{\mu}_n-2\boldsymbol{\mu}_n'\mathbf{A}\mathbf{A}^{-1}\mathbf{b}+\mathbf{b}'\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1}\mathbf{b}).$$

Let $\boldsymbol{\Sigma_n} = \mathbf{A}^{-1}$ and $\mathbf{m}^* = \mathbf{A}^{-1}\mathbf{b}$, then returning to the complete term in the proof gives

$$
\begin{aligned}
&= \exp\left[-\frac{1}{2}\left(\boldsymbol{\mu}_n' \boldsymbol{\Sigma_n}^{-1} \boldsymbol{\mu}_n - 2\boldsymbol{\mu}_n' \boldsymbol{\Sigma_n}^{-1} \mathbf{m}^* + \mathbf{m}^* \boldsymbol{\Sigma_n}^{-1} \mathbf{m}^{*\prime} \right) \right] \\
&= \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_n - \mathbf{m}^*)' \boldsymbol{\Sigma_n}^{-1} (\boldsymbol{\mu}_n - \mathbf{m}^*) \right],
\end{aligned}
\tag{58}
$$

which results in the following posterior distribution for $\boldsymbol{\mu}_n$:

$$
p(\boldsymbol{\mu}_n | \mathbf{Y}, \rho, \sigma_n^{-2}, \mathbf{P}, \boldsymbol{z}, h, \beta) \propto \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_n - \mathbf{m}^*)' \boldsymbol{\Sigma_n}^{-1} (\boldsymbol{\mu}_n - \mathbf{m}^*) \right]
$$

$$
\boldsymbol{\mu}_n | \mathbf{Y}, \rho, \sigma_n^{-2}, \mathbf{P}, \boldsymbol{z}, h, \beta \sim \mathrm{N}\Big(\mathbf{m}^*, \boldsymbol{\Sigma_n} \Big)
$$

where

$$
\boldsymbol{\Sigma_n} = \mathbf{A}^{-1}
$$

$$
\mathbf{m}^* = \mathbf{A}^{-1}\mathbf{b}
$$

$$
\mathbf{A} = \sigma_n^{-2}(1 + \rho W_{nn})^2 \sum_{t=1}^{T} \mathbf{w}(z_t, h)\mathbf{w}(z_t, h)' + [\sigma_n^2 \mathbf{M}]^{-1}
$$

$$
\mathbf{b} = \sigma_n^{-2} \sum_{t=1}^{T} \Big(\mathbf{w}(z_t, h)(1 + \rho W_{nn}) \Big) \Big(y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \Big) + [\sigma_n^2 \mathbf{M}]^{-1} \mathbf{m}.
$$

B.4 Conditional posterior distribution of σ_n^{-2}

This section derives the conditional posterior distribution of σ_n^{-2}:

$$
\begin{aligned}
p(\sigma_n^{-2}|\mathbf{Y},\rho,\boldsymbol{\mu}_n,\mathbf{P},\boldsymbol{z},h,\beta) \propto\ & \pi(\sigma_n^{-2})\mathcal{L}(\rho,\mu_{n0},\mu_{n1},\sigma_n^{-2}\,\boldsymbol{z},h;\mathbf{Y_n}) \\
\propto\ & \sigma_n^{-\nu+2}\exp\Big(-\frac{1}{2}\delta\sigma_n^{-2}\Big) \\
& \times \sigma_n^{-T}\exp\Bigg[-\frac{1}{2}\sigma_n^{-2}\sum_{t=1}^{T}\Big(y_{tn}-\mu_{n0}-\mu_{n1}h_{n,z_t}-\rho\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\Big)^2\Bigg] \\
=\ & \sigma_n^{-\nu-T+2}\exp\Bigg[-\frac{1}{2}\sigma_n^{-2}\Big(\delta+\sum_{t=1}^{T}\Big(y_{tn}-\mu_{n0}-\mu_{n1}h_{n,z_t}-\rho\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\Big)^2\Big)\Bigg] \\
=\ & \sigma_n^{-\nu-T+2}\exp\Bigg[-\frac{1}{2}\sigma_n^{-2}\Big(\delta+\sum_{t=1}^{T}\Big(y_{tn}-\rho\sum_{j=1}^{N}W_{nj}y_{tj}-\mu_{n0}-\mu_{n1}h_{n,z_t}+\rho\sum_{j=1}^{N}W_{nj}(\mu_{j0}+\mu_{j1}h_{j,z_t})\Big)^2\Big)\Bigg] \\
\Leftrightarrow & \\
\sigma_n^{-2}|\mathbf{Y},\rho,\boldsymbol{\mu}_n,\mathbf{P},\boldsymbol{z},h,\beta \sim\ & \Gamma\left(\frac{\nu+T}{2},\frac{\delta+\hat{\delta}}{2}\right),
\end{aligned}
\tag{59}
$$

where $\hat{\delta}=\sum_{t=1}^{T}\Big(y_{tn}-\mu_{n0}-\mu_{n1}h_{n,z_t}-\rho\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\Big)^2$ and the last equality in 59 is written to isolate the term $\rho\sum_{j=1}^{N}W_{nj}(\mu_{j0}+\mu_{j1}h_{j,z_t})$ that is not present in the Spatial Autoregressive Lag (SAL) specification of the model.

B.5 Conditional posterior distribution of ρ

This section derives the conditional posterior distribution of ρ:

$$
\begin{aligned}
p(\rho|\mathbf{Y},\boldsymbol{\mu},\boldsymbol{\Omega},\mathbf{P},\boldsymbol{z},h,\beta) \propto\ & \pi(\rho)\mathcal{L}(\rho,\boldsymbol{\mu},\boldsymbol{\Omega},\boldsymbol{z},h;\mathbf{Y}) \\
\propto\ & \pi(\rho)|\mathbf{I}_N-\rho\mathbf{W}|^T\exp\Bigg[-\frac{1}{2}\sum_{n=1}^{N}\Big(\sigma_n^{-2}\sum_{t=1}^{T}\Big(y_{tn}-\mu_{n0}-\mu_{n1}h_{n,z_t}-\rho\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\Big)^2\Big)\Bigg] \\
\propto\ & \pi(\rho)|\mathbf{I}_N-\rho\mathbf{W}|^T\exp\Bigg[-\frac{1}{2}\rho^2\sum_{n=1}^{N}\Bigg(\sigma_n^{-2}\sum_{t=1}^{T}\Big(\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\Big)^2\Bigg) \\
& +\frac{2}{2}\rho\sum_{n=1}^{N}\Bigg(\sigma_n^{-2}\sum_{t=1}^{T}\Big((y_{tn}-\mu_{n0}-\mu_{n1}h_{n,z_t})\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\Big)\Bigg)\Bigg] \\
=\ & \pi(\rho)\Big(\prod_{n=1}^{N}(1-\rho\gamma_n)\Big)^T\exp\Bigg[-\frac{1}{2}\rho\Big(\rho\mathbf{B}_1-2\mathbf{B}_2\Big)\Bigg].
\end{aligned}
\tag{60}
$$

The last equality follows from the fact that $|\mathbf{I}_N - \rho\mathbf{W}| = \prod_{n=1}^{N}(1-\rho\gamma_n)$ (see discussion in Section 3.2). The terms $\mathbf{B}_1$ and $\mathbf{B}_2$ are given in (61) and are subsequently vectorized.

$$
\begin{aligned}
\mathbf{B}_1 &= \sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\right)^2\right) \\
&= \operatorname{diag}(\boldsymbol{\Omega^{-1}})\begin{bmatrix}
\mathbf{W}_1\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}'\mathbf{W}_1' \\
\mathbf{W}_2\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}'\mathbf{W}_2' \\
\vdots \\
\mathbf{W}_N\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}'\mathbf{W}_N'
\end{bmatrix} \\
\mathbf{B}_2 &= \sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\left((y_{tn}-\mu_{n0}-\mu_{n1}h_{n,z_t})\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\right)\right) \\
&= \operatorname{diag}(\boldsymbol{\Omega^{-1}})\begin{bmatrix}
\mathbf{W}_1\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}}\tilde{\epsilon}_{1,z_1} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}}\tilde{\epsilon}_{1,z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\tilde{\epsilon}_{1,z_T}\end{bmatrix} \\
\mathbf{W}_2\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}}\tilde{\epsilon}_{2,z_1} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}}\tilde{\epsilon}_{2,z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\tilde{\epsilon}_{2,z_T}\end{bmatrix} \\
\vdots \\
\mathbf{W}_N\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}}\tilde{\epsilon}_{N,z_1} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}}\tilde{\epsilon}_{N,z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\tilde{\epsilon}_{N,z_T}\end{bmatrix}
\end{bmatrix}
\end{aligned}
\tag{61}
$$

Given a uniform prior distribution for ρ, (60) is given by:

$$
p(\rho|\mathbf{Y},\boldsymbol{\mu},\boldsymbol{\Omega},\mathbf{P},\boldsymbol{z},h,\beta) \propto \left(\prod_{n=1}^{N}(1-\rho\gamma_n)\right)^T \exp\left[-\frac{1}{2}\rho\Big(\rho\mathbf{B}_1 - 2\mathbf{B}_2\Big)\right]. \tag{62}
$$

Term $\mathbf{B}_1$: $\sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\left(\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})\right)^2\right)$

Vectorizing the most common term $\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t})$

$$
\sum_{j=1}^{N}W_{nj}(y_{tj}-\mu_{j0}-\mu_{j1}h_{j,z_t}) = \mathbf{W}_n(\boldsymbol{y_t}-\boldsymbol{\mu_0}-\boldsymbol{\mu_1}\odot\mathbf{h}_{z_t})
$$

where $\mathbf{W}_n$ is the n^{th} row of $\mathbf{W}$ and $\boldsymbol{y_t}$ is the t^{th} column of the $N\times T$ data matrix $\boldsymbol{y}$.

The term $\Big(\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\Big)^2$ is

$$\Big(\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\Big)^2 = \mathbf{W}_n(\boldsymbol{y_t} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_t})(\boldsymbol{y_t} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_t})'\mathbf{W}_n'$$

The term $\sum_{t=1}^{T}\Big(\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\Big)^2$ is

$$\begin{aligned}\sum_{t=1}^{T}\Big(\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\Big)^2 &= \mathbf{W}_n(\boldsymbol{y_1} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_1})(\boldsymbol{y_1} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_1})'\mathbf{W}_n' \\ &+ \mathbf{W}_n(\boldsymbol{y_2} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_2})(\boldsymbol{y_2} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_2})'\mathbf{W}_n' \\ &\dots \\ &+ \mathbf{W}_n(\boldsymbol{y_T} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_T})(\boldsymbol{y_T} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_T})'\mathbf{W}_n' \\ &= \mathbf{W}_n \begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}'\mathbf{W}_n'\end{aligned}$$

where $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_t}} = \boldsymbol{y_t} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_t}$.

Which gives

$$\sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\Big(\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\Big)^2\right)$$

$$= \begin{bmatrix}\sigma_1^{-2} & \sigma_2^{-2} & \dots & \sigma_N^{-2}\end{bmatrix}\begin{bmatrix}\mathbf{W}_1\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}'\mathbf{W}_1' \\ \mathbf{W}_2\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}'\mathbf{W}_2' \\ \vdots \\ \mathbf{W}_N\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}'\mathbf{W}_N'\end{bmatrix}.$$

Note: The N×T dimensional matrix $\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{z_1} & \tilde{\boldsymbol{\epsilon}}_{z_2} & \dots & \tilde{\boldsymbol{\epsilon}}_{z_T}\end{bmatrix}$ only needs to computed on time for the full term **Term $\mathbf{B}_1$** to be constructed.

Term $\mathbf{B}_2$: $\sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\Big((y_{tn} - \mu_{n0} - \mu_{n1}h_{n,z_t})\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\Big)\right)$.

Again, vectorizing the most common term $\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})$ gives

$$\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t}) = \mathbf{W}_n(\boldsymbol{y_t} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_t}),$$

where $\mathbf{W}_n$ is the n^{th} row of $\mathbf{W}$ and $\boldsymbol{y_t}$ is the t^{th} column of the $N \times T$ data matrix $\boldsymbol{y}$. This gives the following:

$$\begin{aligned}
&\sum_{t=1}^{T}\left((y_{tn} - \mu_{n0} - \mu_{n1}h_{n,z_t})\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\right) = \\
&(y_{1n} - \mu_{n0} - \mu_{n1}h_{n,z_1})\mathbf{W}_n(\boldsymbol{y_1} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_1}) + (y_{2n} - \mu_{n0} - \mu_{n1}h_{n,z_2})\mathbf{W}_n(\boldsymbol{y_2} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_2}) \\
&+ \ldots + (y_{Tn} - \mu_{n0} - \mu_{n1}h_{n,z_T})\mathbf{W}_n(\boldsymbol{y_T} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_T}) \\
&= \mathbf{W}_n\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \ldots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}{}',
\end{aligned}$$

where $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{n}}$ is the nth row of N×T dimensional matrix $\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \ldots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}$ and $\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_t}} = \boldsymbol{y_t} - \boldsymbol{\mu_0} - \boldsymbol{\mu_1} \odot \mathbf{h}_{z_t}$. This gives the complete **Term $\mathbf{B}_2$**:

$$\begin{aligned}
&\sum_{n=1}^{N}\left(\sigma_n^{-2}\sum_{t=1}^{T}\left((y_{tn} - \mu_{n0} - \mu_{n1}h_{n,z_t})\sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1}h_{j,z_t})\right)\right) \\
&= \begin{bmatrix}\sigma_1^{-2} & \sigma_2^{-2} & \ldots & \sigma_N^{-2}\end{bmatrix}\begin{bmatrix}
\mathbf{W}_1\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}}\tilde{\epsilon}_{1,z_1} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}}\tilde{\epsilon}_{1,z_2} & \ldots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\tilde{\epsilon}_{1,z_T}\end{bmatrix} \\
\mathbf{W}_2\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}}\tilde{\epsilon}_{2,z_1} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}}\tilde{\epsilon}_{2,z_2} & \ldots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\tilde{\epsilon}_{2,z_T}\end{bmatrix} \\
\vdots \\
\mathbf{W}_N\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}}\tilde{\epsilon}_{N,z_1} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}}\tilde{\epsilon}_{N,z_2} & \ldots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\tilde{\epsilon}_{N,z_T}\end{bmatrix}
\end{bmatrix}.
\end{aligned} \tag{63}$$

Note: Again the N×T dimensional matrix $\begin{bmatrix}\tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_1}} & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_2}} & \ldots & \tilde{\boldsymbol{\epsilon}}_{\boldsymbol{z_T}}\end{bmatrix}$ only needs to computed one time for the full term **Term $\mathbf{B}_2$** to be constructed.