

University of Rhode Island

DigitalCommons@URI

Open Access Master's Theses

2018

Bayesian Model Averaging of Space Time Car Models with Application to U.S. House Price Forecasting

Caoxin Sun

University of Rhode Island, caoxin@uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Recommended Citation

Sun, Caoxin, "Bayesian Model Averaging of Space Time Car Models with Application to U.S. House Price Forecasting" (2018). *Open Access Master's Theses*. Paper 1181.

<https://digitalcommons.uri.edu/theses/1181>

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

BAYESIAN MODEL AVERAGING OF SPACE TIME CAR MODELS
WITH APPLICATION TO U.S. HOUSE PRICE FORECASTING

BY
CAOXIN SUN

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
STATISTICS

UNIVERSITY OF RHODE ISLAND

2018

MASTER OF SCIENCE THESIS
OF
CAOXIN SUN

APPROVED:

Thesis Committee:

Major Professor Gavino Puggioni

Steffen Ventz

Todd Guilfoos

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2018

ABSTRACT

The housing market has been a significant contribution to U.S. GDP. Forecasting the house price growth rate helps to regulate risks associated with the housing sector and further helps to stabilize the economy. However, due to the volatility in the housing market, forecasting the house price growth rate has been a tough task. In this thesis, we built a conditional autoregressive model incorporated with bayesian model averaging (BMA-CAR) based on quarterly observations from 1976 to 1994 and tested forecasting capability over 1995 to 2012. We extended upon the results of Bork [International Journal of Forecasting, 31, 1 (2015)] to include the effects of spatial autocorrelation but inhibited the allowance for the model and coefficients shifts over time. Our model is based on a hierarchical structure that allows BMA to average out the effects from predictors along with CAR model to account for the remaining spatial structures in the data.

ACKNOWLEDGMENTS

I would like to thank the Computer Science and Statistics Department affording me the financial support studying here as a Master Student; my thesis advisor, Dr. Gavino Puggioni, who helped to shape the framework of my thesis and encouraged me to challenge myself during my struggle time. After taking his Bayesian Statistics class, I decided to pursue a degree in this department. He is the mentor that guided me through the program and lead me into the beautiful world of Statistics; my committee members (Dr. Steffen Ventz and Dr. Todd Guilfoos) who gave me constructive and insightful advice on my thesis writing on model specifications and economics studies; my friends for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last two years.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
TABLE OF CONTENTS	iv
LIST OF FIGURES	vi
LIST OF TABLES	viii
CHAPTER	
1 Introduction	1
1.1 Studying Housing Market	1
1.2 Areal Data and Spatial CAR	3
1.3 Bayesian Model Averaging	4
List of References	5
2 Bayesian Model Averaging of Space Time Conditional Autoregressive Models(BMA-CAR)	8
2.1 Set Up of the Hierarchical Structure	8
2.2 CAR for Spatial Temporal Random Effects ϕ_{kt}	9
2.3 Prior Distributions	11
2.3.1 Model Space Priors	11
2.3.2 Parameter Space Priors	12
2.4 Posterior Computation	14
2.5 Forecasting	15
2.5.1 Bayesian Prediction	15

	Page
2.5.2 MSFE and MSFE ratio	15
List of References	16
3 Application	18
3.1 Data Description	18
3.2 Preliminary Analysis of the Data	20
3.2.1 Neighbor Structure	20
3.2.2 Exploratory Data Analysis	21
3.3 Results	26
3.3.1 Model Selection Results	26
3.3.2 Estimation Results	28
3.3.3 Forecasting Results	30
List of References	35
4 Discussion and Future Work	41
4.1 Performance Gain by Spatial Components and Limitations	41
4.1.1 Model Dimension and Variable Inclusion	41
4.1.2 Limitations of a Global Fixed Effect	42
4.1.3 An Empirical Model	43
4.2 Future Work	44
4.2.1 Data Acquisition	44
4.2.2 Improvement of BMA-CAR Model	45
List of References	46
BIBLIOGRAPHY	47

LIST OF FIGURES

Figure		Page
1	Time Series of real house price growth rate and all the co- variates.	19
2	Neighbor structure used in CAR models.	21
3	Eight BEA regions.	22
4	Time Series of real house price growth rate. 48 states are grouped in the eight BEA regions. Dashed line separates training data from test data. Note that the data was trans- formed by a factor of 400 (see equation 11).	24
5	P-value from Moran’s I test of observed house price growth rate over time. Red dashed line is the 0.05 significant level.	25
6	ACF plot of observed house price growth rate against time. Thick black line is the mean across the 48 states. Two dashed lines are ± 0.16	25
7	Posterior model size distribution after fitting BMA-CAR.	27
8	Posterior model size distribution after fitting BMA-CAR with dilution prior and power of 1.	28
9	p-value from Moran’s I test on residuals over time. Red dashed line is the 0.05 significant level.	29
10	ACF plot of residuals against time. Thick black line is the mean across the 48 states. Two dashed lines are ± 0.16	30
11	Cumulative squared forecast error difference.	32
12	Forecasting results across the 48 states.	35
13	Forecast vs. observed: best predicted states (top 4) vs. worst (bottom 4). Red dashed line represents forecast from BMA-CAR; blue dashed line represents forecast from MEAN benchmark. Blue line is the historic data. Vertical line is the separation of the boom (1995:1-2006:4) and bust period (2007:1-2012:4).	36

Figure	Page	
14	MSFE ratio: best predicted states (top 4) vs. worst (bottom 4). Red dashed line represents the MSFE for BMA-CAR; black line is MSFE for MEAN benchmark. Vertical line represents the separation of the boom and bust period. Notice the different scale on the y-axis.	37
15	Forecast vs. observed: top 4 volatility states (top) vs. bottom 4 (bottom).	38
16	MSFE ratio: top 4 volatility states (top) vs. bottom 4 (bottom).	39
17	Forecasting results across the 48 states.	44

LIST OF TABLES

Table		Page
1	Predictors and Abbreviations	20
2	Average of correlation coefficients within and between regions: New England(NE), Mideast(ME), Great Lakes(GL), Plains(PL), Southwest(SW), Rocky Mountain(RM),Far West(FW).	22
3	Top 5 Most Visited Models	27
4	Top 5 Most Visited Models with dilution prior and power of 1.	28
5	Forecast errors across states	33

CHAPTER 1

Introduction

1.1 Studying Housing Market

Understanding the housing market dynamics are crucial, as the housing sector constitutes a significant share of the GDP and it is the largest component of household wealth in the U.S. [1]. The Bureau of Labor Statistics has estimated in 2010 that roughly 24 percent of the total consumption of American homeowners goes toward housing [2]. Starting from the late 1990s, the U.S. housing market was in a boom period until 2007 when the sub-prime crisis occurred and was followed instantly by a downturn. Modeling and predicting U.S. housing price has received attention from governments, real estate developers and investors. However, it has been a challenging task due to the strong vulnerability of the housing sector to structural changes, macroeconomic policies, regime switching, and market imperfections [1].

For the past four decades, many economists have adopted time series approaches to study the relationship between the U.S. house price and the socio-economic variables. Originally, a large portion of the research pool focused on exploring explanatory variables in the time series regression ([3],[4],[5]) and/or modifying the error terms([6],[7],[8]). Recently, research focused on developing dynamic models that allow for time-varying coefficients. These models could substantially improve prediction by accounting for subsample parameter instability seen in real estate data. Most common methods include, but are not limited to, regime switching models ([9],[10]) and AR, VAR, GARCH models coupled with Kalman Filter techniques. ([11],[1], [12]).

As noted by [13], it is possible that contiguous states may influence each other's housing prices. [2] grouped real house price into 8 Bureau of Economic

Analysis (BEA) regions. The within region correlation is larger than the between region correlation with only one exception. Other studies, such as [14], proved the necessity of using spatial model accounting for dependent residuals arose from only using ordinary least squares. There had been several studies that accounted for spatial autocorrelation, which were summarised in [2]. One major branch of methods is the spatially adapted version of VARS (SpVAR) [15]. SpVAR belongs to the group of Spatial-Temporal Autoregressive Moving Average (STARMA) methods. Seemingly Unrelated Regression (SUR) and error panel data models have also been largely used, though it favors a relatively small number of regions just as STARMA models do. Other authors attempted common correlated effects estimator (CCE)[16][13] and Spatial Dynamic Structural Equation models (SD-SEM)[2] to model common shocks on housing price.

In the very recent study [12], the authors implemented Dynamic Model Selection and Dynamic Model Averaging methods and demonstrated the importance of allowing for both time-varying parameters and model changes. One limitation, however, is that the natural spatial structures existing in the data are not considered in their model. In fact, univariate analysis on each state separately was performed. In this study, we propose to use Conditional Autoregressive (CAR) methods incorporated with Bayesian Model Averaging (BMA) as an alternative forecasting model on the data from [12]. We developed the model in a hierarchical structure that is composed of a fixed effect and a random effect. Fixed effect models the influence from the predictors and the random effect captures the spatial-temporal variations in the data after removing the fixed effect. We hope to achieve better forecasts by taking into consideration the spatial autocorrelation.

1.2 Areal Data and Spatial CAR

Unlike point process where points are all neighbors to each other on a continuous surface, areal data have well defined boundaries and observed data are frequently aggregations within the boundaries or the areal units themselves constitute the units of observation [17]. CAR models are widely used to describe the spatial variation of areal data. CAR models have been extensively used for the analysis of spatial data in diverse areas, such as demography, economics, epidemiology and geography [18].

Modeling spatial autocorrelation for areal data requires the creation of neighbor structures and corresponding weight matrix. There are multiple ways to define "neighbor", such as contiguity-based, graph-based, and k nearest neighbor, after which the neighbor object is converted into a spatial matrix to quantify spatial dependence. Spatial weights matrix \mathbf{W} has elements ω_{ks} that represents the weights of the spatial link between spatial units S_k and S_s . When little is known about the spatial process, a common approach is to take binary representation in which one is for neighbors and zero otherwise [17].

CAR model is a method for smoothing areal data that was originally proposed by [19]. In a CAR model, the spatial component of the center areal unit is seen as conditionally dependent on a weighted average of the spatial components from all the other units, and the weights come from previously created weight matrix. If we use y_k to indicate any observed value of interest, the full conditional distribution for y_k has the following format:

$$[Y_k | y_s, s \neq k] \sim N\left(\sum_s \frac{\omega_{ks} y_s}{\omega_{k+}}, \frac{\tau^2}{\omega_{k+}}\right) \quad (1)$$

where the ω_{k+} is the total number of neighbors of areal unit k , and $\frac{\tau^2}{\omega_{k+}}$ can be viewed as the local variance for areal unit k . The CAR model is often used as a prior for the random effect in the context of a hierarchical model. In a

hierarchical model, we normally use ϕ_k to represent the spatial random effect that failed to be modeled by the fixed effects. Replacing the y_k in Equation 1 by ϕ_k and we would have:

$$[\phi_k | \phi_s, s \neq k] \sim N\left(\sum_s \frac{\omega_{ks}\phi_s}{\omega_{k+}}, \frac{\tau^2}{\omega_{k+}}\right) \quad (2)$$

This model can be denoted as $CAR(\mathbf{W}, \sigma^{-2})$ [20] with a joint distribution as the following by applying Brook's Lemma [21]:

$$[\phi | \mathbf{W}] \propto \exp\left(-\frac{1}{2\sigma^2}\phi^T(\mathbf{D} - \mathbf{W})\phi\right) \quad (3)$$

where \mathbf{D} is the diagonal matrix with elements ω_{k+} . The precision matrix $\Sigma_\phi^{-1} = (\mathbf{D} - \mathbf{W})$ is singular so that the above joint distribution is improper [20]. After some modifications, a proper CAR model has the following covariance [22] :

$$\sigma^2\Sigma_\phi^{-1} = \rho_s(\mathbf{D} - \mathbf{W}) + (1 - \rho_s)\mathbf{I} \quad (4)$$

where $0 \leq \rho_s \leq 1$. If $\rho_s = 1$, CAR model goes back to the improper case as in Equation 1.2. The precision matrix is non-singular if $\Sigma_\phi^{-1} \neq 1$.

1.3 Bayesian Model Averaging

As for the fixed effects, a set of predictors are needed to capture covariates effects. However, it remains a question as to which variables to include in the model; including all covariates adds to computational burden as well as over-fitting. This variable selecting problem has a natural Bayesian solution [23]. Advantages of using BMA include but is not limited to automatic adjustment for multiple comparisons and efficient model space exploration as well as lower forecasting error compared to using one single model [24].

Since the number of variables is relatively small in our case, we can exhaust all possible models and compute their posterior probability in the model space

\mathcal{M} . Suppose a set of K models $\mathcal{M} = M_1, \dots, M_k$ are under consideration for data Y , and θ_k is a vector of unknown parameters that indexes the members of M_k . The probability that M_k in fact generated the data, conditionally on having observed Y , is the following posterior model probability [25]:

$$p(M_k|Y) = \frac{p(Y|M_k)p(M_k)}{\sum_k p(Y|M_k)p(M_k)} \quad (5)$$

where $p(M_k)$ is the prior for model M_k and $p(Y|M_k)$ is the marginal distribution of data under this model that is achieved by integrating out θ_k :

$$\int p(Y|M_k) = \int p(Y|\theta_k, M_k)p(\theta_k|M_k)d\theta_k \quad (6)$$

List of References

- [1] H. S. Guirguis, C. I. Giannikos, and R. I. Anderson, “The us housing market: asset pricing forecasts using time varying coefficients,” *The Journal of real estate finance and economics*, vol. 30, no. 1, pp. 33–53, 2005.
- [2] P. Valentini, L. Ippoliti, and L. Fontanella, “Modeling us housing prices by spatial dynamic structural equation models,” *The Annals of Applied Statistics*, pp. 763–798, 2013.
- [3] D. DiPasquale and W. C. Wheaton, “Housing market dynamics and the future of housing prices,” *Journal of urban economics*, vol. 35, no. 1, pp. 1–27, 1994.
- [4] N. Pain and P. Westaway, “Modelling structural change in the uk housing market: a comparison of alternative house price models,” *Economic Modelling*, vol. 14, no. 4, pp. 587–610, 1997.
- [5] S. Holly and N. Jones, “House prices since the 1940s: cointegration, demography and asymmetries,” *Economic Modelling*, vol. 14, no. 4, pp. 549–565, 1997.
- [6] J. M. Quigley, “A simple hybrid model for estimating real estate price indexes,” *Journal of Housing Economics*, vol. 4, no. 1, pp. 1–12, 1995.
- [7] J. M. Abraham and P. H. Hendershott, “Bubbles in metropolitan housing markets,” National Bureau of Economic Research, Tech. Rep., 1994.
- [8] S. Malpezzi, “A simple error correction model of house prices,” *Journal of housing economics*, vol. 8, no. 1, pp. 27–62, 1999.

- [9] S. Hall, Z. Psaradakis, and M. Sola, “Switching error-correction models of house prices in the united kingdom,” *Economic Modelling*, vol. 14, no. 4, pp. 517–527, 1997.
- [10] G. W. Crawford and M. C. Fratantoni, “Assessing the forecasting performance of regime-switching, arima and garch models of house prices,” *Real Estate Economics*, vol. 31, no. 2, pp. 223–243, 2003.
- [11] J. Clapp and C. Giaccotto, “Evaluating house price forecasts,” *Journal of Real Estate Research*, vol. 24, no. 1, pp. 1–26, 2002.
- [12] L. Bork and S. V. Møller, “Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection,” *International Journal of Forecasting*, vol. 31, no. 1, pp. 63–78, 2015.
- [13] S. Holly, M. H. Pesaran, and T. Yamagata, “A spatio-temporal model of house prices in the usa,” *Journal of Econometrics*, vol. 158, no. 1, pp. 160–173, 2010.
- [14] J. P. LeSage, “Bayesian estimation of spatial autoregressive models,” *International Regional Science Review*, vol. 20, no. 1-2, pp. 113–129, 1997.
- [15] T. H. Kuethe and V. O. Pedde, “Regional housing price cycles: a spatio-temporal analysis using us state-level data,” *Regional studies*, vol. 45, no. 5, pp. 563–574, 2011.
- [16] B. Van Dijk, P. H. Franses, R. Paap, and D. Van Dijk, “Modelling regional house prices,” *Applied Economics*, vol. 43, no. 17, pp. 2097–2110, 2011.
- [17] R. S. Bivand, E. J. Pebesma, V. Gomez-Rubio, and E. J. Pebesma, “Applied spatial data analysis with r.” Springer, 2008, vol. 747248717, ch. 9, pp. 237–268.
- [18] V. De Oliveira, “Bayesian analysis of conditional autoregressive models,” *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 1, pp. 107–133, 2012.
- [19] J. Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236, 1974.
- [20] S. Guha and L. Ryan, “Spatio-temporal analysis of areal data and discovery of neighborhood relationships in conditionally autoregressive models,” 2006.
- [21] D. Brook, “On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems,” *Biometrika*, vol. 51, no. 3/4, pp. 481–483, 1964.

- [22] B. N. Leroux BG, Lei X, “Estimation of disease rates in small areas: a new mixed model for spatial dependence,” in *Statistical models in epidemiology, the environment, and clinical trials*. Springer, 2000, p. 179191.
- [23] P. D. Hoff, *A first course in Bayesian statistical methods*. Springer Science & Business Media, 2009.
- [24] A. Rodriguez and G. Puggioni, “Mixed frequency models: Bayesian approaches to estimation and prediction,” *International Journal of Forecasting*, vol. 26, no. 2, pp. 293–311, 2010.
- [25] H. Chipman, E. I. George, R. E. McCulloch, M. Clyde, D. P. Foster, and R. A. Stine, “The practical implementation of bayesian model selection,” *Lecture Notes-Monograph Series*, pp. 65–134, 2001.

CHAPTER 2

Bayesian Model Averaging of Space Time Conditional Autoregressive Models(BMA-CAR)

2.1 Set Up of the Hierarchical Structure

Suppose the study region covers a set of $k = 1, \dots, K$ non-overlapping areal units $S = S_1, \dots, S_K$ and data are recorded for each time unit for $t = 1, \dots, T$ consecutive time periods. A generalized linear mixed effects model is a convenient candidate for modeling this type of data. The hierarchical structure is as following [1]:

$$Y_{kt} | \mu_{kt} \sim f(y_{kt} | \mu_{kt}, \sigma^2)$$

$$g(\mu_{kt}) = \mathbf{X}_{kt} \boldsymbol{\beta} + \phi_{kt}$$

We define here Y_{kt} as the observed house price growth rate at time t and location k that has in total $K \times T$ rows of observations, and X_{kt} as a vector of $p + 1$ covariates (including the intercept). β is a vector of $p + 1$ that are invariant to space and time changes. ϕ_{kt} has a length of $K \times T$. We expand the second line of the two equations above to show the dimension of each variable as the following:

$$\begin{pmatrix} \mu_{11} \\ \vdots \\ \mu_{K1} \\ \mu_{12} \\ \vdots \\ \mu_{KT} \end{pmatrix} = \begin{pmatrix} x_{11,0} & \dots & x_{11,p} \\ \vdots & & \\ x_{K1,0} & \dots & x_{K1,p} \\ x_{12,0} & \dots & x_{12,p} \\ \vdots & & \\ x_{KT,0} & \dots & x_{KT,p} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \phi_{11} \\ \vdots \\ \phi_{K1} \\ \phi_{12} \\ \vdots \\ \phi_{KT} \end{pmatrix}$$

And we provide here an explanation of the two-stage structure in the above hierarchical model:

- (i) $Y_{kt} | \mu_{kt} \sim f(y_{kt} | \mu_{kt}, \sigma^2)$. At the first stage, the observed value of the

house price growth rate Y_{kt} at specific time t and location k is from a normal distribution with true mean μ_{kt} and observational error σ^2 .

- (ii) $g(\mu_{kt}) = \mathbf{X}_{kt}\boldsymbol{\beta} + \phi_{kt}$. At the second stage, the likelihood was chosen to be Gaussian since the house price growth rate is a continuous response. Therefore, the g function is just the identity link. With identity link, $g(\mu_{kt})$ becomes μ_{kt} and it is approximated by the fixed effects $\mathbf{X}_{kt}\boldsymbol{\beta}$ that captures the influence from predictors plus the spatial-temporal random effects ϕ_{kt} that models the patterns seen in the data after fitting the fixed effects. We assume here that the second stage of the model is a simple linear mixed effects model, that both of the components are additive and enter the model linearly without higher orders or interaction terms.

2.2 CAR for Spatial Temporal Random Effects ϕ_{kt}

In the context of a hierarchical model, we would like to use CAR model as a prior for this random effect ϕ_{kt} . The random effect ϕ_{kt} can be modeled in several methods.

- (i) Linear CAR

$$\phi_{kt} = \beta_1 + \phi_k + (\alpha + \delta_k) \frac{t - \bar{t}}{T}$$

$$\phi_k | \phi_{-k}, \mathbf{W} \sim N \left(\frac{\rho_{int} \sum_{j=1}^K \omega_{kj} \phi_j}{\rho_{int} \sum_{j=1}^K \omega_{kj} + 1 - \rho_{int}}, \frac{\tau_{int}^2}{\rho_{int} \sum_{j=1}^K \omega_{kj} + 1 - \rho_{int}} \right)$$

$$\delta_k | \delta_{-k}, \mathbf{W} \sim N \left(\frac{\rho_{slo} \sum_{j=1}^K \omega_{kj} \delta_j}{\rho_{slo} \sum_{j=1}^K \omega_{kj} + 1 - \rho_{slo}}, \frac{\tau_{slo}^2}{\rho_{slo} \sum_{j=1}^K \omega_{kj} + 1 - \rho_{slo}} \right)$$

In this model, a linear trend with time is assumed. Each areal unit k has its own variation of intercept ϕ_k and slope δ_k from the mean linear trend (intercept of ϕ_1 and slope α) [2]. Both of ϕ_k and δ_k are modeled by CAR prior from the paper [3].

(ii) Time Autoregressive CAR(CARar)

CARar is a spatially autocorrelated autoregressive time series model first introduced by [4]. The single set of random effects $\phi = (\phi_1, \phi_2, \dots, \phi_T)$ is decomposed as:

$$f(\phi_1, \phi_2, \dots, \phi_T) \sim f(\phi_1) \prod_{j=2}^T f(\phi_j | \phi_{j-1})$$

The decomposition above induces temporal autocorrelation by allowing ϕ_j to depend on ϕ_{j-1} . Then, spatial autocorrelation is introduced at ϕ_1 by using Gaussian Markov Random Field (GMRF) prior that is constant over time. The joint prior distribution for ϕ_1 is given by:

$$\phi_1 \sim N(\mathbf{0}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_S)^{-1})$$

where the spatial component is introduced by the variance $\tau^2 \mathbf{Q}(\mathbf{W}, \rho_S)^{-1}$ that corresponds to proper CAR prior. The precision \mathbf{Q} is defined in the previous chapter in the same way as displayed in Equation 4 but with slightly different representation [3]. \mathbf{D} is the diagonal matrix with elements ω_{k+} but can also be written as $\mathbf{W}\mathbf{1}$:

$$\mathbf{Q}(\mathbf{W}, \rho_S) = \rho_S [\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}] + (1 - \rho_S) \mathbf{I} \quad (7)$$

where \mathbf{I} is the $K * K$ identity matrix and $\mathbf{1}$ is a vector of ones. The ρ_S controls the spatial autocorrelation structure and represents λ as in Equation 4: $\rho = 1$ indicates intrinsic CAR prior [5] where the conditional expectation is the average of the random effects from neighbor units; $\rho = 0$ corresponds to independent random effects. Temporal autocorrelation is then induced by:

$$\phi_t | \phi_{t-1} \sim N(\rho_T \phi_{t-1}, \tau^2 \mathbf{Q}(\mathbf{W}, \rho_S)^{-1})$$

where ρ_T is the temporal autoregressive coefficient: $\rho_T = 1$ means strong temporal autocorrelation and in fact the temporal process turns into a random walk; $\rho_T = 0$ leads to temporal independence. In this model, the multivariate temporal autoregressive component is introduced via the mean $\rho_T \phi_{t-1}$ [4].

(iii) Adaptive CAR

This model assumes the autoregressive structure as the previous model CARar. However, it allows for localized spatial structure by modeling the non-zero elements of the neighbor matrix \mathbf{W} as unknown parameters [6]. The adjacent elements $\mathbf{w}^+ = \{w_{ks} | k \sim s\}$ can be estimated and $w_{ks} \in (0, 1)$.

We implemented empirical studies on our dataset using CARlinear, CARar and CARadaptive without model averaging. CARar and CARadaptive gave comparable results and were superior to CARlinear. Because CARar is the simpler method and it gave slightly better forecasting results than CARadaptive, we only used CARar for the final model.

2.3 Prior Distributions

2.3.1 Model Space Priors

Imagine we have M models with model index $m \{1, 2, \dots, M\}$ and in total p covariates. It would be convenient to index each of the 2^p models by the vector $\gamma = (\gamma_1, \dots, \gamma_p)'$, where γ_j is an indicator for inclusion of variable X_j under model M_m . Many Bayesian variable selection implementations have used independent priors of the Bernoulli form [7]:

$$p(M_m) = \eta^{q_\gamma} (1 - \eta)^{(p - q_\gamma)} \tag{8}$$

in which case $q_\gamma = \sum \gamma_j$ the number of non-zero parameters under model M_m and the hyperparameter η is the expected probability that each variable is included. Beta prior is assumed for $\eta \sim \text{Beta}(1, 1)$.

A refinement of the Bernoulli priors is known in the literature as dilution priors. In order to downweight the probability of M_m for the collinearity in X^m , we imposed dilution prior as shown in Equation 9 where R^m is the correlation matrix such that $h(|R^m|) \propto (X^m)^T X^m$. The function h is monotone and it satisfies $h(1) = 1$ and $h(0) = 0$. Common choices of h include $h(a) = a$ and $h(a) = a^{1/2}$ [8]. This gives:

$$p(M_m) \propto h(|R^m|)\eta^{q_\gamma}(1 - \eta)^{(p-q_\gamma)} \quad (9)$$

2.3.2 Parameter Space Priors

We used g-prior to simplify our posteriors for the model averaging process. We would like to introduce the concept of g-prior starting from an ordinary linear regression case.

$$y = \mathbf{X}\beta + \epsilon$$

Assume the goal is to obtain the posterior distribution of β and σ and we do not use g-prior, we would need to sample from the full conditionals using the Markov-Chain Monte-Carlo procedure. If we let β and $1/\sigma$ have the priors as multivariate normal(β_0, Σ_0) and gamma($\nu_0/2, \nu_0\sigma_0^2/2$), we obtain the

following:

$$\begin{aligned}
\beta &\sim \text{multivariate normal}(\mathbf{m}, \mathbf{V}) \\
\mathbf{m} &= (\Sigma_0^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2)^{-1} (\Sigma_0^{-1} \beta_0 + \mathbf{X}^T \mathbf{y} / \sigma^2) \\
\mathbf{V} &= (\Sigma_0^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2)^{-1} \\
\sigma^2 &\sim \text{inverse gamma}(\nu_0 + n] / 2, [\nu_0 \sigma_0^2 + \text{SSR}(\beta)] / 2) \\
\text{SSR}(\beta) &= \sum_{i=1}^n (y_i - \beta^T x_i)^2
\end{aligned}$$

Under a Zellner's g-prior [9] $\beta_0 = \mathbf{0}, \Sigma_0 = g\sigma^2(\mathbf{X}^T \mathbf{X})^{-1}$ where amount of information from n/g observations is assumed for our prior, we can simplify the Gibbs Sampling as the Monte Carlo:

$$\begin{aligned}
\sigma^2 &\sim \text{inverse gamma}(\nu_0 + n] / 2, [\nu_0 \sigma_0^2 + \text{SSR}_g] / 2) \\
\text{SSR}_g &= \mathbf{y}^T \left(\mathbf{I} - \frac{g}{g+1} \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{y} \\
\beta &\sim \text{multivariate normal}(\mathbf{m}, \mathbf{V}) \\
\mathbf{m} &= \frac{g}{g+1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\
\mathbf{V} &= \frac{g}{g+1} \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}
\end{aligned}$$

Now we would like to apply g-prior to our model. The value of g is commonly decided as $g \sim \max(n, p^2)$, where n is the sample size, p is the number of covariates including the intercept. The other parameters have the following assignment of priors:

$$\begin{aligned}
\sigma^2 &\sim \text{Inverse Gamma}(1, 0.01) \\
\beta^m &\sim N(0, g\sigma^2((X^m)^T X^m)^{-1}) \\
\tau^2 &\sim \text{Inverse Gamma}(1, 0.01) \\
\rho_S, \rho_T &\sim \text{Uniform}(0, 1)
\end{aligned}$$

Here the prior for σ^2 has the underlying form of $\sigma^2 \sim \text{Inverse Gamma}(\nu_0/2, \nu_0\sigma_0^2/2)$.

2.4 Posterior Computation

CARBayesST is an extension of R package CARBayes, the latter of which models spatial autocorrelation that remains in the data after the covariates effects are accounted for [10]. CARBayesST is the first dedicated software for spatial-temporal areal unit modeling with conditional autoregressive priors and is capable of capturing temporally changing spatial dynamics [1]. We adopted part of the code from the package and added Bayesian Model Averaging for our model. We used Gibbs Sampling to sample from the full conditional distributions since the closed forms of the posterior distributions are not available. A description of algorithm is as follows.

- (i) Sample the model index $m \in \{1, 2, \dots, M\}$ according to the posterior probabilities $p(M_m|Y)$:

$$p(M_m|Y) = \frac{p(Y|M_m)p(M_m)}{\sum p(Y|M_m)p(M_m)}$$

and $p(Y|M_m)$ is computed by:

$$p(Y|M_m) \propto (1+g)^{-q_1/2} \left(\nu_0\sigma_0^2 + Y^T Y - \frac{g}{g+1} (\hat{\beta}^m)^T ((X^m)^T X^m) \hat{\beta}^m \right)^{-(n+\nu_0)/2}$$

where $\hat{\beta}^m$ is the ordinary least squares estimate under model M_m . Also, here $Y = Y_{obs} - \phi_{kt}$.

- (ii) Sample σ^2 from the marginal:

$$p(\sigma^2|Y) \sim IG((\nu_0 + n)/2, (\nu_0\sigma_0^2 + SSR_g)/2)$$

$$SSR_g = Y^T \left(\mathbf{I} - \frac{g}{g+1} X^m ((X^m)^T X^m)^{-1} X^m \right) Y$$

(iii) Sample β from the full conditional:

$$p(\beta|Y, \sigma) \sim N\left(\frac{g}{g+1}((X^m)^T X^m)^{-1} X^m Y, \frac{g}{g+1} \sigma^2 ((X^m)^T X^m)^{-1}\right)$$

(iv) Update ϕ_{kt} using Metropolis-Hastings

(v) Update ρ_T using R Metropolis-Hastings

(vi) Update τ^2 using R inverse gamma

(vii) Update ρ_S using Metropolis-Hastings

2.5 Forecasting

2.5.1 Bayesian Prediction

Under iteration s and model γ , we set the last time point estimation of ϕ_{kt} in the training set as the initial value for ϕ_t for prediction. Then we produce one step ahead prediction from time point t as following:

$$\begin{aligned} Y_{k,t+1}^{(s)} &\sim N(\mu_{k,t+1}^{(s)}, \sigma^{2(s)}) \\ \mu_{k,t+1}^{(s)} &= (\mathbf{X}_{k,t+1}^m)^{(s)} (\boldsymbol{\beta}^m)^{(s)} + \phi_{k,t+1}^{(s)} \\ \phi_{k,t+1}^{(s)} | \phi_{k,t}^{(s)} &\sim MVN(\rho_T \phi_{k,t}^{(s)}, \tau^{2(s)} \mathbf{Q}(\mathbf{W}, \rho_S)^{-1}) \end{aligned}$$

Multiple steps head predictions are implemented as an extension from one step ahead prediction.

2.5.2 MSFE and MSFE ratio

Model validation is based on mean squared forecasting error (MSFE) on out-of-sample set 1995:1-2012:4, in order to be consistent with [11]. MSFE is derived from squared forecast error (SFE). SFE for model j is computed as:

$$SFE_{k,t,j} = \frac{1}{S} \sum_{s=1}^S (\hat{Y}_{k,t,j}^{(s)} - Y_{k,t})^2$$

where $\hat{Y}_{k,t}^{(s)}$ is the prediction of house price growth rate Y at location k and time t from iteration s and model j ; $Y_{k,t}$ is the observed Y at the same time

and location.

The benchmark model was chosen as expanding window OLS fit with only intercept:

$$\hat{Y}_{MEAN,t+s} = \frac{\sum_{i=1}^{i=t+s-1} Y_i}{t + s - 1}$$

SFE for benchmark mean model is computed as:

$$SFE_{k,t,MEAN} = (\hat{Y}_{k,t,MEAN} - Y_{k,t})^2$$

When checking the prediction performance by state, we introduce the MSFE ratio r that is the ratio between MSFE from model j relative to the MSFE from the benchmark:

$$MSFE_{k,j} = \frac{1}{T} \sum_{t=1995:1}^{2012:4} SFE_{k,t,j}$$

$$MSFE_{k,MEAN} = \frac{1}{T} \sum_{t=1995:1}^{2012:4} SFE_{k,t,MEAN} \quad (10)$$

$$r_k = \frac{MSFE_{k,j}}{MSFE_{k,MEAN}}$$

List of References

- [1] D. Lee, A. Rushworth, and G. Napier, “Spatio-temporal areal unit modelling in r with conditional autoregressive priors using the carbayesst package.”
- [2] L. Bernardinelli, D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini, “Bayesian analysis of spacetime variation in disease risk,” *Statistics in medicine*, vol. 14, no. 21-22, pp. 2433–2443, 1995.
- [3] B. N. Leroux BG, Lei X, “Estimation of disease rates in small areas: a new mixed model for spatial dependence,” in *Statistical models in epidemiology, the environment, and clinical trials*. Springer, 2000, p. 179191.

- [4] A. Rushworth, D. Lee, and R. Mitchell, “A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater london,” *Spatial and spatio-temporal epidemiology*, vol. 10, pp. 29–38, 2014.
- [5] M. Besag J, York J, “A bayesian image restoration, with two applications in spatial statistics,” *Ann Inst Statist Math*, vol. 43, no. 1, pp. 1–20, 1991.
- [6] A. Rushworth, D. Lee, and C. Sarran, “An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 1, pp. 141–157, 2017.
- [7] H. Chipman, E. I. George, R. E. McCulloch, M. Clyde, D. P. Foster, and R. A. Stine, “The practical implementation of bayesian model selection,” *Lecture Notes-Monograph Series*, pp. 65–134, 2001.
- [8] E. I. George *et al.*, “Dilution priors: Compensating for model space redundancy,” in *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown*. Institute of Mathematical Statistics, 2010, pp. 158–165.
- [9] A. Zellner, “On assessing prior distributions and bayesian regression analysis with g-prior distributions,” *Bayesian inference and decision techniques*, 1986.
- [10] D. Lee, “Carbayes version 4.6: An r package for spatial areal unit modelling with conditional autoregressive priors,” *University of Glasgow, Glasgow*, 2017.
- [11] L. Bork and S. V. Møller, “Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection,” *International Journal of Forecasting*, vol. 31, no. 1, pp. 63–78, 2015.

CHAPTER 3

Application

3.1 Data Description

Data are retrieved from the provided appendix of the article [1]. The same data were used in order to make the modeling and forecasting comparable. House price growth rate data were converted by:

$$y_{k,t} = 400 \ln \left(\frac{P_{k,t}}{P_{k,t-1}} \right), k = 1, \dots, 48 \quad (11)$$

where $P_{k,t}$ denotes the level of real house prices in state k and time t . The original dataset consisted of 50 spreadsheets of 10 variables; with each spreadsheet containing the response variable of quarterly measurements (1975:1-2012:4) of real house price growth rate along with the other 9 predictors for each state. We divided the whole dataset into training and testing set. Training period has 3552 observations 48 states multiplied by 74 quarters (1976:03-1994:4); testing period has 3456 rows with 72 quarters of data (1995:1-2012:4). We split our data to stay in consistency with [1]'s methodology and so our forecasting results will be directly comparable.

To account for strong regional difference, both state-level predictors (price-income ratio, unemployment rate, real per capita income growth and labor force growth) and national level ones (30-year mortgage rate, the spread between 10-year and 3-month Treasury rates, industrial production growth, real consumption growth and housing starts) were used. Table 1 summarized the covariates and their abbreviations that are used later in this chapter. Figure 1 included the time series plots of both the dependent variable (house price growth rate) and the independent variables (all the 9 predictors) to show their variations with time.

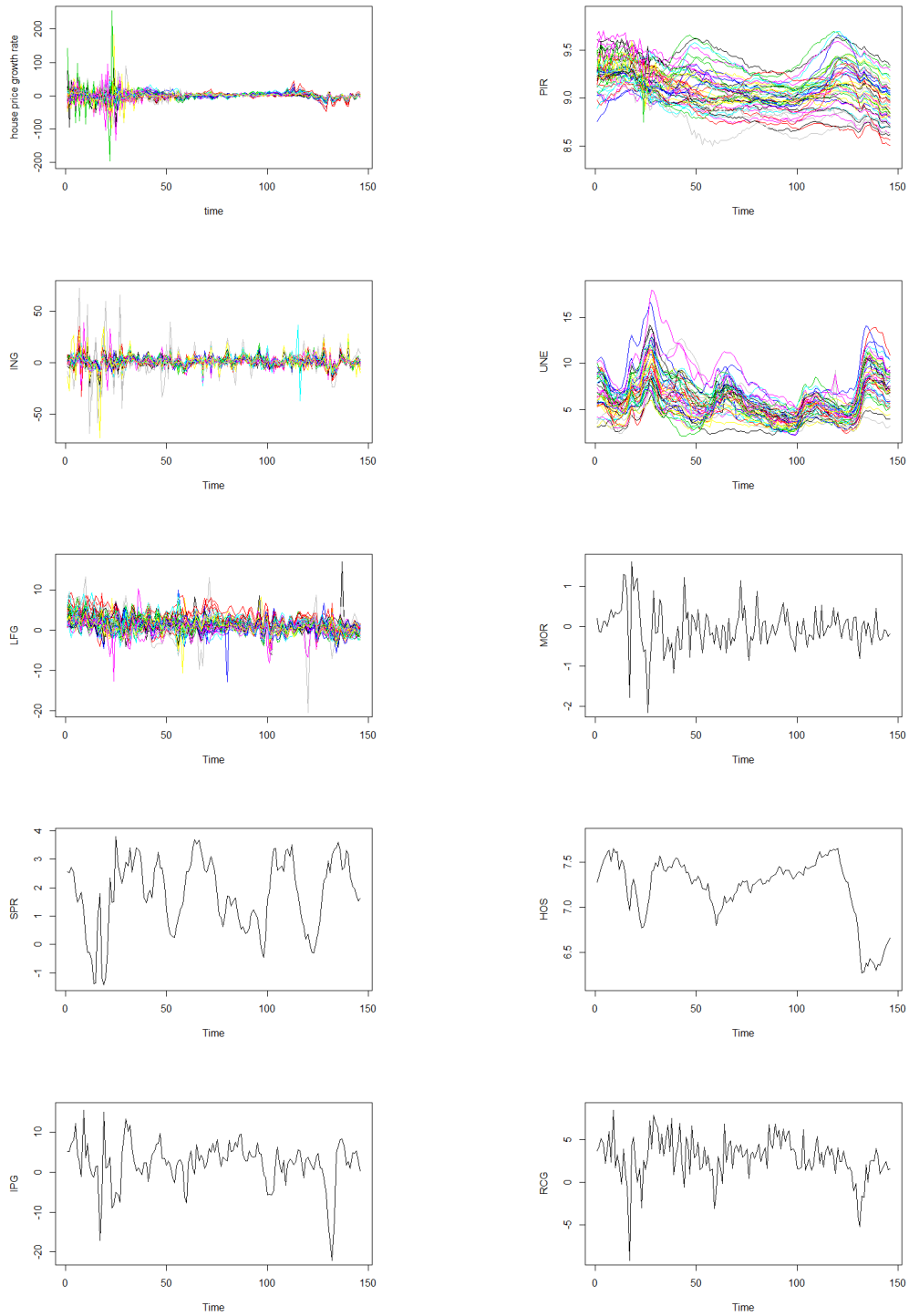


Figure 1: Time Series of real house price growth rate and all the covariates.

Table 1: Predictors and Abbreviations

Level	Variable	Abbreviation
state	price income ratio (in logs)	PIR
state	real per capita income growth (in logs and annualized)	ING
state	unemployment rate	UNE
state	labor force growth (in logs and annualized)	LFG
national	30-year mortgage rate (in first differences)	MOR
national	spread between 10-year and 3-month Treasury rates	SPR
national	housing starts (in logs)	HOS
national	industrial production growth (in logs and annualized)	IPG
national	real consumption growth (in logs and annualized)	RCG

Since Alaska and Hawaii do not share borders with the other states, they would generate vectors of zeros in the spatial neighbor matrix that can lead to computational error. Therefore, these two states were ruled out from our analysis. Modeling and forecasting in this paper are based on the remaining states.

3.2 Preliminary Analysis of the Data

3.2.1 Neighbor Structure

To qualitatively check the spatial associations between the states as well as prepare ourselves with CAR modeling, we create a neighbor structure among the 48 states. Out of the multiple neighbor defining ways, such as contiguity-based, graph-based, and k nearest neighbor, here we chose to use contiguity based method to create our neighbor matrix. We have explored several other neighbor structures while they did not exhibit significant improvement in our final model. Hence, here we skip the discussion on the selection of neighbor structure. Eventually, we came with Queen style contiguity neighbor structure that allows any two states sharing at least one point on the boundary to be neighbors [2]. Neighboring structure used in our model is shown in Figure 2. This neighbor objects described above is then transformed into a binary

weight matrix \mathbf{W} , where $\omega_{ij} = 1$ if j th area is neighbor of i th area; and $\omega_{ij} = 0$ otherwise. This is a conservative approach for creating spatial weight matrix, as little is known about the assumed spatial process.

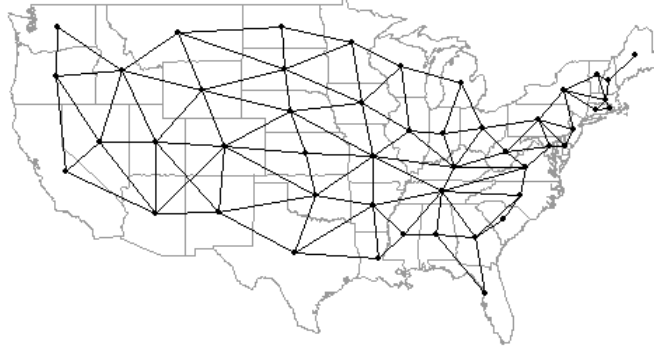


Figure 2: Neighbor structure used in CAR models.

3.2.2 Exploratory Data Analysis

We first tabulated the correlations among the BEA regions and plotted the time series plot for each region to gain the first insight into the spatial and temporal structure in our data. Previous studies have explored the correlations within and between BEA regions to check spatial interactions. The compositions for each BEA region are shown in Figure 3). BY studying the correlation matrix between and within the BEA regions for the duration of the year 1984 to 2011 [3] and the year of 1975 to 2003 [4] house price, it was found that the within region correlations are larger than the between region correlations with few exceptions. A similar feature is shown in our study using data 1975-2012 (see Table 2), that the diagonal correlation coefficients are generally larger than the off-diagonal numbers. Exceptions include Great Lakes - Far

West and Southwest - Far west.

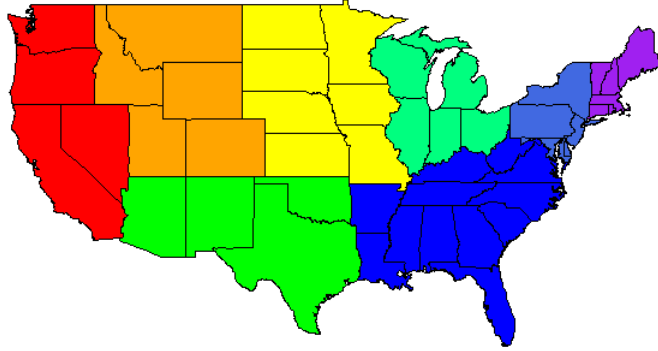


Figure 3: Eight BEA regions.

Table 2: Average of correlation coefficients within and between regions: New England(NE), Mideast(ME), Great Lakes(GL), Plains(PL), Southwest(SW), Rocky Mountain(RM),Far West(FW).

	NE	ME	GL	PL	SE	SW	RM	FW
NE	0.417	-	-	-	-	-	-	-
ME	0.398	0.400	-	-	-	-	-	-
GL	0.289	0.345	0.400	-	-	-	-	-
PL	0.164	0.212	0.323	0.429	-	-	-	-
SE	0.250	0.306	0.372	0.243	0.458	-	-	-
SW	0.147	0.211	0.315	0.261	0.329	0.375	-	-
RM	0.090	0.148	0.339	0.243	0.299	0.374	0.400	-
FW	0.272	0.372	0.449	0.262	0.367	0.383	0.352	0.375

Figure 4 shows that there are similar patterns across the country as well as regional similarities regarding the time series of real house price growth rate (RHPGR). Most states exhibited greater fluctuations in the first 70 quarters, which roughly correspond to the 1975 - 1993 time period. Since then, the housing market growth rate stabilized and possessed a slight upward trend

until near the 110th quarter (2003) when large uphill and downhill movements occurred. Therefore, forecasting over the last half of the time series by studying the first half is not an easy task, since the test set contains features that are fairly distinguishable from the training set.

Next, we quantify the spatial autocorrelation and temporal autocorrelation existing in the data by using Moran's test and time series autocorrelation function (ACF).

Moran's I test is one of the commonly used tests for spatial autocorrelation:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where the observations at areal units are denoted as y_i, y_j and spatial structure as $\omega_{i,j}$ [2]. Moran's I values for each time point was plotted and shown in Figure 5. More than 70% of the time there was significant spatial autocorrelation present. This finding further suggests the necessity of using CAR model to account for the spatial structure.

Autocorrelation function $\rho_{t,s}$ shows the correlation between time points t, s as:

$$\rho_{t,s} = Corr(Y_t, Y_s)$$

The ACF plot that has the correlation coefficient against lags was used to check spatial autocorrelation embedded in the data (Figure 6). The first five lags showed large variation with respect to the temporal autocorrelation coefficient, with a significant proportion appearing above the significance range bounded by ± 0.16 . The mean of the coefficients across states also went above the range for the first five lags, with the first and fifth lag right at the boundary. The ACF plot of the original data exhibits temporal autocorrelation within the data that suggests the necessity of using time series models.

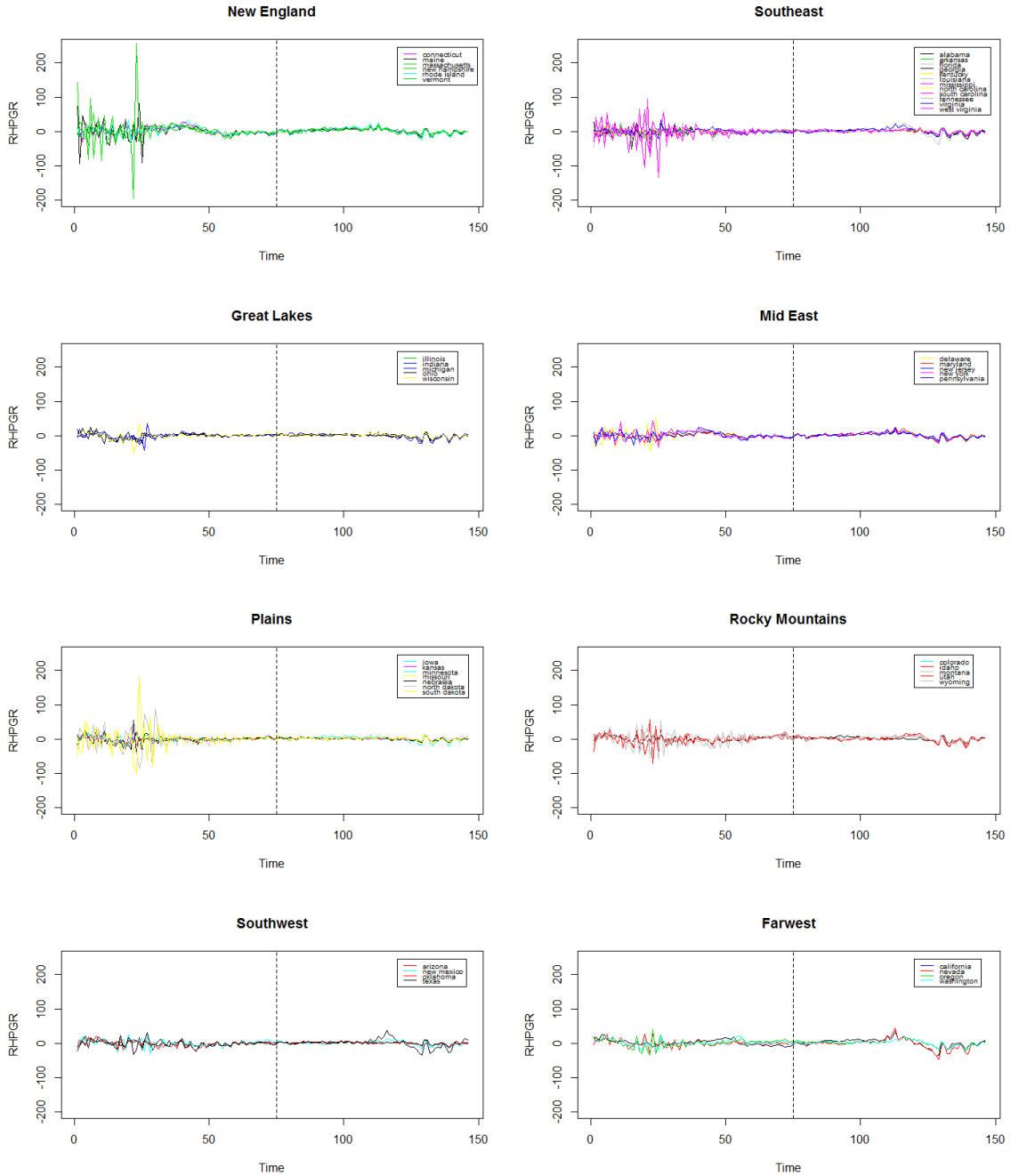


Figure 4: Time Series of real house price growth rate. 48 states are grouped in the eight BEA regions. Dashed line separates training data from test data. Note that the data was transformed by a factor of 400 (see equation 11).

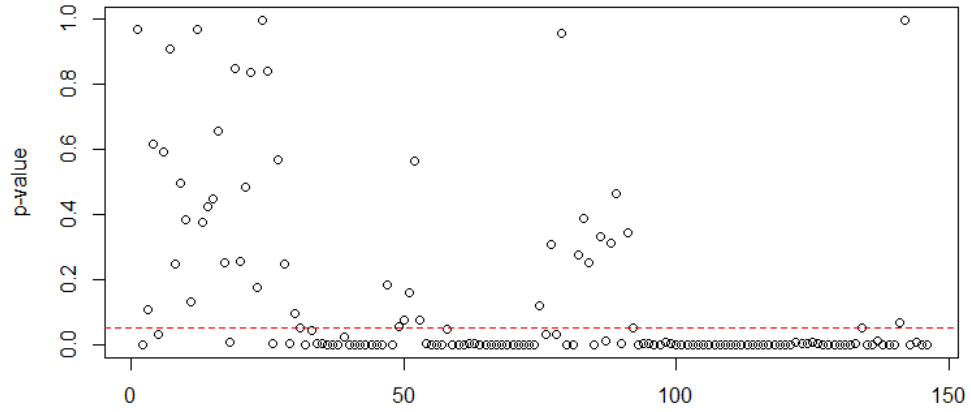


Figure 5: P-value from Moran's I test of observed house price growth rate over time. Red dashed line is the 0.05 significant level.

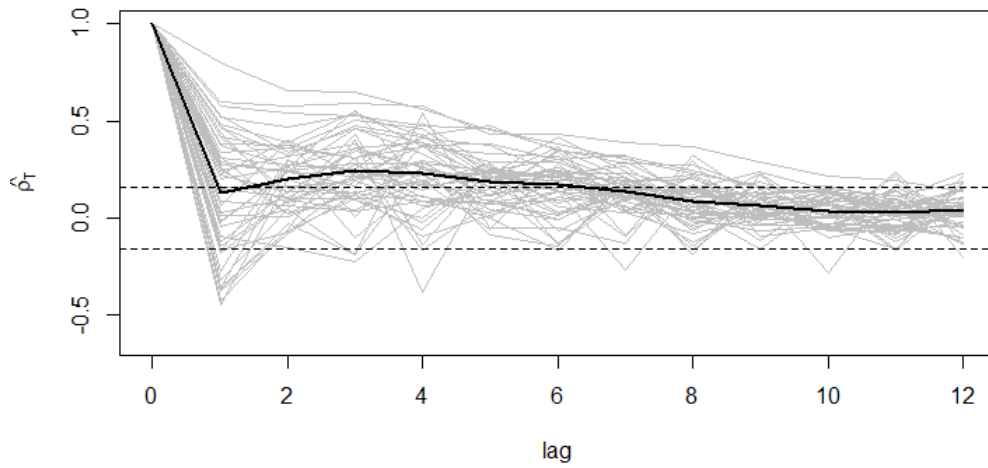


Figure 6: ACF plot of observed house price growth rate against time. Thick black line is the mean across the 48 states. Two dashed lines are ± 0.16 .

3.3 Results

3.3.1 Model Selection Results

We first examined the model selection results by checking the distribution of the selected model sizes. Figure 7 and Figure 8 reveal the distribution of the selected sample size for BMA-CAR without and with dilution prior respectively. The two dilution prior models we tested gave rather similar results thus only one of them is listed here. The two bar charts both reflected the centrality around model sizes of four and five and the reduction of frequency towards the two ends. The latter model (BMA-CAR without dilution prior) sampled models with a size of five slightly more than the latter (BMA-CAR with dilution prior).

We also investigated the variables selected by those most frequently visited models. Table 3 and Table 4 summarize the top five selected models without and with dilution prior $h(\gamma) = \gamma$. Results from the two tables are almost identical, with the most visited models having four to five predictors that evolve around price income ratio, unemployment rate, labor force growth, mortgage rate and housing starts. These five models account for $> 73\%$ of all the models been visited, and in fact, the top two models which are PIR + UNE + LFG + MOR + HOS and UNE + LFG + MOR + HOS alone account for $> 60\%$. The combination of variables of UNE + LFG + MOR + HOS has a significant weight in our model selection process.

[1] plotted out the median value along with the 16th and 84th percentiles of the model sizes foretasted over the testing period across the 50 states. The model sizes exhibited an increasing trend, with the median slowly climbed up from two to four during the span of the forecasting period. The 16th and 84th percentiles band were about one variable away from the median and the

width stayed stable during all time. Therefore, the variations of model sizes across states and time were high, which is not too surprising considering the boom and bust cycle occurred in our forecasting period. Their finding of the variation in model sizes across time and space explained why our model, CAR-BMA that uses global covariates across states and assumes the coefficients are constant over time, did an excellent jobs in prediction in some states and not so well for others. A detailed discussion is left to the next chapter.

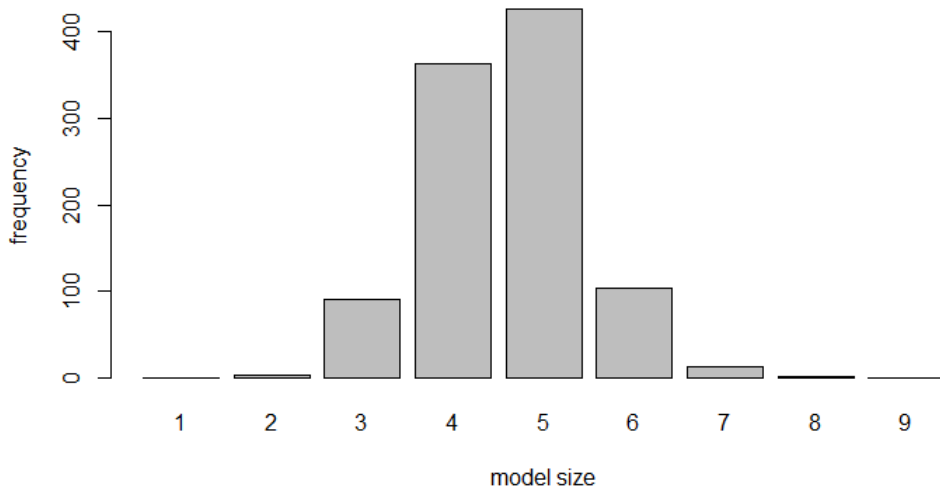


Figure 7: Posterior model size distribution after fitting BMA-CAR.

Table 3: Top 5 Most Visited Models

model	relative frequency
PIR + UNE + LFG + MOR + HOS	0.343
UNE + LFG + MOR + HOS	0.256
UNE + MOR + HOS	0.06
PIR + LFG + MOR + HOS	0.045
PIR + UNE + MOR + HOS	0.044

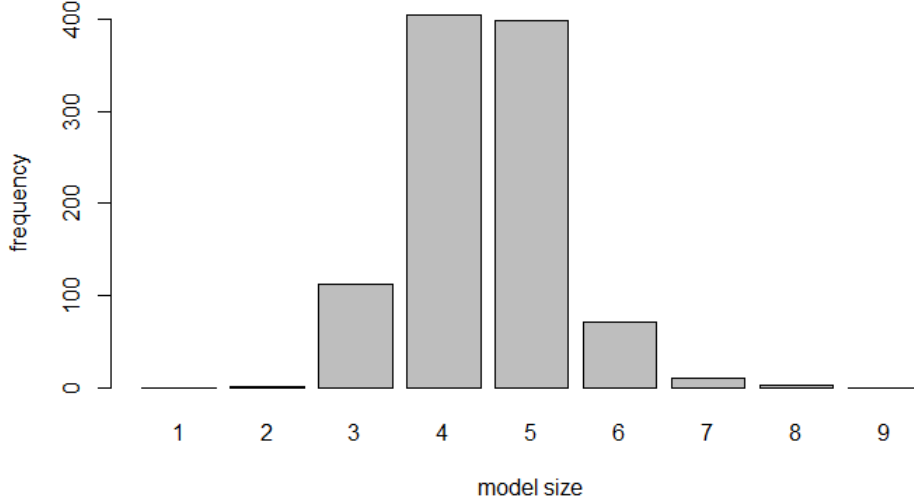


Figure 8: Posterior model size distribution after fitting BMA-CAR with dilution prior and power of 1.

Table 4: Top 5 Most Visited Models with dilution prior and power of 1.

model	relative frequency
PIR + UNE + LFG + MOR + HOS	0.341
UNE + LFG + MOR + HOS	0.286
NUE + MOR + HOS	0.074
PIR + UNE + MOR + HOS	0.063
UNE + LFG + MOR + HOS	0.042

3.3.2 Estimation Results

Moran's I tests were also conducted on the residuals after fitting of BMA-CAR models. The goal is to see if spatial autocorrelation is accounted for and if so, to what extent. BMA-CAR with dilution priors gave comparable residuals as BMA-CAR, so results from BMA-CAR with dilution prior are not included here. As revealed in Figure 9, most of the spatial autocorrelation were removed. In fact, only 11% of the spatial autocorrelation were significant

which is a drastic decline compared to 70% before model fitting. The ACF plot of the residuals is displayed in Figure 10. Except for the line (California) that showed significant coefficients for the first five lags, the other states all exhibited patterns of oscillations centered around the mean 0 after the first lag. Also, most of the coefficients were reduced with regard to the amplitude. A much higher proportion of coefficients were within the bounds when compared to Figure 6. Hence, the temporal autocorrelation originated from the data were largely removed after fitting the BMA-CAR model.

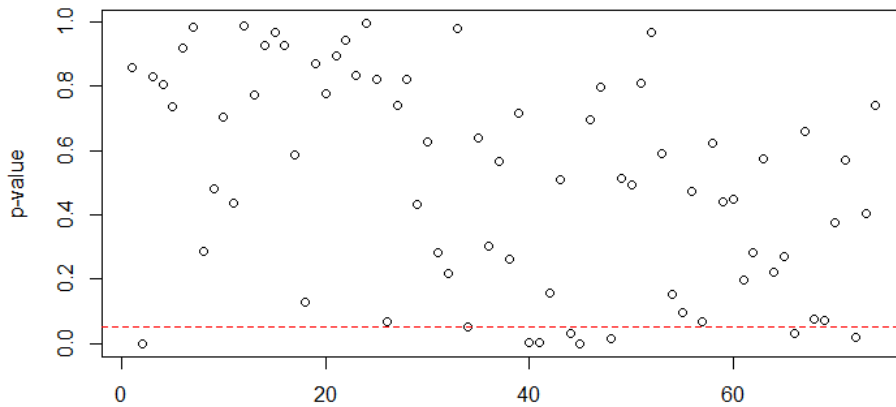


Figure 9: p-value from Moran's I test on residuals over time. Red dashed line is the 0.05 significant level.

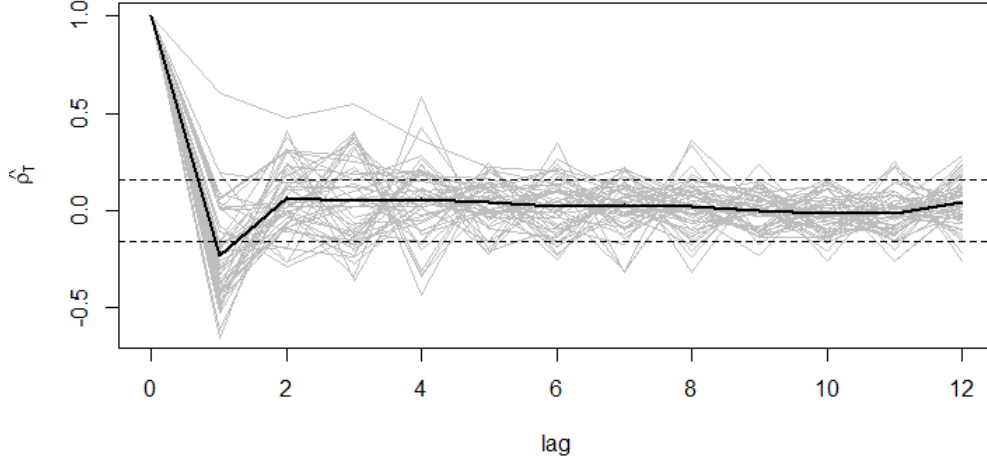


Figure 10: ACF plot of residuals against time. Thick black line is the mean across the 48 states. Two dashed lines are ± 0.16 .

3.3.3 Forecasting Results

We first would like to check the overall forecasting performance over time by using the squared forecasting error difference between the MEAN benchmark and our model. We adopted the method proposed by [5]:

$$CDSFE_{k,t} = \sum_{t=1995:1}^{2012:4} (SFE_{k,t,MEAN} - SFE_{k,t,BMA-CAR})$$

And made a little modification in that we summed CDSFE to each time point instead of using the whole time periods:

$$CDSFE_{k,t} = \sum_{t=1995:1}^t (SFE_{k,t,MEAN} - SFE_{k,t,BMA-CAR})$$

In Figure 11, we plotted out the sum of CDSFE for 48 states as a total:

$$\sum_{k=1}^{48} CDSFE_{k,t}$$

[1] plotted CDSFE against time for DMA, DMS, BMA, EW and AR1. EW is a model that used equal weighting from K OLS regression models; AR1 is

a lag 1 time series model. As pointed out in the paper [1], plotting CDSFE against time reveals the when our model is superior to the MEAN benchmark (positive slope) and those in which the MEAN benchmark predicts better (negative slope). Compared to DMA, DMS, BMA, EW and AR1, our model showed excellent performance for the latter bust period (after the 60th quarter or 2010:Q1) and that is comparable to DMA/DMS and much better than BMA. However, the boom period from the mid 90s to 2006 and the initial housing market meltdown time of 2007-2008, CDSFE from our model was merely better than CDSFE from EW and AR1.

Regardless of the scale of the forecast error, it is worth noting that similar patterns applied to almost all of the CDSFE lines where 2008:4, 2010:1 and 2011:3 experienced significant decline, especially to the best performance models DMA and DMS. [1] did not give an explanation. However, we attribute the observation to policy changes. These three time points are in accordance with the time lines of the Home Affordable Refinance Program (HARP) program. This program was initiated by the U.S. government to help those with mortgage problems refinance their home equities. There were three main time points that marked the initiation of this policy and the modifications, namely HARP 1.0, HARP 2.0 and HARP 3.0, that correspond to loan to the value (LTV) threshold of 105%, 125% and no restriction [6]. Due to the high accordance between the timelines of the HARP program and the performance drops, we believe that the deterioration of the model forecasting performance can be ascribed to policy change.

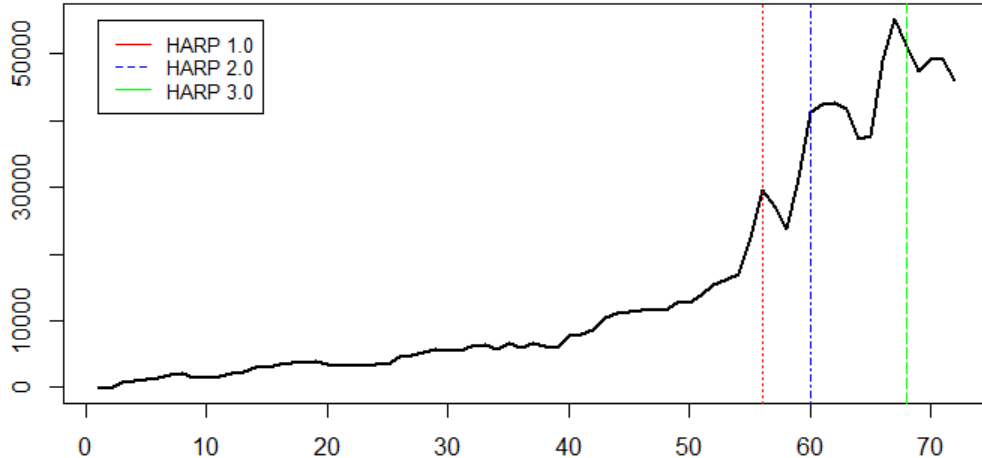


Figure 11: Cumulative squared forecast error difference.

After checking the overall picture for 50 states, we would like to compare the forecasting performance between states. We summarised the MSFE results across the states for four of our models (BMS-CAR, BMA-CAR, BMA-CAR($h(\gamma) = \gamma$), BMA-CAR($h(\gamma) = \gamma^{1/2}$)) and compared with four models (DMA, DMS, BMA, BMS) from [1]. The columns refer to the MSFE ratio (Equation 10) across the states, which are the average, standard deviation, minimum, maximum, and number of states in which BMA-CAR outperformed the benchmark MEAN model. Note that the results from [1] which are DMA, DMA, BMA, BMS in Table 5 were based on 50 states rather than 48. BMS-CAR is a model similar to BMA-CAR, except that only the model with the highest posterior probability (PIR + UNE + LFG + MOR + HOS) was used for modeling.

For the BMA-CAR model, an average MSFE ratio of 0.818 for BMA-CAR as shown in Table 5 is smaller than the reported MSFE ratio for BMA(0.858)

and BMS(0.903), indicating that incorporating spatial dynamics did improve the forecast accuracy compared to solely static Bayesian Model Selection and Bayesian Model Averaging [1]. Adding dilution prior did not improve the forecasting results by a significant amount. BMS-CAR showed poor performance compared to the others. Since BMA-CAR model gave the best forecasting performance and there is no significant difference adding dilution prior, discussion for the rest of this chapter uses only BMA-CAR as an example.

Table 5: Forecast errors across states

Model	Average	Std	Minimum	Maximum	< 1
BMS-CAR	0.838	0.36	0.510	1.60	36/48
BMA-CAR	0.818	0.24	0.505	1.58	40/48
BMA-CAR($h(\gamma) = \gamma$)	0.826	0.24	0.516	1.55	40/48
BMA-CAR($h(\gamma) = \gamma^{1/2}$)	0.833	0.25	0.514	1.57	40/48
DMA	0.751	0.12	0.546	0.961	50/50
DMS	0.796	0.15	0.501	1.165	46/50
BMA	0.858	0.14	0.548	1.177	44/50
BMS	0.903	0.15	0.591	1.188	35/50

To take a closer look at which states generated better and worse forecasts, we use Figure 12 to illustrate the state-wise distribution of the MSFE ratio between our model and the MEAN benchmark.

Most of the worst predicted states were center around the middle of the country while the coastal states tend to be predicted okay. We ranked the states by MSFE ratios and the top four and bottom four states were (Idaho, Georgia, Washington, Delaware) and (Indiana, Nebraska, Iowa, Kentucky) respectively. To discuss the reason behind this forecasting pattern, we would need to first introduce the concept of volatility that came from [7]. Volatility was defined as an indicator of the magnitude of the boom-bust cycle; larger volatility implies a larger cycle. As found by them, housing markets growth volatility and forecasting accuracy have almost a one-to-one relationship and coastal

states tend to have higher volatility than the interior states. The relationship between the volatility and the MSFE level implies better forecast in the interior states than the coastal states.

However, when it comes to MSFE ratio, coastal states tend to have smaller ratios that represent larger performance gain than the MEAN benchmark. As a matter of fact, MSFE ratio is only a relative number that it is not linked to the forecast performance directly. For those interior (stable) states, using the MEAN benchmark is not a bad choice since the historic average may not be too different from the current point. As a result, using very complicated models (such as CAR-BMA) will not cause great gain the performance. In fact, two of the four least volatility states (Iowa, Kentucky see Figure 15) are among the worst MSFE ratio states from our model BMA-CAR (Figure 13). We plotted Figure 15 and Figure 16 to explore the forecasting performance difference between the most and least volatile states. We also plotted out the best performing states and worst performing states from our BMA-CAR model (Figure 13 and 14) to discuss the model forecast performance variation during the boom-and-bust cycle.

In Figure 13, numbers from the forecast were plotted against the observed. In the bottom four figures that show the worst predicted states, BMA-CAR tends to overestimate the growth rate before the recession and underestimate the growth rate after the recession.

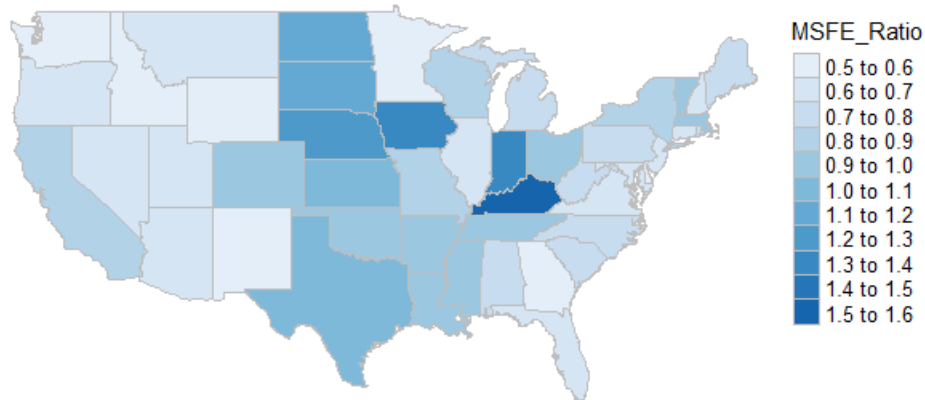


Figure 12: Forecasting results across the 48 states.

List of References

- [1] L. Bork and S. V. Møller, “Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection,” *International Journal of Forecasting*, vol. 31, no. 1, pp. 63–78, 2015.
- [2] R. S. Bivand, E. J. Pebesma, V. Gomez-Rubio, and E. J. Pebesma, “Applied spatial data analysis with r.” Springer, 2008, vol. 747248717, ch. 9, pp. 237–268.
- [3] P. Valentini, L. Ippoliti, and L. Fontanella, “Modeling us housing prices by spatial dynamic structural equation models,” *The Annals of Applied Statistics*, pp. 763–798, 2013.
- [4] S. Holly, M. H. Pesaran, and T. Yamagata, “A spatio-temporal model of house prices in the usa,” *Journal of Econometrics*, vol. 158, no. 1, pp. 160–173, 2010.
- [5] A. Goyal and I. Welch, “Predicting the equity premium with dividend ratios,” *Management Science*, vol. 49, no. 5, pp. 639–654, 2003.
- [6] F. H. F. Agency, “Fhfa, fannie mae and freddie mac announce harp changes to reach more borrowers,” <https://web.archive.org>.

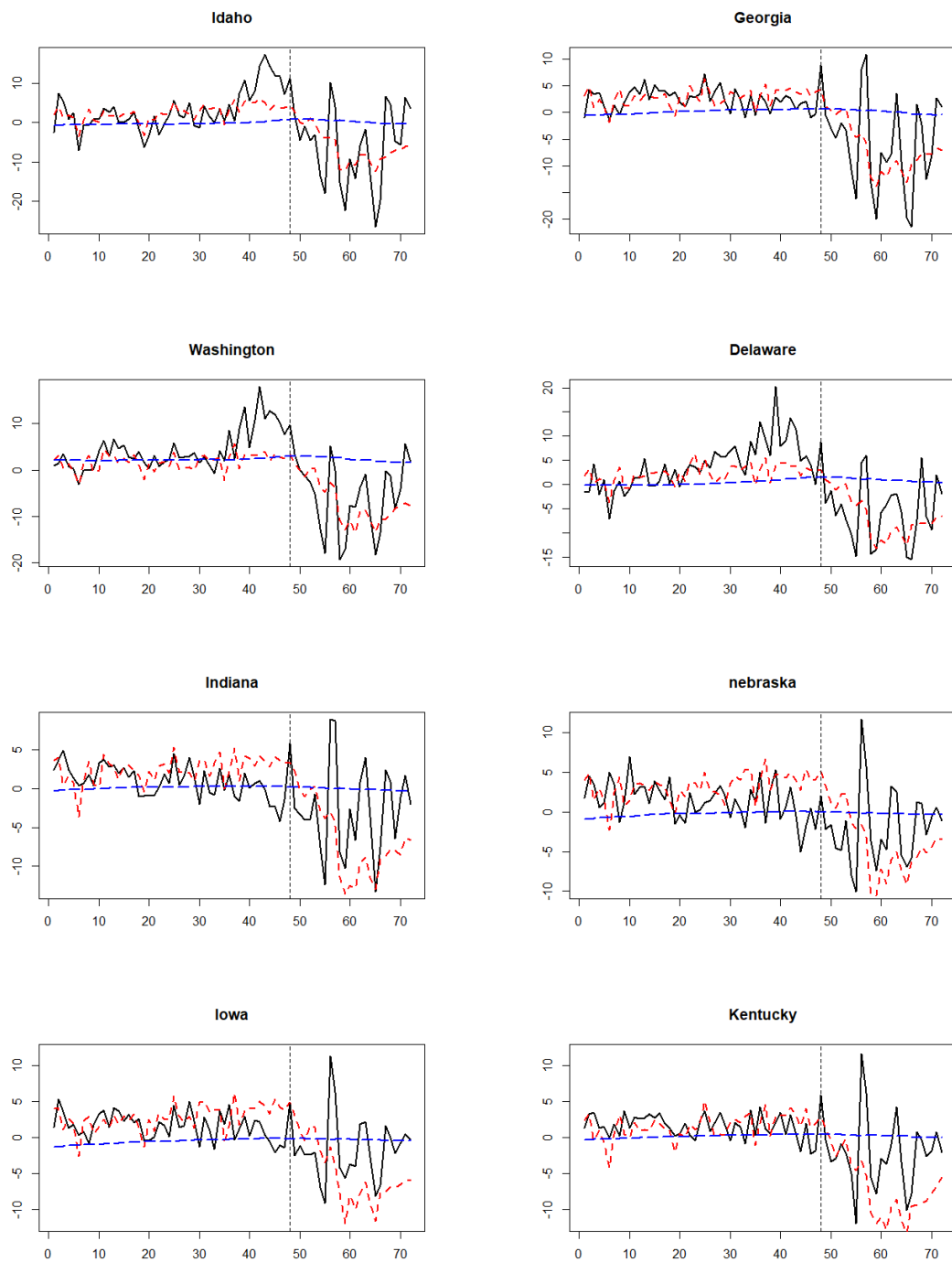


Figure 13: Forecast vs. observed: best predicted states (top 4) vs. worst (bottom 4). Red dashed line represents forecast from BMA-CAR; blue dashed line represents forecast from MEAN benchmark. Blue line is the historic data. Vertical line is the separation of the boom (1995:1-2006:4) and bust period (2007:1-2012:4).

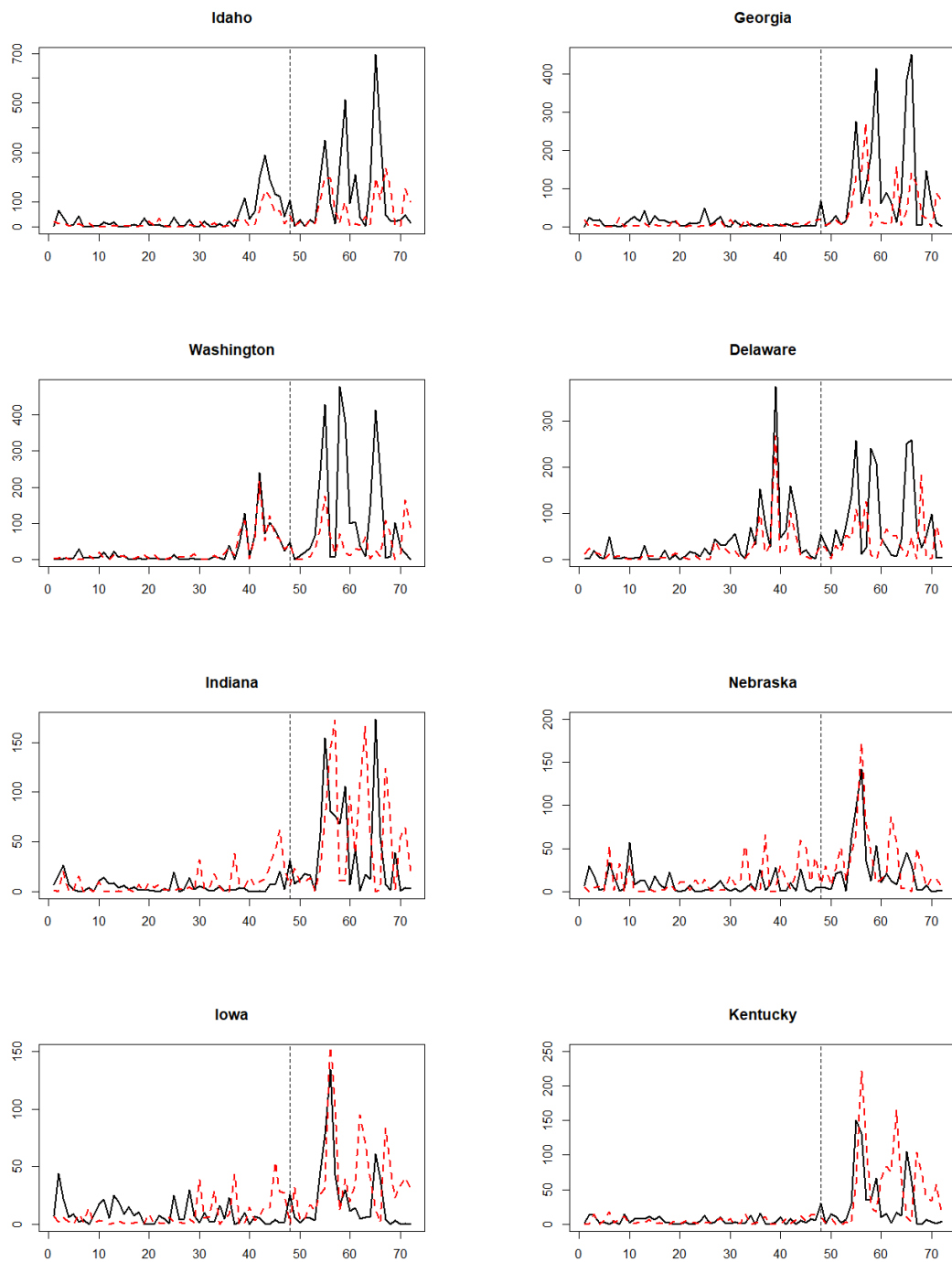


Figure 14: MSFE ratio: best predicted states (top 4) vs. worst (bottom 4). Red dashed line represents the MSFE for BMA-CAR; black line is MSFE for MEAN benchmark. Vertical line represents the separation of the boom and bust period. Notice the different scale on the y-axis.

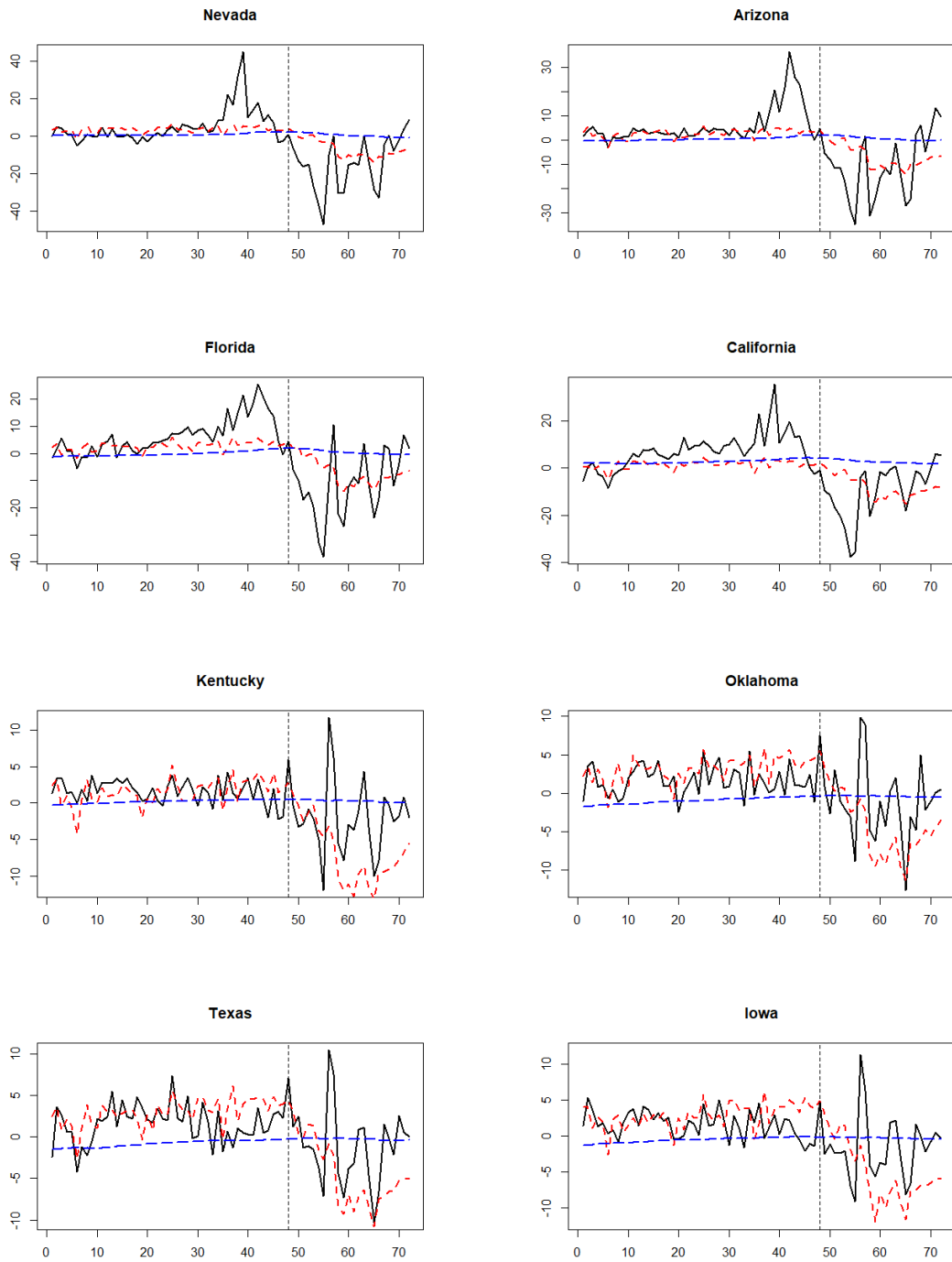


Figure 15: Forecast vs. observed: top 4 volatility states (top) vs. bottom 4 (bottom).

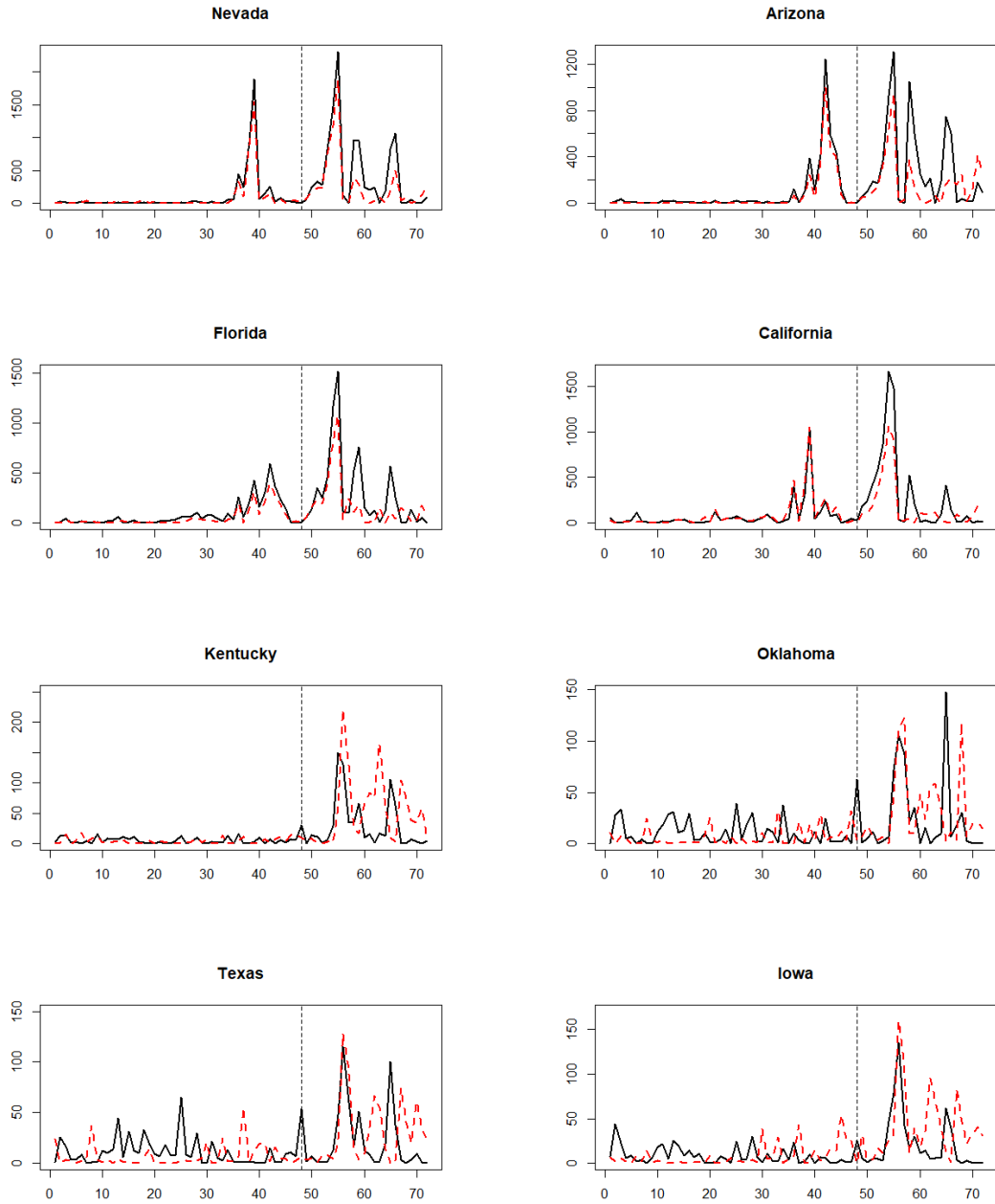


Figure 16: MSFE ratio: top 4 volatility states (top) vs. bottom 4 (bottom).

org/web/20120111193528/http://www.fhfa.gov/webfiles/22721/HARP_release_102411_Final.pdf, [Online; accessed 3-April-2018].

- [7] D. E. Rapach and J. K. Strauss, “Differences in housing price forecastability across us states,” *International Journal of Forecasting*, vol. 25, no. 2, pp. 351–372, 2009.

CHAPTER 4

Discussion and Future Work

4.1 Performance Gain by Spatial Components and Limitations

Here we present a discussion on why our model predicted better in the coastal states and not so well for the interior states. We compare our forecasting results with the forecasting results from [1]. Here we talk about the results in the context of MSFE ratio that represent the forecast gain from using just the MEAN benchmark. Even though the forecasting results from DMA [1] are not the true model (there is no true model anyway), DMA is a more flexible model that allows for both state-wise and time-wise model variations. Moreover, DMA gave smaller forecasting error than BMA-CAR (Table 5) thereby serving as a reasonable benchmark to check with.

4.1.1 Model Dimension and Variable Inclusion

We first start the discussion from the perspective of model dimension. Our posterior distribution for model space favored models of four and five variables. However, discoveries from [1] suggested that many of the low volatility states preferred parsimonious models, that typically two to four variables were needed in their model, and larger model sizes occurred more often around the end of the forecasting period when there was a financial crisis. Consequently, our model potentially selected too many variables for these low volatility states and it leads to overfitting and bad forecasting performance.

We further studied the influence from the choice of variables on forecasting performance. [1] pointed out that the housing market is segmented so that no

single variable drives the whole housing market. They plotted the inclusion probability (median, 16th and 84th percentiles) of each predictor against time and found that the recent economic depression causes model changes. Most of the variables were contained between 30% – 50% across time, with the exception of housing starts and price-income ratio that reached up to 100% right after 2007. In addition, there was a large spread of variable inclusion. For example, housing starts inclusion at the beginning of 2008 ranged from below 25% to near 100%. Thus, in the bust period, housing starts was only an important factor for some states (Arizona and Nevada). Meanwhile, price-income ratio became an essential component in Florida. This high degree of space and time variation in choice of variables near the financial crisis time makes it difficult to predict with globally assumed variables such as in our model. The important drivers, such as housing starts and price-income ratio were selected with high probabilities in our BMA-CAR posterior distributions, with each selected 58% and 99% of the time. High inclusions of housing starts and price-income ratio helped with the performance for certain states that did have these predictors as underlying factors such as Idaho and Georgia (refer to Figure 15 and 16). Nevertheless, containing these predictors suppressed the model performance at states like Iowa and Kentucky.

4.1.2 Limitations of a Global Fixed Effect

To conclude, the overall large model sizes chosen by our model and inclusion of predictors that are unimportant for those stable states give rise to less forecast gain in the interior states than in the coastal states. Our model had globally fixed effects and constant coefficients for each state and time. This restricted the ability to capture the spatial variations regarding model shift as

well as the temporal fluctuations seen in the housing market.

The original analysis from [1] developed house prices for each state individually, while neglecting the fact that house price of the neighboring states have autocorrelation features. By modeling spatial component as random effects, we were able to "borrow strength" across the states to come up with improved estimate for the house price in each [2]. As to the degree of which forecast performance improved, the lower and upper bounds are BMA/BMS and DMA/DMS respectively. That is to say, despite significant improvement compared to BMA/BMS methods due to additional spatial information, CAR-BMA model was not able to beat the model with time-variant features. One of the challenges in our study comes from the great difference seen in the house price growth rate data between the training period and the testing period. House price growth rate underwent significant fluctuations in the 1970s (see Figure reffig:TS 8 regions). The magnitudes of these variations are even larger than those observed near the recent economic recession. Several authors have noted a marked decline in the volatility of real activity and in the volatility and persistence of inflation since the early 1980s [3], [4], [5]. This structural change in the economy may have implications for the housing market.

In addition, assuming constant observation error σ^2 and revolution error τ^2 may not be feasible in our case. The states with high and low volatility showed distinguishable patterns of MSFE ratios, thus including the variations in the volatility in our model can help reduce this MSFE ratio difference and improve the overall forecasts.

4.1.3 An Empirical Model

Although BMA-CAR did not generate better forecasting results than DMA, it is worth noting that we were able to obtain comparable result to

what was achieved by DMA by running a space-time CAR model while just using unemployment, mortgage and consumption as predictors. As revealed in Figure 17, all states except for Nebraska had MSFE ratio smaller than one so that in 47/48 states the MEAN benchmark was beaten. Contrary to forecasting results from DMA or BMA-CAR, results from CAR show generally better predictions near the coast than inland, with the exception of South Dakota and Nebraska.

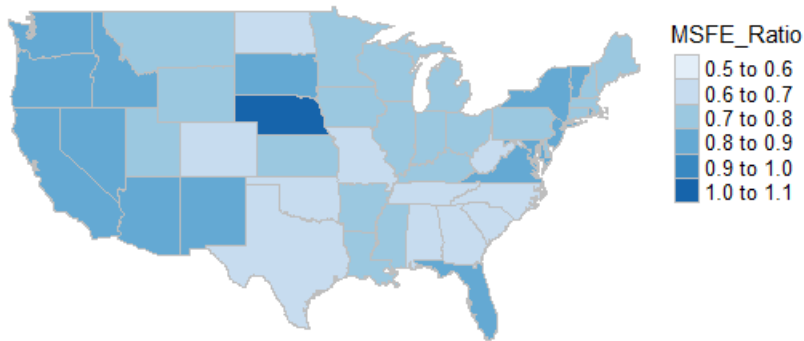


Figure 17: Forecasting results across the 48 states.

4.2 Future Work

4.2.1 Data Acquisition

Our model could be improved at the data acquisition step, in that the data aggregated at the state level may not be reflective on the housing market at smaller scale. By leveraging out the housing data at the state level, we would lose information about the variations in the housing data at smaller scales. For instance, it is not reasonable to assume similar housing market structure in the New York City and the rest of the New York State.

We can instead use data collected at the county level, which is more sim-

ilar to the scale of the housing prices study in real life. Data collected at the metropolitan statistical area (MSA) level is also available and can be a good candidate if the primary interest is to study the housing market in the population dense areas. MSA is a geographical region with a relatively high population density at its core and close economic ties throughout the area. Currently, the United States Office of Management and Budget (OMB) has defined 382 Metropolitan Statistical Areas (MSAs) for the United States [6]. One of the most populous MSA is New York-Newark-Jersey City that spans to three states (New York, New Jersey and Pennsylvania) [7]. Although this MSA covers three states, people tend to live and commute to work within this one area and it makes sense to study them as a whole. Real life applications include real estate investors that use MSA data to study housing trends and population movement [7].

4.2.2 Improvement of BMA-CAR Model

We can improve the forecasts by altering the neighbor structure. Our model currently used geometric definition of neighbors, that the states sharing common borders were called "neighbors". This may not work well for areas such as New England, where the states tend to be very small but highly similar. One alternative way of defining neighbor structure is to use k-nearest neighbors, in which we define each center states having k nearest neighbors based on the distance from their centroids. In this way, we are not limited to the states that are adjacent to each other.

Future research includes creating state-wise variables for each predictor and use Metropolis Hastings algorithm to select a model. Creating a variable for each state allows for variable and coefficient shift between states and further adds to the flexibility of BMA-CAR that assumes global fixed effects. And due

to the large model space(2^{9*48}) to sample from, Metroplis Hastings algorithm is more suitable in this case so that only a portion of the models will be visited.

List of References

- [1] L. Bork and S. V. Møller, “Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection,” *International Journal of Forecasting*, vol. 31, no. 1, pp. 63–78, 2015.
- [2] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical modeling and analysis for spatial data*. Crc Press, 2014.
- [3] O. Blanchard and J. Simon, “The long and large decline in us output volatility,” *Brookings papers on economic activity*, vol. 2001, no. 1, pp. 135–164, 2001.
- [4] M. M. McConnell and G. Perez-Quiros, “Output fluctuations in the united states: What has changed since the early 1980’s?” *American Economic Review*, vol. 90, no. 5, pp. 1464–1476, 2000.
- [5] J. H. Stock and M. W. Watson, “Has the business cycle changed? evidence and explanations,” *Monetary policy and uncertainty: adapting to a changing economy*, pp. 9–56, 2003.
- [6] U. S. O. of Management and Budget, “Omb bulletin no. 15-01: Revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas,” <https://obamawhitehouse.archives.gov/sites/default/files/omb/bulletins/2015/15-01.pdf>, 2015, [Online; accessed 3-April-2018].
- [7] Investopedia, “Definition of ’metropolitan statistical area - msa’,” <https://www.investopedia.com/terms/m/msa.asp>, [Online; accessed 3-April-2018].

BIBLIOGRAPHY

- Abraham, J. M. and Hendershott, P. H., “Bubbles in metropolitan housing markets,” National Bureau of Economic Research, Tech. Rep., 1994.
- Agency, F. H. F., “Fhfa, fannie mae and freddie mac announce harp changes to reach more borrowers,” https://web.archive.org/web/20120111193528/http://www.fhfa.gov/webfiles/22721/HARP_release_102411_Final.pdf, [Online; accessed 3-April-2018].
- Banerjee, S., Carlin, B. P., and Gelfand, A. E., *Hierarchical modeling and analysis for spatial data*. Crc Press, 2014.
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M., “Bayesian analysis of spacetime variation in disease risk,” *Statistics in medicine*, vol. 14, no. 21-22, pp. 2433–2443, 1995.
- Besag, J., “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236, 1974.
- Besag J, York J, M., “A bayesian image restoration, with two applications in spatial statistics,” *Ann Inst Statist Math*, vol. 43, no. 1, pp. 1–20, 1991.
- Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., and Pebesma, E. J., “Applied spatial data analysis with r.” Springer, 2008, vol. 747248717, ch. 9, pp. 237–268.
- Blanchard, O. and Simon, J., “The long and large decline in us output volatility,” *Brookings papers on economic activity*, vol. 2001, no. 1, pp. 135–164, 2001.
- Bork, L. and Møller, S. V., “Forecasting house prices in the 50 states using dynamic model averaging and dynamic model selection,” *International Journal of Forecasting*, vol. 31, no. 1, pp. 63–78, 2015.
- Brook, D., “On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems,” *Biometrika*, vol. 51, no. 3/4, pp. 481–483, 1964.
- Chipman, H., George, E. I., McCulloch, R. E., Clyde, M., Foster, D. P., and Stine, R. A., “The practical implementation of bayesian model selection,” *Lecture Notes-Monograph Series*, pp. 65–134, 2001.
- Clapp, J. and Giaccotto, C., “Evaluating house price forecasts,” *Journal of Real Estate Research*, vol. 24, no. 1, pp. 1–26, 2002.

- Crawford, G. W. and Fratantoni, M. C., “Assessing the forecasting performance of regime-switching, arima and garch models of house prices,” *Real Estate Economics*, vol. 31, no. 2, pp. 223–243, 2003.
- De Oliveira, V., “Bayesian analysis of conditional autoregressive models,” *Annals of the Institute of Statistical Mathematics*, vol. 64, no. 1, pp. 107–133, 2012.
- DiPasquale, D. and Wheaton, W. C., “Housing market dynamics and the future of housing prices,” *Journal of urban economics*, vol. 35, no. 1, pp. 1–27, 1994.
- George, E. I. *et al.*, “Dilution priors: Compensating for model space redundancy,” in *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown*. Institute of Mathematical Statistics, 2010, pp. 158–165.
- Goyal, A. and Welch, I., “Predicting the equity premium with dividend ratios,” *Management Science*, vol. 49, no. 5, pp. 639–654, 2003.
- Guha, S. and Ryan, L., “Spatio-temporal analysis of areal data and discovery of neighborhood relationships in conditionally autoregressive models,” 2006.
- Guirguis, H. S., Giannikos, C. I., and Anderson, R. I., “The us housing market: asset pricing forecasts using time varying coefficients,” *The Journal of real estate finance and economics*, vol. 30, no. 1, pp. 33–53, 2005.
- Hall, S., Psaradakis, Z., and Sola, M., “Switching error-correction models of house prices in the united kingdom,” *Economic Modelling*, vol. 14, no. 4, pp. 517–527, 1997.
- Hoff, P. D., *A first course in Bayesian statistical methods*. Springer Science & Business Media, 2009.
- Holly, S. and Jones, N., “House prices since the 1940s: cointegration, demography and asymmetries,” *Economic Modelling*, vol. 14, no. 4, pp. 549–565, 1997.
- Holly, S., Pesaran, M. H., and Yamagata, T., “A spatio-temporal model of house prices in the usa,” *Journal of Econometrics*, vol. 158, no. 1, pp. 160–173, 2010.
- Investopedia, “Definition of ‘metropolitan statistical area - msa’,” <https://www.investopedia.com/terms/m/msa.asp>, [Online; accessed 3-April-2018].

- Kuethe, T. H. and Pede, V. O., “Regional housing price cycles: a spatio-temporal analysis using us state-level data,” *Regional studies*, vol. 45, no. 5, pp. 563–574, 2011.
- Lee, D., “Carbayes version 4.6: An r package for spatial areal unit modelling with conditional autoregressive priors,” *University of Glasgow, Glasgow*, 2017.
- Lee, D., Rushworth, A., and Napier, G., “Spatio-temporal areal unit modelling in r with conditional autoregressive priors using the carbayesst package.”
- Leroux BG, Lei X, B. N., “Estimation of disease rates in small areas: a new mixed model for spatial dependence,” in *Statistical models in epidemiology, the environment, and clinical trials*. Springer, 2000, p. 179191.
- LeSage, J. P., “Bayesian estimation of spatial autoregressive models,” *International Regional Science Review*, vol. 20, no. 1-2, pp. 113–129, 1997.
- Malpezzi, S., “A simple error correction model of house prices,” *Journal of housing economics*, vol. 8, no. 1, pp. 27–62, 1999.
- McConnell, M. M. and Perez-Quiros, G., “Output fluctuations in the united states: What has changed since the early 1980’s?” *American Economic Review*, vol. 90, no. 5, pp. 1464–1476, 2000.
- of Management, U. S. O. and Budget, “Omb bulletin no. 15-01: Revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas,” <https://obamawhitehouse.archives.gov/sites/default/files/omb/bulletins/2015/15-01.pdf>, 2015, [Online; accessed 3-April-2018].
- Pain, N. and Westaway, P., “Modelling structural change in the uk housing market: a comparison of alternative house price models,” *Economic Modelling*, vol. 14, no. 4, pp. 587–610, 1997.
- Quigley, J. M., “A simple hybrid model for estimating real estate price indexes,” *Journal of Housing Economics*, vol. 4, no. 1, pp. 1–12, 1995.
- Rapach, D. E. and Strauss, J. K., “Differences in housing price forecastability across us states,” *International Journal of Forecasting*, vol. 25, no. 2, pp. 351–372, 2009.
- Rodriguez, A. and Puggioni, G., “Mixed frequency models: Bayesian approaches to estimation and prediction,” *International Journal of Forecasting*, vol. 26, no. 2, pp. 293–311, 2010.

- Rushworth, A., Lee, D., and Mitchell, R., “A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater london,” *Spatial and spatio-temporal epidemiology*, vol. 10, pp. 29–38, 2014.
- Rushworth, A., Lee, D., and Sarran, C., “An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 1, pp. 141–157, 2017.
- Stock, J. H. and Watson, M. W., “Has the business cycle changed? evidence and explanations,” *Monetary policy and uncertainty: adapting to a changing economy*, pp. 9–56, 2003.
- Valentini, P., Ippoliti, L., and Fontanella, L., “Modeling us housing prices by spatial dynamic structural equation models,” *The Annals of Applied Statistics*, pp. 763–798, 2013.
- Van Dijk, B., Franses, P. H., Paap, R., and Van Dijk, D., “Modelling regional house prices,” *Applied Economics*, vol. 43, no. 17, pp. 2097–2110, 2011.
- Zellner, A., “On assessing prior distributions and bayesian regression analysis with g-prior distributions,” *Bayesian inference and decision techniques*, 1986.