ESSAYS ON REGIONAL RECESSIONS, SPATIAL INTERACTIONS AND FORECASTING

by

SERGEI SHIBAEV

A thesis submitted to the
Department of Economics
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
April 2017

Copyright © Sergei Shibaev, 2017
Abstract

This thesis contains three essays spanning the fields of econometrics and empirical macroeconomics. The first essay develops an econometric procedure that enables applied researchers to quantify spatial interactions from panel data where variables exhibit recurrent abrupt shifts in behavior. In empirical macroeconomics, omitting spatial effects is restrictive in many contexts because the units of analysis are regions, the characteristics of which are rarely independent. The second essay employs this methodology to investigate how recessions propagated through small regional economies in the United States from 1990 to 2015. The empirical results identify regions that are potentially at risk of collective economic distress, which is useful for national and regional policy makers. The analysis shows the importance of the spatial (or geographical) dimension in explaining how regional shocks amplify in the economy. The third essay, co-authored with Morten Ø. Nielsen, investigates a unique data set of daily political opinion polls in the United Kingdom from 2010 to 2015. This work explores the forecasting capabilities of the recently developed fractionally cointegrated vector auto-regressive (FCVAR) model. The results show that the FCVAR model delivers superior forecast accuracy relative to a portfolio of existing alternatives. Furthermore, the forecasts generated by the FCVAR model leading into the UK 2015 general election provide a more informative assessment of the current state of public opinion than that suggested by opinion polls.
Chapter 4 of this thesis is an academic research paper written jointly with Morten Ø. Nielsen. This work is forthcoming in the *Journal of the Royal Statistical Society Series A*, under the title “Forecasting daily political opinion polls using the fractionally cointegrated vector auto-regressive model”.
Acknowledgments

I am profoundly indebted and grateful to my supervisors: Morten Ø. Nielsen, for his incredible intellectual mentorship, ceaseless expansion of my horizons and generosity in time and resources, James G. MacKinnon, for his truly invaluable insight, advice and support for my research, and Taylor Jaworski, for his immensely helpful feedback, discussions and encouragement. I would like to sincerely thank them for their rigorous training, knowledge, passion and professionalism, which have always been and will continue to be an inspiration.

I would like to thank Martin Burda, Robert Clark, Christopher Cotton, Alfonso Flores-Lagunes, Allen Head, Jianjian Jin, Søren Johansen, Beverly Lapham, Hashem Pesaran, Michal Popiel, seminar participants at Queen’s University, participants at the Doctoral Workshop in Applied Econometrics (DWAE) 2015, 2016 and participants at the 49th Annual Conference of the Canadian Economics Association for helpful comments and suggestions.

I sincerely thank the Economics Department at Queen’s University and the John Deutsch Institute (JDI) for the Study of Economic Policy for generous financial support.

All errors are my own.
Dedication

To my incredible wife for her vision, devotion, trust and strength.
To my courageous and wise parents for their sacrifices, patience and love.
To my inspiring sister for always believing in me.
To my son and brother, to serve as an example of hard work and perseverance.
To all of my grandparents for their victories.
Contents

Abstract i
Co-Authorship ii
Acknowledgments iii
Dedication iv
Contents v
List of Tables viii
List of Figures x
Chapter 1: Introduction 1

Chapter 2: Measuring Spatial Interactions in the Presence of Common Markov-Switching Components 4
  2.1 Introduction ................................................................. 4
  2.2 Model Framework ...................................................... 9
  2.3 Joint Distribution and Likelihood Function ......................... 17
  2.4 Prior Distributions ..................................................... 19
  2.5 Conditional Distributions ........................................... 22
  2.6 Model Selection ....................................................... 28
  2.7 Estimation ............................................................... 31
  2.8 Conclusion ............................................................. 35

Chapter 3: Recession Propagation in Small Regional Economies 36
  3.1 Introduction ............................................................ 36
  3.2 Model Framework .................................................... 41
    3.2.1 Characterizing regional business cycles ....................... 41
    3.2.2 Endogenous clustering ........................................ 43
B.1.1 BEA Economic Areas ........................................... 154
B.1.2 Testing for spatial correlation ................................. 155
B.1.3 Convergence .................................................... 157
B.1.4 Other parameter estimates ................................. 159
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Formal statement of the model parameters.</td>
</tr>
<tr>
<td>2.2</td>
<td>M-H algorithm with random walk tuning</td>
</tr>
<tr>
<td>2.3</td>
<td>Short summary of Metropolis-within-Gibbs (M-G) algorithm</td>
</tr>
<tr>
<td>3.1</td>
<td>Formal statement of the model parameters.</td>
</tr>
<tr>
<td>3.2</td>
<td>Data summary - Employment shares</td>
</tr>
<tr>
<td>3.3</td>
<td>Spatial weights summary</td>
</tr>
<tr>
<td>3.4</td>
<td>Cluster selection - cross-validation entropy scores</td>
</tr>
<tr>
<td>3.5</td>
<td>Estimated logistic coefficients (posterior means $\hat{\beta}<em>{ki}$) and discrete derivatives $\delta</em>{ki}$</td>
</tr>
<tr>
<td>3.6</td>
<td>Estimated regime transition probabilities (posterior means)</td>
</tr>
<tr>
<td>3.7</td>
<td>Expected duration of expansion and contraction phases (in quarters)</td>
</tr>
<tr>
<td>3.8</td>
<td>Spatial spillovers spanning the period 1990–2015</td>
</tr>
<tr>
<td>3.9</td>
<td>Spatial spillovers spanning the national recessions for the period 1990-2015</td>
</tr>
<tr>
<td>3.10</td>
<td>Industrial Characteristics of BEA EA 146 San Jose-San Francisco-Oakland, CA</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary statistics (data in percentage)</td>
</tr>
<tr>
<td>4.2</td>
<td>Summary statistics (logit transformed data)</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Proportion of regions with negative employment growth</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>BEA Economic Areas</td>
<td>52</td>
</tr>
<tr>
<td>3.3</td>
<td>Regional covariates - employment shares (%)</td>
<td>56</td>
</tr>
<tr>
<td>3.4</td>
<td>Contiguity weighting example</td>
<td>59</td>
</tr>
<tr>
<td>3.5</td>
<td>Spatial Model Posterior cluster affiliation probabilities based on exogenous industrial composition variables alone [posterior means of ( \exp(\mathbf{x}'<em>{nk}\beta_k)/(1+\exp(\mathbf{x}'</em>{nk}\beta_k)) )]</td>
<td>63</td>
</tr>
<tr>
<td>3.6</td>
<td>Spatial Model Posterior cluster affiliation probabilities updated with observed regional employment growth [posterior means of ( h_{nk} )]</td>
<td>63</td>
</tr>
<tr>
<td>3.7</td>
<td>Geographical concentrations in Cluster 1</td>
<td>64</td>
</tr>
<tr>
<td>3.8</td>
<td>Geographical concentrations in Cluster 2</td>
<td>65</td>
</tr>
<tr>
<td>3.9</td>
<td>Posterior probabilities of aggregate regime indicator ( z_t )</td>
<td>71</td>
</tr>
<tr>
<td>3.10</td>
<td>Posterior draws of ( \rho ): ( \hat{\rho} = 0.72 ) (red line)</td>
<td>75</td>
</tr>
<tr>
<td>4.1</td>
<td>Time series plots of data 2010-05-14 – 2015-05-06</td>
<td>94</td>
</tr>
<tr>
<td>4.2</td>
<td>Serial dependence structure</td>
<td>95</td>
</tr>
<tr>
<td>4.3</td>
<td>Forecasts for the first training set</td>
<td>118</td>
</tr>
<tr>
<td>4.4</td>
<td>Time series plot of estimated common stochastic trend and UKIP data</td>
<td>124</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

The first two essays in this thesis are motivated by the growing importance of characterizing two types of phenomena frequently observed in many macroeconomic and financial variables; namely, regime shifting and spatial interactions. Theoretical econometricians, statisticians and applied researchers have made large strides forward inventing new methods, improving existing modeling techniques and exploring difficult questions to better understand changes in regime and spatial effects. The objective of this work is to develop a modeling framework that can provide better insights for applied researchers by jointly and explicitly modeling both of these phenomena, making a methodological and empirical contribution to this area of research.

The first phenomenon, that of regime shifting, relates to the important nonlinearities in variables that can be due to the economy or specific markets transitioning between specific phases. Examples are variables whose fluctuations correlate with financial markets being bearish/bullish or the economy expanding/contracting; see Piger (2009) and Hamilton (2016) for comprehensive discussions. Modeling this type of behavior in data sets with large cross-section and time series dimensions is difficult. Recent advances in econometrics due to Hamilton and Owyang (2012) allow
such behavior to be characterized from panel data, but lack the ability to explicitly consider the connections between cross-sectional units. This leads to the second phenomenon, frequently encountered in empirical macroeconomics, where omitting spatial effects is restrictive in many contexts because the units of analysis are regions, the characteristics of which are rarely independent. Recent evidence on the importance of the spatial (or geographical) dimension, in the context of business cycles, has been found in the work of Fogli, Hill, and Perri (2015).

The essay in Chapter 2 is a methodological piece that develops an econometric procedure generalizing the model of Hamilton and Owyang (2012) to allow for spatial interactions to be explicitly modeled. The proposed model allows important nonlinearities and spatial effects to be jointly and endogenously characterized from panel data. The essay in Chapter 3 uses the modeling framework of Chapter 2 to investigate the business cycle characteristics of small regions in the United States from 1990 to 2015. The vast majority of empirical regional business cycle analysis of the United States analyzes large regional divisions such as the eight BEA regions or the 48 lower US states. This work argues and delivers evidence that analyzing smaller geographical units is more informative for characterizing how recessions spread geographically in the United States.

The third essay in this thesis explores the forecasting capabilities of the recently developed fractionally cointegrated vector auto-regressive (FCVAR) model of Johansen (2008) and Johansen and Nielsen (2012). The analysis investigates a unique data set of daily political opinion polls in the United Kingdom from 2010 to 2015 that is contained within a single government regime. In addition to forecasting, the analysis leverages the FCVAR methodology to estimate the long-run equilibrium relations
between the vote share support for all political parties in the data set.

The data set that is studied has several advantageous properties for modeling and forecasting opinion polls. The first advantage is that all time series in the data set span a single government regime, which is desirable because it is free of election cycles. Therefore, unlike quarterly and monthly frequency opinion poll time series, recovering the underlying dependence structure in these data attributed to public opinion does not require fitting political cycles. The second attractive property is that this data set is part of a recent and on-going survey, making it very relevant for forecasting indicators of political support. This allows the empirical investigation of the FCVAR methodology to consider an application to the recent 2015 UK general election.

The main body of this thesis is organized as follows. Chapter 2 develops all aspects of the proposed econometric model, including the estimation and model selection procedures. Chapter 3 uses the methodology developed in the previous chapter to conduct an empirical investigation of recession propagation through small regional economies in the United States from 1990 to 2015. Chapter 4 contains a research paper, co-authored with Morten Ø. Nielsen, on forecasting daily opinion polls in the United Kingdom using the latest advances in multivariate fractional times series methods. Appendix A contains details on estimation algorithms and mathematical proofs, and Appendix B contains supplementary tables and figures.
Chapter 2

Measuring Spatial Interactions in the Presence of
Common Markov-Switching Components

2.1 Introduction

This chapter makes a contribution to econometric methods by developing a model capable of estimating spatial interactions in a class of regime-switching models. 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. The recent work of Hamilton and Owyang (2012) has provided researchers with a framework for inferring common Markov-switching components from panel data, which opens the door for many interesting empirical investigations. This chapter aims to improve on their modeling framework with the objective of providing better insights in applied research. To the best knowledge of the author, this work is the first to propose a model capable of explicitly modeling spatial interactions when dealing with common Markov-switching components in panel data.
2.1. INTRODUCTION

Markov-switching models are a popular choice in economics for characterizing important non-linearities observed in economic and financial data; refer to Piger (2009) and Hamilton (2016) for a comprehensive overview of applications in economics. Examples of topics and variables that frequently rely on regime-switching dynamics are: business cycles, inflation, interest rates, volatility and monetary policy actions. The methodology is popular because many financial and macroeconomic time series variables exhibit changes in behavior over long periods. The behavior of a series between these episodes of change can often be reasonably approximated through a linear time series model, but not over the entire observed period. One approach to dealing with this occurrence is to build a non-linear model that restricts the model to be linear in each regime (i.e. piecewise linear). This approach is central to most regime-switching models encountered in time series econometrics, where the phases of linearity are referred to as regimes, and the mechanism of dynamic change is captured through a probability law governing the transitions between regimes. Therefore, the change in regime is itself a random variable, given that it cannot realistically be assumed to be deterministic (see discussion in Hamilton (1994, 1989)). The methodology takes its name from the fact that it describes the probability law governing the changes in regime by a Markov chain over a discrete finite number of states (regimes).

The modeling choices in this chapter are heavily motivated by empirical applications to studying macroeconomic regimes, regime shifting and shock propagation. Hamilton (2016) and Ramey (2016) offer a comprehensive overview of frontier research on macroeconomic regimes and shocks. The model proposed in this chapter contributes to this growing literature by allowing researchers to quantify spatial interactions between unobserved shocks in the economy, in the presence of common
Markov-switching components. In empirical macroeconomics, omitting spatial effects is restrictive in many contexts because the units of analysis are regions, the characteristics of which are rarely independent. Recent evidence on the importance of the spatial dimension of business cycles can be found in the empirical work of Fogli, Hill, and Perri (2015).

The proposed model is an extension of the recently developed regime-switching model of Hamilton and Owyang (2012). 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.

Building on the original framework leverages an important advantage of the methodology, the ability to endogenously group cross-sectional units that exhibit similar behavior (e.g., synchronicities in abrupt shifts or similar characteristics). Importantly, this endogenous grouping mechanism does not restrict groups to be mutually exclusive, which is convenient for characterizing regimes at different points in time. In the context of recessions, 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 original model’s mechanism can learn about regional synchronicities through time allowing for such occurrences, while accounting for the importance of region-specific characteristics to supplement the characterization of similarities and differences between regions.

The endogenous grouping mechanism discussed in this chapter relates to two literatures focused on clustering methodologies. The first spans the fields of empirical
2.1. INTRODUCTION


The second literature spans the fields of theoretical statistics and econometrics. Alternatives to the procedure of Hamilton and Owyang (2012), the latent-class clustering of Paap and van Dijk (2005) and the well known k-means clustering algorithms are Bayesian clustering methods from the mixture modeling literature. Examples are Liu, Zhang, Palumbo, and Lawrence (2003), Heller and Ghahramani (2005) and Lau and Green (2007). Liu, Zhang, Palumbo, and Lawrence (2003) propose a dimension reduction procedure to precede the fitting of Gaussian mixtures, known to be a computationally expensive clustering strategy for high dimensional data. Heller and Ghahramani (2005) develop an algorithm for hierarchical clustering that uses a model-based criterion to decide on merging clusters, which they show has advantages
over distance-based clustering algorithms. Lau and Green (2007) develop a general formulation for Bayesian model-based clustering.

Central to most clustering procedures is the detection and choice of the number of clusters. This chapter will discuss two objective model selection procedures that can be used to favor and discriminate between spatial model specifications with varying numbers of idiosyncratic clusters (or groupings). Both approaches restrict the number of clusters to be finite. This assumption is appropriate when the researcher is interested in transitions between recurrent regimes; for example, to characterize economic variables that follow recurrent cycles, such as GDP and employment growth rates. In this context, the proposed spatial model is strongly motivated for empirical work when cross-sectional dependence is either observed or suspected in panel data comprised of variables that exhibit recurrent abrupt changes in behavior.

In applications where the researcher is not comfortable with transitions only between recurrent regimes (e.g. forecasting applications), they may consider the hidden Markov model (HMM) or infinite HMM (IHMM) methodologies as alternatives. These models allow for both recurring regimes and new regimes, the latter capturing structure change. HMMs are time series generalizations of mixture models and IHMMs are HMMs with countably infinitely many states. The recent work of Song (2012) and Maheu and Yang (2016) leverage this inherent ability by designing IHMMs to capture the dynamics of U.S. interest rates. Alternatively, an example of an HMM structure is the Bayesian estimation and prediction procedure of Pesaran, Pettenuzzo, and Timmermann (2006) for forecasting time series subject to multiple structural breaks. IHMMs were introduced by Beal, Ghahramani, and Rasmussen (2002). Teh, Jordan, Beal, and Blei (2006) showed that IHMMs can be derived from
hierarchical Dirichlet processes and Bratières, Gael, Vlachos, and Ghahramani (2010) provide parallel and distributed implementations.

The remainder of this chapter is structured as follows. Section 2.2 describes the model specification. Section 2.3 discusses the joint distribution of all random variables in the model and the likelihood function. Section 2.4 formally states all prior distributions. The conditional distributions for all model parameters are given in Section 2.5, with the discussion focused on the population parameters affected by the introduction of the spatial parameter. Section 2.6 describes the procedure used for model selection. Section 2.7 outlines and discusses the estimation algorithm and implementation of the model. Supplementary details regarding the inference and estimation procedures are discussed in Appendix A.

2.2 Model Framework

This section describes the model in two parts. The first part describes the baseline framework of Hamilton and Owyang (2012) for inferring common Markov-switching components in panel data. The second part describes the proposed model specification for characterizing spatial interactions, which extends the original framework.

Let $Y$ denote a $T \times N$ dimensional matrix containing an observed panel data set being investigated by the researcher, where $T$ is the time series dimension and $N$ is the cross-section dimension. Furthermore, suppose this data set exhibits recurrent abrupt fluctuations in the time series variables, perhaps due to the fact that they correlate with financial markets being bearish/bullish or the economy expanding/contracting. Hamilton and Owyang (2012) propose the following model specification for characterizing such recurrent abrupt shifts in multiple time series variables:
\[ y_t | \{ z_t = k \} = \mu_0 + \mu_1 \odot h_k + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Omega), \]  

(2.1)

where \( y_t = (y_{1t}, \ldots, y_{Nt})' \) is a vector of dependent variables for the \( N \) cross-sectional units at time \( t \), \( k = 1, \ldots, K \) is the regime (or cluster) index where \( K \) is the total number of regimes and \( z_t \) is an aggregate indicator, \( z_t \in \{1, 2, \ldots, K\} \), indicating which of the \( K \) clusters is active at time \( t \). The sequence \( \{z_1, z_2, \ldots, z_T\} \) is generated using the two step filter of Hamilton (1994). Each cluster \( k \) has an associated \( N \times 1 \) state vector \( h_k = (h_{1k} \ldots h_{Nk})' \), where the \( n^{th} \) element is unity when unit \( n \) is associated with cluster \( k \), and is zero otherwise, and \( \odot \) is the Hadamard (element-by-element) product. The state vector \( h_k \) is driven by fixed covariates, \( x_n \). The model in (2.1) postulates that, for a given \( k \), the observations are temporally uncorrelated. Conditional on \( h_1, h_2, \ldots, h_K \), the standard Markov-switching framework applies. \( \Omega \) is diagonal, with \( \varepsilon_1 \ldots \varepsilon_T \) assumed to be independent.

In cases where the applied researcher possesses meaningful knowledge regarding a particular configuration of cross-sectional units, he or she can impose such configurations \textit{a priori} for a subset of the cluster affiliation vectors \( h_1, h_2, \ldots, h_K \). A basic example of such a restriction is imposing \( h_1 \) to be a vector of ones, indicating that all units belong to that grouping, leaving the remaining \( K - 1 \) group configurations to be learned from the data. It is convenient for applied research that these clusters (or groupings) are not restricted to be mutually exclusive, since the aggregate regime indicator identifies at which point in time each cluster was active. This would not be the case if one was to rely on the k-means clustering methodology for defining groups, in which case groups would be mutually exclusive.

Whether unit \( n \) belongs to cluster \( k \) when \( z_t = k \) is determined by the observed synchronicities of that unit’s time series variable, \( y_{nt} \), with the the other units in the
2.2. MODEL FRAMEWORK

cluster, and the similarities of that unit’s fixed covariates, $x_n$, to the fixed covariates of the other units in the cluster. The model first identifies the probability that $n$ belongs to cluster $k$ based only on the fixed regional covariates, computed from

$$
\Pr(h_{nk} = 1 | \beta_k) = \frac{\exp(x_{nk}'\beta_k)}{1 + \exp(x_{nk}'\beta_k)}.
$$

The model then updates this probability with the observed fluctuations in the dependent variables. For simplicity it is assumed that the same covariates influence each cluster. In practice this assumption is appropriate when the researcher wishes to rely on similarities between characteristics across the cross-sectional dimension to drive the clustering mechanism. Such a setting is common in empirical macroeconomics, when the same fixed characteristics such as industrial composition are used to control for similarities between geographical units.

Extending the model specification in (2.1) is motivated by two observations. The first is that, to be operational, the model 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. This assumption cannot be relaxed directly. Therefore, attention is focused on alleviating its restrictiveness while at the same time providing meaningful insights for empirical applications.

The second motivating observation is that in many empirical macroeconomic applications the cross-sectional units of analysis are regions. When dealing with geographical units, such as regional divisions of a country, cross section observations are
2.2. MODEL FRAMEWORK

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.

The argument in this chapter 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 $\Omega$) in the original 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. In the context of regional business cycle analysis, the proposed approach postulates that every region in the economy has ties to one or more other regions, stemming from specific 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 chapter 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) to cross-sectional unit(s) can be spatially correlated with the shocks of other units. The proposed SAE version of the original regime-switching model is

\[
y_t | \{z_t = k\} = \mu_0 + \mu_1 \odot h_k + \varepsilon_t \\
\varepsilon_t = \rho W \varepsilon_t + u_t, \quad u_t \sim N(0, \Omega).
\]

(2.3)

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 \( \rho \) indicates that shocks are expected to be higher (lower), if on average, shocks to other cross-sectional units are high. \( W \) is a row-standardized matrix of spatial weights, and each row of \( W \), \( w_i \), dictates the spatial dependency of unit \( i \) to all other units. \( \rho \) does not depend on the aggregated regime indicator \( z_t \), because it captures the magnitude of spatial correlation that cannot be explained by the clustering mechanism, which only considers coincident abrupt shifts as the source of correlation in \( y_t \).

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. The SAE specification adds only a single parameter to be estimated, albeit one that plays a very important role through the fixed exogenous spatial weighting metric embodied by the matrix \( \mathbf{W} \) when it is fixed and exogenous. 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. Another advantage of the SAE specification, is that it provides a framework for applied researchers to quantify the magnitude of spatial dependence, through the spatial parameter \( \rho \), conditional on the spatial weighting matrix \( \mathbf{W} \). The specification of \( \mathbf{W} \), which defines the connections between cross-sectional units, determines the interpretation of \( \rho \).

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

\[
\begin{align*}
\mathbf{y}_t \mid \{z_t\} &= \mu_0 + \mu_1 \odot \mathbf{h}_k + \mathbf{\varepsilon}_t, \\
\mathbf{\varepsilon}_t &= \mathbf{\Psi} \mathbf{\varepsilon}_t + \mathbf{u}_t, \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}), \\
\mathbf{\Psi} &= \begin{bmatrix}
\rho_1 & 0 & \ldots & 0 \\
0 & \rho_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \rho_N
\end{bmatrix}
\end{align*}
\] (2.4)

The model in (2.4) quantifies cross-section unit specific degrees of spatial spillovers.

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

\[
\begin{align*}
\mathbf{y}_t \mid \{z_t\} &= \rho \mathbf{W} \mathbf{y}_t \mid \{z_t\} + \mu_0 + \mu_1 \odot \mathbf{h}_k + \mathbf{\varepsilon}_t, \quad \mathbf{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}).
\end{align*}
\] (2.5)
This chapter 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 $\rho$. 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 $\rho$ as capturing the spatial regional interactions in the macroeconomy.

The proposed model (2.3) nests the non-spatial (restricted) model (2.1) as a special case $\rho = 0$. Therefore, the posterior density of the spatial parameter $\rho$ provides a natural framework to test the hypothesis that the spatial parameter is different from zero for a given spatial weighting structure embodied by $W$. 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, Monfort, Renault, and Trognon (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.

In all three of the discussed spatial specifications, the dependence structure is based on a fixed metric embodied by the spatial weighting matrix \( W \). 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 chapter considers fixed specifications for \( W \) 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 of being a substantially more parsimonious approach than relaxing the diagonal assumption for the variance-covariance matrix of the model’s error vector.
2.3 Joint Distribution and Likelihood Function

This section discusses the joint distribution and the likelihood function. 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.

The joint distribution for the data (Y), population parameters (ρ, μ, Ω, β), variables for the dynamic change mechanism (regime transition probabilities, P) and latent variables (z, h) is given by (2.6). To be consistent with the compact notation in Hamilton and Owyang (2012), the cluster affiliation variables are grouped in H = {h, ξ, λ}, where ξ and λ are auxiliary parameters used in the Bayesian binary and multinomial regression of Holmes and Held (2006). ξ and λ in H = {h, ξ, λ} are integrated out as they affect the likelihood function only through the value of h and are only relevant as auxiliary parameters to facilitate generation of posterior values of β. The spatial weighting matrix W is suppressed as an argument in the notation because it is fixed.

\[ p(Y, ρ, μ, Ω, P, z, h, β) = p(Y|ρ, μ, Ω, z, h)p(z|P)p(ρ)p(μ, Ω)p(P)p(h|β)p(β) \] (2.6)

The compact expression for the arguments of the joint distribution in (2.6) 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 (2.6). Y is the T \times N matrix of the dependent variables. ρ is a scalar parameter measuring the degree of spatial dependence. μ is a N \times 2 dimensional matrix of the switching constant coefficients, where each row, μ_n, is defined as μ_n = [μ_n0, μ_n1]. The N \times N dimensional error variance-covariance matrix, Ω, is diagonal with N distinct elements: \text{diag}(Ω) =
2.3. JOINT DISTRIBUTION AND LIKELIHOOD FUNCTION

\((\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)\). The aggregate regime indicator, \(z\), 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 \(t\). \(p\) is the matrix of transition probabilities that governs the mechanism of dynamic change between regimes in the model. \(h\) is a set of vectors, \(h = \{h_1, h_2, \ldots, h_{K-f}\}\), where each vector \(h_k = (h_{1k}, h_{2k}, \ldots, h_{Nk})'\) determines which regions belong to cluster \(k\). \(f\) indicates the number of state vectors that are fixed and exogenous. This allows practitioners to impose groupings that they wish to control for in the empirical application. An example of this was discussed in Section 2.2. \(\beta\) is the set of all logistic coefficient vectors: \(\beta = \{\beta_1, \beta_2, \ldots, \beta_{K-f}\}\), where \(\beta_k = (1, \beta_{k1}, \ldots, \beta_{kP_k})'\) and \(P_k\) is the number of covariates explaining the cluster affiliations.

Prior to deriving the likelihood function, it is convenient to note that the model (2.3) can be written equivalently for the \(n^{th}\) observation at time \(t\) as

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

(2.7)

where the spatial term is factored into two additive terms to separate out the interaction of the spatial weights with the switching constant, \(\mu_{j0} + \mu_{j1} h_{j,z_t}\).

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

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

The term \(|I_N - \rho W|\) 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
\[ y_t = \mu_0 + \mu_1 \odot h_k + (I_N - \rho W)^{-1} u_t. \]  \hspace{1cm} (2.8)

Since the error covariance matrix is diagonal for \( u_t \), there exists a vector of homoskedastic random errors \( v_t \)

\[ v_t = \Omega^{-\frac{1}{2}} u_t. \]  \hspace{1cm} (2.9)

Substituting (2.9) into equation (2.8) gives

\[ \Omega^{-\frac{1}{2}} (I_N - \rho W)(y_t - \mu_0 - \mu_1 \odot h_k) = v_t. \]  \hspace{1cm} (2.10)

The error terms \( v_t \) have a known distribution but are unobserved. Therefore it is necessary to introduce a Jacobian term to derive the joint distribution for \( y_t \) from the joint distribution of these error terms through the relationships in (2.10). The Jacobian for the transformation of the random variables vector \( v_t \) into the random variables vector \( y_t \) is

\[ J = \det \left( \frac{\partial v_t}{\partial y_t} \right) = |\Omega^{-\frac{1}{2}}||I_N - \rho W|. \]  \hspace{1cm} (2.11)

This section defined the functional form of the likelihood with the spatial parameter \( \rho \). An important implication of the Jacobian term defined in (2.11) is that its presence will require an augmentation to the MCMC estimation procedure (sampling algorithm), which will be discussed in Section 2.7.

2.4 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 2.1 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 $\mu$ and $\Omega$ are standard for Markov-switching models (see Kim and Nelson, 1999). For every cross-sectional unit, $n$, parameters $\mu_n$ are assigned a Normal prior distribution (Equation 2.12). The Inverse Gamma distribution is specified for $\sigma_n^2$, the $n^{th}$ diagonal element of the diagonal variance-covariance matrix $\Omega$, which implies a Gamma prior for the precision, $\sigma_n^{-2}$ (Equation 2.13).

$$
\pi(\mu_n | \sigma_n) \propto |\sigma_n^2 M|^{-0.5} \exp \left[ -\frac{1}{2} (\mu_n - m)' [\sigma_n^2 M]^{-1} (\mu_n - m) \right] \quad (2.12)
$$

$$
\pi(\sigma_n^{-2}) \propto \sigma_n^{-\nu+2} \exp \left( -\frac{1}{2} \delta \sigma_n^{-2} \right) \quad (2.13)
$$

Prior elicitation follows the original framework of Hamilton and Owyang (2012). The $\beta_k$ coefficients for the logistic clustering procedure adopt a Normal prior distribution, $\pi(\beta_k) \sim N(b_k, B_k)$. Each column of the transition probability matrix, $P$, adopts a Dirichlet prior, $\pi(P_p) \sim D(0)$, which is improper. This implies that draws from the posterior distribution of $P_p$ will result in a transition matrix $P$ that has columns that sum to one, contrary to the more standard way of reporting $P$ with rows summing to one. The priors used for the latent variables $z, h, \xi, \lambda$ are given in the cluster affiliation category in Table 2.1, 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 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 $\rho$. 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 (2.3) enters the joint prior distribution $\pi(\rho, \mu, \Omega)$. Following the spatial econometrics literature – LeSage and Pace (2009, p.129) – the priors for $\mu_n$ and $\sigma_n^{-2}$ are assumed to be independent of the spatial parameter $\rho$, which is constant across all $N$ regions. The joint prior is given as

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

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 $\rho$ 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 $\pi(\rho) \sim U(\gamma_{\text{min}}^{-1}, \gamma_{\text{max}}^{-1})$ or $\pi(\rho) \sim U(-1, 1)$, where $\gamma_{\text{min}}$ and $\gamma_{\text{max}}$ represent the minimum and maximum eigenvalues of the spatial weight matrix $W$. Other than restricting $\rho$ to lie in a bounded interval, these distributions are uninformative with respect to $\rho$ and assign an equal probability to any realization of $\rho$ on that interval. These two choices are appropriate to use when the magnitude and significance of $\rho$ is of particular interest. If $\hat{\rho} \neq 0$ (estimated by the posterior mean) and $\rho$ has a high sign certainty probability\(^1\) then there are significant spatial interactions along

\(^1\)The probability mass on the same side of zero as the posterior mean.
the cross-sectional dimension. Otherwise, with the restriction of \( \rho = 0 \), the model is that of Hamilton and Owyang (2012).

2.5 Conditional Distributions

The conditional distribution of a given parameter in the model, denoted as \( \theta \), is given by

\[
P(\theta|Y, \rho, P, z, h, \beta) = \frac{p(Y, \rho, \mu, \Omega, P, z, h, \beta)}{\int p(Y, \rho, \mu, \Omega, P, z, h, \beta) d\theta},
\]

(2.15)

where \( \theta \) is an element of the set \( \{\rho, \mu, \Omega, P, z, h, \beta\} \).

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 \( \mu, \Omega \)**

The conditional posterior distributions of each individual parameter \( \mu_n \) and \( \sigma_n^{-2} \) are derived from

\[
p(\mu_n, \sigma_n^{-2}|Y, \rho, P, z, h, \beta) \propto \pi(\mu_n, \sigma_n^{-2}) \sigma_n^{-T} \exp \left\{ -\frac{1}{2} \sigma_n^{-2} \sum_{t=1}^{T} \left( y_{tn} - \mu_{n0} - \mu_{n1} h_{n,zt} \right)^2 \right\}.
\]

Both conditional posterior distributions for \( \mu_n \) and \( \sigma_n^{-2} \) are affected by \( \rho \), 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.
2.5. **CONDITIONAL DISTRIBUTIONS**

Table 2.1: Formal statement of the model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distributions</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switching Constant</td>
<td>( \pi \left( \frac{\mu_0}{\mu_1} \right) \sim N(m, \sigma^2 M) )</td>
<td>( m = \begin{pmatrix} 1 \ -2 \end{pmatrix}, \quad M = I_2 )</td>
</tr>
<tr>
<td>Variance of errors</td>
<td>( \pi(1/\sigma_n^2) \sim \Gamma(\nu/2, \delta/2) )</td>
<td>( \nu = 0, \quad \delta = 0 )</td>
</tr>
<tr>
<td>Spatial dependence</td>
<td>( \pi(\rho) \sim U(-1, 1) )</td>
<td></td>
</tr>
<tr>
<td>Spatial weight matrix</td>
<td>( W = \begin{bmatrix} w_{11} &amp; \cdots &amp; w_{1N} \ \vdots &amp; \ddots &amp; \vdots \ w_{N1} &amp; \cdots &amp; w_{NN} \end{bmatrix} )</td>
<td>( w_{ij} ) are row standardized spatial weights s.t. ( \sum_{i=1}^{N} w_{ni} = 1, \forall n )</td>
</tr>
<tr>
<td>Cluster affiliation</td>
<td>( \pi(h_{nk}) = \begin{cases} \frac{1}{1 + \exp(x'<em>{nk}/\beta_k)}, &amp; \text{if } h</em>{nk} = 0 \ \exp(x'<em>{nk}/\beta_k), &amp; \text{if } h</em>{nk} = 1 \end{cases} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( h_{nk} = \begin{cases} 1, &amp; \text{if } \xi_{nk} &gt; 0 \ 0, &amp; \text{otherwise} \end{cases} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \pi(\xi_{nk}</td>
<td>\beta_k, \lambda_{nk}) \sim N(x'<em>{nk}/\beta_k, \lambda</em>{nk}) )</td>
</tr>
<tr>
<td></td>
<td>( \pi(\beta_k) \sim N(b_k, B_k) )</td>
<td>( b = 0_p, \quad B = 0.5 I_p )</td>
</tr>
<tr>
<td></td>
<td>( \pi(\lambda_{nk}</td>
<td>\xi_{nk}, \beta_k) \sim \text{GIG}\left(\frac{1}{2}, 1, \frac{1}{2}r_{nk}^2\right) )</td>
</tr>
<tr>
<td>GIG ( \equiv \text{Generalized Inverse Gaussian} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition probabilities</td>
<td>( \pi(P_p) \sim D(\alpha) ) (Dirichlet)</td>
<td>( \alpha = 0 )</td>
</tr>
</tbody>
</table>

Notes: With the exception of the spatial dependence and spatial weight matrix, prior elicitation follows the framework of Hamilton and Owyang (2012). The priors for the variance of errors and the transition probabilities are improper.
The conditional posterior distribution of $\mu_n$ is (see Appendix A.0.2 for proof)

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

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

where

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

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

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

$$\mathbf{b} = \sigma_n^{-2} \sum_{t=1}^{T} \left( w(z_t, h)(1 + \rho W_m) \right) \left( ytn - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \right) + [\sigma_n^2 M]^{-1} \mathbf{m}.$$
where
\[ B_1 = \sum_{n=1}^{N} \left( \sigma_n^{-2} \sum_{t=1}^{T} \left( \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t})^2 \right) \right), \]
\[ B_2 = \sum_{n=1}^{N} \left( \sigma_n^{-2} \sum_{t=1}^{T} \left( (y_{tn} - \mu_{n0} - \mu_{n1} h_{n,z_t}) \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \right) \right). \]

The distribution in (2.17) has no known standard form. This has implications for the sampling procedure required to estimate the model, which is discussed in Section 2.7 that describes the MCMC sampling algorithm for \( \rho \) and for the full model. The distribution in (2.17), like the more standard SAL and SAE univariate models in the spatial econometrics literature (see LeSage and Pace (2009) for SAL example), is non-zero for almost all values of \( \rho \) on the real line.

### Conditional distribution of transition probability matrix \( P \)

Conditional on \( H \) and \( z \) the inference is that of a standard \( K \)-state Markov switching process. The conditional posterior distribution does not involve \( \rho \).

\[ p(P|Y, \rho, \mu_n, \Omega, H) \propto \pi(z|P)\pi(P) \]

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

\[ \alpha^*_{pq} = \frac{\sum_{t=2}^{T} \delta(z_{t-1} = p, z_t = q)}{\sum_{t=2}^{T} \delta(z_{t-1} = p)}, \quad (2.18) \]

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 $z, h, \xi, \lambda$

The aggregate regime indicator, $z$, for which $z_t \in \{1, 2, \ldots, K\}$, designates which regime is in active state at $t$. The conditional posterior distribution of $z$ is computed using the two step filter of Hamilton (1994). $\rho$ only enters the forecast error computation (i.e. the non-constant portion of the likelihood). 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.

$$p(z|Y, \rho, \mu, \Omega, H, \beta) \propto \mathcal{L}(\rho, \mu, \Omega, z, h; Y)\pi(z|P) \quad (2.19)$$

Let $H = \{h, \xi, \lambda\}$. For the set $h = \{h_1, h_2, \ldots, h_{K-f}\}$, each vector $h_k = (h_{1k} h_{2k} \ldots h_{Nk})'$ determines which regions belong to cluster $k$. For $h_k$, introducing $\rho$ affects the likelihood function.

$$p(h_k|Y, H^k, \rho, \mu, \Omega, P, \beta) \propto \mathcal{L}(\rho, \mu, \Omega, z, h; Y)\pi(h_k|\beta_k)$$

$$= \prod_{n=1}^{N} \mathcal{L}(\rho, \mu_{n0}, \mu_{n1}, \sigma_n^{-2}, z, h_{nk}, h^k; Y_n)\pi(h_{nk}|\beta_k), \quad (2.20)$$

where $H^k$ is the set of all elements belonging to the cluster affiliation parameters not associated with cluster $k$. That is, $H^k = \{h_j, \xi_j, \lambda_j : j = 1, \ldots, K-f; j \neq k\}$ and $h^{[k]} = \{h_{ni} : i = 1, \ldots, K-f; i \neq k\}$.

Each individual $h_{nk}$ indicates the affiliation of cross-section unit $n$ to cluster $k$ independently across cross-section units - follows from (2.15) - and is drawn from

$$\Pr(h_{nk} = 1|Y_n, h^{[k]}, \rho, \mu_n, \sigma_n^{-2}, P, z, \beta_k)$$

$$= \frac{\mathcal{L}(h_{nk} = 1, h^{[k]}, \rho, \mu_n, \sigma_n^{-2}, z; Y_n)\Pr(h_{nk} = 1|\beta_k)}{\sum_{i=0}^{1} \mathcal{L}(h_{nk} = i, h^{[k]}, \rho, \mu_n, \sigma_n^{-2}, z; Y_n)\Pr(h_{nk} = i|\beta_k)}, \quad (2.21)$$
where Pr($h_{nk} = i|\beta_k$) is computed from

$$
\Pr(h_{nk} = i|\beta_k) = \begin{cases} 
\frac{1}{1+\exp(x_{nk}'\beta_k)} & \text{if } i = 0 \\
\frac{\exp(x_{nk}'\beta_k)}{1+\exp(x_{nk}'\beta_k)} & \text{if } i = 1
\end{cases}.
$$

(2.22)

Nothing is affected by the inclusion of $\rho$ for the auxiliary parameters $\xi$ and $\lambda$, the outcomes of which influence the generation of $h_{nk}$. Hamilton and Owyang (2012) introduce $\xi$ and $\lambda$ to generate $h_{nk}$ following the auxiliary variable approaches to Bayesian binary and multinomial regression of Holmes and Held (2006)\(^2\). As in Hamilton and Owyang (2012), the conditional posterior distribution of $\xi_k$ is

$$
p(\xi_k|Y, h_k, H^k, \rho, \mu, \Omega, P, z, \beta) = \pi(\xi_k|h_k, \beta_k) = \prod_{n=1}^{N} \pi(\xi_{nk}|h_{nk}, \beta_k).
$$

(2.23)

Each element $\xi_{nk}$ is computed from $\xi_{nk} = x_{nk}'\beta_k - \log(u^{-1}_{nk} - 1)$, with $u^{-1}$ computed from

$$
u^{-1} = \begin{cases} 
\frac{1}{1+\exp(x_{nk}'\beta_k)}u_{nk} & \text{if } h_{nk} = 0 \\
\exp(x_{nk}'\beta_k) + \frac{1}{1+\exp(x_{nk}'\beta_k)}u_{nk} & \text{if } h_{nk} = 1
\end{cases}.
$$

(2.24)

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

The conditional posterior distribution of $\lambda_k$ is

$$
p(\lambda_k|Y, \xi_k, h_k, H^k, \rho, \mu, \Omega, P, z, \beta) = \pi(\lambda_k|h_k, \beta_k) \propto \pi(\xi_k|h_k, \beta_k)\pi(\lambda_k) = \prod_{n=1}^{N} \pi(\xi_{nk}|\lambda_{nk}, \beta_k)\pi(\lambda_{nk}).
$$

(2.25)

Following Holmes and Held (2006), each element $\lambda_{nk}$ is a draw from $\lambda_{nk} \sim \text{GIG}(\frac{1}{2}, 1, r_{nk}^2)$

where $r_{nk}^2$ is calculated from\(^3\)

$$
r_{nk} = \xi_{nk} - x_{nk}'\beta_k.
$$

(2.26)

\(^2\)The feasibility of their approach for logistic regression with auxiliary variables is based on the observations of Andrews and Mallows (1974)

\(^3\)Generalized Inverse Gaussian distribution
2.6. MODEL SELECTION

Conditional distribution of population parameter $\beta$

$$p(\beta|Y, \rho, \mu, \Omega, P, z, H) = \prod_{k=1}^{\kappa} p(\beta_k|\xi_k, \lambda_k)$$ (2.27)

The conditional posterior distribution is just a standard Normal regression model for each $\beta_k$. $\rho$ has no impact here. $\kappa = K - f$ is the number of clusters. The conditional posterior of $\beta_k$ is given by

$$\beta_k|Y, \mu_n, \Omega, P, z, H \sim N\left(b^*_k, B^*_k\right),$$

with a mean $b^*_k$ and variance $B^*_k$

$$B^*_k = \left(B_k^{-1} + X_k V_k^{-1}X_k'\right)^{-1}$$

$$b^*_k = B^*_k \left(B_k^{-1}b_k + X_k V_k^{-1}\xi_k\right)$$ (2.28)

which is a standard Normal regression of the form

$$\xi_k = X_k ^\prime \beta_k + \varepsilon_k, \quad \varepsilon_k \sim N(0, V_k)$$

$$V_k = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$$ (2.29)

2.6 Model Selection

This section discusses objective model selection procedures that can be used to specify the spatial model. Two candidate procedures are discussed: $R$–fold cross validation and Bayes Factors (BF).

In the non-spatial model of Hamilton and Owyang (2012), model selection is required to determine the number of clusters. The authors rely on a cross-validation procedure for selection, which computes a quasi-out-of-sample score for every model specification. This score is built as follows. The full data set is segmented into $R$
equal-sized blocks, denoted as $r = 1, \ldots, R$. One of the $r$ blocks is retained as a validation set, and the remaining $R - 1$ blocks serve as a training set. The model is estimated using the training set, and the estimated model is tested on the omitted validation set. This procedure repeats $R$ times, as 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 spatial model in this chapter adds another component, the spatial weighting matrix, $W$, that could in principle be specified according to an objective criterion when there is insufficient guidance for which exogenous weighting structure to specify for $W$.

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

$$
\text{Score} = \frac{1}{M} \sum_{m=1}^{M} \sum_{r=1}^{R} \sum_{t=t_r}^{t_{r+1}-1} \left( \log|\Omega^{[r,m]}| + (y_t - y_t^f)'(\Omega^{[r,m]})^{-1}(y_t - y_t^f) \right),
$$

(2.30)

where $M$ is the number of sampling iterations, $R$ is the number of equal blocks, $\Omega^{[r,m]}$ 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, $y_t = (y_{1t}, \ldots, y_{Nt})'$ is the vector of employment growth rates for $N$ regions at date $t$, $y_t^f$ is the forecast of $y_t$ conditional on the aggregate indicator $z_t^{[r,m]}$, and $(y_t - y_t^f)$ is the forecast error vector.

An alternative approach for model selection is to rely on Bayes Factors (BF) to discriminate between models, which is prevalent in econometric models that use Bayesian posterior inference. 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
2.6. MODEL SELECTION

\[ BF_{ij} = \frac{m_i(Y)}{m_j(Y)} = \frac{\int L_i(\Phi; Y)\pi_i(\Phi)d\Phi}{\int L_j(\Phi; Y)\pi_j(\Phi)d\Phi}, \]

where \( \Phi = \{\rho, \mu, \Omega, z, H, \beta\} \). BF has the advantage of penalizing more heavily parameterized models, which is not relevant for comparing weighting matrix structures that are fixed and exogenous, 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 (2.31). 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 BF 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 \( Y \) (the normalization constant of the posterior density) denoted as \( m(Y) \) and given by

\[ m(Y) = \frac{L(\rho, \mu, \Omega, z, H, \beta; Y)\pi(\rho, \mu, \Omega, z, H, \beta)}{p(\rho, \mu, \Omega, z, H, \beta|Y)}. \]

which Chib (1995) 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

\[ \ln \hat{m}(Y) = \ln L(\rho^*, \mu^*, \Omega^*, z^*, H^*, \beta^*; Y) + \ln \pi(\rho^*, \mu^*, \Omega^*, z^*, P^*, H^*, \beta^*) - \ln p(\rho^*, \mu^*, \Omega^*, z^*, P^*, H^*, \beta^*|Y), \]

which requires the evaluation of the log-likelihood function, prior and the joint posterior distribution (third term in (2.33)) for a given set \( \{\rho^*, \mu^*, \Omega^*, z^*, P^*, H^*, \beta^*\} \). The computational challenge is due to the fact that the joint posterior functional...
2.7. **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 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 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 (2.33) would base model selection on the estimated BF given by

\[
\hat{B}_{ij} = \exp\left(\ln \hat{m}_i(Y) - \ln \hat{m}_j(Y)\right)
\]

Both cross-validation and Bayes Factors have advantages and disadvantages. Due to the complicated 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 for selection, which explains why the measure was not employed by Hamilton and Owyang (2012). The drawback of cross-validation is that it is a computationally expensive procedure.

2.7 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 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
2.7. ESTIMATION

Table 2.2: M-H algorithm with random walk tuning

Step 1: Initialize the draw. Example: \( \rho^{(1)} = 0 \) and \( c^{(1)} = 0.1 \).

Step 2: For \( j = 1, \ldots, N_{\text{burn-in}}, N_{\text{burn-in}} + 1, \ldots, N_{\text{burn-in}} + N_{\text{keep}} \)

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

\[
\rho^* = \rho^{(j)} + c^{(j)} u^{(j)}
\]  

(2.35)

where \( u^{(j)} \) is a draw from \( u \sim N(0, 1) \).

(b) If \( |\rho^*| \geq 1 \), the candidate value is rejected.

(c) If \( |\rho^*| < 1 \), the candidate value is accepted as \( \rho^{(j+1)} = \rho^* \) with probability:

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

(2.36)

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

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

(2.37)

parameter \( \rho \) requires sampling from a distribution of unknown form (refer to Section 2.5). The sampling algorithm for \( \rho \) 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, Shankar, and Rahman (2002) for a comprehensive discussion. 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 2.2. 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 2.2 is nested within 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 2.3, which suppresses all conditioning variables, hyperparameters and functional forms. After initializing parameters (Step 0), the samplers iteratively draw following Steps 1 through 6 for \( j = 1, \ldots, N^{\text{burn-in}}, N^{\text{burn-in}} + 1, \ldots, N^{\text{burn-in}} + N^{\text{keep}} \), where \( N^{\text{burn-in}} \) is the specified number of burn-in iterations that are discarded and \( N^{\text{keep}} \) is the number of iterations retained for inference as posterior draws.
Table 2.3: Short summary of Metropolis-within-Gibbs (M-G) algorithm

| Step 0: | Initialize all parameters |
| Step 1: | Draw cluster |
| (a) $\beta^{(j+1)}_k, k = 1, \ldots, K - f$ |
| (b) $h^{(j+1)}_{nk}, k = 1, \ldots, K - f$ and $n = 1, \ldots, N$ |
| (c) $\zeta^{(j+1)}_{nk}, k = 1, \ldots, K - f$ and $n = 1, \ldots, N$ |
| (d) $\lambda^{(j+1)}_{nk}, k = 1, \ldots, K - f$ and $n = 1, \ldots, N$ |
| Step 2: | Draw $\mu^{(j+1)}_n, n = 1, \ldots, N$ |
| Step 3: | Draw $\sigma^{-2(j+1)}_n, n = 1, \ldots, N$ |
| Step 4: | Draw $\rho^{(j+1)}$ using the M-H algorithm defined in Table 2.2. |
| Step 5: | Draw aggregate regime indicator, $z^{(j+1)}_t, t = 1, 2, \ldots, T$ |
| Step 6: | Draw the transition probabilities, $P^{(j+1)}$ |

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

A complete and detailed outline of the full sampling procedure, M-G algorithm, for estimating the model is given in Appendix A.0.1.

The implementation of the model requires two types of restrictions to be imposed. This is necessary to ensure the model is identified and is not affected by the well-known label switching problem. These restrictions follow directly from the original non-spatial model of Hamilton and Owyang (2012). The first restriction is to normalize $\mu_{n1} < 0$. The second restriction is to rule out transitions between the idiosyncratic clusters, by restricting $p_{ij} = 0$ for all $i \neq j$ when $i$ and $j$ are both cluster regimes. In addition to these restrictions proposed by Hamilton and Owyang (2012), this
chapter proposes that a formal MCMC output convergence test is conducted to ensure that the label switching problem has not surfaced. The test diagnostic used should compare the identification of parameters across and within separate MCMC runs. The diagnostic of Gelman and Rubin (1992) is a good candidate, as it measures the potential scale from continuing the MCMC run further, and is constructed based on within and between chain variability of posterior parameter draws.

2.8 Conclusion

This chapter developed a new econometric framework that enables applied researchers to measure spatial interactions in the presence of common Markov-switching components. The modeling choices were strongly motivated by the view that omitting spatial effects is restrictive in macroeconomic applications because the units of analysis are often geographical units, the characteristics of which are rarely independent. The proposed methodology has a broad range of applications in situations where cross-sectional dependence is either observed or suspected in panel data, comprised of variables that exhibit recurrent abrupt changes in behavior. The model’s structure provides a convenient setting for endogenously characterizing contagion and propagation dynamics. Areas of economic research that stand to directly benefit from these abilities are empirical applications to studying macroeconomic regimes, regime shifting and shock propagation. In addition to the baseline spatial model, a more general model specification for characterizing spatial interactions was proposed. Researchers are encouraged to consider extending and adapting the specification to the specific needs of their empirical study, and the methodological contribution in this chapter provides a building block for such developments.
Chapter 3

Recession Propagation in Small Regional Economies

3.1 Introduction

This chapter makes a contribution to empirical macroeconomics. The econometric model developed in Chapter 2 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\(^2\). To the best knowledge of the author, this work is the first to analyze

\(^1\)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.
regional business cycles in small regional economies in the United States using a multivariate regime-switching model.

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 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 chapter 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, Dreger, and Kholodilin (2011), Fogli, Hill, and Perri (2015) and Beraja, Hurst, and Ospina (2016). Notably, the work of Beraja, Hurst, and Ospina (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, Franses, Paap, and van Dijk (2007) and Hamilton and Owyang (2012).
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 Rieberman (2005), Artis, Dreger, and Kholodilin (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, Piger, and Wall (2005) and Owyang, Piger, Wall, and Wheeler (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 chapter 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 chapter’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 chapter 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
3.1. INTRODUCTION

Rickman (2005) and van Dijk, Franses, Paap, and van Dijk (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, Franses, Paap, and van Dijk (2007) uses a latent-class clustering to identify co-movements between regions in the Netherlands. The paper uses 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, Dreger, and Kholodilin (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, Hill, and Perri (2015) uses 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 chapter.

The remainder of the chapter is structured as follows. Section 3.2 describes the
model and its structure for characterizing regional business cycles, regional clusters and spatial dependence. Section 3.3 discusses the data and spatial weighting structure considered in the empirical investigation, including alternative measures. The empirical results are presented and analyzed in Section 3.4 including model selection procedure for identifying the number of regional clusters and a discussion of some policy implications using the results. Section 3.5 concludes.

3.2 Model Framework

This section discusses the model in three parts. Section 3.2.1 motivates the use of Markov-switching models to characterize regional business cycles and explains the model specification of Hamilton and Owyang (2012). Section 3.2.2 describes the endogenous clustering mechanism. Section 3.2.3 motivates and describes the extended model specification which introduces a spatial dependence component into the existing model.

3.2.1 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,
3.2. MODEL FRAMEWORK

\begin{equation}
y_t = \mu_0 + \mu_1 s_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2),
\end{equation}

\begin{equation}
s_t = \begin{cases} 
  1, & \text{if recession,} \\
  0, & \text{if expansion.}
\end{cases}
\end{equation}

Model (3.1) 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, Piger, and Wall (2005) and Owyang, Piger, Wall, and Wheeler (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

\begin{equation}
y_t = \mu_0 + \mu_1 \odot s_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Omega),
\end{equation}

where \( y_t = (y_{1t}, \ldots, y_{Nt})' \) is a vector of employment growth rates for \( N \) regions at date \( t \), \( s_t = (s_{1t}, \ldots, s_{Nt})' \) is a vector of recession (contraction phase) indicators \( (s_{nt} = 1 \text{ when region } n \text{ is in recession and } s_{nt} = 0 \text{ when region } n \text{ is in expansion}) \), and \( \odot \) is the Hadamard (element-by-element) product. The \( n^{th} \) element of the \( N \times 1 \) vector \( \mu_0 + \mu_1 \) is the average employment growth in region \( n \) during recession and
3.2. MODEL FRAMEWORK

the $n^{th}$ element of the $N \times 1$ vector $\mu_0$ 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 3.3.

3.2.2 Endogenous clustering

For the empirical application in Hamilton and Owyang (2012) which looks at the lower 48 US states, the model defined in (3.2) 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 (3.2), 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, $P$. The model is written as

$$y_t \mid \{z_t = k\} = \mu_0 + \mu_1 \odot h_k + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Omega), \quad (3.3)$$

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 $h_k = (h_{1k} \ldots h_{Nk})'$, where the $n^{th}$ element is unity when region $n$ is associated with the cluster $k$, and is zero otherwise. Conditional on $h_1, h_2, \ldots, h_K$, the standard Markov-switching framework applies.

Following Hamilton and Owyang (2012)’s inference procedure for the configurations of the cluster affiliation vectors $h_1, h_2, \ldots, h_K$, two clusters are imposed a priori

\(^3\)The binary recession indicator $s_{nt}$ for region $n \in \{1, 2, \ldots, N\}$ implies $\eta = 2^N$ distinct regimes.
3.2. MODEL FRAMEWORK

to capture nationwide recessions and expansions. The nationwide expansion cluster is $h_K$ (a column of zeros) where every region is in expansion when the aggregate indicator $z_t = K$, and the nationwide recession cluster is $h_{K-1}$ (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 $h_1, h_2, \ldots, h_{K-2}$ and economic downturns that spread nationwide, determined by clusters $h_{K-1}, h_K$.

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 $x_n$, 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

$$
\Pr(h_{nk} = 1|\beta_k) = \frac{\exp(x_n'\beta_k)}{1 + \exp(x_n'\beta_k)}.
$$

The model then updates this probability with observed regional employment growth patterns. If the probability in (3.4) 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 3.3. 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.
3.2. MODEL FRAMEWORK

3.2.3 Spatial spillovers

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.

The analysis will employ the spatial model developed in Chapter 2, which explicitly characterizes spatial interaction along the cross-sectional dimension with a spatial autoregressive error (SAE) component. 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

$$y_t | \{z_t = k\} = \mu_0 + \mu_1 \odot h_k + \varepsilon_t$$
$$\varepsilon_t = \rho W \varepsilon_t + u_t, \quad u_t \sim N(0, \Omega).$$

(3.5)

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 $\rho$ indicates that shocks are expected to be higher (lower), if on average, shocks to neighboring regions are high. $W$ is a row-standardized matrix of spatial weights, and each row of $W$, $w_i$, 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 3.3.3.

The more general model specification in (2.4) of Chapter 2 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 (2.4), the more parsimonious SAE specification in (3.5) with a single common spatial parameter is more desirable for addressing the questions posed on the nature of spatial interactions.

The empirical application in Section 3.4 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 model. The latter will be shown to be strongly robust to various specifications of the model.

3.2.4 Prior elicitation

This section discusses the modeling choices regarding the prior distributions assigned to the parameters in the model, including a formal statement of the model parameters. Table 3.1 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 $\mu$ and $\Omega$ are standard for Markov-switching models (see Kim and Nelson, 1999). The parameters $\mu_n$, which characterize the growth rates for region $n$ during regimes of recession and expansion are assigned a Normal prior distribution (Equation 3.6). The Inverse Gamma distribution is specified for $\sigma^2_n$, the $n^{th}$ diagonal element of the diagonal variance-covariance matrix $\Omega$, which implies a Gamma prior for the precision, $\sigma^{-2}_n$ (Equation 3.7).

\[
\pi(\mu_n|\sigma_n) \propto \sigma_n^2 M^{-0.5} \exp \left[ -\frac{1}{2} (\mu_n - m)' [\sigma_n^2 M]^{-1} (\mu_n - m) \right] \quad (3.6)
\]

\[
\pi(\sigma_n^{-2}) \propto \sigma_n^{-\nu+2} \exp \left( -\frac{1}{2} \delta \sigma_n^{-2} \right) \quad (3.7)
\]
## 3.2. MODEL FRAMEWORK

Table 3.1: Formal statement of the model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior distributions</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average employment growth in region ( n )</td>
<td>( \pi \left( \mu_{n0}, \mu_{n1} \right) \sim N(m, \sigma^2M) )</td>
<td>( m = \begin{pmatrix} 1 \ -2 \end{pmatrix}, \ M = I_2 )</td>
</tr>
<tr>
<td>Variance of errors</td>
<td>( \pi(1/\sigma^2_n) \sim \Gamma(\nu/2, \delta/2) )</td>
<td>( \nu = 0, \ \delta = 0 )</td>
</tr>
<tr>
<td>Spatial dependence</td>
<td>( \pi(\rho) \sim U(-1, 1) )</td>
<td></td>
</tr>
<tr>
<td>Spatial weight matrix</td>
<td>( W = \begin{bmatrix} w_{11} &amp; \ldots &amp; w_{1N} \ \vdots &amp; \ddots &amp; \vdots \ w_{N1} &amp; \ldots &amp; w_{NN} \end{bmatrix} )</td>
<td>( w_{ij} ) are row standardized spatial weights s.t. ( \sum_{i=1}^{N} w_{ni} = 1, \ \forall n )</td>
</tr>
<tr>
<td>Cluster affiliation</td>
<td>( \pi(h_{nk}) = \begin{cases} \frac{1}{1 + \exp(x'<em>{nk} \beta_k)}, &amp; \text{if } h</em>{nk} = 0 \ \frac{\exp(x'<em>{nk} \beta_k)}{1 + \exp(x'</em>{nk} \beta_k)}, &amp; \text{if } h_{nk} = 1 \end{cases} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( h_{nk} = \begin{cases} 1, &amp; \text{if } \xi_{nk} &gt; 0 \ 0, &amp; \text{otherwise} \end{cases} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \pi(\xi_{nk}</td>
<td>\beta_k, \lambda_{nk}) \sim N(x'<em>{nk} \beta_k, \lambda</em>{nk}) )</td>
</tr>
<tr>
<td></td>
<td>( \pi(\beta_k) \sim N(b_k, B_k) )</td>
<td>( b = 0_p, \ B = 0.5I_p )</td>
</tr>
<tr>
<td></td>
<td>( \pi(\lambda_{nk}</td>
<td>\xi_{nk}, \beta_k) \sim \text{GIG}\left(\frac{1}{2}, 1, r^2_{nk}\right) )</td>
</tr>
<tr>
<td>GIG ( \equiv ) Generalized Inverse Gaussian</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition probabilities</td>
<td>( \pi(P_p) \sim D(\alpha) ) (Dirichlet)</td>
<td>( \alpha = 0 )</td>
</tr>
</tbody>
</table>

Notes: With the exception of the spatial dependence and spatial weight matrix, prior elicitation follows the framework of Hamilton and Owyang (2012). The priors for the variance of errors and the transition probabilities are improper.
The $\beta_k$ coefficients for the logistic clustering procedure adopt a Normal prior distribution, $\pi(\beta_k) \sim N(b_k, B_k)$. Each column of the transition probability matrix, $P$, adopts a diffuse Dirichlet prior, $\pi(P_p) \sim D(0)$. The priors used for the latent variables $z, h, \xi, \lambda$ are given in the cluster affiliation category in Table 3.1, 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).

The spatial parameter, $\rho$, is assigned a uniform prior distribution $\pi(\rho) \sim U(-1, 1)$. This assumption allows $\rho$ to be interpreted as a spatial correlation coefficient and is otherwise uninformative, assigning an equal probability to any realization of $\rho$ on the $(-1, 1)$ interval. This prior assumption is appropriate to use when the magnitude and significance of $\rho$ is of particular interest. In the context of this chapter’s empirical investigation, the estimation results for $\rho$ are of great importance. Furthermore, the assumption allows for $\rho = 0$, that if supported empirically (e.g. $\hat{\rho} = 0$) would favor the nested original (non-spatial) specification. Otherwise, $\hat{\rho} \neq 0$ (estimated by the posterior mean) and a high sign certainty probability$^4$ would support the spatial specification given that significant spatial interactions between regions are estimated.

3.3 Data and spatial weighting

This section describes the choice of aggregation level, the data sets, and the spatial weighting structure. Section 3.3.1 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 3.3.2 describes the regional covariates to be used in the clustering mechanism. Lastly,

$^4$The probability mass on the same side of zero as the posterior mean.
Section 3.3.3 defines the main specification of the spatial weighting structure and existing alternatives, some of which will serve as robustness checks for the estimation results.

### 3.3.1 County-level employment – BEA Economic Areas

The goal of this chapter 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 3.1. 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.

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 is chosen for the analysis because it is the most recent definition and roughly falls at the midpoint of the period analyzed by the empirical investigation; from 1990 to 2015. Relying on the geographical detail provided by the 1995 classification may be problematic for the studied period, since the BEA used historical data preceding 1995 to establish the socioeconomic homogeneity between counties. It is in this context that the 2004 classification is expected to provide a more accurate assessment of homogeneity across contiguous counties for the period 1990 to 2015. The 2004 classification defines 177 contiguous groups of counties that represent relevant regional markets surrounding metropolitan and micropolitan statistical areas in the United States, while the 1995 classification defined 172 groups. See Figure 3.2 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 B.1 in Appendix B.1.

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 censuses. Specifically, these 3106 geographical units consist of: 3001 contiguous counties, 64 Louisiana Parishes, 41 independent cities in Virginia and the District of Columbia. 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 program of the US Census Bureau. The data enter the model as annualized quarter-over-quarter growth rates.
Figure 3.2: BEA Economic Areas

(a) Counties

(b) Overlay with state borders
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. 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 chapter 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.

### 3.3.2 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 3.2 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, industry-specific characteristics using employment shares provide an intuitive geographical/spatial comparison of each regional cluster. This can be done by analyzing the higher and lower concentrations of specific industries employment composition in each grouping. Four of the covariates are analogs to those considered at the state level. 

---

5The 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)
### 3.3. DATA AND SPATIAL WEIGHTING

<table>
<thead>
<tr>
<th>Industry</th>
<th>NAICS</th>
<th>County-level data</th>
<th>Aggregated variable description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oil &amp; gas extraction</td>
<td>211</td>
<td>QCEW 1990-2014</td>
<td>177</td>
</tr>
<tr>
<td>Mining &amp; quarrying</td>
<td>212</td>
<td>QCEW 1990-2014</td>
<td>177 average (%) share</td>
</tr>
<tr>
<td>Construction</td>
<td>23</td>
<td>QCEW 1990-2014</td>
<td>177 of total employment</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>33-33</td>
<td>QCEW 1990-2014</td>
<td>177 ↓</td>
</tr>
<tr>
<td>Financial &amp; insurance</td>
<td>52</td>
<td>QCEW 1990-2014</td>
<td>177 (%) share</td>
</tr>
</tbody>
</table>

**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.

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 3.3. 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 employees are considered.

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. This is a result of the model’s structure, which restricts that there can only be a single value for each cross-sectional unit and covariate pair.
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.

3.3.3 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 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, \( \rho \), compared to the original model. However, the computational intensity of the MCMC estimation algorithm depends on the sparsity of the spatial weighting matrix \( W \).

This chapter advocates the use of a spatial weighting structure based on shared physical borders between economic areas. This contiguity weighting is defined in (3.8) and is illustrated with an example in Figure 3.4.
Figure 3.3: 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
3.3. DATA AND SPATIAL WEIGHTING

Contiguity-based weighting matrix

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

\( \gamma_{ij} = \begin{cases} 
1, & \text{if region } j \text{ shares a common border with region } i \\
0, & \text{otherwise}
\end{cases} \) \hspace{1cm} (3.8)

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 \( W \) 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 \( W \) in comparison to well-known existing alternatives. This can be seen in Table 3.3, which shows how many non-zero weights various \( W \) specifications have for the 177 BEA economic areas. A contiguity-based weighting imposes a lighter computational load on the model while leveraging the inherent spatial similarities accounted for by the BEA economic area classification.
3.3. DATA AND SPATIAL WEIGHTING

Table 3.3: Spatial weights summary

<table>
<thead>
<tr>
<th></th>
<th>$W_{\text{contig}}$</th>
<th>$W_{400\text{km}}$</th>
<th>$W_{600\text{km}}$</th>
<th>$W_{800\text{km}}$</th>
<th>$W_{\text{inv-dist}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td># of weights</td>
<td>912</td>
<td>1882</td>
<td>3926</td>
<td>6552</td>
<td>31152</td>
</tr>
</tbody>
</table>

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

Geography-based restrictions have been supported in empirical factor models 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 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, Hill, and Perri (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 $W$ 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 $W$:

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)

$$
\gamma_{ij} = \begin{cases} 
1, & \text{if region } j \text{ is the largest trading partner of region } i \\
0, & \text{otherwise}
\end{cases}
$$

\(^1\)Edges connecting bordering economic areas outside of the California state border are suppressed to illustrate the spatial weighting for the specific state.
Figure 3.4: Contiguity weighting example

(a) California and surrounding 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
3.4 Empirical results and discussion

This section presents and discusses the results of the empirical investigation. Section 3.4.1 outlines the selection of idiosyncratic clusters using a cross-validation procedure. Section 3.4.2 analyzes the regional composition of each idiosyncratic cluster grouping. Section 3.4.3 analyzes how each phase of the regional (non-nationwide) and nationwide business cycles propagate in the economy. Section 3.4.4 discusses the assessment of spatial regional interactions, including a robustness analysis. Section 3.4.5 discusses some policy implications based on the findings.
3.4. EMPIRICAL RESULTS AND DISCUSSION

3.4.1 Selection of idiosyncratic clusters

Model selection will rely on $R$-fold cross validation scores for choosing the number of idiosyncratic clusters, $\kappa = K - 2$, the same selection procedure used in the empirical investigation of Hamilton and Owyang (2012). Because approximating the marginal likelihood function for the model is a difficult task, implementing a cross-validation procedure is more straightforward than relying on approximate Bayes Factors (BF) for selection and explains why the measure was not employed by Hamilton and Owyang (2012).

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 for the higher dimensional model specifications (e.g. $K = 8$). Convergence diagnostics are discussed in detail in Appendix B.1.3. Cross-validation is conducted using $R = 5$ folds for equal MCMC sequence lengths and considers up to six ($\kappa = 6$) clusters.

Cross-validation results are reported in Table 3.4. 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 with two idiosyncratic clusters ($\kappa = 2$, four regimes in total). A formal assessment of MCMC output convergence for this specification is given in Appendix B.1.3. The results show that all 1322 parameters converge, indicating that the target posterior distributions based on the retained
3.4. EMPIRICAL RESULTS AND DISCUSSION

Table 3.4: Cluster selection - cross-validation entropy scores

<table>
<thead>
<tr>
<th>Clusters ($\kappa$)</th>
<th>Spatial Model</th>
<th>Restricted Model ($\rho = 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Block 1</td>
<td>Block 2</td>
</tr>
<tr>
<td>1</td>
<td>6856.0</td>
<td>8455.4</td>
</tr>
<tr>
<td>2</td>
<td>6847.5</td>
<td>8434.6</td>
</tr>
<tr>
<td>3</td>
<td>6845.4</td>
<td>8460.7</td>
</tr>
<tr>
<td>4</td>
<td>6864.0</td>
<td>8462.6</td>
</tr>
<tr>
<td>5</td>
<td>6847.4</td>
<td>8469.9</td>
</tr>
<tr>
<td>6</td>
<td>6851.3</td>
<td>8457.4</td>
</tr>
</tbody>
</table>

Notes: The aggregate scores for each number of clusters. 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 $N_{\text{burn-in}} = 250000$ burn-in iterations and $N_{\text{keep}} = 25000$ samples retained for inference.

MCMC draws are reliable for inference and there is very little potential scale reduction from continuing the MCMC algorithm.

3.4.2 Regional spatial clusters

The mechanism of endogenous clustering is best illustrated in the following two steps. The first step groups units based only on the exogenous industrial composition variables that capture fixed regional characteristics. This provides a geographical illustration (see Figure 3.5) 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 3.6 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 the hierarchal
3.4. EMPIRICAL RESULTS AND DISCUSSION

Figure 3.5: Spatial Model
Posterior cluster affiliation probabilities based on exogenous industrial composition variables alone [posterior means of $\exp(x'_{nk}\beta_k)/(1 + \exp(x'_{nk}\beta_k))]$

(a) Cluster 1
(b) Cluster 2

Figure 3.6: Spatial Model
Posterior cluster affiliation probabilities updated with observed regional employment growth [posterior means of $h_{nk}$]

(a) Cluster 1
(b) Cluster 2

Comparing Figures 3.6(a) and 3.7(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 3.6. Regions with a higher than 0.5 probability of being affiliated with a cluster are deemed to be strongly affiliated to that grouping, while 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 3.7 and 3.8.

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.
3.4. EMPIRICAL RESULTS AND DISCUSSION

Figure 3.8: Geographical concentrations in Cluster 2

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 small regional divisions into which clusters.

Let a discrete derivative for the industrial characteristic $i$ in cluster $k$ be denoted as $\delta_{ki}$. 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, $\delta_{ki}$ 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

$$
x_q' = [1 \ x_1 \ x_2 \ ... \ x_i - s_{x_i} \ ... \ \bar{x}_{P_k-1} \ \bar{x}_{P_k}]',
$$

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

$$
x_s' = [1 \ x_1 \ x_2 \ ... \ x_i + s_{x_i} \ ... \ \bar{x}_{P_k-1} \ \bar{x}_{P_k}]'.
$$

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

$$
\hat{\beta}_k = [1 \ \hat{\beta}_{k1} \ \hat{\beta}_{k2} \ ... \ \hat{\beta}_{ki} \ ... \ \hat{\beta}_{kP_k-1} \ \hat{\beta}_{kP_k}]',
$$

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

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

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

$$
\delta_{ki} = \Pr(h_{sk} = 1|\hat{\beta}_k, x_s) - \Pr(h_{qk} = 1|\hat{\beta}_k, x_q)
= \frac{\exp(x_s'\hat{\beta}_k)}{1+\exp(x_s'\hat{\beta}_k)} - \frac{\exp(x_q'\hat{\beta}_k)}{1+\exp(x_q'\hat{\beta}_k)}
= \exp(x_s'\hat{\beta}_k) - \frac{\exp(x_q'\hat{\beta}_k)}{1+\exp(x_q'\hat{\beta}_k)}
$$

(3.9)
3.4. EMPIRICAL RESULTS AND DISCUSSION

A positive (negative) value of $\delta_{ki}$ 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 3.5: Estimated logistic coefficients (posterior means $\hat{\beta}_{ki}$) and discrete derivatives $\delta_{ki}$

<table>
<thead>
<tr>
<th>Sector</th>
<th>$\hat{\beta}_{1i}$</th>
<th>$\delta_{1i}$</th>
<th>$\hat{\beta}_{2i}$</th>
<th>$\delta_{2i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.055</td>
<td>0.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(53)</td>
<td>(51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>-0.208*</td>
<td>-0.501*</td>
<td>-0.115*</td>
<td>-0.087*</td>
</tr>
<tr>
<td></td>
<td>(81)</td>
<td>(79)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finance &amp; insurance</td>
<td>0.443*</td>
<td>0.219*</td>
<td>0.104</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(83)</td>
<td>(58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mining &amp; quarrying</td>
<td>-0.097</td>
<td>-0.021</td>
<td>-0.067</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(55)</td>
<td>(53)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oil &amp; gas extraction</td>
<td>0.138</td>
<td>0.019</td>
<td>0.220</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(58)</td>
<td>(62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small firms</td>
<td>0.031</td>
<td>0.089</td>
<td>-0.011</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(57)</td>
<td>(43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>-0.215*</td>
<td>-0.124*</td>
<td>-0.162</td>
<td>-0.027</td>
</tr>
<tr>
<td></td>
<td>(70)</td>
<td>(62)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 $N_{\text{burn-in}} = 250000$ burn-in iterations and $N_{\text{keep}} = 25000$ samples retained for inference.

Table 3.5 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 3.5 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 $\delta_{ki}$ 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 characterizing the distinctions between regions in the data. This cluster closely resembles the oil and gas producing regions, which is evident when the labor concentrations in the oil and gas sector in Figure 3.3 are compared to the geographical concentrations in cluster 2 in Figure 3.8. The evidence in Table 3.5 confirms 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 ($\hat{\beta}_{24} = 0.220$).

### 3.4.3 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 3.9 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
3.4. EMPIRICAL RESULTS AND DISCUSSION

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 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 3.9 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 were not sufficient to designate a nationwide contraction in all 177 areas, but as the crisis progressed its 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 3.9 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 the cluster affiliation probability. The last panel of Figure 3.9 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 for negative co-movements in employment growth patterns. Figure 3.9 also shows that both cluster probabilities rise following each NBER recession period. This is not a result of the nationwide contraction phase probability decreasing, and suggests that the regions affiliated with the clusters are slower to exit recessions.

The estimates associated with the mechanism of dynamic change in the model provide important information for understanding the business cycle phases. The estimated transition matrix is given in Table 3.6. 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) phases more (less) likely to be followed by a phase of expansion (contraction). The last two columns of Table 3.6 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
Figure 3.9: Posterior probabilities of aggregate regime indicator $z_t$

Notes: Shaded regions indicate NBER recessions. Posterior means computed for MCMC sequence based on $N_{\text{burn-in}} = 250000$ burn-in iterations and $N_{\text{keep}} = 25000$ samples retained for inference.
Table 3.6: Estimated regime transition probabilities (posterior means)

<table>
<thead>
<tr>
<th></th>
<th>From nationwide expansion</th>
<th>From nationwide contraction</th>
<th>From cluster 1 contraction</th>
<th>From cluster 2 contraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>To nationwide expansion</td>
<td>0.86</td>
<td>0.10</td>
<td>0.25</td>
<td>0.40</td>
</tr>
<tr>
<td>To nationwide contraction</td>
<td>0.03</td>
<td>0.72</td>
<td>0.35</td>
<td>0.21</td>
</tr>
<tr>
<td>To cluster 1 contraction</td>
<td>0.03</td>
<td>0.10</td>
<td>0.40</td>
<td>0</td>
</tr>
<tr>
<td>To cluster 2 contraction</td>
<td>0.08</td>
<td>0.08</td>
<td>0</td>
<td>0.39</td>
</tr>
</tbody>
</table>

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.

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 3.7. 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 3.7: Expected duration of expansion and contraction phases (in quarters)

<table>
<thead>
<tr>
<th></th>
<th>Nationwide expansion</th>
<th>Nationwide contraction</th>
<th>Cluster 1 in contraction</th>
<th>Cluster 2 in contraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of quarters</td>
<td>7.14</td>
<td>3.57</td>
<td>1.67</td>
<td>1.64</td>
</tr>
</tbody>
</table>

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

### 3.4.4 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 B.1.2, Figures B.2 and B.3. For each spatial weighting, the vector of employment growth rates, $y_t$, 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 $\rho$, which alleviates the restrictive assumption of a diagonal variance-covariance matrix for the error vector. The posterior draws of the spatial parameter $\rho$ are shown in Figure 3.10. Inference is drawn on the entire surface of the posterior distribution for which the draws of $\rho$ 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 $\rho$ 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 3.10 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 $\rho$ 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 $W$, 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 $W$ to assess how robust the finding of a strong positive spatial interaction is to various spatial weights. The robustness results are given in Table 3.8. The estimates of $\rho$ for each spatial weighting are given in an increasing order of the number of weights defined in each $W$ matrix for the 177 geographical units. The most sparse matrix is the contiguity weighting, $W_{\text{contig}}$, with 912 non-zero weights. The least sparse matrix is the symmetric inverse distance weighting, $W_{\text{inv-dist}}$, 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.
3.4. EMPIRICAL RESULTS AND DISCUSSION

Figure 3.10: Posterior draws of $\rho$: $\hat{\rho} = 0.72$ (red line)

![Graph of posterior draws](image)

Table 3.8: Spatial spillovers spanning the period 1990–2015

<table>
<thead>
<tr>
<th></th>
<th>$W_{\text{contig}}$</th>
<th>$W_{400\text{km}}$</th>
<th>$W_{600\text{km}}$</th>
<th>$W_{800\text{km}}$</th>
<th>$W_{\text{inv-dist}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}$</td>
<td>0.72</td>
<td>0.77</td>
<td>0.85</td>
<td>0.89</td>
<td>0.94</td>
</tr>
<tr>
<td>(0.47,0.97)</td>
<td>(0.58,0.97)</td>
<td>(0.72,0.98)</td>
<td>(0.78,0.99)</td>
<td>(0.88,0.99)</td>
<td></td>
</tr>
<tr>
<td>Number of weights</td>
<td>(max = 31152)</td>
<td>912</td>
<td>1882</td>
<td>3926</td>
<td>6552</td>
</tr>
</tbody>
</table>

Notes: The posterior means of $\rho$ are obtained by estimating the full model with two ($\kappa = 2$) idiosyncratic clusters on different spatial weighting matrices $W$. $W_{\text{contig}}$ is the contiguity weighting, $W_{400\text{km}}$, $W_{600\text{km}}$ and $W_{800\text{km}}$ are 400, 600 and 800 kilometer distance band weightings, and $W_{\text{inv-dist}}$ is the symmetric inverse distance weighting. The 90 percent equal-tailed credible intervals are given in parentheses.
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, Hill, and Perri (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 chapter 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 3.9 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
Table 3.9: Spatial spillovers spanning the national recessions for the period 1990-2015

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_{\text{contig}}$</td>
<td>912</td>
<td>0.70</td>
<td>0.70</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.44,0.97)</td>
<td>(0.42,0.96)</td>
<td>(0.66,0.98)</td>
</tr>
<tr>
<td>$W_{400\text{km}}$</td>
<td>1882</td>
<td>0.75</td>
<td>0.72</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.53,0.97)</td>
<td>(0.47,0.97)</td>
<td>(0.71,0.98)</td>
</tr>
<tr>
<td>$W_{600\text{km}}$</td>
<td>3926</td>
<td>0.85</td>
<td>0.81</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.71,0.98)</td>
<td>(0.64,0.98)</td>
<td>(0.82,0.99)</td>
</tr>
<tr>
<td>$W_{800\text{km}}$</td>
<td>6552</td>
<td>0.89</td>
<td>0.86</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.79,0.98)</td>
<td>(0.73,0.98)</td>
<td>(0.87,0.99)</td>
</tr>
<tr>
<td>$W_{\text{inv-dist}}$</td>
<td>31152</td>
<td>0.94</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.88,0.99)</td>
<td>(0.86,0.99)</td>
<td>(0.93,0.99)</td>
</tr>
</tbody>
</table>

Notes: The posterior means of $\rho$ 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 $W$. $W_{\text{contig}}$ is the contiguity weighting, $W_{400\text{km}}$, $W_{600\text{km}}$, and $W_{800\text{km}}$ are 400, 600 and 800 kilometer distance band weightings, and $W_{\text{inv-dist}}$ is the symmetric inverse distance weighting. The 90 percent equal-tailed credible intervals are given in parentheses.

Regional markets was amplified during the crisis and in the period that followed. This interpretation stems from the spatial component structure of the model, which may be capturing, at the regional level, how different the aggregate shock affecting the economy was for this cycle. This amplification is robust to other spatial weightings.

### 3.4.5 Policy implications

This section discusses some of the policy implications based on the empirical findings. For exposition, the discussion will use BEA Economic Area 146 San Jose-San Francisco.
Francisco-Oakland, CA as a working example. Suppose you are a regional policy maker or researcher interested in a specific United States county or small regional area – BEA Economic Area 146. 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 regions 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. 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?”

(Q3) “Over the last several decades, when was the degree of spatial spillovers between regions the highest?”

Providing insights for (Q1) requires the spatial model’s structure for inferring
common Markov-switching components. (Q1) requires explicit modeling of the synchronicities in the variable that fluctuates over the business cycle, which cannot be realistically assumed to be deterministic for small regional economies. The spatial component in the model gives insights into (Q2) and (Q3), which are posed to draw parallels to questions typically posed in applications of spatial econometric models; see Fogli, Hill, and Perri (2015) for an application example of a standard spatial lag regression model.

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 3.4.4 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 1990s and early 2000s 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 B.1 and 3.7. It is important to refer to Figures 3.6(a) and 3.6(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 3.6(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 region’s characteristics and follow the discrete derivative analysis in Section 3.4.2. The industrial characteristics of this region are given in Table 3.10, 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 3.6(b). This means that co-movements in observed employment growth for the period 1990–2015 are more important than fixed characteristics 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, the answer for the policy-relevant question (Q1) posed earlier would be: Yes, counties surrounding San Francisco, 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
3.4. EMPIRICAL RESULTS AND DISCUSSION

Table 3.10: Industrial Characteristics of BEA EA 146
San Jose-San Francisco-Oakland, CA

<table>
<thead>
<tr>
<th></th>
<th>Manufacturing</th>
<th>Finance &amp; insurance</th>
<th>Mining &amp; quarrying</th>
<th>Oil &amp; gas extraction</th>
<th>Small firms</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>EA 146</td>
<td>13.86</td>
<td>4.62</td>
<td>0.02</td>
<td>0.00</td>
<td>48.06</td>
<td>5.53</td>
</tr>
<tr>
<td>$\bar{x}_i$</td>
<td>15.86</td>
<td>4.32</td>
<td>0.18</td>
<td>0.12</td>
<td>47.31</td>
<td>5.93</td>
</tr>
<tr>
<td>$\bar{x}<em>i + s</em>{x_i}$</td>
<td>22.71</td>
<td>5.67</td>
<td>0.77</td>
<td>0.49</td>
<td>55.00</td>
<td>7.50</td>
</tr>
</tbody>
</table>

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 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 chapter.

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.

3.5 Conclusion

This chapter has employed a new framework for measuring spatial interactions when estimating macroeconomic regimes and regime shifts 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. 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 spatial methodology developed in Chapter 2. 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.
Chapter 4

Forecasting Daily Political Opinion Polls Using the
Fractionally Cointegrated Vector Auto-regressive Model

Written jointly with Morten Ø. Nielsen (Queen’s University and CREATES)
Forthcoming in the Journal of the Royal Statistical Society: Series A

4.1 Introduction

In this chapter we investigate the forecasting performance of the recently developed fractionally cointegrated vector autoregressive (FCVAR) model of Johansen (2008) and Johansen and Nielsen (2012) relative to a portfolio of competing models at various forecast horizons. The FCVAR model generalizes the concept of cointegration, and in particular generalizes Johansen’s (1995) cointegrated VAR (CVAR) model to fractionally integrated time series, and hence allows estimating long-run equilibrium relationships between fractional time series.

The FCVAR model is a very recently developed statistical model, and it is therefore of particular interest to examine the gains this model can deliver for the purposes
of forecasting. The choice of data set for applying the model should reflect a current and relevant issue for forecasting. A prominent example is the desire to predict political support and election vote share outcomes. This chapter addresses this task by applying the FCVAR model to a novel data set which is comprised of polling results of political support in the United Kingdom for the period 2010–2015 at the business-daily observation frequency. The fractional integration behavior of political opinion polling data has been well established in the literature, albeit for time series at lower frequencies (monthly, quarterly), e.g. Box-Steffensmeier and Smith (1996), Byers, Davidson, and Peel (1997, 2000, 2002), Dolado, Gonzalo, and Mayoral (2002, 2003), and Jones, Nielsen, and Popiel (2014). For a more general reference on fractional integration methods in political time series data, see Box-Steffensmeier and Tomlinson (2000) and Lebo, Walker, and Clarke (2000); both in the special issue of Electoral Studies edited by Lebo and Clarke (2000). It therefore appears natural to apply a fractional time series model such as the FCVAR to model and forecast political opinion polls.

The industry standard for measuring the current state of political support is through opinion polling. The demand for polling and survey methodology is largely driven by the client’s desire to form an accurate understanding of the current state of opinion on a particular question. The poll evidence then serves as an input into the decision making process. When polls are conducted at regular intervals, it seems natural to use a statistical model to extract the full potential of the information contained in these time series of poll results by using them to forecast public opinion beyond the most recent poll date. However, long time series of poll data are scarce, and, to the best of our knowledge, all previous studies that have analyzed time series
of political opinion polls have used data observed at the monthly frequency or lower, see, e.g., above references. Authors of these studies have noted that an ideal data set would have all observations contained within a single government regime spanning only one political cycle, while providing a large enough sample to conduct meaningful statistical analysis. The data set used in this chapter fully satisfies both desired properties: it spans the entire UK political cycle following the 2010 UK general election, it is conducted at a high observation frequency (business-daily), and it is very recent and on-going, and hence very relevant also for forecasting poll standings which can be viewed as the predicted vote shares for each political party in an election.

The long time series provided in our data set facilitates forecast accuracy evaluation using several forecast evaluation procedures. In particular, it allows the formation of a large number of training sets from which each statistical model can produce forecasts. We apply two standard procedures to assess forecast accuracy: the rolling window and recursive forecasting schemes. The main distinction between the two schemes is how they select the training sets used for estimating the models. The rolling window scheme uses a fixed training set length (commonly referred to as a window) that moves across the data set, and the recursive scheme uses an expanding training set length with a fixed start date. The portfolio of models we consider consists of eight statistical models, four of which are variants of the FCVAR model. These are then evaluated on their forecasting ability relative to a group of four popular competing models. Among the latter, the CVAR model serves as the main multivariate benchmark model. The simple ARFIMA$(0,d,0)$ and the more general ARFIMA$(p,d,q)$ models serve as the fractional univariate benchmarks, where the former was found by, e.g., Byers, Davidson, and Peel (1997) and Dolado, Gonzalo,
and Mayoral (2002), to fit (monthly) UK polling data well. Finally, we include the ARMA(\(p,q\)) model as the classical univariate benchmark. Forecast accuracy is assessed at seven out-of-sample forecast horizons: 1, 5, 10, 15, 20, 25, and 50 steps ahead.

The forecasting analysis in this chapter shows that the FCVAR model delivers valuable gains in predicting political support. Both forecasting schemes agree on this finding. The accuracy of forecasts generated by the FCVAR model is better than all multivariate and univariate models in the portfolio, and overall the four variants of the FCVAR model are ranked as the four top performing models. Not only do they perform better relative to the other models, but the forecasting performance of all FCVAR variants is within very close range of each other. When compared to the multivariate benchmark model, the FCVAR model substantially outperforms the CVAR model in 56 of 56 cases, and the relative forecast improvement is highest at the 15–50 steps ahead forecast horizons, where the root mean squared forecast error (RMSFE) of the FCVAR model is up to 20% lower than that of the CVAR benchmark model. Previous literature, as cited above, has documented the superiority of fractional (ARFIMA) models for forecasting polling data. Compared to this more difficult benchmark, the RMSFE of the FCVAR model is as much as 15% lower, and the advantage of the FCVAR model again appears to be increasing with the forecast horizon.

As an empirical application, we apply the FCVAR model to the full data set, comprising observations until the day before the 2015 UK general election. We first consider estimation of the model, with interpretations of both the estimated cointegrating relations and estimated common stochastic trend. It appears that the latter
4.2 Fractional integration, polling data, and summary statistics

In important early contributions, Box-Steffensmeier and Smith (1996) and Byers, Davidson, and Peel (1997, 2002) show that political popularity, as measured by public opinion polls, can be modeled as fractional time series processes. The fractional (or fractionally integrated or just integrated) time series models are based on the fractional difference operator,

\[ \Delta^d X_t = \sum_{n=0}^{\infty} \pi_n (-d) X_{t-n}, \quad (4.1) \]
where the fractional coefficients $\pi_n(u)$ are defined in terms of the binomial expansion 
\[(1 - z)^{-u} = \sum_{n=0}^{\infty} \pi_n(u) z^n, \text{ i.e.},\]
\[
\pi_0(u) = 1 \text{ and } \pi_n(u) = \frac{u(u + 1) \cdots (u + n - 1)}{n!}.
\]

For details and for many intermediate results regarding this expansion and the fractional coefficients, see, e.g., Johansen and Nielsen (2016, Appendix A). Efficient calculation of fractional differences, which we apply in our analysis, is discussed in Jensen and Nielsen (2014).

With the definition of the fractional difference operator in (4.1), a time series $X_t$ is said to be fractional of order $d$, denoted $X_t \in I(d)$, if $\Delta^d X_t$ is fractional of order zero, i.e. if $\Delta^d X_t \in I(0)$. The latter property can be defined in the frequency domain as having spectral density that is finite and non-zero near the origin or in terms of the linear representation coefficients if the sum of these is non-zero and finite, see, e.g., (Johansen and Nielsen, 2012, p. 2672). An example of an $I(0)$ process is the stationary and invertible ARMA model.

The standard reasoning for political opinion poll series being fractional relies on Robinson’s (1978) and Granger’s (1980) aggregation argument, and can briefly be described as follows. Suppose individual level voting or polling behavior is governed by the (possibly binary) autoregressive process

\[x_{i,t} = \delta_{i,1} + \delta_{i,2} x_{i,t-1} + u_{i,t}, \]

where $i = 1, \ldots, N$ denotes individuals and $t = 1, 2, \ldots$ as usual denotes time. The important point here is that the autoregressive coefficients $\delta_{i,2}$ differ across individuals. Some individuals have coefficients $\delta_{i,2} \approx 0$ and are referred to as “floating” voters,
whereas others have coefficients $\delta_{i,2} \approx 1$ and are referred to as “committed” voters.\(^1\)

If it is assumed that the distribution of $\delta_{i,2}$ across individuals in the population follows a Beta$(u, v)$ distribution, then the aggregate vote share or polling share $X_t = N^{-1} \sum_{i=1}^{N} x_{i,t}$ is fractionally integrated of order $d = 1 - v$ when $N$ is large, i.e., $X_t \in I(1 - v)$. For more details, see Box-Steppensmeier and Smith (1996) or Byers, Davidson, and Peel (1997, 2002).

The above theoretical argument in favor of modeling opinion poll data as fractional time series has been supported in empirical work by a large number of authors. For example, Box-Steppensmeier and Smith (1996) estimate fractional models for US data, Byers, Davidson, and Peel (1997) and Dolado, Gonzalo, and Mayoral (2002) analyze UK data, Dolado, Gonzalo, and Mayoral (2003) analyze Spanish data, Byers, Davidson, and Peel (2000) analyze data for eight countries, and Jones, Nielsen, and Popiel (2014) analyze Canadian data. All find strong evidence in support of fractional integration with estimates of $d$ around $0.6 - 0.8$. In addition, Byers, Davidson, and Peel (2007) analyze an updated version of the sample in Byers, Davidson, and Peel (1997) and show that the change to phone interviews had no effect on estimates of $d$ and did not appear to constitute a structural break.

The aggregate polling data set we analyze is from the on-going YouGov daily poll of voting intention for political parties in the United Kingdom. Each business day survey participants are asked the question:

“If there were a general election tomorrow, which party would you vote for? Conservative, Labour, Liberal Democrat, Scottish Nationalist/Plaid

\(^1\)“Floating” voters are defined as those who do not have a strong alliance to one party and may be more easily swayed by current events, media, etc., and “committed” voters, on the other hand, are those who consistently vote for a particular party and are less inclined to change their voting preference.
4.2. FRACTIONAL INTEGRATION, POLLING DATA, AND SUMMARY STATISTICS

Table 4.1: Summary statistics (data in percentage)

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
<th>Kurt</th>
<th>Start date</th>
<th>End date</th>
<th>Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conservative</td>
<td>34.67</td>
<td>3.29</td>
<td>27</td>
<td>44</td>
<td>0.76</td>
<td>2.89</td>
<td>2010/05/14</td>
<td>2015/05/06</td>
<td>1227</td>
</tr>
<tr>
<td>Labour</td>
<td>39.39</td>
<td>3.32</td>
<td>30</td>
<td>45</td>
<td>-0.44</td>
<td>2.33</td>
<td>2010/05/14</td>
<td>2015/05/06</td>
<td>1227</td>
</tr>
<tr>
<td>Lib. Dem.</td>
<td>9.41</td>
<td>1.87</td>
<td>5</td>
<td>21</td>
<td>1.42</td>
<td>7.75</td>
<td>2010/05/14</td>
<td>2015/05/06</td>
<td>1227</td>
</tr>
<tr>
<td>UKIP</td>
<td>11.65</td>
<td>2.82</td>
<td>5</td>
<td>19</td>
<td>-0.18</td>
<td>2.30</td>
<td>2012/04/16</td>
<td>2015/05/06</td>
<td>771</td>
</tr>
<tr>
<td>Green</td>
<td>3.40</td>
<td>1.75</td>
<td>1</td>
<td>10</td>
<td>0.91</td>
<td>2.85</td>
<td>2012/06/18</td>
<td>2015/05/06</td>
<td>729</td>
</tr>
</tbody>
</table>

Notes: The table presents summary statistics for the polling data (expressed in percentages). The statistics presented are the sample mean, standard deviation, minimum, maximum, skewness, kurtosis, start and end dates, and number of observations.

Cymru, some other party, would not vote, don’t know?”

If the respondent replied “some other party”, he/she would then be presented with a list of prompted alternatives, at which point they would be able to select United Kingdom Independence Party, Green Party, and so on.

This poll, and hence the data series, is business-daily and began on May 14th, 2010 (so shortly after the 55th UK general election of 2010 held on May 6th). With the next general election held on May 7th, 2015, this on-going survey provides a long series of polling data all within the tenure of a single government regime. The results presented in this chapter use May 6th, 2015, as the end date, which was the last day the poll was conducted before the election, for a total of 1227 business-daily observations. Previous empirical studies of political support have analyzed monthly and quarterly data spanning several decades and election cycles. Thus, this daily frequency data set is particularly attractive for estimating models within a single election cycle.

Our analysis focuses on the three major political parties in the United Kingdom:

<sup>2</sup>Starting April 7th, 2015, i.e. for the last month before the 2015 election, YouGov changed their polling frequency to daily, including non-business days. We ignore this minor change, as well as the break in polling over the Christmas holiday, and in our analysis we treat all observations as equi-distant as is standard.
the Conservative Party (CP) and the Liberal Democrats (LD), which together constitute the British government over the sample period (coalition formed on May 12th, 2010), and the Labour Party (LP)—the official opposition. Until April 15th, 2012, YouGov reported all outcomes from “some other party” in the residual time series, so that no distinction was made between, e.g., United Kingdom Independence Party (UKIP or just IP) and the Green Party (GP).

On April 16th, 2012, YouGov changed the way they reported the outcomes of their polls in their UK Polling Report, and started reporting the United Kingdom Independence Party as a separate time series (rather than it being included in the residual category). On June 18th, 2012, they also started reporting the Green Party as a separate series rather than as part of the residual category. This facilitates extending the analysis to four or even five political parties, albeit for a substantially shorter data set that spans only the second half of the 2010 to 2015 political cycle in the UK. However, unlike the three major political parties, neither the UKIP nor the Green Party are stated explicitly as choices in the survey question posed to the poll participants as quoted above, and for a survey respondent to indicate that they wish to vote for either of these parties, they must first choose “some other party” after which they are presented with a list of prompted alternatives, at which point they are able to select UKIP, Green Party, Scottish Nationalist/Plaid Cymru, and so on. Although we include the UKIP and the Green Party in our empirical application to the 2015 UK general election, this characteristic of the survey, together with the substantially shorter sample size, leads us to exclude these two parties from our main forecasting analysis. In Table 4.1 we present some summary statistics for the polling data, where these are given in percentage vote shares.
4.2. FRACTIONAL INTEGRATION, POLLING DATA, AND SUMMARY STATISTICS

Table 4.2: Summary statistics (logit transformed data)

<table>
<thead>
<tr>
<th>Series</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
<th>Kurt</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>−0.63</td>
<td>0.14</td>
<td>−0.99</td>
<td>−0.24</td>
<td>0.67</td>
<td>2.81</td>
</tr>
<tr>
<td>LP</td>
<td>−0.43</td>
<td>0.14</td>
<td>−0.84</td>
<td>−0.20</td>
<td>−0.50</td>
<td>2.41</td>
</tr>
<tr>
<td>LD</td>
<td>−2.28</td>
<td>0.20</td>
<td>−2.94</td>
<td>−1.32</td>
<td>0.50</td>
<td>4.59</td>
</tr>
<tr>
<td>IP</td>
<td>−2.05</td>
<td>0.29</td>
<td>−2.94</td>
<td>−1.45</td>
<td>−0.59</td>
<td>2.54</td>
</tr>
<tr>
<td>GP</td>
<td>−3.47</td>
<td>0.51</td>
<td>−4.59</td>
<td>−2.19</td>
<td>0.21</td>
<td>2.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELW(m)</th>
<th>(m = T^{0.6})</th>
<th>(m = T^{0.7})</th>
<th>(m = T^{0.8})</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>0.79</td>
<td>0.70</td>
<td>0.65</td>
</tr>
<tr>
<td>LP</td>
<td>0.88</td>
<td>0.71</td>
<td>0.64</td>
</tr>
<tr>
<td>LD</td>
<td>0.85</td>
<td>0.66</td>
<td>0.64</td>
</tr>
<tr>
<td>IP</td>
<td>0.75</td>
<td>0.69</td>
<td>0.62</td>
</tr>
<tr>
<td>GP</td>
<td>0.73</td>
<td>0.63</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Notes: The table presents summary statistics for the logit transform of the polling data. The start and end dates are the same as in Table 4.1, as are the statistics presented, with the addition of \(\text{ELW}(m)\), which denotes the exact local Whittle estimator of Shimotsu and Phillips (2005) with bandwidth parameter \(m\), whose asymptotic standard error is \((4m)^{-1/2}\).

The analysis proceeds after converting the polling data to log-odds. This is done to map variables on the unit interval into the real line, in order to use error terms with unbounded support in our models (for more details and background, see e.g. Byers, Davidson, and Peel (1997)). The log-odds or logit transformation for a variable \(Y_t \in (0, 1)\) is

\[
y_t = \log \left( \frac{Y_t}{1 - Y_t} \right),
\]

where \(Y_t\) is the original series and \(y_t\) is the logit transformed series with support \((−\infty, \infty)\). Table 4.2 presents summary statistics for the logit transformed data. The original data and the logit transform of the data are shown in Figures 4.2(a) and 4.2(b), respectively.

As mentioned earlier, the fractional integration characteristic of political opinion polling data has been well established in the literature for monthly and quarterly data. To add to this body of literature, we computed the sample autocorrelation functions and estimated spectral density functions of each of the three series, and we display these in Figures 4.3(a) and 4.3(b), respectively. For a fractionally integrated time series, we would expect that the sample autocorrelation functions decay very...
Figure 4.1: Time series plots of data 2010-05-14 – 2015-05-06

(a) Percentage

(b) Logit transform

Note: Black line is Conservative (CP), red line is Labour (LP), blue line is Liberal Democrats (LD), purple line is UKIP (IP), and green line is the Green Party (GP).
4.2. FRACTIONAL INTEGRATION, POLLING DATA, AND SUMMARY STATISTICS

Figure 4.2: Serial dependence structure

(a) Sample autocorrelation functions

(b) Estimated spectral density functions

slowly (hyperbolically, as opposed to exponentially) and that the estimated spectral density functions have mass concentrated near the origin. See, e.g., Baillie (1996) for a review covering these properties. Indeed, both these features appear clearly in Figure 4.2.

Finally, for each univariate series we have computed semiparametric estimates of the fractional integration parameter, \(d\), that do not rely on the specification of a parametric model or lag structure. Specifically, we computed the exact local Whittle (ELW) estimates of Shimotsu and Phillips (2005), which are displayed in the last three columns in Table 4.2 for three different choices of bandwidth parameter, \(m = T^{0.6}\), \(m = T^{0.7}\), and \(m = T^{0.8}\) (when this is not an integer, the integer part of the result
is used). In each case, the asymptotic standard error of the estimate is \((4m)^{-1/2}\), so, for example, when \(m = T^{0.6}\) the asymptotic standard error of the estimate is 0.059 for the first three series where \(T = 1227\). The results from the ELW estimates suggest that each series is fractionally integrated with a fractional integration parameter that is statistically significantly less than one.

More generally, the ELW estimates in Table 4.2 are in line with estimates from the literature, e.g., Box-Steffensmeier and Smith (1996), Byers, Davidson, and Peel (1997, 2000), Dolado, Gonzalo, and Mayoral (2002, 2003), and Jones, Nielsen, and Popiel (2014), where estimates around 0.6 – 0.8 are commonly found for polling data at the monthly and quarterly frequencies. An important property of fractional processes is self-similarity, in other words that the autocorrelation structure is independent of sampling frequency (unlike exponential decay models such as those of the ARMA type). The fact that the fractional parameters estimated here using daily data are generally very close to estimates obtained using monthly and quarterly data is thus another important reason for favouring the fractional approach. The evidence presented here clearly shows fractional integration characteristics for political opinion polls at the daily frequency, as suggested by both the self-similarity property and the theoretical (aggregation-based) arguments discussed earlier.

4.3 Statistical methodology: FCVAR model

Our analysis applies the FCVAR model of Johansen (2008) and Johansen and Nielsen (2012). This model is a generalization of Johansen’s (1995) CVAR model to allow for fractionally integrated and fractionally cointegrated time series.
4.3. STATISTICAL METHODOLOGY: FCVAR MODEL

4.3.1 Variants of the FCVAR model and interpretations

For a time series $Y_t$ of dimension $p$, the well-known CVAR model with a so-called “restricted constant” term is given in error correction form as

$$
\Delta Y_t = \alpha(\beta'Y_{t-1} + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \varepsilon_t = \alpha L(\beta'Y_t + \rho') + \sum_{i=1}^{k} \Gamma_i L^i \Delta Y_t + \varepsilon_t, \quad (4.4)
$$

where, as usual, $\varepsilon_t$ is a $p$-dimensional independent and identically distributed error term with mean zero and covariance matrix $\Omega$. The simplest way to derive the FCVAR model from the CVAR model is to replace the difference and lag operators, $\Delta$ and $L = 1 - \Delta$, in (4.4) by their fractional counterparts, $\Delta^b$ and $L_b = 1 - \Delta^b$, respectively, and apply the resulting model to $Y_t = \Delta^d X_t$. We then obtain

$$
\Delta^d X_t = \alpha \Delta^d L_b(\beta'X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L^i_b X_t + \varepsilon_t, \quad (4.5)
$$

where $\Delta^d$ is the fractional difference operator, and $L_b = 1 - \Delta^b$ is the fractional lag operator.

Model (4.5) nests Johansen’s (1995) CVAR model in (4.4) as the special case $d = b = 1$. Some of the parameters are well-known from the CVAR model and these have the usual interpretations also in the FCVAR model. The most important of these are the long-run parameters $\alpha$ and $\beta$, which are $p \times r$ matrices with $0 \leq r \leq p$, and $\rho$, which is an $r$-vector. The rank $r$ is termed the cointegration, or sometimes cofractional, rank. The columns of $\beta$ constitute the $r$ cointegration (cofractional) vectors, such

---

3In principle, the restricted constant term should be included as $\rho' \pi_t(1)$, where $\pi_t(1)$ denotes the coefficient in (4.1). This is mathematically convenient, but makes no difference in terms of the practical implementation because the infinite summation in (4.1) needs to be truncated in practice.

4Both the fractional difference and fractional lag operators are defined in terms of their binomial expansion in the lag operator, $L$, as in (4.1). Note that the expansion of $L_b$ has no term in $L^0$ and thus only lagged disequilibrium errors appear in (4.5).
that $\beta'X_t$ are the cointegrating combinations of the variables in the system, i.e. the long-run equilibrium relations. The parameters in $\alpha$ are the adjustment or loading coefficients which represent the speed of adjustment towards equilibrium for each of the variables. The restricted constant term $\rho$ is interpreted as the mean level of the long-run equilibria $\beta'X_t$ when these are stationary. The short-run dynamics of the variables are governed by the parameters $($\Gamma_1, \ldots, \Gamma_k$) in the autoregressive augmentation.

The FCVAR model has two additional parameters compared with the CVAR model, namely the fractional parameters $d$ and $b$. Here, $d$ denotes the fractional integration order of the observable time series, while the parameter $b$ determines the degree of fractional cointegration, i.e. the reduction in fractional integration order of $\beta'X_t$ compared to $X_t$ itself. Both fractional parameters are estimated jointly with the other parameters, see Section 4.3.2. The FCVAR model (4.5) thus has the same main structure as the standard CVAR model (4.4), in that it allows for modeling of both cointegration and adjustment towards equilibrium, but is more general since it accommodates fractional integration and cointegration.

We note that the fractional difference as defined in (4.1) is an infinite series, but any observed sample will include only a finite number of observations. This makes calculation of the fractional differences as defined in (4.1) impossible. In practice, therefore, the summation in (4.1) would need to be truncated at $n = t - 1$, and the bias introduced by application of such a truncation is analyzed by Johansen and Nielsen (2016) using higher-order expansions in a simpler model. They show, albeit in a simpler model, that this bias can be avoided by including a level parameter, $\mu$, that shifts each of the series by a constant. We follow this suggestion and also
4.3. STATISTICAL METHODOLOGY: FCVAR MODEL

consider the unobserved components formulation

\[ X_t = \mu + X_0^t, \quad \Delta^d X_0^t = \alpha \Delta^{d-b} L_b \beta' X_0^t + \sum_{i=1}^{k} \Gamma_i \Delta^{d} L_b^i X_0^t + \varepsilon_t, \quad (4.6) \]

from which we easily derive the model

\[ \Delta^d (X_t - \mu) = \alpha \beta' \Delta^{d-b} L_b (X_t - \mu) + \sum_{i=1}^{k} \Gamma_i \Delta^{d} L_b^i (X_t - \mu) + \varepsilon_t. \quad (4.7) \]

The formulation (4.7) includes the restricted constant, which may be obtained as \( \rho' = -\beta' \mu \). More generally, the level parameter \( \mu \) in (4.7) is meant to accommodate a non-zero starting point for the first observation on the process, i.e., for \( X_1 \); see Johansen and Nielsen (2016).

Our forecasting analysis applies the versions of the FCVAR model given in (4.5) and (4.7) and we provide comparisons with the CVAR model in (4.4) as our multivariate benchmark model. Following the work of Jones, Nielsen, and Popiel (2014) we also consider the sub-models obtained by setting \( d = b \) in (4.5) and (4.7), which results in disequilibrium errors that are \( I(0) \). Thus, the four variants of the FCVAR model that we consider are

1. FCVAR\(_{d,b,\rho}\): model (4.5) with restricted constant \( \rho \) and fractional parameters \( d \) and \( b \),
2. FCVAR\(_{d,b,\mu}\): model (4.7) with level parameter \( \mu \) and fractional parameters \( d \) and \( b \),
3. FCVAR\(_{d=b,\rho}\): model (4.5) with restricted constant \( \rho \) and fractional parameter \( d = b \),
4. FCVAR\(_{d=b,\mu}\): model (4.7) with level parameter \( \mu \) and fractional parameter \( d = b \).
4.3. STATISTICAL METHODOLOGY: FCVAR MODEL

In each model the fractional parameters are estimated as described in the next sub-section, possibly with the restriction \( d = b \) imposed as appropriate.

4.3.2 Maximum likelihood estimation

The models (4.5) and (4.7) are estimated by conditional maximum likelihood. It is assumed that a sample of length \( T + N \) is available on \( X_t \), where \( N \) denotes the number of observations used for conditioning, for details see Johansen and Nielsen (2012, 2016). For the standard CVAR model, the arguments are well-known and conditioning on the first \( N \geq k + 1 \) observations leads to reduced rank regression estimation. For the FCVAR model, we proceed similarly by maximizing the conditional log-likelihood function

\[
\log L_T(\lambda) = -\frac{T p}{2} (\log(2\pi) + 1) - \frac{T}{2} \log \det \left\{ \sum_{t=N+1}^{T+N} \epsilon_t(\lambda)\epsilon_t(\lambda)^\prime \right\},
\]

where the residuals are defined as

\[
\epsilon_t(\lambda) = \Delta^d X_t - \alpha \Delta^{d-b} L_b(\beta' X_t + \rho') - \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t, \quad \lambda = (d, b, \alpha, \beta, \Gamma_i, \rho),
\]

for model (4.5), and hence also for sub-models of model (4.5), such as that obtained under the restriction \( d = b \). For model (4.7) the residuals are

\[
\epsilon_t(\lambda) = \Delta^d (X_t - \mu) - \alpha \beta' \Delta^{d-b} L_b(X_t - \mu) - \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i (X_t - \mu), \quad \lambda = (d, b, \alpha, \beta, \Gamma_i, \mu),
\]

and similarly for sub-models of model (4.7).

It is shown in Johansen and Nielsen (2012) how, for fixed \( (d, b) \), the estimation of model (4.5) reduces to regression and reduced rank regression as in Johansen (1995).
In this way the parameters \((\alpha, \beta, \Gamma_i, \rho)\) can be concentrated out of the likelihood function, and numerical optimization is only needed to optimize the profile likelihood function over the two fractional parameters, \(d\) and \(b\). In model (4.7) we can similarly concentrate the parameters \((\alpha, \beta, \Gamma_i)\) out of the likelihood function resulting in numerical optimization over \((d, b, \mu)\), thus making the estimation of model (4.7) somewhat more involved numerically than that of model (4.5).

The asymptotic analysis of the FCVAR model is provided in Johansen and Nielsen (2012). For model (4.5), Johansen and Nielsen (2012) show that asymptotic theory is standard when \(b < 0.5\), and for the case \(b > 0.5\) asymptotic theory is non-standard and involves fractional Brownian motion of type II. Specifically, when \(b > 0.5\), Johansen and Nielsen (2012) show that under i.i.d. errors with suitable moment conditions, the conditional maximum likelihood parameter estimates \((\hat{d}, \hat{b}, \hat{\alpha}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k)\) are asymptotically Gaussian, while \((\hat{\beta}, \hat{\rho})\) are locally asymptotically mixed normal. These results allow asymptotically standard (chi-squared) inference on all parameters of the model, including the cointegrating relations and orders of fractionality, using quasi-likelihood ratio tests. As in the CVAR model, see Johansen (1995), the same results hold for the same parameters in the model (4.7), whereas the asymptotic distribution theory for the remaining parameter, \(\mu\), is currently unknown.

Likelihood ratio (trace-type) tests for cointegration rank can be calculated as well, and hypotheses on the cointegration rank can be tested in the same way as in the CVAR model. In the FCVAR model, the asymptotic distribution of the tests for cointegration rank depends on the unknown (true value of the) scalar parameter \(b\), which complicates empirical analysis compared to the CVAR model. However, the distribution can be simulated on a case-by-case basis, or the computer programs by
MacKinnon and Nielsen (2014) can be used to obtain either critical values or $P$ values
for the rank test. The calculation of maximum likelihood estimators and test statistics
is discussed in detail in Johansen and Nielsen (2012) and Nielsen and Popiel (2016),
with the latter providing Matlab computer programs that we apply in our empirical
analysis.

4.3.3 Forecasting from the FCVAR model

We now discuss how to forecast the (logit transformed) polling data, that is $X_t$, from the FCVAR model (since the CVAR model is a special case obtained as $d = b = 1$, forecasts from that model are derived in the same way). Because the model
is autoregressive, the best (minimum mean squared error) linear predictor takes a
simple form and is relatively straightforward to calculate. We first note that

\[ \Delta^d(X_{t+1} - \mu) = X_{t+1} - \mu - (X_{t+1} - \mu) + \Delta^d(X_{t+1} - \mu) = X_{t+1} - \mu - L_d(X_{t+1} - \mu) \]

and then rearrange (4.7) as

\[ X_{t+1} = \mu + L_d(X_{t+1} - \mu) + \alpha \beta^\prime \Delta^{d-b} L_b(X_{t+1} - \mu) + \sum_{i=1}^k \Gamma_i \Delta^d L_i^b(X_{t+1} - \mu) + \varepsilon_{t+1}, \quad (4.11) \]

Since $L_b = 1 - \Delta^b$ is a lag operator, so that $L_b^i X_{t+1}$ is known at time $t$ for $i \geq 1$, this
equation can be used as the basis to calculate forecasts from the model.

We let conditional expectation given the information set at time $t$ be denoted
$E_t(\cdot)$, and the best (minimum mean squared error) linear predictor forecast of any
variable $Z_{t+1}$ given information available at time $t$ be denoted $\hat{Z}_{t+1|t} = E_t(Z_{t+1})$.
Clearly, we then have that the forecast of the innovation for period $t + 1$ at time $t$ is
\[ \hat{\varepsilon}_{t+1|t} = E_t(\varepsilon_{t+1}) = 0, \]
and $\hat{X}_{t+1|t}$ is then easily found from (4.11). Inserting coefficient
estimates based on data available up to time $t$, denoted\(^5\) $(\hat{d}, \hat{b}, \hat{\mu}, \hat{\beta}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k)$, we have that

$$
\hat{X}_{t+1|t} = \hat{\mu} + L_{\hat{d}}(X_{t+1} - \hat{\mu}) + \hat{\alpha}\hat{\beta}'\Delta L_{\hat{b}}^\hat{d}(X_{t+1} - \hat{\mu}) + \sum_{i=1}^{k} \hat{\Gamma}_i \Delta L_{\hat{b}}^i(X_{t+1} - \hat{\mu}). \tag{4.12}
$$

This defines the one-step ahead forecast of $X_{t+1}$ given information at time $t$.

Multi-period ahead forecasts can be generated recursively. That is, to calculate the $h$-step ahead forecast, we first generalize (4.12) as

$$
\hat{X}_{t+j|t} = \hat{\mu} + L_{\hat{d}}(\hat{X}_{t+j|t} - \hat{\mu}) + \hat{\alpha}\hat{\beta}'\Delta L_{\hat{b}}(\hat{X}_{t+j|t} - \hat{\mu}) + \sum_{i=1}^{k} \hat{\Gamma}_i \Delta L_{\hat{b}}^i(\hat{X}_{t+j|t} - \hat{\mu}), \tag{4.13}
$$

where $\hat{X}_{s|t} = X_s$ for $s \leq t$. Then forecasts are calculated recursively from (4.13) for $j = 1, 2, \ldots, h$ to generate $h$-step ahead forecasts, $\hat{X}_{t+h|t}$.

Clearly, one-step ahead and $h$-step ahead forecasts for the model (4.5) with a restricted constant term instead of the level parameter can be calculated entirely analogously. We will apply the forecasts (4.12) and (4.13) for both models (4.5) and (4.7) in our analysis below for several forecast horizons, $h$.

4.4 Forecasting analysis

In this section we present and discuss our main forecasting analysis. In the first subsection we briefly discuss some preliminary estimation results to introduce and compare the different variants of the FCVAR model, and the next two subsections then present the forecasting procedure and the corresponding results.

\(^5\)To emphasize that these estimates are based on data available at time $t$, they could be denoted by a subscript $t$. However, to avoid cluttering the notation we omit this subscript and let it be understood in the sequel.
4.4. FORECASTING ANALYSIS

4.4.1 Preliminary estimation results

In Table 4.3 we present some preliminary estimation results for the three-party data set that includes CP, LP, and LD. In Section 4.5.1 we consider a more detailed analysis of estimation results from the FCVAR model for a five-party data set, but for now the results in Table 4.3 will serve as illustrations of the FCVAR model variants.

Each panel of Table 4.3 shows FCVAR estimation results for one of the four variants considered. Specifically, Panel A shows results for model (4.5) with two fractional parameters, \((d, b)\), and a restricted constant term, \(\rho\), Panel B shows results for model (4.7) with two fractional parameters, \((d, b)\), and the level parameter, \(\mu\), while Panels C and D show the corresponding results for the models with only one fractional parameter, \(d = b\). In addition, the multivariate Portmanteau Q-test for serial correlation up to order six in the residuals is reported as \(Q_{\hat{e}}\) and the maximized log-likelihood value is reported as \(\log L\). Standard errors are in parentheses below \(\hat{d}\) and \(\hat{b}\) and \(P\) values are in parentheses below \(Q_{\hat{e}}\).

Since it is infeasible to present results for all the many models and training sets on which our forecasting analysis is based, the results in Table 4.3 are based simply on the full sample of size \(T + N = 1227\). In the models with a restricted constant term in Panels A and C, the first \(N = 20\) observations are used as initial values in the estimation, while in the models with a level parameter in Panels B and D, we follow Johansen and Nielsen (2016) and set \(N = 0\).

For each model we initially chose the lag order by the Bayesian information criterion, and then conducted cointegration rank tests. The model was then estimated and the residuals tested for white noise (multivariate Portmanteau Q-test reported), which was not rejected for three of the four models with a lag order of \(k = 1\) and
4.4. FORECASTING ANALYSIS

Table 4.3: Preliminary estimation results: three parties

<table>
<thead>
<tr>
<th>Panel</th>
<th>FCVAR model, $k = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>FCVAR$_{d,b,\rho}$ model, $k = 1$</td>
</tr>
<tr>
<td></td>
<td>$\hat{d} = 0.774, \hat{b} = 0.094, \hat{\alpha} = \begin{bmatrix} 0.532 &amp; -0.810 \ 0.405 &amp; 0.034 \ 5.517 &amp; 10.748 \end{bmatrix}$, $\hat{\beta} = \begin{bmatrix} 1.000 &amp; 0.000 \ 0.000 &amp; 1.000 \ 2.247 &amp; -1.962 \end{bmatrix}$, $\hat{\rho} = [0.019 -1.536]$, $\hat{\Gamma} = \begin{bmatrix} -6.011 &amp; 0.970 &amp; -2.485 \ -0.250 &amp; -5.686 &amp; -0.684 \ -5.875 &amp; -10.224 &amp; 2.080 \end{bmatrix}$, $Q_{\hat{\varepsilon}} = 50.111$, $\log \mathcal{L} = 4898.335$</td>
</tr>
<tr>
<td>B</td>
<td>FCVAR$_{d,b,\mu}$ model, $k = 1$</td>
</tr>
<tr>
<td></td>
<td>$\hat{d} = 0.624, \hat{b} = 0.273, \hat{\alpha} = \begin{bmatrix} 0.009 &amp; -0.156 \ 0.448 &amp; 0.647 \ 0.064 &amp; -0.354 \end{bmatrix}$, $\hat{\beta} = \begin{bmatrix} 1.000 &amp; 0.000 \ 0.000 &amp; 1.000 \ 0.947 &amp; -0.748 \end{bmatrix}$, $\hat{\mu} = \begin{bmatrix} -0.396 \ -0.660 \ -1.597 \end{bmatrix}$, $\hat{\Gamma} = \begin{bmatrix} -1.278 &amp; 0.283 &amp; -0.033 \ -0.367 &amp; -2.105 &amp; 0.074 \ -0.308 &amp; 0.478 &amp; -1.924 \end{bmatrix}$, $Q_{\hat{\varepsilon}} = 57.148$, $\log \mathcal{L} = 4965.089$</td>
</tr>
<tr>
<td>C</td>
<td>FCVAR$_{d=b,\rho}$ model, $k = 1$</td>
</tr>
<tr>
<td></td>
<td>$\hat{d} = 0.627, \hat{\alpha} = \begin{bmatrix} -0.024 &amp; -0.024 \ -0.013 &amp; 0.005 \ -0.100 &amp; 0.132 \end{bmatrix}$, $\hat{\beta} = \begin{bmatrix} 1.000 &amp; 0.000 \ 0.000 &amp; 1.000 \ 0.370 &amp; -0.922 \end{bmatrix}$, $\hat{\rho} = [1.955 -1.652]$, $\hat{\Gamma} = \begin{bmatrix} -0.552 &amp; 0.054 &amp; 0.029 \ 0.049 &amp; -0.587 &amp; 0.031 \ 0.049 &amp; -0.031 &amp; -0.554 \end{bmatrix}$, $Q_{\hat{\varepsilon}} = 76.692$, $\log \mathcal{L} = 4875.513$</td>
</tr>
<tr>
<td>D</td>
<td>FCVAR$_{d=b,\mu}$ model, $k = 1$</td>
</tr>
<tr>
<td></td>
<td>$\hat{d} = 0.572, \hat{\alpha} = \begin{bmatrix} -0.022 &amp; -0.018 \ 0.079 &amp; 0.055 \ -0.058 &amp; -0.048 \end{bmatrix}$, $\hat{\beta} = \begin{bmatrix} 1.000 &amp; 0.000 \ 0.000 &amp; 1.000 \ 2.403 &amp; -3.679 \end{bmatrix}$, $\hat{\mu} = \begin{bmatrix} -0.411 \ -0.653 \ -1.574 \end{bmatrix}$, $\hat{\Gamma} = \begin{bmatrix} -0.497 &amp; 0.075 &amp; 0.027 \ -0.032 &amp; -0.652 &amp; 0.022 \ -0.061 &amp; 0.119 &amp; -0.694 \end{bmatrix}$, $Q_{\hat{\varepsilon}} = 60.163$, $\log \mathcal{L} = 4962.473$</td>
</tr>
</tbody>
</table>

Notes: The table shows estimation results for the four variations of the FCVAR model. The multivariate Portmanteau Q-test for serial correlation up to order six in the residuals is reported as $Q_{\hat{\varepsilon}}$ and the maximized log-likelihood value is reported as $\log \mathcal{L}$. Standard errors are in parentheses below $\hat{d}$ and $\hat{b}$ and the $P$ value is in parenthesis below $Q_{\hat{\varepsilon}}$. The sample size is $T + N = 1227$, and the first $N = 20$ ($N = 0$) observations are used as initial values in the models with a restricted constant term (level parameter).
cointegration rank of $r = 2$. For the FCVAR$_{d=b,\rho}$ model, the Q-test $P$ value is 0.023, but after adding an additional lag the model estimation failed to converge to sensible parameter values, so we chose to maintain the $k = 1$ structure as chosen by the Bayesian information criterion, and we conclude that the model appears to be correctly specified.

We note that the FCVAR models impose the same value of the fractional integration parameter, $d$, for each time series in the system. This restriction can be tested in a local Whittle or exact local Whittle framework, see Robinson and Yajima (2002) and Nielsen and Shimotsu (2007), respectively. Both require the selection of several bandwidth parameters and can be quite sensitive to these. We applied the latter methodology because it allows for nonstationary values of $d$, and we were not able to reject the hypothesis of equality of the $d$ parameters for the three series using a range of bandwidth parameters (unreported $T_0$ statistics are all less than 0.84 in the notation of Nielsen and Shimotsu (2007)). Furthermore, we note that if the $d$ parameters for each of the univariate time series were in fact different, imposing the same $d$ in the FCVAR model would be an important source of mis-specification, which would lead to neglected serial correlation (of the fractional integration type) in the residuals. Since the residuals in our estimated models do not display such strong signs of serial correlation, this does not appear to be an issue.

We will discuss and interpret estimated parameters in detail for a larger model in Section 4.5.1 below, in particular $\hat{\alpha}$ and $\hat{\beta}$. For now, we note from Table 4.3 that $\hat{d}$ is very strongly significantly different from one and that $\hat{b}$ is strongly significantly different from both zero and one in all models. This suggests very clearly that the FCVAR model is more appropriate for this data than the non-fractional CVAR model.
since the latter has $d = b = 1$ imposed. Comparing across the models, it appears that
the estimates of $d$ are fairly close, ranging from 0.57 to 0.77, whereas the estimates
of $b$ are quite different in the models with $d \neq b$ in Panels A and B of Table 4.3.

Further comparison across models leads to consideration of the likelihood ratio
test statistic for the null hypothesis that $d = b$. That is, for the models with a
restricted constant term (and $N = 20$), we can test the null of the model in Panel
C against the alternative of the model in Panel A, and for the models with a level
parameter (and $N = 0$), we can test the null of the model in Panel D against the
alternative of the model in Panel B. In the first case, the likelihood ratio test statistic
is 45.644, and in the second case, the likelihood ratio test statistic is 5.232. In both
cases, this is asymptotically chi-squared distributed with one degree of freedom, so
the conclusions of these tests differ somewhat and consequently we proceed with the
consideration of all four models.

4.4.2 Forecasting methodology

The four variants of the FCVAR model presented above are evaluated on their fore-
casting ability relative to a group of popular competing models. The CVAR model
(4.4) serves as the multivariate benchmark model. The main univariate benchmark
is the ARFIMA$(p, d, q)$ model,

\[ A(L) \Delta^d (X_t - \mu) = B(L) \varepsilon_t, \quad (4.14) \]

where $A(L)$ and $B(L)$ are the autoregressive and moving average polynomials, satisfying standard regularity conditions. A special case of (4.14) is the simple ARFIMA$(0, d, 0)$
model, which is also included because it was found by, e.g., Byers, Davidson, and Peel
(1997) and Dolado, Gonzalo, and Mayoral (2002), to fit (monthly) UK polling data
4.4. FORECASTING ANALYSIS

well. Finally, another special case of (4.14) is the standard ARMA($p, q$) model, which we include as the classical univariate benchmark. Estimation of the univariate models is by minimization of the conditional sum-of-squares, see Box and Jenkins (1970) for ARMA models and Hualde and Robinson (2011) and Nielsen (2015) for ARFIMA models, while forecasting is done using the best (minimum mean squared error) linear predictor which appears standard for these models.

The forecasting procedure applies the standard rolling window and recursive forecasting schemes to examine forecasting accuracy. The main distinction between the two schemes is how they select the training sets used for estimating each model to produce forecasts. The rolling window scheme uses a fixed training set length (usually referred to as a window) that moves across the data set. The recursive scheme uses an expanding training set length with a fixed start date at the beginning of the data set. In order to assess the forecasting capability of each model, it is necessary to generate predictions from a sufficiently large number of training sets used to estimate each model, and it is preferable that each training set is long enough for reliable estimation and forecasting.

For the rolling window scheme, each statistical model uses training sets with a fixed length of $T + N = 600$ observations, approximately half the length of the data set. For the recursive scheme, the first training set includes $T + N = 600$ observations and each subsequent training set includes one more observation, until the last training set which has $T + N = 1227 - h$ observations, where $h$ is the forecast horizon. This implies that for both forecasting schemes, the total number of training sets is $1227 - 600 - h + 1 = 628 - h$, and the first training set is the same for both procedures. For the FCVAR models with a restricted constant term we use the first
$N = 20$ observations of each training set as initial values, and for the FCVAR models with the level parameter we follow Johansen and Nielsen (2016) and use $N = 0$ initial values. The CVAR model applies estimation conditional on $N = k + 1$ initial values, such that maximum likelihood estimation reduces to reduced rank regression.

The forecasting programs use all applicable model specification criteria and tests consistently across both the multivariate and univariate model types, and all rejection rules for statistical hypothesis testing are conducted at the five percent level of significance. For all models, the model specification is based on the very first training set, and the same specification is then applied to all training sets. That is, we maintain the same lag orders and cointegration ranks for all training sets, but all the parameters of the models are re-estimated for each training set before forecasts are calculated.

The multivariate specifications involve first selecting the lag order, $k$. Lag order selection is initially based on the Bayesian information criterion (BIC). Given the lag order, cointegration rank tests are performed, which determine the number of cointegrating relations, $r$, for each model. The multivariate model is then estimated using these values of $k$ and $r$. In the next step, the program performs a multivariate Portmanteau Q-test for white noise up to order six on the residuals. If the white noise test rejects, then an additional lag is added and the rank test, estimation, and white noise test are repeated in sequence until the program fails to reject white noise for the residuals. The univariate specification differs from the multivariate only in that two lag orders, $p$ and $q$, need to be selected conditional on the univariate white noise test failing to reject.

Following the completion of the specification algorithm, the forecasting program
estimates the model for all training sets and uses the estimated model parameters to generate $h$-step ahead forecasts for each time series. All multivariate and univariate models considered generate forecasts recursively, see Section 4.3.3. The ARFIMA($0,d,0$) model is included in the portfolio due to its popularity for political opinion poll data. Its fixed lag orders make it the only model in the portfolio with lag orders not determined by a decision rule in the forecasting program.

Forecasts are generated for seven out-of-sample horizons, $h$: 1, 5, 10, 15, 20, 25, and 50 steps ahead. As mentioned above, the number of training sets is different for each forecast horizon, and we denote this number $M_h$. The models are ranked based on the multivariate (system) root mean squared forecast error,

$$\text{RMSFE}_{\text{sys}} = \sqrt{\frac{1}{pM_h} \sum_{i=1}^{p} \sum_{j=1}^{M_h} \left( \hat{X}_{i,T_j+h|T_j} - X_{i,T_j+h} \right)^2},$$  \hspace{1cm} (4.15)

as well as the univariate root mean squared forecast errors for each series,

$$\text{RMSFE}_i = \sqrt{\frac{1}{M_h} \sum_{j=1}^{M_h} \left( \hat{X}_{i,T_j+h|T_j} - X_{i,T_j+h} \right)^2},$$  \hspace{1cm} (4.16)

where, in both cases, $h$ denotes the forecast horizon, $p = 3$ is the number of series, i.e., the dimension of the multivariate system, $M_h = 628 - h$ is the number of training sets, and $T_j$ is the terminal date of training set $j$. The individual $\text{RMSFE}_i$ ($i = \text{CP, LP, LD}$) measures the typical magnitude of forecast errors for each individual time series, while the $\text{RMSFE}_{\text{sys}}$ measures the typical magnitude of all forecast errors produced by each model.
4.4.3 Forecasting results

This section discusses the forecast performance results and concludes with several figures of forecasts generated by all models in the portfolio.

Tables 4.4 and 4.5 report the RMSFE\(_i\) \((i = \text{CP, LP, LD})\) and RMSFE\(_{\text{sys}}\) values for the rolling and recursive schemes, respectively. Numbers in parentheses are the corresponding rankings at each individual forecast horizon. The models are ranked based on the RMSFE\(_{\text{sys}}\) because for each model it provides a single measurement of forecast accuracy for all three time series. We note that the ARFIMA\((p,d,q)\) model specifies both lag orders to be zero, i.e. \(p = q = 0\), for all three series, so that the results for the ARFIMA\((p,d,q)\) and ARFIMA\((0,d,0)\) models are identical, and therefore we do not report the ARFIMA\((0,d,0)\) results. Given the results from the literature cited in Section 4.2 above, this univariate model specification is not surprising. We also note that the ARMA\((p,q)\) model specifies \((p,q) = (0,1)\) for all three series by a very slim margin over \((p,q) = (1,0)\) in terms of the BIC; the forecast performance (unreported) with \((p,q) = (1,0)\) is qualitatively very similar to that reported with \((p,q) = (0,1)\).

We begin the assessment of the forecast accuracy with a discussion of the one-step ahead forecasts. This seems natural prior to examining performance at other subjectively selected horizons that may be of interest in any given application. In the context of political opinion polls, one can easily imagine the relevance of forecasting poll standings or vote shares at many different horizons.

According to the recursive scheme, all four variants of the FCVAR model outperform all competing models at the one-step ahead forecasting horizon. According to
4.4. FORECASTING ANALYSIS

Table 4.4: Root mean squared forecast errors – rolling window forecast scheme

<table>
<thead>
<tr>
<th>Model</th>
<th>Series</th>
<th>1 step</th>
<th>5 step</th>
<th>10 step</th>
<th>15 step</th>
<th>20 step</th>
<th>25 step</th>
<th>50 step</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCVAR_{d,b,\rho}</td>
<td>CP</td>
<td>0.0562</td>
<td>0.0613</td>
<td>0.0637</td>
<td>0.0656</td>
<td>0.0678</td>
<td>0.0704</td>
<td>0.0781</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0499</td>
<td>0.0524</td>
<td>0.0569</td>
<td>0.0609</td>
<td>0.0647</td>
<td>0.0669</td>
<td>0.0690</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1134</td>
<td>0.1161</td>
<td>0.1233</td>
<td>0.1298</td>
<td>0.1358</td>
<td>0.1415</td>
<td>0.1650</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0785 (1)</td>
<td>0.0816 (2)</td>
<td>0.0866 (2)</td>
<td>0.0910 (2)</td>
<td>0.0953 (2)</td>
<td>0.0996 (2)</td>
<td>0.1176 (2)</td>
</tr>
<tr>
<td>FCVAR_{d,b,\mu}</td>
<td>CP</td>
<td>0.0561</td>
<td>0.0608</td>
<td>0.0632</td>
<td>0.0656</td>
<td>0.0677</td>
<td>0.0702</td>
<td>0.0781</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0505</td>
<td>0.0549</td>
<td>0.0607</td>
<td>0.0669</td>
<td>0.0723</td>
<td>0.0781</td>
<td>0.1030</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1174</td>
<td>0.1241</td>
<td>0.1329</td>
<td>0.1400</td>
<td>0.1445</td>
<td>0.1499</td>
<td>0.1663</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0785 (1)</td>
<td>0.0813 (1)</td>
<td>0.0858 (1)</td>
<td>0.0901 (1)</td>
<td>0.0941 (1)</td>
<td>0.0987 (1)</td>
<td>0.1162 (1)</td>
</tr>
<tr>
<td>FCVAR_{d}</td>
<td>CP</td>
<td>0.0569</td>
<td>0.0629</td>
<td>0.0664</td>
<td>0.0684</td>
<td>0.0708</td>
<td>0.0729</td>
<td>0.0782</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0506</td>
<td>0.0525</td>
<td>0.0569</td>
<td>0.0604</td>
<td>0.0645</td>
<td>0.0687</td>
<td>0.0908</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1174</td>
<td>0.1241</td>
<td>0.1329</td>
<td>0.1400</td>
<td>0.1445</td>
<td>0.1499</td>
<td>0.1663</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0808 (4)</td>
<td>0.0859 (4)</td>
<td>0.0918 (4)</td>
<td>0.0965 (4)</td>
<td>0.1002 (4)</td>
<td>0.1041 (3)</td>
<td>0.1183 (3)</td>
</tr>
<tr>
<td>CVAR_{\rho}</td>
<td>CP</td>
<td>0.0605</td>
<td>0.0656</td>
<td>0.0682</td>
<td>0.0731</td>
<td>0.0783</td>
<td>0.0828</td>
<td>0.1039</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0502</td>
<td>0.0540</td>
<td>0.0597</td>
<td>0.0653</td>
<td>0.0705</td>
<td>0.0828</td>
<td>0.1056</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1200</td>
<td>0.1260</td>
<td>0.1242</td>
<td>0.1326</td>
<td>0.1379</td>
<td>0.1450</td>
<td>0.1799</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0786 (3)</td>
<td>0.0829 (3)</td>
<td>0.0888 (3)</td>
<td>0.0952 (3)</td>
<td>0.1002 (4)</td>
<td>0.1076 (4)</td>
<td>0.1345 (5)</td>
</tr>
<tr>
<td>ARFIMA(p,d,q)</td>
<td>CP</td>
<td>0.0578</td>
<td>0.0623</td>
<td>0.0664</td>
<td>0.0700</td>
<td>0.0730</td>
<td>0.0758</td>
<td>0.0860</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0531</td>
<td>0.0638</td>
<td>0.0746</td>
<td>0.0834</td>
<td>0.0909</td>
<td>0.0977</td>
<td>0.1252</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1158</td>
<td>0.1244</td>
<td>0.1335</td>
<td>0.1410</td>
<td>0.1466</td>
<td>0.1521</td>
<td>0.1707</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0808 (4)</td>
<td>0.0884 (5)</td>
<td>0.0963 (5)</td>
<td>0.1029 (5)</td>
<td>0.1081 (5)</td>
<td>0.1132 (5)</td>
<td>0.1319 (4)</td>
</tr>
<tr>
<td>ARMA(p,q)</td>
<td>CP</td>
<td>0.0899</td>
<td>0.1167</td>
<td>0.1178</td>
<td>0.1187</td>
<td>0.1187</td>
<td>0.1190</td>
<td>0.1207</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.1109</td>
<td>0.1582</td>
<td>0.1605</td>
<td>0.1628</td>
<td>0.1650</td>
<td>0.1673</td>
<td>0.1781</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1627</td>
<td>0.1947</td>
<td>0.1957</td>
<td>0.1971</td>
<td>0.1983</td>
<td>0.1997</td>
<td>0.2049</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.1250 (7)</td>
<td>0.1597 (7)</td>
<td>0.1612 (7)</td>
<td>0.1627 (7)</td>
<td>0.1640 (7)</td>
<td>0.1653 (7)</td>
<td>0.1715 (7)</td>
</tr>
</tbody>
</table>

Notes: The overall performance of each model is measured by the root mean square forecast error of the entire multivariate system. The ARFIMA(0, d, 0) model is not included because the ARFIMA(p, d, q) model specifies both lag orders to zero for all three series. The ARMA(p, q) model specifies (p, q) = (0, 1) for all three series. Numbers in parentheses are the corresponding rankings at each individual forecast horizon. The number 1 rank is assigned to the best performing model and the number 7 rank is assigned to the worst performing model. Results are based on h-step ahead forecasts produced using 628-h training sets of length 600.

the rolling window scheme, three of the four FCVAR variants outperform all competing models, and the FCVAR_{d=b,\rho} model is tied with the ARFIMA(p, d, q) model. Both forecasting schemes rank a variant of the FCVAR model with two fractional parameters as the top performing model. The rolling window scheme ranks the FCVAR_{d,b,\mu} and FCVAR_{d,b,\rho} models first (tied), while the recursive scheme ranks the FCVAR_{d,b,\rho} model first and the FCVAR_{d=b,\rho} model second. An important observation is that the third and fourth ranked FCVAR specifications are very close in performance to
4.4. FORECASTING ANALYSIS

Table 4.5: Root mean squared forecast errors – recursive window forecast scheme

<table>
<thead>
<tr>
<th>Model</th>
<th>Series</th>
<th>1 step</th>
<th>5 step</th>
<th>10 step</th>
<th>15 step</th>
<th>20 step</th>
<th>25 step</th>
<th>50 step</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCVAR(_{d,b,\rho})</td>
<td>CP</td>
<td>0.0561</td>
<td>0.0616</td>
<td>0.0642</td>
<td>0.0667</td>
<td>0.0697</td>
<td>0.0727</td>
<td>0.0839</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0505</td>
<td>0.0548</td>
<td>0.0611</td>
<td>0.0671</td>
<td>0.0726</td>
<td>0.0784</td>
<td>0.1062</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1123</td>
<td>0.1144</td>
<td>0.1197</td>
<td>0.1251</td>
<td>0.1292</td>
<td>0.1342</td>
<td>0.1520</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0781</td>
<td>0.0814</td>
<td>0.0860</td>
<td>0.0906</td>
<td>0.0946</td>
<td>0.0991</td>
<td>0.1175</td>
</tr>
<tr>
<td>FCVAR(_{d,b,\mu})</td>
<td>CP</td>
<td>0.0563</td>
<td>0.0619</td>
<td>0.0653</td>
<td>0.0688</td>
<td>0.0728</td>
<td>0.0765</td>
<td>0.0905</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0499</td>
<td>0.0552</td>
<td>0.0564</td>
<td>0.0602</td>
<td>0.0640</td>
<td>0.0686</td>
<td>0.0939</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1144</td>
<td>0.1193</td>
<td>0.1267</td>
<td>0.1338</td>
<td>0.1387</td>
<td>0.1439</td>
<td>0.1614</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0793</td>
<td>0.0835</td>
<td>0.0895</td>
<td>0.0956</td>
<td>0.1015</td>
<td>0.1082</td>
<td>0.1390</td>
</tr>
<tr>
<td>FCVAR(_{d,\rho})</td>
<td>CP</td>
<td>0.0563</td>
<td>0.0617</td>
<td>0.0646</td>
<td>0.0672</td>
<td>0.0701</td>
<td>0.0732</td>
<td>0.0846</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0507</td>
<td>0.0550</td>
<td>0.0612</td>
<td>0.0673</td>
<td>0.0729</td>
<td>0.0786</td>
<td>0.1058</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1144</td>
<td>0.1193</td>
<td>0.1267</td>
<td>0.1338</td>
<td>0.1387</td>
<td>0.1439</td>
<td>0.1614</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0792</td>
<td>0.0838</td>
<td>0.0894</td>
<td>0.0948</td>
<td>0.0991</td>
<td>0.1037</td>
<td>0.1217</td>
</tr>
<tr>
<td>CVAR(_{\rho})</td>
<td>CP</td>
<td>0.0615</td>
<td>0.0684</td>
<td>0.0712</td>
<td>0.0697</td>
<td>0.0731</td>
<td>0.0766</td>
<td>0.0887</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0551</td>
<td>0.0595</td>
<td>0.0667</td>
<td>0.0828</td>
<td>0.0926</td>
<td>0.1017</td>
<td>0.1388</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1208</td>
<td>0.1222</td>
<td>0.1386</td>
<td>0.1427</td>
<td>0.1500</td>
<td>0.1572</td>
<td>0.1789</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0845</td>
<td>0.0878</td>
<td>0.0979</td>
<td>0.1066</td>
<td>0.1145</td>
<td>0.1218</td>
<td>0.1469</td>
</tr>
<tr>
<td>ARFIMA((p, d, q))</td>
<td>CP</td>
<td>0.0585</td>
<td>0.0627</td>
<td>0.0668</td>
<td>0.0706</td>
<td>0.0739</td>
<td>0.0770</td>
<td>0.0881</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.0517</td>
<td>0.0566</td>
<td>0.0634</td>
<td>0.0692</td>
<td>0.0744</td>
<td>0.0792</td>
<td>0.0991</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1182</td>
<td>0.1241</td>
<td>0.1359</td>
<td>0.1468</td>
<td>0.1560</td>
<td>0.1651</td>
<td>0.2067</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.0818</td>
<td>0.0867</td>
<td>0.0948</td>
<td>0.1022</td>
<td>0.1085</td>
<td>0.1147</td>
<td>0.1389</td>
</tr>
<tr>
<td>ARMA((p, q))</td>
<td>CP</td>
<td>0.1105</td>
<td>0.1503</td>
<td>0.1511</td>
<td>0.1516</td>
<td>0.1515</td>
<td>0.1517</td>
<td>0.1523</td>
</tr>
<tr>
<td></td>
<td>LP</td>
<td>0.1157</td>
<td>0.1630</td>
<td>0.1706</td>
<td>0.1719</td>
<td>0.1731</td>
<td>0.1744</td>
<td>0.1808</td>
</tr>
<tr>
<td></td>
<td>LD</td>
<td>0.1714</td>
<td>0.2124</td>
<td>0.2135</td>
<td>0.2150</td>
<td>0.2165</td>
<td>0.2180</td>
<td>0.2249</td>
</tr>
<tr>
<td></td>
<td>System</td>
<td>0.1354</td>
<td>0.1792</td>
<td>0.1803</td>
<td>0.1815</td>
<td>0.1824</td>
<td>0.1834</td>
<td>0.1884</td>
</tr>
</tbody>
</table>

Notes: The overall performance of each model is measured by the root mean square forecast error of the entire multivariate system. The ARFIMA\((0, d, 0)\) model is not included because the ARFIMA\((p, d, q)\) model specifies both lag orders to zero for all three series. The ARMA\((p, q)\) model specifies \((p, q) = (0, 1)\) for all three series. Numbers in parentheses are the corresponding rankings at each individual forecast horizon. The number 1 rank is assigned to the best performing model and the number 7 rank is assigned to the worst performing model. Results are based on \(h\)-step ahead forecasts produced using 628-\(h\) training sets of length= 600, 601, \ldots, 1227−\(h\).

The top performing variant, showing that the reliability of one-step ahead forecasts generated by the FCVAR model is robust to the number of fractional parameters and type of deterministic terms used in the specification, at least for this data set.

The longer forecast horizons considered, 5, 10, 15, 20, 25 and 50 steps ahead, deliver results that are in agreement with the findings for the one-step ahead horizon.

The model rankings across all forecast horizons determine that the accuracy of both short, medium and long term forecasts generated by the FCVAR model is better than
the other models in the portfolio. This can be seen from the fact that for 14 of 14 cases (1 to 50 steps ahead in both the rolling and the recursive schemes), the top two performing models are always variants of the FCVAR model and three of the top four models are always variants of the FCVAR model. Furthermore, with the exception of one forecast horizon (50 steps ahead) in both schemes, all four variants of the FCVAR model are ranked as the top four models. The exception occurs when only one variant of the FCVAR model underperforms by a small margin relative to the ARFIMA model. Overall, this evidence provides strong support for the application of the FCVAR model for forecasting next day (one-step), next week (5-steps), and all the way up to ten weeks ahead (50-steps) poll standings.

The results for both forecasting schemes suggest that the FCVAR model with two fractional parameters produces the smallest average forecast errors. A variant of the FCVAR model with two fractional parameters always outperforms both sub-models, the FCVAR\(_{d=b,p}\) and FCVAR\(_{d=b,\mu}\), although only by very small margins. The recursive scheme determines the FCVAR\(_{d,b,p}\) model as the absolute favorite at all forecast horizons, while the rolling window scheme ranks the FCVAR\(_{d,b,\mu}\) model as the favorite at all forecast horizons greater than one-step, and ranks the FCVAR\(_{d,b,\mu}\) and FCVAR\(_{d,b,p}\) models as the top two performing models for next day forecasting.

Thus, the model rankings strongly suggest that the forecasting accuracy of the FCVAR is better than both the fractional benchmark model (ARFIMA) and the multivariate benchmark model (CVAR). To assess the degree of relative performance, Table 4.6 reports the RMSFE percentage change,

\[
100 \left( \frac{\text{RMSFE}_{\text{sys}}(\text{FCVAR})}{\text{RMSFE}_{\text{sys}}(\text{ARFIMA})} - 1 \right),
\]  

(4.17)
4.4. FORECASTING ANALYSIS

Table 4.6: Percentage change in RMSFE$_{sys}$: FCVAR vs. ARFIMA($p, d, q$)

<table>
<thead>
<tr>
<th>Model</th>
<th>1 step</th>
<th>5 step</th>
<th>10 step</th>
<th>15 step</th>
<th>20 step</th>
<th>25 step</th>
<th>50 step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: rolling scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCVAR$_{d,b,\rho}$</td>
<td>-2.79</td>
<td>-7.68</td>
<td>-10.07</td>
<td>-11.53</td>
<td>-11.88</td>
<td>-12.05</td>
<td>-10.87</td>
</tr>
<tr>
<td>FCVAR$_{d,b,\mu}$</td>
<td>-2.90</td>
<td>-8.05</td>
<td>-10.87</td>
<td>-12.39</td>
<td>-12.92</td>
<td>-12.77</td>
<td>-11.89</td>
</tr>
<tr>
<td>FCVAR$_{d=b,\rho}$</td>
<td>-0.01</td>
<td>-2.88</td>
<td>-4.62</td>
<td>-6.24</td>
<td>-7.41</td>
<td>-8.05</td>
<td>-10.28</td>
</tr>
<tr>
<td>FCVAR$_{d=b,\mu}$</td>
<td>-2.77</td>
<td>-6.26</td>
<td>-7.82</td>
<td>-7.48</td>
<td>-7.31</td>
<td>-4.94</td>
<td>2.01</td>
</tr>
<tr>
<td>Panel B: recursive scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCVAR$_{d,b,\mu}$</td>
<td>-3.01</td>
<td>-3.66</td>
<td>-5.58</td>
<td>-6.44</td>
<td>-6.48</td>
<td>-5.68</td>
<td>0.08</td>
</tr>
<tr>
<td>FCVAR$_{d=b,\rho}$</td>
<td>-3.15</td>
<td>-3.35</td>
<td>-5.71</td>
<td>-7.26</td>
<td>-8.66</td>
<td>-9.61</td>
<td>-12.41</td>
</tr>
<tr>
<td>FCVAR$_{d=b,\mu}$</td>
<td>-2.98</td>
<td>-4.22</td>
<td>-6.58</td>
<td>-7.72</td>
<td>-8.43</td>
<td>-8.34</td>
<td>-6.53</td>
</tr>
</tbody>
</table>

Notes: Negative values favor the FCVAR model. Results are based on $h$-step ahead forecasts produced using $628-h$ training sets of length 600 (rolling scheme) or length = 600, 601, ..., 1227 − $h$ (recursive scheme).

of the FCVAR model relative to the ARFIMA model for the rolling and recursive schemes in Panels A and B, respectively. Negative values favor the FCVAR model and positive values favor the ARFIMA model.

Previous literature, as cited earlier, has extensively documented the superiority of fractional (ARFIMA) models for modeling and forecasting polling data. Compared to this important benchmark, Table 4.6 shows that the RMSFE of the FCVAR model is as much as 13% lower for the rolling scheme and 15% lower for the recursive scheme. In 54 of 56 cases in Table 4.6, the multivariate fractional model outperforms the univariate fractional model and in 16 of 56 cases, the FCVAR model delivers more than a 10% reduction in the RMSFE$_{sys}$ relative to the ARFIMA model. The gains at the longer horizons are more pronounced, and the gains appear to be larger for the FCVAR models with two fractional parameters. These results provide strong evidence that in addition to being able to model fractionally integrated time series, a property
4.4. FORECASTING ANALYSIS

Table 4.7: Percentage change in RMSFE\textsubscript{sys}: FCVAR vs. CVAR

<table>
<thead>
<tr>
<th>Model</th>
<th>1 step</th>
<th>5 step</th>
<th>10 step</th>
<th>15 step</th>
<th>20 step</th>
<th>25 step</th>
<th>50 step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: rolling scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCVAR\textsubscript{d,b,ρ}</td>
<td>−6.24</td>
<td>−7.80</td>
<td>−11.50</td>
<td>−13.60</td>
<td>−14.43</td>
<td>−14.48</td>
<td>−13.10</td>
</tr>
<tr>
<td>FCVAR\textsubscript{d,b,μ}</td>
<td>−6.34</td>
<td>−8.18</td>
<td>−12.29</td>
<td>−14.44</td>
<td>−15.45</td>
<td>−15.19</td>
<td>−14.10</td>
</tr>
<tr>
<td>FCVAR\textsubscript{d=b,ρ}</td>
<td>−3.56</td>
<td>−3.01</td>
<td>−6.14</td>
<td>−8.42</td>
<td>−10.10</td>
<td>−10.60</td>
<td>−12.52</td>
</tr>
<tr>
<td>FCVAR\textsubscript{d=b,μ}</td>
<td>−6.22</td>
<td>−6.39</td>
<td>−9.29</td>
<td>−9.64</td>
<td>−10.00</td>
<td>−7.58</td>
<td>−0.55</td>
</tr>
<tr>
<td>Panel B: recursive scheme</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCVAR\textsubscript{d,b,ρ}</td>
<td>−7.53</td>
<td>−7.32</td>
<td>−12.13</td>
<td>−15.07</td>
<td>−17.43</td>
<td>−18.66</td>
<td>−20.00</td>
</tr>
<tr>
<td>FCVAR\textsubscript{d,b,μ}</td>
<td>−6.08</td>
<td>−4.92</td>
<td>−8.53</td>
<td>−10.32</td>
<td>−11.39</td>
<td>−11.17</td>
<td>−5.36</td>
</tr>
<tr>
<td>FCVAR\textsubscript{d=b,ρ}</td>
<td>−6.23</td>
<td>−4.61</td>
<td>−8.65</td>
<td>−11.11</td>
<td>−13.45</td>
<td>−14.88</td>
<td>−17.18</td>
</tr>
<tr>
<td>FCVAR\textsubscript{d=b,μ}</td>
<td>−6.06</td>
<td>−5.47</td>
<td>−9.50</td>
<td>−11.55</td>
<td>−13.24</td>
<td>−13.68</td>
<td>−11.61</td>
</tr>
</tbody>
</table>

Notes: Negative values favor the FCVAR model. Results are based on \(h\)-step ahead forecasts produced using 628-\(h\) training sets of length 600 (rolling scheme) or length= 600, 601, \ldots, 1227 − \(h\) (recursive scheme).

shared by both the ARFIMA and FCVAR models, the FCVAR model’s ability to allow for fractionally cointegrated time series results in better forecast accuracy.

Similarly, Table 4.7 reports the RMSFE percentage change of the FCVAR model relative to the CVAR model. Again, this table shows improved forecast accuracy for all variants of the FCVAR model relative to the CVAR model. In 56 of 56 cases in Table 4.7, all four variants of the FCVAR model perform better than the CVAR model. For the FCVAR models with a restricted constant, the improvements in performance increase substantially as the forecasting horizon increases. The recursive scheme shows 15%, 17%, 19% and 20% improvement for the 15, 20, 25 and 50 step ahead horizons, attained by the FCVAR model with two fractional parameters and a restricted constant. The rolling scheme shows up to 14% improvement for the FCVAR models with a restricted constant, and up to 15% improvement for the FCVAR models with a level parameter. In 30 of 56 cases, the FCVAR model delivers more than a
10% reduction in the RMSFE$_{sys}$ relative to the CVAR model. Overall the fractional models clearly outperform the non-fractional models, with the gains becoming more pronounced at longer forecast horizons.

To conclude the forecast comparison, examples of forecasts generated by all models in the portfolio are presented in Figure 4.3. The presented forecasts are generated using the first training set, which is common to both the rolling and recursive forecasting schemes. The figure shows the last 9 observations in the training set, followed by the out-of-sample observations and forecasts beginning at the 10$^{th}$ observation and continuing up to the longest horizon considered, $h = 50$, for the CP, LP and LD series in separate graphs. Panel (a) shows forecasts of the logit transformed series, while in Panel (b) all series (and forecasts) are transformed back to percentage vote shares to make interpretations easier. The figure thus provides an illustration of how all models forecast political support as measured by daily opinion polls. In Panel (a) we also include 90% confidence bands shown using slightly thinner lines.$^6$

The forecasts for this particular training set are in agreement with the evidence presented in the model ranking exercise. An interesting observation, which is present in results from other training sets as well, is that even though the CVAR forecasts exhibit more dynamics in the short run (which is a result of a higher lag order selected compared to the FCVAR), this does not translate into more accurate predictions for short horizons. In particular, as the short run dynamics die out for all multivariate models, it is evident that fractional cointegration generates point forecasts that are closer to the subsequently realized observations. These conclusions are strongly supported by the reported confidence bands in Figure 4.4(a), which show that forecasts

$^6$Following the advice of a referee, these were simply calculated from the moving-average representation of the models ignoring estimation uncertainty.
Figure 4.3: Forecasts for the first training set

(a) Forecasts of logit transformed series

(b) Forecasts of vote shares in percentage

Notes: The training set is the first window, which is shared by the rolling window and recursive window forecasting schemes. Panel (a) shows forecasts of logit transformed series, where 90% confidence bands are shown using with slightly thinner lines, and Panel (b) shows forecasts of vote shares in percentage. The ARFIMA(0, d, 0) model is not included because the ARFIMA(p, d, q) model specifies both lag orders to zero for all three series. The ARMA(p, q) model specifies (p, q) = (0, 1) for all three series.
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Based on the FCVAR models are much more accurate than those based on the non-fractional CVAR model. This is especially true at the longer horizons, as one might have expected. Finally, we note that all variants of the FCVAR produce similar predictions, and these are reasonably close to the realized data series in all three panels, while the other models in the portfolio only perform well in some cases.

4.5 Empirical application to the 2015 UK general election

In this section we present an application to the 2015 UK general election, which was deemed the most unpredictable election in decades in the media. Opinion poll agencies predicted a hung parliament, but ended up significantly underestimating the Conservative Party vote share; the party that won the election with a majority representation in Parliament. Note, however, that vote shares cannot be mapped into election outcomes in the context of the UK election process. This can be seen from the fact that the realized vote share for the Conservative Party was 36.8%, but the party won 330 out of 650 constituencies in the country; an outcome that the predicted vote share of the previous-day YouGov opinion poll, 34%, does not exclude.

The three political parties represented in the full data set, spanning the entire duration of the survey, are the three major political parties in the UK that have historically had the most representation in government by a strong margin over other parties running in the election. However, as the 2015 general election has shown, other parties not in the top three (as measured by representation in parliament or by opinion polls) can be important players. In this election, the UKIP and the Green Party received 12.7% and 3.8% vote shares, in particular. As discussed earlier, on April 16th, 2012, and June 18th, 2012, respectively, YouGov changed the way they
reported the outcomes of their polls and started reporting the UKIP and the Green Party as separate time series (rather than being included in the residual category). Therefore, in this empirical application to the 2015 UK general election, we apply the multivariate and univariate models to both the case of three political parties (based on the full sample spanning the entire 2010 to 2015 political cycle) and to the case of either four or five political parties (based on shorter data sets spanning only the second half of the 2010 to 2015 political cycle).\footnote{There were also some regional parties with non-negligible vote shares, but we do not include these in our analysis because they seem to compete on a different basis and with a somewhat different agenda.}

A key strength of our data set for the purpose of statistical modeling and forecasting, and hence for vote share prediction, is that the observed time series are contained within one political cycle. This should allow application of a relatively simple statistical model, and the previous analysis has shown strong support for the FCVAR model for this task.

### 4.5.1 Estimation results prior to the election

In the first part of this empirical application, we analyze and interpret the estimated model coefficients more carefully. To this end, we consider the full five-party data set consisting of \( CP_t, LP_t, LD_t, IP_t, \) and \( GP_t \), but for a reduced sample size covering June 18\textsuperscript{th}, 2012, to May 6\textsuperscript{th}, 2015, for a total of \( T+N = 729 \) observations, which is the period where observations are available for all five parties. The increased dimension of the model tends to cause problems in the multi-dimensional numerical optimization required for the FCVAR models with either two fractional parameters or with the level parameter, and for this reason we consider only the FCVAR\textsubscript{d=b,ρ} model for the full data set, since this model requires only one-dimensional numerical optimization.
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Table 4.8: FCVAR\(_{d=b,\rho}\) estimation results: five parties

Model:
\[
\Delta \hat{d} = \hat{\alpha} \left( \hat{\beta}' \Delta \hat{d} L_{\hat{b}} + \hat{\rho}' \right) + \hat{\Gamma}_1 \Delta \hat{d} L_{\hat{b}}^2 + \hat{\Gamma}_2 \Delta \hat{d} L_{\hat{b}}^2 + \hat{\varepsilon}_t
\]

Parameters:
\[
\hat{\alpha} = \begin{bmatrix}
-0.147 & 0.140 & 0.063 & 0.046 \\
0.019 & -0.232 & 0.070 & -0.069 \\
-0.038 & 0.145 & -0.425 & 0.071 \\
-0.012 & -0.300 & 0.018 & -0.083 \\
-0.872 & -1.549 & -0.752 & -0.335
\end{bmatrix}, \quad \hat{\beta} = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 1.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 1.000 \\
0.170 & -0.604 & 0.510 & 2.760
\end{bmatrix}
\]
\[
\hat{d} = 0.813, \quad \hat{\rho} = \begin{bmatrix}
1.068 \\
-0.555 \\
3.872 \\
7.821
\end{bmatrix}
\]
\[
Q_{\hat{\varepsilon}} = 166.098, \quad \log \mathcal{L} = 3632.605
\]

Notes: The table shows FCVAR estimation results for model (4.5) with one fractional parameter, \(d = b\), and a restricted constant term, \(\rho\). The multivariate Portmanteau Q-test for serial correlation up to order six in the residuals is reported as \(Q_{\hat{\varepsilon}}\) and the maximized log-likelihood value is reported as \(\log \mathcal{L}\). The standard error is in parenthesis below \(\hat{d}\) and the \(P\) value is in parenthesis below \(Q_{\hat{\varepsilon}}\). The sample size is \(T + N = 729\) and the first \(N = 20\) observations are used as initial values.

and thus remains feasible.

As usual, the estimation begins with lag length selection. The BIC first suggests \(k = 1\), but for this choice the Portmanteau Q-test rejects the null of no serial correlation in the residuals with a \(P\) value of 0.000. Consequently, we increase the lag length to \(k = 2\) for which the Q-test \(P\) value is 0.17. Next, the LR test for cointegration rank (the trace test) produces \(P\) values of 0.051 and 0.977 for \(r = 3\) and \(r = 4\), respectively, and we proceed with the specification \(r = 4\). As before, all results are conditional upon \(N = 20\) initial values.

The estimation results for the FCVAR\(_{d=b,\rho}\) model for the five-party data set are
presented in Table 4.8. We focus our interpretations on the long-run cointegration parameters, $\alpha$ and $\beta$.

To interpret the estimated cointegrating relations in $\hat{\beta}$, we find it convenient to re-normalize them on the coefficient for the Conservatives. That is, we re-normalize $\beta$ such that

$$
\beta = \begin{bmatrix}
-1 & -1 & -1 & -1 \\
\beta_{\text{LP}} & 0 & 0 & 0 \\
0 & \beta_{\text{LD}} & 0 & 0 \\
0 & 0 & \beta_{\text{IP}} & 0 \\
0 & 0 & 0 & \beta_{\text{GP}}
\end{bmatrix}.
$$

We note that this is simply a re-normalization because a similar rotation of the $\alpha$ matrix implies that the product $\alpha\beta'$, and hence the likelihood, is unchanged. The normalization in (4.18) implies that each cointegrating relation takes the form

$$
\text{CP}_t = \beta_S S_t \text{ for } S \in \{\text{LP, LD, IP, GP}\}.
$$

Thus, the normalization of $\beta$ given in (4.18) seems more straightforward to interpret because it directly relates the vote shares of each party to that of a main party, CP, rather than that of GP.

Applying the normalization in (4.18) to $\hat{\beta}$ in Table 4.8, we find

$$
\hat{\beta}_{\text{LP}} = -0.282, \quad \hat{\beta}_{\text{LD}} = 0.334, \quad \hat{\beta}_{\text{IP}} = 0.062, \quad \hat{\beta}_{\text{GP}} = -0.170.
$$

It is now clear that the government parties, CP and LD, move together in the long-run, although movements in the LD poll share are only about 1/3 of those in the CP poll share. It also seems that the IP poll share moves in the same direction as CP, but again with a coefficient less than one. On the other hand, the poll shares of both opposition parties, LP and GP, move in the opposite direction of the government
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Another relevant interpretation of the model can be obtained from the permanent-transitory (PT) decomposition of Gonzalo and Granger (1995) applied to the FC-VAR model. For any matrix $Q$, we define $Q_\perp$ such that $Q_\perp = Q_\perp Q = Q_\perp Q_\perp = 0$. Then, according to the PT decomposition, $X_t$ may be decomposed into a transitory (stationary) part, $\beta'X_t$, and a permanent part, $W_t = \alpha'_\perp X_t$, using the identity $\beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp + \alpha (\beta' \alpha)^{-1} \beta' = I_p$, which implies that

$$X_t = (\beta_\perp (\alpha'_\perp \beta_\perp)^{-1} \alpha'_\perp + \alpha (\beta' \alpha)^{-1} \beta')X_t = A_1 W_t + A_2 \beta'X_t. \quad (4.19)$$

Here, $W_t$ is the common permanent component of $X_t$, or the common stochastic trend(s). In the case of poll shares, $W_t$ is interpreted as the long-run dominant theme(s) of public opinion, in the sense that information that does not affect $W_t$ will not have a permanent effect on the poll shares given in $X_t$. On the other hand, $\beta'X_t$ is the transitory or stationary component of $X_t$, and is interpreted as information that does not have a permanent effect on the poll shares in $X_t$.

We first calculate the estimated common stochastic trend, $\hat{W}_t = \hat{\alpha}'_\perp X_t$, which is plotted in Figure 4.4 together with the UKIP data series. Interestingly, the common stochastic trend does not appear to reflect the traditional right-left political spectrum, but rather seems to follow the UKIP series very closely. Thus, we next test the hypothesis that the common stochastic trend is in fact $W_t = I P_t$, i.e. that $\alpha_\perp = [0 \ 0 \ 0 \ 1 \ 0]'$. The corresponding mirror hypothesis on $\alpha$ is that the fourth row is equal to zero, so that the test is seen to have four degrees of freedom. The LR statistic is 9.753 with a $P$ value of 0.045 in the asymptotic $\chi^2_4$-distribution. This $P$ value is formally less than 5%, although not by much. Thus, one could make a case for either rejecting the null hypothesis or not. In any case, re-estimating the
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Figure 4.4: Time series plot of estimated common stochastic trend and UKIP data series

Note: Orange line is estimated common stochastic trend ($\hat{W}_t$) and purple line is UKIP ($IP_t$).

A model with the restrictions imposed (results not reported) yields very similar results to those presented in Table 4.8 for the unrestricted model, and we shall continue with the unrestricted model.

Whether there is an exact match between the common stochastic trend ($W_t$) and the UKIP vote share ($IP_t$) or not, i.e. whether imposing the restriction that $W_t = IP_t$ or not, the interpretation of the common stochastic trend seen in Figure 4.4 is that the main theme of political debate in the UK during this time period from June 18th, 2012, to May 6th, 2015, has revolved around the independence question in relation to the European Union, at least in terms of long-run movements of poll shares. We proceed to calculate the coefficient on $W_t$ in (4.19), which yields

$$\hat{A}_1 = [0.076 \ -0.269 \ 0.228 \ 1.230 \ -0.446]^\prime$$

for the unrestricted model (i.e., without imposing $W_t = IP_t$). For the restricted model (imposing $W_t = IP_t$), the results are qualitatively similar.
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Interpreting $W_t$ as a measure of the strength of Euro-skepticism in public opinion, the estimated coefficients in $\hat{A}_1$ suggest that, as Euro-skepticism gains ground and $W_t$ increases, this leads to a large increase in popularity of the UKIP. However, when $W_t$ increases, the government parties (CP and LD) also increase in popularity, whereas the opposition parties (LP and GP) decrease in popularity.\(^8\)

4.5.2 Predicting vote shares of the election

The UKIP and Green Party poll series exhibit an important caveat, which is that unlike the three major political parties, the UKIP and the Green Party are not stated explicitly as choices in the survey question posed to the poll participants. In the forecasting analysis we do not include the Green Party because doing so would further reduce the sample size and the increased dimension of the model tends to cause problems in the multi-dimensional numerical optimization required for the FCVAR models with either two fractional parameters or with the level parameter. Thus, in the forecasting analysis we consider either three parties (CP\(_t\), LP\(_t\), and LD\(_t\)) or four parties (including also IP\(_t\)).

We next present predictions of all models in the portfolio leading into the 56\(^{th}\) UK general election held on May 7\(^{th}\), 2015. Specifically, Tables 4.9 (three parties) and 4.10 (four parties) show the opinion poll and the election vote share predictions of each model one month, one week and one day preceding the election day, as well as the final election vote share outcome. The predicted values represent forecasts of the poll standings which can be viewed as predicted vote shares for each political party. These tables also provide a convenient pairwise comparison of vote share predictions.

\(^8\)Recall that the poll shares are logit transformed. Because of this nonlinear transformation, absolute magnitudes of the coefficients are difficult to interpret, and, in particular, there is no requirement that the coefficients sum to zero.
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

across models and political parties.

Relative to the poll predictions on May 6th, 2015, which underestimated the vote share of the Conservative Party and overestimated the vote shares for Labour and the Liberal Democrats, the forecasts in both Table 4.9 and Table 4.10 show that all models in the portfolio predicted a fall in the vote share for the Liberal Democrats, in agreement with the realized vote share. The three-party forecasts in Table 4.9 generated on the day before the election (Panel C) show that (i) all models predicted vote shares very similar to the poll predictions for the Conservative Party and the Labour Party (with the exception of the ARMA prediction for the Labour Party), (ii) the two FCVAR models with a level parameter and the CVAR model predicted vote shares closest to the election outcomes for the Labour Party, and (iii) the multivariate models predicted vote shares closest to the election outcomes for the Liberal Democrats. The four-party forecasts in Table 4.10 generated on the day before the election (Panel C) show that (i) variants of the FCVAR model predicted vote shares that are closest to the election outcome for the Conservative Party, (ii) with the exception of the ARMA model all models show similar predictions for the Labour Party, (iii) the multivariate models predicted vote shares closest to the election outcomes for the Liberal Democrats, and (iv) with the exception of the CVAR model all models show similar predictions for the UKIP.

As shown in Tables 4.9 and 4.10, the opinion polls severely underestimated the Conservative Party vote shares in the election and similarly overestimated the Labour Party vote shares. Since all statistical models considered here rely solely on the opinion poll series for information, it is not surprising that for political parties for which the election result strongly deviated from the opinion poll prediction, the statistical
4.5. **EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION**

Table 4.9: Vote share prediction and results for the 2015 UK general election: three parties

<table>
<thead>
<tr>
<th></th>
<th>Conservative</th>
<th>Labour</th>
<th>Liberal Democrats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: 1 month before election</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poll</td>
<td>33.00</td>
<td>35.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Forecasts FCVAR(_d,b,ρ)</td>
<td>33.41</td>
<td>34.59</td>
<td>7.56</td>
</tr>
<tr>
<td>FCVAR(_d,b,μ)</td>
<td>33.38</td>
<td>33.76</td>
<td>7.66</td>
</tr>
<tr>
<td>FCVAR(_d=b,ρ)</td>
<td>33.15</td>
<td>34.65</td>
<td>7.19</td>
</tr>
<tr>
<td>FCVAR(_d=b,μ)</td>
<td>33.22</td>
<td>33.67</td>
<td>7.51</td>
</tr>
<tr>
<td>CVAR(_ρ)</td>
<td>32.68</td>
<td>32.91</td>
<td>7.23</td>
</tr>
<tr>
<td>ARFIMA(0, d, 0)</td>
<td>34.15</td>
<td>35.07</td>
<td>8.38</td>
</tr>
<tr>
<td>ARFIMA(p, d, q)</td>
<td>34.31</td>
<td>34.69</td>
<td>8.13</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>34.63</td>
<td>39.45</td>
<td>9.30</td>
</tr>
<tr>
<td><strong>Panel B: 1 week before election</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poll</td>
<td>34.00</td>
<td>35.00</td>
<td>8.00</td>
</tr>
<tr>
<td>Forecasts FCVAR(_d,b,ρ)</td>
<td>33.81</td>
<td>34.65</td>
<td>8.00</td>
</tr>
<tr>
<td>FCVAR(_d,b,μ)</td>
<td>33.72</td>
<td>34.35</td>
<td>8.10</td>
</tr>
<tr>
<td>FCVAR(_d=b,ρ)</td>
<td>33.64</td>
<td>34.58</td>
<td>7.69</td>
</tr>
<tr>
<td>FCVAR(_d=b,μ)</td>
<td>33.64</td>
<td>34.24</td>
<td>8.04</td>
</tr>
<tr>
<td>CVAR(_ρ)</td>
<td>33.57</td>
<td>33.79</td>
<td>7.70</td>
</tr>
<tr>
<td>ARFIMA(0, d, 0)</td>
<td>34.04</td>
<td>34.79</td>
<td>8.43</td>
</tr>
<tr>
<td>ARFIMA(p, d, q)</td>
<td>34.07</td>
<td>34.59</td>
<td>8.38</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>34.61</td>
<td>39.36</td>
<td>9.27</td>
</tr>
<tr>
<td><strong>Panel C: 1 day before election</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poll</td>
<td>34.00</td>
<td>34.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Forecasts FCVAR(_d,b,ρ)</td>
<td>33.95</td>
<td>34.02</td>
<td>8.68</td>
</tr>
<tr>
<td>FCVAR(_d,b,μ)</td>
<td>33.89</td>
<td>33.77</td>
<td>8.90</td>
</tr>
<tr>
<td>FCVAR(_d=b,ρ)</td>
<td>33.98</td>
<td>34.05</td>
<td>8.53</td>
</tr>
<tr>
<td>FCVAR(_d=b,μ)</td>
<td>33.83</td>
<td>33.77</td>
<td>8.90</td>
</tr>
<tr>
<td>CVAR(_ρ)</td>
<td>33.81</td>
<td>33.66</td>
<td>8.65</td>
</tr>
<tr>
<td>ARFIMA(0, d, 0)</td>
<td>33.89</td>
<td>34.11</td>
<td>9.23</td>
</tr>
<tr>
<td>ARFIMA(p, d, q)</td>
<td>33.86</td>
<td>33.98</td>
<td>9.04</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>34.16</td>
<td>37.07</td>
<td>9.80</td>
</tr>
<tr>
<td><strong>2015 election result</strong></td>
<td>36.80</td>
<td>30.50</td>
<td>7.90</td>
</tr>
</tbody>
</table>

Notes: The table shows the opinion poll and the election vote share predictions of each model one month (Panel A), one week (Panel B), and one day (Panel C) preceding the election day. The last row shows the election vote share outcomes. Results are for three political parties using the full data set spanning the entire political cycle from May 14\(^{th}\), 2010 to May 6\(^{th}\), 2015. For all three series, the ARFIMA(p, d, q) model specifies (p, q) = (1, 0) and the ARMA(p, q) model specifies (p, q) = (0, 1).
### 4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Table 4.10: Vote share prediction and results for the 2015 UK general election: four parties

<table>
<thead>
<tr>
<th>Panel A: 1 month before election</th>
<th>Conservative</th>
<th>Labour</th>
<th>Liberal Democrats</th>
<th>UKIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poll</td>
<td>33.00</td>
<td>35.00</td>
<td>8.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Forecasts FCVAR&lt;sub&gt;d,b,ρ&lt;/sub&gt;</td>
<td>33.42</td>
<td>34.25</td>
<td>7.56</td>
<td>13.97</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d,b,μ&lt;/sub&gt;</td>
<td>34.99</td>
<td>34.34</td>
<td>8.12</td>
<td>12.23</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d=ρ&lt;/sub&gt;</td>
<td>34.16</td>
<td>34.14</td>
<td>7.14</td>
<td>13.43</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d=μ&lt;/sub&gt;</td>
<td>34.06</td>
<td>33.88</td>
<td>7.00</td>
<td>13.32</td>
</tr>
<tr>
<td>CVAR&lt;sub&gt;ρ&lt;/sub&gt;</td>
<td>32.99</td>
<td>33.57</td>
<td>6.81</td>
<td>14.64</td>
</tr>
<tr>
<td>ARFIMA(0, d, 0)</td>
<td>33.00</td>
<td>35.67</td>
<td>8.01</td>
<td>12.75</td>
</tr>
<tr>
<td>ARFIMA(p, d, q)</td>
<td>33.00</td>
<td>35.67</td>
<td>7.93</td>
<td>12.99</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>32.59</td>
<td>38.81</td>
<td>8.84</td>
<td>11.26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: 1 week before election</th>
<th>Conservative</th>
<th>Labour</th>
<th>Liberal Democrats</th>
<th>UKIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poll</td>
<td>34.00</td>
<td>35.00</td>
<td>8.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Forecasts FCVAR&lt;sub&gt;d,b,ρ&lt;/sub&gt;</td>
<td>33.71</td>
<td>34.81</td>
<td>8.03</td>
<td>12.97</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d,b,μ&lt;/sub&gt;</td>
<td>34.46</td>
<td>34.60</td>
<td>8.34</td>
<td>12.45</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d=ρ&lt;/sub&gt;</td>
<td>34.37</td>
<td>34.71</td>
<td>7.92</td>
<td>12.82</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d=μ&lt;/sub&gt;</td>
<td>34.36</td>
<td>34.60</td>
<td>8.34</td>
<td>12.45</td>
</tr>
<tr>
<td>CVAR&lt;sub&gt;ρ&lt;/sub&gt;</td>
<td>32.84</td>
<td>34.83</td>
<td>7.58</td>
<td>14.13</td>
</tr>
<tr>
<td>ARFIMA(0, d, 0)</td>
<td>33.30</td>
<td>35.18</td>
<td>8.21</td>
<td>12.53</td>
</tr>
<tr>
<td>ARFIMA(p, d, q)</td>
<td>33.30</td>
<td>35.18</td>
<td>8.24</td>
<td>12.53</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>32.60</td>
<td>38.68</td>
<td>8.82</td>
<td>11.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: 1 day before election</th>
<th>Conservative</th>
<th>Labour</th>
<th>Liberal Democrats</th>
<th>UKIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poll</td>
<td>34.00</td>
<td>34.00</td>
<td>10.00</td>
<td>12.00</td>
</tr>
<tr>
<td>Forecasts FCVAR&lt;sub&gt;d,b,ρ&lt;/sub&gt;</td>
<td>33.82</td>
<td>34.33</td>
<td>8.58</td>
<td>12.58</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d,b,μ&lt;/sub&gt;</td>
<td>34.28</td>
<td>34.07</td>
<td>8.86</td>
<td>12.51</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d=ρ&lt;/sub&gt;</td>
<td>34.15</td>
<td>34.26</td>
<td>8.48</td>
<td>12.53</td>
</tr>
<tr>
<td>FCVAR&lt;sub&gt;d=μ&lt;/sub&gt;</td>
<td>34.28</td>
<td>34.07</td>
<td>8.86</td>
<td>12.51</td>
</tr>
<tr>
<td>CVAR&lt;sub&gt;ρ&lt;/sub&gt;</td>
<td>32.97</td>
<td>33.86</td>
<td>7.22</td>
<td>14.63</td>
</tr>
<tr>
<td>ARFIMA(0, d, 0)</td>
<td>33.58</td>
<td>34.19</td>
<td>8.98</td>
<td>12.20</td>
</tr>
<tr>
<td>ARFIMA(p, d, q)</td>
<td>33.58</td>
<td>34.19</td>
<td>8.86</td>
<td>12.21</td>
</tr>
<tr>
<td>ARMA(p, q)</td>
<td>32.90</td>
<td>36.57</td>
<td>9.43</td>
<td>11.56</td>
</tr>
</tbody>
</table>

2015 election result 36.80 30.50 7.90 12.70

Notes: The table shows the opinion poll and the election vote share predictions of each model one month (Panel A), one week (Panel B), and one day (Panel C) preceding the election day. The last row shows the election vote share outcomes. Results are for four political parties using the subsample of the data set spanning the second half of the political cycle from April 16<sup>th</sup>, 2012 to May 6<sup>th</sup>, 2015. For the CP and LP series, the ARFIMA(p, d, q) model specifies (p, q) = (0, 0) and the ARMA(p, q) model specifies (p, q) = (0, 1). For the LD and IP series, the ARFIMA(p, d, q) model specifies (p, q) = (1, 0) and the ARMA(p, q) model specifies (p, q) = (0, 1).
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Figure 4.5: Forecasts over 50 polling days leading into the election: three parties

(a) Forecasts of logit transformed series

(b) Forecasts of vote shares in percentage

Notes: Each subfigure shows the evolution of forecasts by different models variants for the May 7th, 2015 general election. Panel (a) shows forecasts of logit transformed series, where 90% confidence bands are shown using slightly thinner lines, and Panel (b) shows forecasts of vote shares in percentage. The forecasts are calculated starting with the data available 50 polling days prior to the election and continue to May 6th, 2015, the day before the election. The results use the full data set spanning the entire political cycle (May 14th, 2010 to May 6th, 2015).
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

Figure 4.6: Forecasts over 50 polling days leading into the election: four parties

(a) Forecasts of logit transformed series

(b) Forecasts of vote shares in percentage

Notes: Each subfigure shows the evolution of forecasts by different models variants for the May 7th, 2015 general election. Panel (a) shows forecasts of logit transformed series, where 90% confidence bands are shown using slightly thinner lines, and Panel (b) shows forecasts of vote shares in percentage. The forecasts are calculated starting with the data available 50 polling days prior to the election and continue to May 6th, 2015, the day before the election. The results use the subsample of the data set spanning the second half of the political cycle (April 16th, 2012 to May 6th, 2015).
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

models find it difficult to predict the high realized vote shares for the Conservative Party and low realized vote shares for the Labour Party. It is in this context that the evolution of the predictions serves to complement opinion polls, since as the most recent poll data became available, forecasters would naturally obtain forecasts for the election.

To illustrate how the analysis presented in this chapter complements the industry standard of using the latest opinion poll as an indicator of future political support, we examine the evolution of the model predictions as the observed data approaches the election day on May 7th, 2015, which allows the analyst to monitor the dynamics of the opinion poll and the model predictions. The intuitive appeal of conducting this exercise, is to mimic a real life scenario where an analyst is tasked with analyzing this data set for forecasting purposes leading up to the election. In such a setting, as every day drew nearer to the election, the survey updated and provided the analyst with an additional data point, at which point the natural thing for him or her to do would be to update the models considered and produce forecasts onto the election day. By repeating this exercise for every survey update, the analyst would have produced a tracked series of forecasts by the day before the election. Figures 4.5 and 4.6 show forecasts by all models over 50 polling days leading into the election, for the three-party and the four-party analysis, respectively. That is, at time $h$ polling days prior to the election, each figure shows the $h$-step ahead predictions from each model. Panel (a) of each figure shows forecasts of the logit transformed series, where 90% confidence bands are shown using slightly thinner lines, and Panel (b) of each figure shows forecasts of vote shares in percentage. With these plots, we can analyze the dynamics of model forecasts leading into a particular date of interest, in this case the
4.5. EMPIRICAL APPLICATION TO THE 2015 UK GENERAL ELECTION

election day, and compare both across models and with the actual daily poll series.

For the three-party analysis in Figure 4.5, the evolution of forecasts for all FCVAR model variants show an upward trend in the support for the Conservative Party leading into the election, and a downward trend for the Labour Party for one week leading into the election. These two trends project the correct direction for the realized vote shares in the election. For the Liberal Democrats, the strong upward trends in the predicted vote share, as shown by nearly all models, tend to follow the opinion poll and is therefore suspect to exhibit the same tendency as political polls, in that they tend to over-represent support for smaller political parties when compared to the election vote share outcomes and the representation in government.

For all models, the confidence bands in Figure 4.6(a) become narrower as the election approaches, which is expected because the forecast horizon shortens. For all the multivariate models, the election outcomes for both the Conservative Party and the Labour Party lie outside the confidence bands for most of the time period considered in the figure. For the univariate models, the confidence bands are wider and include the Conservative Party and Labour Party election outcomes for a large part of the sampling period (except the ARMA model for LP). For the Liberal Democrats the confidence bands of all models include the election outcome throughout the time period in the figure.

In the four-party analysis in Figure 4.6, the realized vote share for the UKIP, which was ranked third by the political opinion polls, was very close to that prescribed by the opinion poll. As a result, most models in the portfolio appear to perform well. However, the evolution of the FCVAR model predictions appear to converge somewhat more closely to the realized vote share for the UKIP. In particular, the CVAR model
overshoots the election result for the UKIP. For both the three-party and the four-party analyses, we find that the FCVAR model predictions are on average closer to the realized vote shares than their analogs from the CVAR model.

The confidence bands in Figure 4.7(a) are broadly in agreement with those from Figure 4.6(a) for the Conservatives, Labour, and Liberal Democrats, with the exception that the univariate forecast bands now also exclude the election outcomes for the Conservatives and Labour. For the UKIP we find that the election result is within the confidence bands for all models throughout the time period considered in the figure. However, noting the different scaling of the second axis for the IP plots for fractional and non-fractional models, respectively, it is apparent that the overprediction of the election result by the CVAR model is also accompanied by a somewhat wider confidence band compared with the FCVAR model variants.

In general, the forecasting results from this empirical application show how modeling time series of political opinion polls using the FCVAR model, which is strongly favored by the model forecast comparisons, can complement the industry standard of basing predictions solely on the most recent opinion poll (e.g., the poll standings on the day preceding the election day) and provide a more informative assessment of the current state of public opinion. Specifically, in the case of the 2015 UK general election, the FCVAR model forecasts provide additional information on party support compared with the hung parliament prediction of the opinion poll.

4.6 Concluding remarks

This chapter has examined the forecasting performance of the fractionally cointegrated vector autoregressive (FCVAR) model of Johansen (2008) and Johansen and
Nielsen (2012) relative to a portfolio of competing models at several forecast horizons. The model was applied in the context of predicting political support in the form of opinion polls; a very relevant topic in the context of forecasting. The analysis used a novel data set of daily polling of political support in the United Kingdom over the period 2010–2015. The analysis has shown how statistical modeling of time series of public opinion polls can improve forecasting public opinion beyond the most recent poll date. This complements the industry standard for measuring the current state of political support through opinion polling, and contributes to decision making processes that rely on poll evidence as inputs.

Specifically, the forecasting analysis has shown that the FCVAR model delivers valuable gains in predicting political support. The accuracy of both short, medium, and long term forecasts generated by the FCVAR model is better than all multivariate and univariate models in the portfolio. Indeed, overall, the four variants of the FCVAR model are the top performing models. Not only do they perform better relative to the other models, but the forecasting performance of all FCVAR model variants are within close range of each other. When compared to both the fractional benchmark model (ARFIMA) and the multivariate benchmark model (CVAR), the FCVAR model significantly outperforms the benchmark at all forecast horizons and the gains are more pronounced at the longer forecast horizons, where the root mean squared forecast error is up to 15% lower than the fractional benchmark and 20% lower than the multivariate benchmark. Overall, the evidence provides strong support for the application of the FCVAR model relative to both univariate fractional models and multivariate non-fractional models. Fractional cointegration substantially improves forecast accuracy, and the gains become more pronounced at longer forecast horizons.
4.6. CONCLUDING REMARKS

More generally, it is of interest to consider how generalizable our results are to other data sets. The arguments in Section 4.2 certainly suggest that all polling data may be best modeled as fractional time series. This in turn suggests that the FCVAR model would be applicable to all such data, and that the FCVAR model would produce superior forecasts for such data. The important self-similarity property of fractional processes suggests that the observed superior forecast performance is driven by the structure of the FCVAR model and not the business-daily observation frequency of the data. This implies that similar forecasting gains would be expected for polling data at varying observation frequencies. Furthermore, the arguments in Section 4.2 suggest that our results may be generalizable to all data that have an aggregation structure, such as, in political science, any polling data, voting data, government support data, partisan indicators, etc. A thorough investigation of whether this is in fact the case is beyond the scope of this chapter, and we leave this interesting issue for future research. However, we do note that the very general applicability of fractional integration models is in line with the literature, see e.g. Box-Steffensmeier and Tomlinson (2000) and Lebo, Walker, and Clarke (2000).

In an empirical application to the 2015 UK general election, we have discussed the FCVAR model estimation results, with interpretations of both the estimated cointegrating relations and estimated common stochastic trend. It appears that the latter can be interpreted as a measure of Euro-skepticism, rather than an indicator of the more traditional left-right political spectrum, reflecting public opinion and debate in the sampling period which was to a great extent focused on the European Union question. The forecasts generated by the FCVAR model leading into the election appear to provide a more informative assessment of the current state of
public opinion on electoral support than the hung parliament prediction of the opinion polls. Specifically, when three political parties are modeled over the full election cycle, the FCVAR model predicts the correct direction for the realized vote shares in the election for the Conservative and Labour parties. When four political parties are modeled over the shorter available data set spanning the second half of the election cycle, the predictions of the FCVAR models are closer to the realized vote shares than their analogs from the CVAR model, and in particular the FCVAR models appear to converge more closely to the realized vote share for the UKIP.
Bibliography


**Nielsen, M. Ø., and M. K. Popiel** (2016): “A Matlab program and user’s guide for the fractionally cointegrated VAR model,” QED working paper 1330, Queen’s University.


Appendix A

A.0.1 Complete MCMC algorithm

The complete Metropolis-within-Gibbs sampling algorithm, corresponding to the compact outline of the algorithm given in Table 2.3 is:

**Step 0:** Initialize all parameters

**Step 1:** Draw cluster

(a) Draw $\beta_{k}^{(j+1)}$ from $\beta_{k}|Y_{n}, \mu, \Omega, P, z, H \sim N\left(b_{k}^{*}, B_{k}^{*}\right)$ with a mean $b_{k}^{*}$ and variance $B_{k}^{*}$ calculated from:

$$
B_{k}^{*(j+1)} = \left(B_{k}^{-1} + X_{k}V_{k}^{-1(j)}X_{k}'\right)^{-1}
$$

$$
b_{k}^{*(j+1)} = B_{k}^{*(j+1)}\left(B_{k}^{-1}b_{k} + X_{k}V_{k}^{-1(j)}\xi_{k}^{(j)}\right)
$$

which is a standard Normal regression of the form

$$
\xi_{k} = X_{k}\beta_{k} + \varepsilon_{k}, \quad \varepsilon_{k} \sim N(0, V_{k})
$$

$$
V_{k} = \text{diag}(\lambda_{1k}, \lambda_{2k}, \ldots, \lambda_{Nk})
$$

(b) Draw $h_{nk}^{(j+1)}$, the affiliation of region $n$ to cluster $k$ independently across regions (follows from (2.15)) from
\[
\Pr(h^{(j+1)}_{nk} = 1|Y_n, h^{[k]}, \rho^{(j)}, \mu^{(j)}, \sigma^{2(j)}, P^{(j)}, z^{(j)}, \beta^{(j+1)}_k) = \frac{L(h^{(j+1)}_{nk} = 1, h^{[k]}, \rho^{(j)}, \mu^{(j)}, \sigma^{2(j)}, z^{(j)}; Y_n) \Pr(h^{(j+1)}_{nk} = 1|\beta^{(j+1)}_k)}{\sum_{i=0}^{L} L(h^{(j+1)}_{nk} = i, h^{[k]}, \rho^{(j)}, \mu^{(j)}, \sigma^{2(j)}, z^{(j)}; Y_n) \Pr(h^{(j+1)}_{nk} = i|\beta^{(j+1)}_k)}
\]

(A.3)

where \(\Pr(h^{(j+1)}_{nk} = i|\beta^{(j+1)}_k)\) is computed from

\[
\Pr(h^{(j+1)}_{nk} = i|\beta^{(j+1)}_k) = \begin{cases} 
\frac{1}{1 + \exp(x'_{nk}\beta^{(j+1)}_k)} & \text{if } i = 0 \\
\frac{\exp(x'_{nk}\beta^{(j+1)}_k)}{1 + \exp(x'_{nk}\beta^{(j+1)}_k)} & \text{if } i = 1
\end{cases}
\]

(A.4)

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

(c) Draw \(\xi^{(j+1)}_{nk} = x'_{nk}\beta^{(j+1)}_k - \log(u^{-1(j+1)} - 1)\) where \(u^{-1(j+1)}\) is computed from

\[
u^{-1(j+1)} = \begin{cases} 
\frac{1}{1 + \exp(x'_{nk}\beta^{(j+1)}_k)}u^*_{nk} & \text{if } h^{(j+1)}_{nk} = 0 \\
\frac{1}{1 + \exp(x'_{nk}\beta^{(j+1)}_k)} + \frac{1}{1 + \exp(x'_{nk}\beta^{(j+1)}_k)}u^*_{nk} & \text{if } h^{(j+1)}_{nk} = 1
\end{cases}
\]

(A.5)

where \(u^*_{nk}\) is a draw from \(u \sim U[0, 1]\)

(d) Draw \(\lambda^{(j+1)}_{nk}\) from \(\lambda_{nk} \sim \text{GIG}\left(\frac{1}{2}, 1, r^2_{nk}\right)\) where \(r^2_{nk}\) is calculated from\(^1\):

\[
r_{nk} = \xi^{(j+1)}_{nk} - x'_{nk}\beta^{(j+1)}_k
\]

(A.6)

Step 2: Draw \(\mu^{(j+1)}_{nk}\) from \(\mu_{nk}|Y, \rho, \sigma_n^2, P, z, h, \beta \sim N(m^*, \Sigma_n)\) with a mean \(m^* = A^{-1}b\) and variance \(\Sigma_n = A^{-1}\) calculated from:

\(^1\)Generalized Inverse Gaussian distribution
\[ A = \sigma_n^{-2(j)}(1 + \rho(j)W_{nn})^2 \sum_{t=1}^{T} w(z_t^{(j)}, h^{(j+1)}) w(z_t^{(j)}, h^{(j+1)})' + [\sigma_n^{2(j)}M]^{-1} \]

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

(A.7)

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

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

\[ \hat{\delta} = \sum_{t=1}^{T} \left( y_{tn} - \mu_{n0}^{(j+1)} - \mu_{n1}^{(j+1)} h_{n,z_t^{(j)}}^{(j+1)} - \rho(j) \sum_{i \neq n} W_{nj}(y_{ti} - \mu_{i0}^{(j+1)} - \mu_{i1}^{(j+1)} h_{i,z_t^{(j)}}^{(j)}) \right) \]

\[ - \rho(j) W_{nn}(y_{tn} - \mu_{n0}^{(j+1)} - \mu_{n1}^{(j+1)} h_{n,z_t^{(j)}}^{(j+1)})^2 \]

(A.8)

**Step 4:** Draw \( \rho^{(j+1)} \) using the M-H algorithm defined in Table 2.2.

**Step 5:** Draw the aggregate regime indicator, \( z^{(j+1)} \) (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, \( P^{(j+1)} \), by independently drawing every column \( P_p^{(j+1)} \) from \( P_p \sim D(\alpha_p^*) \) with the \( q^{th} \) hyperparameter, \( \alpha_{pq}^* \), of vector \( \alpha_p^* \) calculated as
\[ \alpha^{(j+1)}_{pq} = \frac{\sum_{t=2}^{T} \delta(z(j+1)_{t-1} = p, z(j+1)_t = q)}{\sum_{t=2}^{T} \delta(z(j+1)_{t-1} = p)} \]  
\[ (A.9) \]

the fraction of times regime \( p \) is followed by regime \( q \) in the drawn sequence \( \{z^{(j+1)}_1, z^{(j+1)}_2, \ldots, z^{(j+1)}_T\} \).

A.0.2 Conditional posterior distribution of \( \mu \)

This section derives the conditional posterior distribution of \( \mu \). The key to deriving the conditional posterior distribution of \( \mu_n \) is the multivariate completion of squares or ellipsoidal rectification identity:

\[ u'Au - 2\alpha'u = (u - A^{-1}\alpha)'A(u - A^{-1}\alpha) - u'A^{-1}\alpha. \]  
\[ (A.10) \]

The conditional posterior distribution of \( \mu \) is:

\[
p(\mu_n|Y, \rho, \sigma_n^{-2}, P, z, h, \beta) \propto \pi(\mu_n|\sigma_n^{-2})L(\rho, \mu_n, \sigma_n^{-2} z, h; Y_n)
\]
\[
= |\sigma_n^2 M|^{-0.5} \exp \left\{ - \frac{1}{2} (\mu_n - m)'[\sigma_n^2 M]^{-1}(\mu_n - m) \right\}
\]
\[
\times \sigma_n^{-T}|I_N - \rho W|^T \exp \left\{ - \frac{1}{2} \sum_{n=1}^{N} (\sigma_n^{-2} \sum_{t=1}^{T} (y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} - \mu_n w(z_t, h)) + \rho \sum_{j=1}^{N} W_{nj}(\mu_{j0} + \mu_{j1} h_j, z_t))^2 \right\}
\]
\[
\times \exp \left\{ - \frac{1}{2} (\mu_n'[\sigma_n^2 M]^{-1} \mu_n - 2\mu_n'[\sigma_n^2 M]^{-1} m + m'[\sigma_n^2 M]^{-1} m) \right\}
\]
\[
\times \exp \left\{ - \frac{1}{2} \sigma_n^{-2} \left( \sum_{t=1}^{T} (y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj})^2 \right) \right\}
\]
\[
- 2 \sum_{t=1}^{T} \left( \mu_n' w(z_t, h) + \rho \sum_{j=1}^{N} W_{nj} (\mu_{j0} + \mu_{j1} h_{j,z_t}) \right) \left( y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \right) + \sum_{t=1}^{T} \left( \mu_n' w(z_t, h) + \rho \sum_{j=1}^{N} W_{nj} (\mu_{j0} + \mu_{j1} h_{j,z_t}) \right)^2 \right) \Bigg) \Bigg].
\]

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

\[
\propto \exp \left\{ -\frac{1}{2} \left( \mu_n' [\sigma_n^2 M]^{-1} \mu_n - 2 \mu_n' [\sigma_n^2 M]^{-1} m + m' [\sigma_n^2 M]^{-1} m \right) \right\} \times \exp \left\{ -\frac{1}{2} \sigma_n^{-2} \left( -2 \sum_{t=1}^{T} \left( \mu_n' w(z_t, h)(1 + \rho W_{nn}) \right) \left( y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \right) + \sum_{t=1}^{T} \left( \mu_n' w(z_t, h)(1 + \rho W_{nn}) \right)^2 \right) \right\} = \exp \left\{ -\frac{1}{2} \left( \mu_n' [\sigma_n^2 M]^{-1} \mu_n - 2 \mu_n' [\sigma_n^2 M]^{-1} m + m' [\sigma_n^2 M]^{-1} m \right) \right\} \times \exp \left\{ -\frac{1}{2} \sigma_n^{-2} \left( -2 \sum_{t=1}^{T} \left( \mu_n' w(z_t, h)(1 + \rho W_{nn}) \right) \left( y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \right) + \left( \mu_n' (1 + \rho W_{nn})^2 \sum_{t=1}^{T} w(z_t, h) \right) w(z_t, h)' \mu_n \right) \right\} \right\}
\]
\[ \propto \exp \left\{ -\frac{1}{2} \left( \mu_n' [\sigma_n^2 M]^{-1} \mu_n + \mu_n' \sigma_n^{-2} (1 + \rho W_m)^2 \sum_{t=1}^{T} w(z_t, h) w(z_t, h)' \mu_n \right. \right. \\
- \left. \left. 2 \mu_n' [\sigma_n^2 M]^{-1} m - 2 \mu_n' \sigma_n^{-2} \sum_{t=1}^{T} \left( w(z_t, h)(1 + \rho W_m) \right) \left( y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \right) \right) \right\} \]

Collecting terms before completing the multivariate square gives

\[ = \exp \left\{ -\frac{1}{2} \left( \mu_n' \left[ \sigma_n^{-2} (1 + \rho W_m)^2 \sum_{t=1}^{T} w(z_t, h) w(z_t, h)' + [\sigma_n^2 M]^{-1} \right] \mu_n \right. \right. \\
- \left. \left. 2 \mu_n' \sigma_n^{-2} \sum_{t=1}^{T} \left( w(z_t, h)(1 + \rho W_m) \right) \left( y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} \right) + [\sigma_n^2 M]^{-1} m \right) \right\} \].

Considering the term in the exponential

\[ -\frac{1}{2} (\mu_n' A \mu_n - 2 \mu_n' b) \]

and introducing a term to complete the square, along with using \( I = AA^{-1} \), gives

\[ -\frac{1}{2} (\mu_n' A \mu_n - 2 \mu_n' AA^{-1} b + b' A^{-1} AA^{-1} b) \].
Let $\Sigma_n = A^{-1}$ and $m^* = A^{-1}b$, then returning to the complete term in the proof gives

$$= \exp \left\{ -\frac{1}{2} \left( \mu'_n \Sigma_n^{-1} \mu_n - 2 \mu'_n \Sigma_n^{-1} m^* + m^* \Sigma_n^{-1} m^* \right) \right\}$$

$$= \exp \left\{ -\frac{1}{2} (\mu_n - m^*)' \Sigma_n^{-1} (\mu_n - m^*) \right\},$$

which results in the following posterior distribution for $\mu_n$:

$$p(\mu_n | Y, \rho, \sigma^{-2}_n, P, z, h, \beta) \propto \exp \left\{ -\frac{1}{2} (\mu_n - m^*)' \Sigma_n^{-1} (\mu_n - m^*) \right\}$$

$$\mu_n | Y, \rho, \sigma^{-2}_n, P, z, h, \beta \sim N\left( m^*, \Sigma_n \right)$$

where

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

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

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

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

A.0.3 Conditional posterior distribution of $\sigma^{-2}_n$

This section derives the conditional posterior distribution of $\sigma^{-2}_n$:
\[ p(\sigma_n^{-2}|Y, \rho, \mu_n, P, z, h, \beta) \propto \pi(\sigma_n^{-2}) \mathcal{L}(\rho, \mu_{n0}, \mu_{n1}, \sigma_n^{-2} z, h; Y_n) \]

\[ \propto \sigma_n^{-\nu + 2} \exp \left\{ -\frac{1}{2} \delta \sigma_n^{-2} \right\} \]

\[ \times \sigma_n^{-T} \exp \left\{ -\frac{1}{2} \sigma_n^{-2} \sum_{t=1}^{T} \left( y_{tn} - \mu_{n0} - \mu_{n1} h_{n,z_t} - \rho \sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \right)^2 \right\} \]

\[ = \sigma_n^{-\nu - T + 2} \exp \left\{ -\frac{1}{2} \sigma_n^{-2} \left( \delta + \sum_{t=1}^{T} \left( y_{tn} - \mu_{n0} - \mu_{n1} h_{n,z_t} - \rho \sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \right)^2 \right) \right\} \]

\[ = \sigma_n^{-\nu - T + 2} \exp \left\{ -\frac{1}{2} \sigma_n^{-2} \left( \delta + \sum_{t=1}^{T} \left( y_{tn} - \rho \sum_{j=1}^{N} W_{nj} y_{tj} - \mu_{n0} - \mu_{n1} h_{n,z_t} \right) \right) \right) \]

\[ + \rho \sum_{j=1}^{N} W_{nj}(\mu_{j0} + \mu_{j1} h_{j,z_t})^2 \right\} \]

\[ \Leftrightarrow \]

\[ \sigma_n^{-2}|Y, \rho, \mu_n, P, z, h, \beta \sim \Gamma \left( \frac{\nu + T}{2}, \frac{\delta + \hat{\delta}}{2} \right). \]

(A.12)

where \( \hat{\delta} = \sum_{t=1}^{T} \left( y_{tn} - \mu_{n0} - \mu_{n1} h_{n,z_t} - \rho \sum_{j=1}^{N} W_{nj}(y_{tj} - \mu_{j0} - \mu_{j1} h_{j,z_t}) \right)^2 \) and the last equality in A.12 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.

**A.0.4 Conditional posterior distribution of \( \rho \)**

This section derives the conditional posterior distribution of \( \rho \):
\[ p(\rho | Y, \mu, \Omega, P, z, h, \beta) \propto \pi(\rho) \mathcal{L}(\rho, \mu, \Omega, z, h; Y) \]
\[
\propto \pi(\rho) |I_N - \rho W|^T \exp \left\{ - \frac{1}{2} \sum_{n=1}^{N} \left( \sum_{t=1}^{T} \left( y_{tn} - \mu_{n0} - \mu_{n1} h_{nzt} \right)^2 \right) \right\} \\
- \rho \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{jzt})^2 \right) \right\} \\
\propto \pi(\rho) |I_N - \rho W|^T \exp \left\{ - \frac{1}{2} \rho^{2} \sum_{n=1}^{N} \left( \sum_{t=1}^{T} \left( \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{jzt}) \right)^2 \right) \right\} \\
+ \frac{2}{2} \rho \sum_{n=1}^{N} \left( \sum_{t=1}^{T} (y_{tn} - \mu_{n0} - \mu_{n1} h_{nzt}) \right) \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{jzt}) \left\{ - \frac{1}{2} \rho (\rho B_1 - 2B_2) \right\}.
\] (A.13)

The last equality follows from the fact that \( |I_N - \rho W| = \prod_{n=1}^{N} (1 - \rho \gamma_n) \) (see discussion in Section 2.3). The terms \( B_1 \) and \( B_2 \) are given as

\[ B_1 = \sum_{n=1}^{N} \left( \sum_{t=1}^{T} \left( \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{jzt}) \right)^2 \right) \]
\[
= \text{diag}(\Omega^{-1}) \begin{bmatrix} W_1 \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} & \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} & \cdots & \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} \\
W_2 \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} & \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} & \cdots & \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} \\
\vdots & \vdots & \ddots & \vdots \\
W_N \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} & \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} & \cdots & \tilde{\epsilon}_{z_1} \tilde{\epsilon}_{z_2} \cdots \tilde{\epsilon}_{z_T} \end{bmatrix} \]

\[ B_2 = \sum_{n=1}^{N} \left( \sum_{t=1}^{T} \left( \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{n0} - \mu_{n1} h_{nzt}) \right) \right) \sum_{j=1}^{N} W_{nj} (y_{tj} - \mu_{j0} - \mu_{j1} h_{jzt}) \left\{ - \frac{1}{2} \rho (\rho B_1 - 2B_2) \right\}.
\] (A.14)

where \( \tilde{\epsilon}_{zt} = y_t - \mu_0 - \mu_1 \odot h_{zt} \), and \( \tilde{\epsilon}_{nzt} = y_{tn} - \mu_{n0} - \mu_{n1} h_{nzt} \).
Appendix B

B.1 Figures, Tables and Convergence

B.1.1 BEA Economic Areas

Figure B.1: BEA Economic Areas

(a) Surrounding statistical areas
B.1.2 Testing for spatial correlation

Figure B.2: Moran’s I Statistic: Positive (negative) values indicate positive (negative) spatial autocorrelation
Figure B.3: Moran’s I Test for Spatial Correlation: P-values ($H_0$: no spatial autocorrelation)
B.1.3 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 ($\theta$) 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 $\theta$ should be close to one, indicating that further MCMC draws are not needed to improve our inference about the posterior distribution of $\theta$.

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

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

(B.1)

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

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

(B.2)

where $\theta_{mn}$ 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,

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

(B.3)

which overestimates the population variance $\sigma^2_{\theta}$, while $Var^w$ underestimates $\sigma^2_{\theta}$, both converge in expectation as $n \rightarrow \infty$. 
The GR diagnostic is given by

\[ GR = \sqrt{\frac{\hat{V}}{\text{Var}_w d} d - 2}, \]  

(B.4)

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

\[ \hat{V} = \hat{\sigma}_\theta^2 + \frac{\text{Var}_b}{MN_{\text{keep}}}, \]  

(B.5)

and \( d \) are the degrees of freedom estimated as

\[ d = \frac{2\hat{V}^2}{\text{Var}_w \hat{V}}. \]  

(B.6)

Table B.1: GR convergence diagnostic results

<table>
<thead>
<tr>
<th>Parameter Group</th>
<th>Parameter count</th>
<th>All GR &lt; 1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho )</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>177</td>
<td>Y</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>177</td>
<td>Y</td>
</tr>
<tr>
<td>( \Omega )</td>
<td>177</td>
<td>Y</td>
</tr>
<tr>
<td>( z )</td>
<td>408</td>
<td>Y</td>
</tr>
<tr>
<td>( P )</td>
<td>14</td>
<td>Y</td>
</tr>
<tr>
<td>( h_1 )</td>
<td>177</td>
<td>Y</td>
</tr>
<tr>
<td>( h_2 )</td>
<td>177</td>
<td>Y</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>7</td>
<td>Y</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>7</td>
<td>Y</td>
</tr>
</tbody>
</table>

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

The results in Table B.1 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.

B.1.4 Other parameter estimates
### Table B.2: Regional average employment growth during expansions – estimated coefficients $\mu_{n0}$ (posterior means)

<table>
<thead>
<tr>
<th>Region</th>
<th>$\mu_{n0}$</th>
<th>Region</th>
<th>$\mu_{n0}$</th>
<th>Region</th>
<th>$\mu_{n0}$</th>
<th>Region</th>
<th>$\mu_{n0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>47</td>
<td>0.45</td>
<td>93</td>
<td>0.31</td>
<td>138</td>
<td>0.40</td>
</tr>
<tr>
<td>2</td>
<td>0.33</td>
<td>48</td>
<td>0.14</td>
<td>94</td>
<td>0.59</td>
<td>139</td>
<td>0.28</td>
</tr>
<tr>
<td>3</td>
<td>0.33</td>
<td>49</td>
<td>1.02</td>
<td>95</td>
<td>0.79</td>
<td>140</td>
<td>0.86</td>
</tr>
<tr>
<td>4</td>
<td>0.18</td>
<td>50</td>
<td>0.41</td>
<td>96</td>
<td>0.22</td>
<td>141</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>0.42</td>
<td>51</td>
<td>0.49</td>
<td>97</td>
<td>0.26</td>
<td>142</td>
<td>1.49</td>
</tr>
<tr>
<td>6</td>
<td>0.60</td>
<td>52</td>
<td>0.40</td>
<td>98</td>
<td>0.70</td>
<td>143</td>
<td>0.19</td>
</tr>
<tr>
<td>7</td>
<td>0.58</td>
<td>53</td>
<td>0.45</td>
<td>99</td>
<td>0.53</td>
<td>144</td>
<td>1.05</td>
</tr>
<tr>
<td>8</td>
<td>0.98</td>
<td>54</td>
<td>0.20</td>
<td>100</td>
<td>0.22</td>
<td>145</td>
<td>1.02</td>
</tr>
<tr>
<td>9</td>
<td>0.77</td>
<td>55</td>
<td>1.06</td>
<td>101</td>
<td>0.85</td>
<td>146</td>
<td>0.91</td>
</tr>
<tr>
<td>10</td>
<td>1.57</td>
<td>56</td>
<td>0.92</td>
<td>102</td>
<td>0.22</td>
<td>147</td>
<td>0.58</td>
</tr>
<tr>
<td>11</td>
<td>0.16</td>
<td>57</td>
<td>2.34</td>
<td>103</td>
<td>0.17</td>
<td>148</td>
<td>1.26</td>
</tr>
<tr>
<td>12</td>
<td>2.76</td>
<td>58</td>
<td>0.49</td>
<td>104</td>
<td>1.70</td>
<td>149</td>
<td>1.14</td>
</tr>
<tr>
<td>13</td>
<td>0.14</td>
<td>59</td>
<td>0.33</td>
<td>105</td>
<td>0.49</td>
<td>150</td>
<td>0.30</td>
</tr>
<tr>
<td>14</td>
<td>0.99</td>
<td>60</td>
<td>0.36</td>
<td>106</td>
<td>0.43</td>
<td>151</td>
<td>0.27</td>
</tr>
<tr>
<td>15</td>
<td>0.30</td>
<td>61</td>
<td>1.03</td>
<td>107</td>
<td>1.60</td>
<td>152</td>
<td>0.90</td>
</tr>
<tr>
<td>16</td>
<td>2.11</td>
<td>62</td>
<td>0.74</td>
<td>108</td>
<td>0.18</td>
<td>153</td>
<td>0.30</td>
</tr>
<tr>
<td>17</td>
<td>0.43</td>
<td>63</td>
<td>0.21</td>
<td>109</td>
<td>0.80</td>
<td>154</td>
<td>0.43</td>
</tr>
<tr>
<td>18</td>
<td>0.34</td>
<td>64</td>
<td>0.86</td>
<td>110</td>
<td>2.36</td>
<td>155</td>
<td>1.26</td>
</tr>
<tr>
<td>19</td>
<td>1.09</td>
<td>65</td>
<td>0.22</td>
<td>111</td>
<td>1.41</td>
<td>156</td>
<td>0.58</td>
</tr>
<tr>
<td>20</td>
<td>1.84</td>
<td>66</td>
<td>0.18</td>
<td>112</td>
<td>0.82</td>
<td>157</td>
<td>0.90</td>
</tr>
<tr>
<td>21</td>
<td>0.80</td>
<td>67</td>
<td>0.41</td>
<td>113</td>
<td>0.45</td>
<td>158</td>
<td>0.15</td>
</tr>
<tr>
<td>22</td>
<td>0.12</td>
<td>68</td>
<td>0.54</td>
<td>114</td>
<td>0.23</td>
<td>159</td>
<td>1.02</td>
</tr>
<tr>
<td>23</td>
<td>1.12</td>
<td>69</td>
<td>1.11</td>
<td>115</td>
<td>0.85</td>
<td>160</td>
<td>0.27</td>
</tr>
<tr>
<td>24</td>
<td>0.33</td>
<td>70</td>
<td>0.56</td>
<td>116</td>
<td>1.34</td>
<td>161</td>
<td>0.26</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>71</td>
<td>0.67</td>
<td>117</td>
<td>0.66</td>
<td>162</td>
<td>0.06</td>
</tr>
<tr>
<td>26</td>
<td>1.03</td>
<td>72</td>
<td>0.16</td>
<td>118</td>
<td>0.75</td>
<td>163</td>
<td>0.92</td>
</tr>
<tr>
<td>27</td>
<td>0.29</td>
<td>73</td>
<td>1.58</td>
<td>119</td>
<td>0.82</td>
<td>164</td>
<td>0.46</td>
</tr>
<tr>
<td>28</td>
<td>0.12</td>
<td>75</td>
<td>1.25</td>
<td>120</td>
<td>0.75</td>
<td>165</td>
<td>0.19</td>
</tr>
<tr>
<td>29</td>
<td>1.35</td>
<td>76</td>
<td>0.29</td>
<td>121</td>
<td>1.20</td>
<td>166</td>
<td>0.34</td>
</tr>
<tr>
<td>30</td>
<td>1.08</td>
<td>77</td>
<td>0.37</td>
<td>122</td>
<td>0.52</td>
<td>167</td>
<td>0.30</td>
</tr>
<tr>
<td>31</td>
<td>0.47</td>
<td>78</td>
<td>0.70</td>
<td>123</td>
<td>0.94</td>
<td>168</td>
<td>0.98</td>
</tr>
<tr>
<td>32</td>
<td>0.51</td>
<td>79</td>
<td>1.07</td>
<td>124</td>
<td>0.26</td>
<td>169</td>
<td>0.59</td>
</tr>
<tr>
<td>33</td>
<td>0.99</td>
<td>80</td>
<td>0.18</td>
<td>125</td>
<td>1.12</td>
<td>170</td>
<td>0.51</td>
</tr>
<tr>
<td>34</td>
<td>0.15</td>
<td>81</td>
<td>0.17</td>
<td>126</td>
<td>0.41</td>
<td>171</td>
<td>0.25</td>
</tr>
<tr>
<td>35</td>
<td>1.73</td>
<td>82</td>
<td>0.47</td>
<td>127</td>
<td>0.17</td>
<td>172</td>
<td>0.87</td>
</tr>
<tr>
<td>36</td>
<td>0.88</td>
<td>83</td>
<td>0.26</td>
<td>128</td>
<td>1.75</td>
<td>173</td>
<td>0.27</td>
</tr>
<tr>
<td>37</td>
<td>0.52</td>
<td>84</td>
<td>0.66</td>
<td>129</td>
<td>0.15</td>
<td>174</td>
<td>0.82</td>
</tr>
<tr>
<td>38</td>
<td>0.69</td>
<td>85</td>
<td>0.42</td>
<td>130</td>
<td>0.94</td>
<td>175</td>
<td>0.42</td>
</tr>
<tr>
<td>39</td>
<td>1.01</td>
<td>86</td>
<td>1.03</td>
<td>131</td>
<td>0.97</td>
<td>176</td>
<td>0.33</td>
</tr>
<tr>
<td>40</td>
<td>0.50</td>
<td>87</td>
<td>0.24</td>
<td>132</td>
<td>0.36</td>
<td>177</td>
<td>1.17</td>
</tr>
<tr>
<td>41</td>
<td>1.76</td>
<td>88</td>
<td>0.66</td>
<td>133</td>
<td>1.14</td>
<td>178</td>
<td>0.19</td>
</tr>
<tr>
<td>42</td>
<td>0.23</td>
<td>89</td>
<td>0.90</td>
<td>134</td>
<td>0.48</td>
<td>179</td>
<td>0.21</td>
</tr>
<tr>
<td>43</td>
<td>0.10</td>
<td>90</td>
<td>0.92</td>
<td>135</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>1.63</td>
<td>91</td>
<td>0.91</td>
<td>136</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>0.62</td>
<td>92</td>
<td>2.61</td>
<td>137</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>0.77</td>
<td>70</td>
<td>0.56</td>
<td>129</td>
<td>1.75</td>
<td>173</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Notes: Region number corresponds to the BEA code associated with that economic area. $\kappa = 2$. All posterior draws of $\mu_{n0}$ lie on the same side of zero as their respective posterior means.
Table B.3: Regional average employment growth during recessions – estimated coefficients $\mu_n + \mu_{n1}$ (posterior means)

<table>
<thead>
<tr>
<th>Region</th>
<th>$\mu_n + \mu_{n1}$</th>
<th>Region</th>
<th>$\mu_n + \mu_{n1}$</th>
<th>Region</th>
<th>$\mu_n + \mu_{n1}$</th>
<th>Region</th>
<th>$\mu_n + \mu_{n1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-0.81</td>
<td>47.</td>
<td>-1.76</td>
<td>93.</td>
<td>-0.82</td>
<td>138.</td>
<td>-1.02</td>
</tr>
<tr>
<td>2.</td>
<td>-0.19</td>
<td>48.</td>
<td>-0.52</td>
<td>94.</td>
<td>-0.47</td>
<td>139.</td>
<td>-0.73</td>
</tr>
<tr>
<td>3.</td>
<td>-1.28</td>
<td>49.</td>
<td>0.55</td>
<td>95.</td>
<td>0.13</td>
<td>140.</td>
<td>-0.03</td>
</tr>
<tr>
<td>4.</td>
<td>-0.13</td>
<td>50.</td>
<td>-1.09</td>
<td>96.</td>
<td>-0.26</td>
<td>141.</td>
<td>-0.29</td>
</tr>
<tr>
<td>5.</td>
<td>-0.50</td>
<td>51.</td>
<td>-0.41</td>
<td>97.</td>
<td>-1.43</td>
<td>142.</td>
<td>-0.48</td>
</tr>
<tr>
<td>6.</td>
<td>-0.34</td>
<td>52.</td>
<td>-0.85</td>
<td>98.</td>
<td>-0.58</td>
<td>143.</td>
<td>-0.54</td>
</tr>
<tr>
<td>7.</td>
<td>0.17</td>
<td>53.</td>
<td>-0.58</td>
<td>99.</td>
<td>-0.35</td>
<td>144.</td>
<td>0.39</td>
</tr>
<tr>
<td>8.</td>
<td>0.34</td>
<td>54.</td>
<td>-0.25</td>
<td>100.</td>
<td>-0.81</td>
<td>145.</td>
<td>0.16</td>
</tr>
<tr>
<td>9.</td>
<td>-0.46</td>
<td>55.</td>
<td>0.58</td>
<td>101.</td>
<td>0.19</td>
<td>146.</td>
<td>-2.20</td>
</tr>
<tr>
<td>10.</td>
<td>-0.71</td>
<td>56.</td>
<td>-0.03</td>
<td>102.</td>
<td>-1.32</td>
<td>147.</td>
<td>0.04</td>
</tr>
<tr>
<td>11.</td>
<td>-0.85</td>
<td>57.</td>
<td>1.73</td>
<td>103.</td>
<td>-0.73</td>
<td>148.</td>
<td>0.13</td>
</tr>
<tr>
<td>12.</td>
<td>0.07</td>
<td>58.</td>
<td>-1.01</td>
<td>104.</td>
<td>0.81</td>
<td>149.</td>
<td>0.25</td>
</tr>
<tr>
<td>13.</td>
<td>-0.48</td>
<td>59.</td>
<td>-1.14</td>
<td>105.</td>
<td>-0.98</td>
<td>150.</td>
<td>-1.04</td>
</tr>
<tr>
<td>14.</td>
<td>0.10</td>
<td>60.</td>
<td>-0.87</td>
<td>106.</td>
<td>-1.07</td>
<td>151.</td>
<td>-0.53</td>
</tr>
<tr>
<td>15.</td>
<td>-1.02</td>
<td>61.</td>
<td>0.10</td>
<td>107.</td>
<td>-0.53</td>
<td>152.</td>
<td>-1.59</td>
</tr>
<tr>
<td>16.</td>
<td>0.16</td>
<td>62.</td>
<td>0.12</td>
<td>108.</td>
<td>-1.15</td>
<td>153.</td>
<td>-0.50</td>
</tr>
<tr>
<td>17.</td>
<td>-0.06</td>
<td>63.</td>
<td>-0.27</td>
<td>109.</td>
<td>-0.84</td>
<td>154.</td>
<td>-0.56</td>
</tr>
<tr>
<td>18.</td>
<td>-0.23</td>
<td>64.</td>
<td>-0.28</td>
<td>110.</td>
<td>0.25</td>
<td>155.</td>
<td>0.45</td>
</tr>
<tr>
<td>19.</td>
<td>0.48</td>
<td>65.</td>
<td>-1.04</td>
<td>111.</td>
<td>0.21</td>
<td>156.</td>
<td>-1.29</td>
</tr>
<tr>
<td>20.</td>
<td>-0.14</td>
<td>66.</td>
<td>-1.53</td>
<td>112.</td>
<td>-0.95</td>
<td>157.</td>
<td>-0.73</td>
</tr>
<tr>
<td>21.</td>
<td>-1.30</td>
<td>67.</td>
<td>-1.22</td>
<td>113.</td>
<td>-0.58</td>
<td>158.</td>
<td>-1.00</td>
</tr>
<tr>
<td>22.</td>
<td>-0.68</td>
<td>68.</td>
<td>-1.51</td>
<td>114.</td>
<td>-0.76</td>
<td>159.</td>
<td>-0.25</td>
</tr>
<tr>
<td>23.</td>
<td>-0.04</td>
<td>69.</td>
<td>-0.33</td>
<td>115.</td>
<td>-0.14</td>
<td>160.</td>
<td>-1.07</td>
</tr>
<tr>
<td>24.</td>
<td>-0.27</td>
<td>70.</td>
<td>-0.05</td>
<td>116.</td>
<td>0.18</td>
<td>161.</td>
<td>-0.33</td>
</tr>
<tr>
<td>25.</td>
<td>-0.10</td>
<td>71.</td>
<td>-0.43</td>
<td>117.</td>
<td>-0.73</td>
<td>162.</td>
<td>-0.72</td>
</tr>
<tr>
<td>26.</td>
<td>-0.22</td>
<td>72.</td>
<td>-1.24</td>
<td>118.</td>
<td>-1.03</td>
<td>163.</td>
<td>-0.61</td>
</tr>
<tr>
<td>27.</td>
<td>-0.50</td>
<td>73.</td>
<td>0.53</td>
<td>119.</td>
<td>-0.10</td>
<td>164.</td>
<td>-1.42</td>
</tr>
<tr>
<td>28.</td>
<td>-0.25</td>
<td>75.</td>
<td>-0.14</td>
<td>120.</td>
<td>-0.53</td>
<td>165.</td>
<td>-0.78</td>
</tr>
<tr>
<td>29.</td>
<td>-0.03</td>
<td>76.</td>
<td>-0.47</td>
<td>121.</td>
<td>-0.39</td>
<td>166.</td>
<td>-0.61</td>
</tr>
<tr>
<td>30.</td>
<td>-1.08</td>
<td>77.</td>
<td>-0.35</td>
<td>122.</td>
<td>-0.79</td>
<td>167.</td>
<td>-0.43</td>
</tr>
<tr>
<td>31.</td>
<td>-1.06</td>
<td>78.</td>
<td>-0.19</td>
<td>123.</td>
<td>-0.30</td>
<td>168.</td>
<td>-0.66</td>
</tr>
<tr>
<td>32.</td>
<td>-0.60</td>
<td>79.</td>
<td>-0.79</td>
<td>124.</td>
<td>-0.75</td>
<td>169.</td>
<td>-0.47</td>
</tr>
<tr>
<td>33.</td>
<td>0.53</td>
<td>80.</td>
<td>-0.52</td>
<td>125.</td>
<td>-0.16</td>
<td>170.</td>
<td>-1.14</td>
</tr>
<tr>
<td>34.</td>
<td>-1.03</td>
<td>81.</td>
<td>-0.55</td>
<td>126.</td>
<td>-0.73</td>
<td>171.</td>
<td>-1.38</td>
</tr>
<tr>
<td>35.</td>
<td>-0.34</td>
<td>82.</td>
<td>-0.97</td>
<td>127.</td>
<td>-0.81</td>
<td>172.</td>
<td>0.04</td>
</tr>
<tr>
<td>36.</td>
<td>0.29</td>
<td>83.</td>
<td>-0.75</td>
<td>128.</td>
<td>-0.47</td>
<td>173.</td>
<td>-0.07</td>
</tr>
<tr>
<td>37.</td>
<td>-0.75</td>
<td>84.</td>
<td>-0.68</td>
<td>129.</td>
<td>-0.55</td>
<td>174.</td>
<td>-0.58</td>
</tr>
<tr>
<td>38.</td>
<td>-0.44</td>
<td>85.</td>
<td>0.10</td>
<td>130.</td>
<td>-0.80</td>
<td>175.</td>
<td>-0.34</td>
</tr>
<tr>
<td>39.</td>
<td>0.15</td>
<td>86.</td>
<td>-0.02</td>
<td>131.</td>
<td>-1.08</td>
<td>176.</td>
<td>-0.22</td>
</tr>
<tr>
<td>40.</td>
<td>-0.21</td>
<td>87.</td>
<td>-0.24</td>
<td>132.</td>
<td>-0.69</td>
<td>177.</td>
<td>-0.41</td>
</tr>
<tr>
<td>41.</td>
<td>-0.78</td>
<td>88.</td>
<td>0.18</td>
<td>133.</td>
<td>-0.14</td>
<td>178.</td>
<td>-1.18</td>
</tr>
<tr>
<td>42.</td>
<td>-0.81</td>
<td>89.</td>
<td>0.28</td>
<td>134.</td>
<td>0.20</td>
<td>179.</td>
<td>-0.82</td>
</tr>
<tr>
<td>43.</td>
<td>-1.36</td>
<td>90.</td>
<td>-0.33</td>
<td>135.</td>
<td>-0.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>-1.22</td>
<td>91.</td>
<td>-0.50</td>
<td>136.</td>
<td>-0.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>-0.19</td>
<td>92.</td>
<td>1.20</td>
<td>137.</td>
<td>-0.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Region number corresponds to the BEA code associated with that economic area. $\kappa = 2$. All posterior draws of $\mu_{n1}$ lie on the same side of zero as their respective posterior means.
Table B.4: Estimated $\sigma_n^2$ (posterior means)

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

Notes: Region number corresponds to the BEA code associated with that economic area. $\kappa = 2$. 

B.1. FIGURES, TABLES AND CONVERGENCE