Essays on Hierarchical Bayesian Estimation of Spatio-Temporal Economic Models

by

Oleksiy Tokovenko

(Under the direction of Lewell F. Gunter)

Abstract

This dissertation contributes to the economic literature dedicated to the analysis of the complex spatio-temporal models in the Bayesian hierarchical setting. The present research consists of two empirical applications in the areas of international trade and agricultural yield prediction.

The first empirical study focuses on the problem of estimating the effect of social barriers to trade. I develop a Bayesian social relation gravity model that properly accounts for a two-stage decision process and the interactive nature of bilateral trade. The model uses the distance based threshold specification and assumes invariance to the labeling of equations. The estimated effects of the fundamental variables are found to be mostly consistent with the results reported traditionally in the gravity model research literature. Formal tests based on the Savage-Dickey density ratio show the significance of effects of governance levels on international trade and suggest asymmetry of the social costs of trade.

The second empirical study presents a spatio-temporal statistical model of agricultural yield prediction based on spatial mixtures of distributions. The proposed method combines several hierarchical and sequential Bayesian estimation procedures that allow the general problem to be addressed with a series of simpler tasks, providing the required flexibility. The spatial correlation hypothesis is studied by comparing the alternative models using

the posterior predictive criterion under a squared loss function. Findings indicate that the proposed approach has a better ability to predict agricultural yields and is able to correctly suggest the possibility of a significant decrease in yields level. The premium rates for crop insurance suggested by the normal county level model are found to be significantly lower than the premium rates calculated based on the spatial mixtures model, especially for the years and locations where unfavorable events actually occurred.

INDEX WORDS: Hierarchical model, Bayesian estimation, Gravity model, Spatio-temporal model, Mixtures of distributions, Yield prediction

Essays on Hierarchical Bayesian Estimation of Spatio-Temporal Economic Models

by

Oleksiy Tokovenko

B.S., Bila Tserkva State Agrarian University, 2001

M.S., The University of Georgia, 2004

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2010

Essays on Hierarchical Bayesian Estimation of Spatio-Temporal Economic Models

by

Oleksiy Tokovenko

Approved:

Major Professor: Lewell F. Gunter

Committee: Jeffrey H. Dorfman
William D. Lastrapes

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
June 2010

Dedication

Dedicated to my family

Acknowledgments

First of all I would like to express my deepest gratitude to my major professor, Dr. Lewell F. Gunter, for his constant support and understanding throughout my graduate studies. I am grateful to Dr. Jeffrey H. Dorfman for his professional advice in both academic matters and the career choice and for his wise guidance through the world of Bayesian econometrics. I would like to thank Dr. William D. Lastrapes for sharing his professional experience with me as the teacher and advisor.

I am especially thankful for all the faculty and staff of the Department of Agricultural and Applied Economics and the University of Georgia. I would also like to thank the Department of Agricultural and Applied Economics at the University of Georgia for the financial support throughout the whole program of study.

I am indebted to my colleagues and friends for being there for me both when I am happy and when I am desperate.

Finally I would like to give my the most sincere gratitude to my parents, Vasyl and Tetyana Tokovenko for their undying love and support and for their willingness to sacrifice everything for my happiness. Thank you for having not a slightest doubt in my ability to excel. None of this would be possible without you being by my side.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Hierarchical regression models often become a useful tool for studying complex economic problems when there are different sets of information available on different levels of measurement. The general structure of the Bayesian hierarchical model includes a sequence of conditional distributions that comprise the levels of the model hierarchy. The model parameters at each level are assumed to be correlated and to be a sample from a common prior distribution, which in turn is described by the previous stage model parameters, called hyperparameters, that are assigned their own prior distribution, known as a hyperprior distribution. In theory there could be as many stages and correspondingly as many sets of hyperparameters and hyperdistributions as required. However, it is suggested that there is no reason to include more than two levels of hierarchy of a given structure as a practical consideration since the quality of inference about the model parameters decreases at each additional level while the interpretation of the results become increasingly difficult (Gill 2002). As an example, the posterior distribution for a simple two-stage model can be described as

$$p(\theta, \alpha, \beta|y) \propto \prod_{i=1}^{K} \Big[p(y|\theta_i)p(\theta_i|\alpha, \beta)\Big]p(\alpha)p(\beta) \tag{1.1}$$

where $p(y|\theta_i)$ is the model likelihood, $p(\theta_i|\alpha, \beta)$ is the prior distribution on θ_i, while $p(\alpha)$ and $p(\beta)$ are the hyperpriors. Correspondingly, θ and α and β are the sets of model parameters and hyperparameters. The hierarchical model with more stages will involve multiple nests of parameter sets, however the general result will be similar to (1.1) in its structure. Note that, as pointed out in Ntzoufras (2009), the term hierarchical models refers less to a specific type of Bayesian model but rather to "a general set of modeling principles". Robert (2007, p.

461) concludes "...that hierarchical modelings are indeed included in the Bayesian paradigm, and therefore that this approach enjoys the general optimality properties of the Bayesian approach...". Strictly speaking, a classical Bayesian model that combines the data likelihood $p(y|\theta)$ and the prior information about the model parameters in a form of a prior distribution $p(\theta)$ can be seen directly as a hierarchical model with only one level of hierarchy.

The special structure of hierarchical models may increase efficiency in the analysis of datasets with complex relations between the observed units. Such datasets are often naturally multilevel and applying the regular non-hierarchical regression methods may be inappropriate in these cases. As discussed in Gelman *et al.* (2004), the more parsimonious models cannot provide an adequate fit, while the models with too many parameters can fit the data arbitrarily well, producing, however, inferior predictions of new observations. The hierarchical approach becomes an alternative to both extremes by incorporating an additional correlation structure to the parameter space to provide the necessary flexibility and explanatory power to the regression models while avoiding the problem of overfitting the data.

The important benefit of using the hierarchical structure in the regression analysis beyond the quality of model fit consideration is the possibility to specify the uncertainty about both the model parameters and the restrictions on them through the hyperparameters of the lower level of the model hierarchy. Justifying the use of multilevel Bayesian modeling Robert (2007, p. 464) argues that in this setting the hierarchical approach offers "... a compromise between the Jeffreys noninformative distributions, which are diffuse but sometimes difficult to use or justify, and the conjugate distributions, which have limited subjective justification but are analytically tractable". By transferring the parameter uncertainty to the hyperparameter level the hierarchical approach usually improves the robustness of the estimation results to the choice of prior distributions as it essentially implies the Bayesian averaging of the results obtained from the conjugate analysis of the set of underlying regression models defined by the hyperparameter space (see Berger (1985) and Robert (2007) for more details on hierarchical priors and Bayesian robustness).

Despite their generally complex structure, the multilevel composition of the Bayesian hierarchical models makes them particularly convenient to estimate using iterative MCMC methods. Considering an example of a simple hierarchical model in (1.1), the Gibbs sampler algorithm (see, e.g. Koop, Poirier and Tobias 2007) requires drawing the posterior values of θ_i, α and β from three sets of posterior conditional distributions:

$$\theta_i \sim p(\theta_i|y, \alpha, \beta) \tag{1.2}$$

$$\alpha \sim p(\alpha|\{\theta\}_i^K, \beta) \tag{1.3}$$

$$\beta \sim p(\beta|\{\theta\}_i^K, \alpha) \tag{1.4}$$

where (1.2) uses the assumption of conditional independence of θ_i, while (1.3) and (1.4) reflect the fact that $\{\theta\}_i^K$ is sufficient for α and β (Rossi, Allenby and McCulloch 2005). More discussions on inference in Bayesian hierarchical models and their various applications can be found in Gill (2002), Koop, Poirier and Tobias (2007), Ntzoufras (2009) and Rossi, Allenby and McCulloch (2005).

This dissertation presents two applications of the Bayesian hierarchical approach to regression analysis. Chapter 2 is dedicated to the roles of social interactions and barriers in international agricultural trade. To facilitate an empirical study we develop a Bayesian social relations gravity model that represents a most common class of hierarchical models based on the inclusion of random effects. Modeling the individual and pair specific effects allows one to describe a complex correlation structure in the observed repeated responses resulting from social interactions between countries that are an inevitable part of international trade. Chapter 3 proposes a new approach to the agricultural yield prediction using a spatio-temporal model. This approach allows for nonnormality and the spatial correlation of yields. Bayesian mixture modeling involves the inclusion of augmentation variables which is a property shared by another class of popular hierarchical models referred to as a latent variable models. It is worth noticing that the social relation gravity model of international trade introduced in Chapter 2 also employs data augmentation as a part of the estimation method. Besides this fact, both proposed model applications are essentially spatio-temporal, where

the gravity equations share this property through location specific effects that define the origin-destination class of spatial econometric interaction models (LeSage and Pace 2009).

The arguments considered in this chapter suggest that the Bayesian hierarchical approach is indeed a very general method of modeling complex data generating processes that can be applied to the analysis of a wide variety of economic problems in an efficient manner. However, despite its theoretically and empirically appealing features, it does have its drawbacks and limitations, as discussed in Chapter 4 with respect to the theoretical and empirical results of studies presented in this dissertation.

Bibliography

[1] Berger, J.O. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2nd Edition. Springer-Verlag, New York, NY.

[2] Gelman, A., J.B. Carlin, H.S. Stern and D.B. Rubin. 2004. Bayesian Data analysis, 2nd edition. Chapman & Hall/CRC, New York, NY.

[3] Gill, J. 2002. *Bayesian Methods: A Social and Behavioral Approach*, 1st edition. Chapman & Hall/CRC, New York, NY.

[4] Koop, G., D.J. Poirier and J.L. Tobias. 2007. *Bayesian Econometric Methods*. Cambridge University Press, New York, NY.

[5] LeSage, J.P. and R.K. Pace. 2009. *Introductoin to Spatial Econometrics*. Chapman & Hall/CRC, New Yokr, N.Y.

[6] Ntzoufras, I. 2009. *Bayesian Modeling Using WinBUGS.* John Wiley and Sons, Inc., Hoboken, NJ.

[7] Robert, C.P. 2007. *The Bayesian Choice*. Springer-Verlag, New York, NY.

[8] Rossi, P.E., G.M. Allenby and R. McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley and Sons, Inc., Hoboken, NJ.

Chapter 2

A Bayesian Social Relations Gravity Model of International Trade

2.1 Introduction

The gravity model has been successfully used in empirical work to analyze the effects of various international economic events by modeling trade flows between countries and predicting potential outcomes of policy changes. Recent applications include e.g. the analysis of trade agreements (Rose 2004, Baier and Bergstrand 2007, Grant and Lambert 2008), border effects (Anderson and van Wincoop 2003) and custom union effects (Bun and Klaassen 2007). The classical gravity equation relates the volume of bilateral trade to the size of partner countries and the distance between them. Anderson and van Wincoop (2003) and Baier and Bergstrand (2007) employ a theoretically consistent gravity model that tracks back to Anderson (1979). However, the empirical form of the model is also widely employed due to the simplicity of estimation and intuitive appeal.

A significant focus of both empirical and theoretical research is placed upon investigating the sensitivity of analytical results to the gravity equation specification, including the infamous "zero gravity" problem associated with the basic log-linear formulation of the model (see e.g. Kandogan 2007). The latter issue proved to be a notorious source of inconsistencies and bias in economic inference produced by the gravity model, and generally requires applying the solution proposed by Eaton and Tamura (1994) in a form of threshold Tobit model or by Santos Silva and Tenreyro (2006) who suggest to use the Pseudo-Poisson Liekli-hood estimation. Besides the technical difficulties, the increased complexity of gravity models in recent years is also associated with recognition of the importance of non-economic factors for empirical international trade research. These factors can create additional trade barriers

that are different from the transportation or other directly measured costs of trade and generate burdens that can be associated with overcoming multilateral resistance between trading partners. Among other works, Ranjan and Tobias (2007) discuss country-specific measures of contract enforcement as a potential type of contracting costs, while Ward and Hoff (2007) focus on the effect of political and institutional variables on international commerce.

The objective of this study is to extend the empirical gravity model by taking into account the social structure of the international trade process and introducing a measure of social costs of trade that is assumed to be asymmetric. The first part of the problem requires developing a statistical model that represents exchange volumes as the result of pairwise interactions between trading partners, as discussed in the methodology section. Further, we propose to include the estimated social distance between countries in the model to measure associated costs of contracting and negotiations.

2.2 Model

We employ the Bayesian social relations (BSR) model of Gill and Swartz (2001, 2007) that was originally designed for statistical analysis of round-robin interaction data obtained from social science experiments. Bilateral trade volumes can be naturally presented as such type of data by elaborating on an actual round-robin design of the gravity trade model (e.g., see Ward and Hoff 2007). Given this observation BSR model can be generalized in a discrete-continuous data round-robin design that expresses the bilateral trade flows in the form of the paired responses as the result of interaction of partner countries in a trade process. The proposed generalization employs the threshold Tobit methodology developed by Ranjan and Tobias (2007) who in turn provide the Bayesian treatment to the original Eaton and Tamura (1994) framework to address a zero trade problem. The complete model used in our study

can be described by the following set of equations:

$$\ln(y^*_{ijt} + \tau d_{ij}) = \alpha_t + X_{ijt}\beta_1 + a_i + X_{it}\beta_2 + b_j + X_{jt}\beta_3 + c_{ij} + \epsilon_{ijt} \tag{2.1}$$

$$\ln(y^*_{jit} + \tau d_{ji}) = \alpha_t + X_{jit}\beta_1 + a_j + X_{jt}\beta_2 + b_i + X_{it}\beta_3 + c_{ji} + \epsilon_{jit} \tag{2.2}$$

where

$$y_{ijt} = \begin{cases} y^*_{ijt}, & \text{if} \quad y^*_{ijt} > 0; \\ 0, & \text{if} \quad -\tau d_{ij} < y^*_{ijt} \leq 0 \end{cases} \quad \text{and} \quad y_{jit} = \begin{cases} y^*_{jit}, & \text{if} \quad y^*_{jit} > 0; \\ 0, & \text{if} \quad -\tau d_{ji} < y^*_{jit} \leq 0 \end{cases} \tag{2.3}$$

In the model above y_{ijt} and y_{jit} represent actual trade flows between countries i and j at time period t with $\tau d_{ij} > 0$ denoting the trading pair specific threshold value, where d_{ij} stands for the geographic distance between country i and j. Ranjan and Tobias (2007) suggests that the variable $W^*_{ijt} = y^*_{ijt} + \tau d_{ijt}$ can be interpreted as the desired level of trade assumed to have at least a small positive value, implying that countries are willing to trade in general. The model (2.1) – (2.3) can be rewritten as follows

$$\ln(W^*_{ijt}) = \alpha_t + X_{ijt}\beta_1 + a_i + X_{it}\beta_2 + b_j + X_{jt}\beta_3 + c_{ij} + \epsilon_{ijt} \tag{2.4}$$

$$\ln(W^*_{jit}) = \alpha_t + X_{jit}\beta_1 + a_j + X_{jt}\beta_2 + b_i + X_{it}\beta_3 + c_{ji} + \epsilon_{jit} \tag{2.5}$$

where

$$y_{ijt} = \begin{cases} W^*_{ijt} - \tau d_{ijt}, & \text{if} \quad W^*_{ijt} > \tau d_{ijt}; \\ 0, & \text{if} \quad 0 < W^*_{ijt} \leq \tau d_{ijt} \end{cases} \tag{2.6}$$

$$y_{jit} = \begin{cases} W^*_{jit} - \tau d_{ijt}, & \text{if} \quad W^*_{jit} > \tau d_{ijt}; \\ 0, & \text{if} \quad 0 < W^*_{jit} \leq \tau d_{ijt} \end{cases} \tag{2.7}$$

This reparametrization suggests that the actual trade occurs and is equal to $y^*_{ijt} = W^*_{ijt} - \tau d_{ijt}$ and $y^*_{jit} = W^*_{jit} - \tau d_{jit}$, correspondingly, when the desired amount of trade is greater than the expected losses from the distance related costs. On the contrary, if the desired amount of trade is not enough to cover such losses the countries choose not to trade.

The righthand side of the model equations (2.1) and (2.2) includes matrices X_{it} and X_{jt} of country i and j specific factors. We construct each X_{it} and X_{jt} matrix to have the same set of

characteristics, described by the following variables: $\ln(GDP_{it})$, $\ln(POP_{it})$, $\ln(AGR/POP_{it})$, $\ln(SEA/POP_{it})$ and UNL_{it}, but allow the effect of these characteristics to be asymmetric and to depend on whether the corresponding country stands as the exporter or importer in relation to a given trade flow, with the effect of interest being captured by the regression coefficient β_1 or β_2, respectively. The variables $\ln(GDP_{it})$ and $\ln(POP_{it})$ denote the logs of the country i gross domestic product and its population size at a given time period t that conventionally serve as measures of the country size. Since the primary interest of this study is international trade with the products of agriculture and fishery we choose to include the industry specific variables $\ln(AGR/POP_{it})$ and $\ln(SEA/POP_{it})$ that stand for the logs of the area of agricultural land and the length of exclusive fishery zone calculated per capita to approximate a relative factor endowment for the country i at a given time period t. Finally, UNL_{it} is a binary variable that takes a value of 1 if the official language of country i is also one of the UN official languages and 0 otherwise. The symmetric effect of interaction is captured by the matrix X_{ijt} containing the trading partners' common characteristics, such as the log of geographic distance between the country i and j, sharing a common border and official language and being members of a trade agreement, measured by variables $\ln(DIST_{ij})$, $CONT_{ij}$, $LANG_{ij}$ and TA_{ij}, respectively. Note that for the our case $X_{ijt} = X_{jit}$, with the first three characteristics being fixed over time.

For further analysis, following Ranjan and Tobias (2007) we choose to use the trading countries' aggregate governance indicators (GI) as a measure of contract enforcement and the social costs of trade in the following three nested models:

M0: No GI's are included in the model

M1: Quantity $|GI_{it} - GI_{jt}|$ is included in X_{ijt} and X_{jit} as a measure of distance between countries i and j in the space of social indicators.

M2: The corresponding levels of GI_{it} and GI_{jt} are included in X_{it} and X_{jt} in addition to the model $M1$

The model $M0$ is the classical gravity model that includes only fundamental explanatory variables. The model $M1$ naturally assumes the symmetric component of the social cost function, where only the absolute value of the difference in governance level between two trading partners is being used to explain the effect of the social variables on trade outcome. In this case it does not matter which country has the higher score and whether this country generates or receives the trade flow. The model $M2$ introduces the measures of the governance level GI_{it} and GI_{jt} in asymmetric fashion, to differentiate the trading partners roles and their relative scores.

The model is completed by specifying error components a, b, and c that define exporter, importer, and pair specific effects, respectively,

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \sim MVN(0, \Sigma_{ab}) \quad \begin{pmatrix} c_{ij} \\ c_{ji} \end{pmatrix} \sim MVN(0, \Sigma_c) \quad \text{and} \quad \begin{pmatrix} \epsilon_{ijt} \\ \epsilon_{jit} \end{pmatrix} \sim MVN(0, \Sigma_\epsilon) \tag{2.8}$$

where

$$\Sigma_{ab} = \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{bmatrix} \quad \Sigma_c = \sigma_c^2 \begin{bmatrix} 1 & \rho_c \\ \rho_c & 1 \end{bmatrix} \quad \text{and} \quad \Sigma_\epsilon = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho_\epsilon \\ \rho_\epsilon & 1 \end{bmatrix} \tag{2.9}$$

Note that the diagonal elements of Σ_ϵ are identical as defined by the invariant model. For compactness of notation define $Q_{ijt} = [X_{ijt} X_{it} X_{jt}]$, $Q_{jit} = [X_{jit} X_{jt} X_{it}]$ and $\beta = \{\beta_1, \beta_2, \beta_3\}$. Then (2.1) and (2.2) become

$$\ln(y^*_{ijt} + \tau d_{ij}) = \alpha_t + Q_{ijt}\beta + a_i + b_j + c_{ij} + \epsilon_{ijt} \tag{2.10}$$

$$\ln(y^*_{jit} + \tau d_{ji}) = \alpha_t + Q_{jit}\beta + a_j + b_i + c_{ji} + \epsilon_{jit} \tag{2.11}$$

2.3 Estimation

Following the bivariate SUR Tobit model specification in (2.1) – (2.9) we define the likelihood function of the latent data trade pattern between partner countries i and j at time t as

$$L(y^*_{ijt}, y^*_{jit}; \Gamma, Q_{ijt}, Q_{jit}) = |J_{ijt}| MVN(y^*_{ijt}, y^*_{jit}; \Gamma, Q_{ijt}, Q_{jit}) \tag{2.12}$$

for $y_{ijt} > -\tau d_{ij}$ and $y_{jit} > -\tau d_{ji}$, where J_{ijt} is the Jacobian of transformation from ϵ_{ijt} and ϵ_{jit} to y^*_{ijt} and y^*_{jit}

$$J_{ijt} = \frac{1}{(y_{ijt} + \tau d_{ij})(y_{jit} + \tau d_{ji})} \tag{2.13}$$

and Γ denotes the set of all model parameters. Correspondingly, $\Gamma_{-\beta}$ defines the set of all model parameters excluding β. Marginalizing (2.12) over the latent data yields the SUR Tobit likelihood of the standard form. Following Huang (2001) (case of $p = 2$) we derive the likelihood function for each pair of observations y_{ijt} and y_{jit} according to one of four possible regimes defined by the model as

$$L(y_{ijt}, y_{jit}; \Gamma, Q_{ijt}, Q_{jit}) = L(y^*_{ijt}, y^*_{jit}; \Gamma, Q_{ijt}, Q_{jit}) \tag{2.14}$$

if $y_{ijt} > 0$ and $y_{jit} > 0$,

$$L(y_{ijt}, y_{jit}; \Gamma, Q_{ijt}, Q_{jit}) = \int_{-\infty}^{\ln(\tau d_{ij})} L(y^*_{ijt}, y^*_{jit}; \Gamma, Q_{ijt}, Q_{jit}) dy^*_{ijt} \tag{2.15}$$

if $y_{ijt} = 0$ and $y_{jit} > 0$,

$$L(y_{ijt}, y_{jit}; \Gamma, Q_{ijt}, Q_{jit}) = \int_{-\infty}^{\ln(\tau d_{ij})} L(y^*_{ijt}, y^*_{jit}; \Gamma, Q_{ijt}, Q_{jit}) dy^*_{jit} \tag{2.16}$$

if $y_{ijt} > 0$ and $y_{jit} = 0$, and

$$L(y_{ijt}, y_{jit}; \Gamma, Q_{ijt}, Q_{jit}) = \int_{-\infty}^{\ln(\tau d_{ij})} \int_{-\infty}^{\ln(\tau d_{ij})} L(y^*_{ijt}, y^*_{jit}; \Gamma, Q_{ijt}, Q_{jit}) dy^*_{ijt} dy^*_{jit} \tag{2.17}$$

if there is no trade between countries i and j at time period t, i.e. both $y_{ijt} = 0$ and $y_{jit} = 0$. The observed data likelihood function is then

$$L(Y; \Gamma, Q_{ijt}, Q_{jit}) = \prod_{i=1}^{n} \prod_{j>i}^{n} L(y_{ijt}, y_{jit}; \Gamma, Q_{ijt}, Q_{jit}) \tag{2.18}$$

The following general form of the prior distribution can be considered to complete the model

$$p(\Gamma) = p(\theta, \phi) \left(\prod_{i=1}^{n} \prod_{j>i} p(c_{ij}, c_{ji} | \Sigma_c) \right) p(\Sigma_c) p(\Sigma_\epsilon) p(\tau) \tag{2.19}$$

where $\theta = \{\alpha, \beta\}$ and $\phi = \{a, b\}$.

Note, that the special form of covariance matrices of error terms and pair specific random effects makes it difficult to apply the conventional linear regression estimation methods directly, since it requires imposing equality restrictions on the diagonal elements of Σ_c and Σ_ϵ. Adopting the more efficient Wong (1982) approach to the invariant normal model we introduce the following linear transformation

$$u_{ijt} = y^*_{ijt} + y^*_{jit}, \quad \psi_{ij} = c_{ij} + c_{ji} \quad \text{and} \quad \delta_{ijt} = \epsilon_{ijt} + \epsilon_{jit} \tag{2.20}$$

$$v_{ijt} = y^*_{ijt} - y^*_{jit}, \quad \xi_{ij} = c_{ij} - c_{ji} \quad \text{and} \quad \gamma_{ijt} = \epsilon_{ijt} - \epsilon_{jit} \tag{2.21}$$

The new model is

$$u_{ijt} = 2\alpha_t + (Q_{ijt} + Q_{jit})\beta + a_i + a_j + b_i + b_j + \psi_{ij} + \delta_{ijt} \tag{2.22}$$

$$v_{jit} = (Q_{ijt} - Q_{jit})\beta + a_i - a_j - b_i + b_j + \xi_{ji} + \gamma_{jit} \tag{2.23}$$

where δ_{ijt} and γ_{jit} are zero-mean independent error terms with variances $\sigma^2_\delta = 2\sigma^2_\epsilon(1+\rho_\epsilon)$ and $\sigma^2_\gamma = 2\sigma^2_\epsilon(1-\rho_\epsilon)$. The generated pair specific random effects ψ_{ij} and ξ_{ji} are also independent by transformation, with variances $\sigma^2_\psi = 2\sigma^2_c(1+\rho_c)$ and $\sigma^2_\xi = 2\sigma^2_c(1-\rho_c)$.

We choose the following general form of prior distribution for the parameters of transformed model

$$p(\Gamma) = p(\theta,\phi)\left(\prod_{i=1}^{n}\prod_{j>i}^{n} p(\xi_{ij}|\sigma^2_\xi)p(\psi_{ij}|\sigma^2_\psi)\right)p(\sigma^2_\xi)p(\sigma^2_\psi)p(\sigma^2_\delta)p(\sigma^2_\gamma)p(\tau) \tag{2.24}$$

The prior distribution on θ and ϕ parameters is taken to be the multivariate normal such that

$$p(\theta,\phi) = MVN\left(\begin{pmatrix}\theta_0\\ \phi_0\end{pmatrix}, \begin{pmatrix}\Sigma_{\theta_0} & 0\\ 0 & \Sigma_{\phi_0}\end{pmatrix}\right) \tag{2.25}$$

where θ_0 and ϕ_0 denote prior mean quantities and Σ_{θ_0} and Σ_{ϕ_0} denote prior covariances of θ and ϕ, respectively. Note that by construction the inverse of Σ_{ϕ_0} is given by $(\Sigma_{ab}\otimes I_n)^{-1} = \Sigma^{-1}_{ab}\otimes I_n$ so that

$$\Sigma^{-1}_{\phi_0} = \begin{pmatrix} (\sigma^2_b/\triangle)I_n & -(\sigma^2_{ab}/\triangle)I_n \\ -(\sigma^2_{ab}/\triangle)I_n & (\sigma^2_a/\triangle)I_n \end{pmatrix} \tag{2.26}$$

where $\triangle = \sigma_a^2\sigma_b^2 - \sigma_{ab}^2$. Based on the original model assumptions the transformed random effects parameters ψ_{ij} and ξ_{ij} are set to have zero-mean normal prior distributions of the form

$$p(\psi_{ij}) = N(0, \sigma_\psi^2) \quad \text{and} \quad p(\xi_{ij}) = N(0, \sigma_\xi^2) \tag{2.27}$$

We assign the inverted gamma prior distributions for variance parameters σ_ξ^2, σ_ψ^2, σ_δ^2 and σ_γ^2 as

$$p(\sigma_\delta^2) = IG(e_{1\delta}, e_{2\delta}) \tag{2.28}$$

$$p(\sigma_\gamma^2) = IG(e_{1\gamma}, e_{2\gamma}) \tag{2.29}$$

$$p(\sigma_\xi^2) = IG(k_{1\xi}, k_{2\xi}) \tag{2.30}$$

$$p(\sigma_\psi^2) = IG(k_{1\psi}, k_{2\psi}) \tag{2.31}$$

The model specification is completed by assigning the following normal prior distribution to the threshold parameter τ

$$p(\tau) = TN_{[0,\infty)}(\mu_\tau, \sigma_\tau^2) \tag{2.32}$$

The posterior distribution for model parameters is obtained by combining the likelihood function (2.18) with the priors in (2.25), (2.27) – (2.32). We fit the model using a Gibbs sampler with data augmentation where the posterior simulations are conducted by iteratively drawing according to Steps 1 – 4 below.

Step 1: $\tau, y^* | \Gamma_{-\tau,y^*}, y$

Ranjan and Tobias (2007) suggest that due to high degrees of autocorrelation between latent data y^* and threshold parameter τ these objects should be grouped together in a single block. Then the draws from the corresponding joint posterior distribution of τ and y^* can be obtained using the method of composition by recognizing the following identity:

$$p(\tau, y^* | \Gamma_{-\tau,y^*}, y) = p(y^* | \Gamma_{-y^*}, y) p(\tau | \Gamma_{-\tau,y^*}, y) \tag{2.33}$$

The second component of $p(\tau, y^*|\Gamma_{-\tau,y^*}, y)$ has the following form

$$p(\tau|\Gamma_{-\tau,y^*}, Y) \propto p(\tau) \prod_{i=1}^{n} \prod_{j>i}^{n} L(y_{ijt}, y_{jit}; \Gamma, Q_{ijt}, Q_{jit}) \tag{2.34}$$

Unfortunately, as in Ranjan and Tobias (2007), the distribution of τ is not of any standard form, therefore the draws from its conditional posterior cannot be obtained in a direct way. However, since τ is a scalar, a number of different methods could be used to sample from (2.34) indirectly. Due to the high costs of computing the likelihood and difficulties of finding a good approximating density for (2.34) we have chosen to employ the random-walk chain Metropolis-Hastings sub-step (see, e.g., Koop, Porier and Tobias (2007)) over the discrete approximation method and the independence chain Metropolis-Hastings algorithm to generate draws from the targeted posterior distribution. The algorithm generates proposed draw of τ' given the current value of the parameter τ and the realization of the increment random variable χ according to the following random walk process

$$\tau' = \tau + \chi \tag{2.35}$$

where χ is chosen to be normal as a common and convenient option. Equation (2.35) implies that the candidate draw is generated from the univariate normal density $N(\tau, s^2)$ and then accepted with the probability

$$\Pr(\tau', \tau) = \min \left[\frac{p(\tau = \tau'|\Gamma_{-\tau,y^*}, Y)}{p(\tau = \tau|\Gamma_{-\tau,y^*}, Y)}, 1 \right]. \tag{2.36}$$

It is important to select the value of parameter s such that efficient sampling is done. A common practice for the scalar case is to keep the acceptance rate for τ around 50%.

Given the sampled value of the threshold variable applying the data augmentation scheme to SUR Tobit model in hand requires the simulation of the censored data from a truncated multivariate normal distribution. This task can be efficiently accomplished by generating each augmenting variable conditional on the value of the realization of the other response variable in the system (see Huang (2001) for discussion of the general case for p equations).

Using the properties of the bivariate invariant normal distribution the complete posterior conditional distribution for $z_{ijt} = \ln(y^*_{ijt} + \tau d_{ij})$ is defined as follows

$$p(z_{ijt}|\Gamma_{-z_{ijt}}, Y) \sim \begin{cases} TN_{(-\infty,\ln(\tau d_{ij})]}(\mu_{ijt} + \rho_\epsilon(z_{jit} - \mu_{jit}), \sigma^2_\epsilon(1-\rho_\epsilon)) & \text{if} \quad y_{ijt} = 0; \\ \ln(y_{ijt} + \tau d_{ij}) & \text{if} \quad y_{ijt} > 0, \end{cases} \tag{2.37}$$

where

$$\mu_{ijt} = \alpha_t + Q_{ijt}\beta + a_i + b_j + c_{ij} \quad \text{and} \quad \mu_{jit} = \alpha_t + Q_{jit}\beta + a_j + b_i + c_{ji}. \tag{2.38}$$

The density $TN_{(-\infty,\ln(\tau d_{ij})]}(\mu, \sigma^2)$ denotes the conditional univariate normal distribution with mean μ and variance σ^2 truncated to the interval $(-\infty, \ln(\tau d_{ij})]$. Note that the expression for the posterior distribution of z_{ijt} in (2.37) is similar to the result of Ranjan and Tobias (2007, Eq.19) derived for a case of independence of error terms. The corresponding conditional draw of z_{jit} can be obtained by sampling from (2.37) with i and j indices interchanged in the expression.

Step 2: $\theta, \phi|\Gamma_{-\theta,\phi}, Y$

Since many of the covariates in Q_{ij} and Q_{ji} are time invariant or close to being such the problem of estimating parameters θ and ϕ becomes equivalent to the problem of estimating the parameters of overparametrized two-way interaction ANOVA model. It can be shown that the estimated values of θ depend on the choice of country and pair specific random effects. To make the proposed model identified it is thus necessary to impose a number of constraints on the parameter space. Ntzoufras (2009) suggests two particularly useful restrictions, the corner and sum-to-zero, out of which we choose to apply the former one. For this study the corner constraint implies that the effects a_i, b_j and c_{ij} are all set to zero, for some pair of indices i^* and j^* such that $i^* \neq j^*$. Then exporter i^* and importer j^* become a reference trading pair and the corresponding restricted random effects define the baseline bilateral trade level to be $\alpha_t + Q_{ijt}\beta$ for $t = 1, \ldots, T$. The conventional choice

of reference levels in analysis of variance models is the first or the last one, given the order of levels. However, it is more reasonable to select the countries that are close to being the average exporter and importer, correspondingly, as the reference trading partners, since all the inference about the model parameters is made relative to the baseline levels. The proposed parametrization not only makes estimation feasible but also provides a way of interpreting each of the random effect parameters. To demonstrate this, let i and j denote the indices of the reference exporting and importing country, respectively. Calculating the mean difference between the baseline desired bilateral trade and the export from a given country k and a reference importer j using the equation

$$\begin{aligned} E(\ln(y^*_{kjt} + \tau d_{kj}) - \ln(y^*_{jit} + \tau d_{ij})) &= \alpha_t + Q_{kjt}\beta + a_k + b_j + c_{kj} - \alpha_t - Q_{ijt}\beta - a_i - b_j - c_{ij} \\ &= (Q_{kjt} - Q_{ijt})\beta + a_k + c_{kj} \end{aligned} \tag{2.39}$$

suggests that the sum of random effects $a_k + c_{kj}$ define the portion of this difference unexplained by the change in covariate values in Q_{kjt} and Q_{ijt}. By analogy with the previous result, the sum of random effects $b_k + c_{ik}$ account for the change in import levels from the reference exporter i that are due to the difference in unobserved country and pair specific characteristics of trading partners i, j and k, as the following relation shows

$$\begin{aligned} E(\ln(y^*_{ikt} + \tau d_{ik}) - \ln(y^*_{jit} + \tau d_{ij})) &= \alpha_t + Q_{ikt}\beta + a_i + b_k + c_{ik} - \alpha_t - Q_{ijt}\beta - a_i - b_j - c_{ij} \\ &= (Q_{ikt} - Q_{ijt})\beta + b_k + c_{ik} \end{aligned} \tag{2.40}$$

Note, that the corner restriction used in this study is just enough for the identification purposes. A variety of other parameterizations are also available under both corner and sum-to-zero constraints that provide different interpretations of the country specific effects (see Ntzoufras 2009, Sec. 5.4 for details on applied analysis of variance models in Bayesian setting).

Let $\eta_{ijt} = \psi_{ij} + \delta_{ijt}$ and $\lambda_{jit} = \xi_{ji} + \gamma_{jit}$ be composite error terms. Stacking equations (2.22) and (2.23) over t and ij pairs yields a model in vector representation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 2i_m \otimes I_T & Q_u & A & A \\ 0 & Q_v & B & -B \end{pmatrix} \begin{pmatrix} \theta \\ \phi \end{pmatrix} + \begin{pmatrix} \eta \\ \lambda \end{pmatrix} \tag{2.41}$$

where Q_u, Q_v, A and B are the appropriate design matrices. The associated error term covariance matrix is

$$\Omega = \begin{pmatrix} I_m \otimes \Sigma_\eta & 0 \\ 0 & I_m \otimes \Sigma_\lambda \end{pmatrix} \tag{2.42}$$

where $\Sigma_\eta = E(\eta\eta^T) = \sigma_\delta^2 I_T + \sigma_\psi^2 i_T i_T^T$ and $\Sigma_\lambda = E(\lambda\lambda^T) = \sigma_\gamma^2 I_T + \sigma_\xi^2 i_T i_T^T$. Note, that the block-diagonal structure of Ω defined by a transformation described above allows us to obtain a simple expression for its partitioned inverse given by

$$\Omega^{-1} = \begin{pmatrix} I_m \otimes \Sigma_\eta^{-1} & 0 \\ 0 & I_m \otimes \Sigma_\lambda^{-1} \end{pmatrix}. \tag{2.43}$$

The result in (2.43) allows a significantly lower costs of computing the joint posterior density of regression coefficients θ and ϕ.

Given the multivariate normal prior distribution in (2.25) the joint posterior distribution of θ and ϕ is defined by the conventional protocol for Bayesian linear regression models with the general covariance matrix as

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} \sim MVN(Gg, G) \tag{2.44}$$

where G is $p \times p$ matrix such that

$$G = \left(W_u^T (I_m \otimes \Sigma_\eta^{-1}) W_u + W_v^T (I_m \otimes \Sigma_\lambda^{-1}) W_v + \begin{pmatrix} \Sigma_{\theta_0}^{-1} & 0 \\ 0 & \Sigma_{\phi_0}^{-1} \end{pmatrix} \right)^{-1} \tag{2.45}$$

and g is $p \times 1$ vector such that

$$g = W_u^T (I_m \otimes \Sigma_\eta^{-1}) u + W_v^T (I_m \otimes \Sigma_\lambda^{-1}) v + \begin{pmatrix} \Sigma_{\theta_0}^{-1} \theta_0 \\ \Sigma_{\phi_0}^{-1} \phi_0 \end{pmatrix} \tag{2.46}$$

where p is the number of elements in θ vector. The form of posterior in (2.44) – (2.46) implies that parameters θ, ϕ and random effects ψ and ξ are sampled in a single block. Such an approach is generally preferred since it may improve the mixing of the chain in the Gibbs sampling procedure in case of the correlated parameters as discussed in Koop, Poirier and Tobias (2007, Chapter 12) and works well when mT is relatively small. However as the number of countries included in the sample and the length of the data grow employing the general random effect type of estimating scheme becomes increasingly difficult or even infeasible due to the stability issues associated with the inverse operation in (2.45) and the amount of memory required for computing and storing G (remember that $m = n(n-1)/2$, i.e. the size of G increases exponentially in the size of the sample of trade partners, n). In this case an alternative though probably less efficient procedure can be used. Assuming the same multivariate normal prior distribution (2.25) appropriate draws of θ and ϕ are obtained from the following joint posterior distribution conditional on the values of random effects ψ and ξ generated in the previous iteration of the Gibbs algorithm

$$\begin{pmatrix} \theta \\ \phi \end{pmatrix} \sim MVN(Hh, H) \tag{2.47}$$

where H is $p \times p$ matrix such that

$$H = \left(W_u^T W_u/\sigma_\delta^2 + W_v^T W_v/\sigma_\gamma^2 + \begin{pmatrix} \Sigma_{\theta_0}^{-1} & 0 \\ 0 & \Sigma_{\phi_0}^{-1} \end{pmatrix} \right)^{-1} \tag{2.48}$$

and h is $p \times 1$ vector such that

$$h = W_u^T(u - \psi \otimes i_T)/\sigma_\delta^2 + W_v^T(v - \xi \otimes i_T)/\sigma_\gamma^2 + \begin{pmatrix} \Sigma_{\theta_0}^{-1}\theta_0 \\ \Sigma_{\phi_0}^{-1}\phi_0 \end{pmatrix} \tag{2.49}$$

Note, that expressions in (2.47) – (2.49) are similar to the result obtained by Hoff (2005). The corresponding corner parameter restrictions are imposed by removing the appropriate columns of matrices A and B in (2.41). The random effects ψ and ξ are then sampled from their full conditional posterior distributions as described in Step 3 under either of two

proposed strategies.

Step 3: $\psi, \xi | \Gamma_{-\psi,\xi}, Y$

Let $\hat{u}_{ijt} = 2\mu_t + (Q_{ijt} + Q_{jit})\beta + a_i + a_j + b_i + b_j$ and $\hat{v}_{ijt} = (Q_{ijt} + Q_{jit})\beta + a_i - a_j - b_i + b_j$. Assuming the normal priors (2.27) the random effects ψ_{ij} and ξ_{ij} are drawn from their respective posterior conditional distributions defined as

$$\psi_{ij} \sim N(Dd, D) \tag{2.50}$$

where

$$D = (T/\sigma_\delta^2 + 1/\sigma_\psi^2)^{-1} \quad \text{and} \quad d = \sum_{t=1}^{T} (u_{ijt} - \hat{u}_{ijt})/\sigma_\delta^2 \tag{2.51}$$

and

$$\xi_{ij} \sim N(Rr, R) \tag{2.52}$$

where

$$R = (T/\sigma_\gamma^2 + 1/\sigma_\xi^2)^{-1} \quad \text{and} \quad r = \sum_{t=1}^{T} (v_{ijt} - \hat{v}_{ijt})/\sigma_\gamma^2. \tag{2.53}$$

The obtained values of ψ_{ij} and ξ_{ij} should be transformed to the original model parameters c_{ij} and c_{ji}, that can be analyzed and used in Step 1 of the algorithm. This can be done using the relations established by (2.20) and (2.21) as follows

$$c_{ij} = \frac{\psi_{ij} + \xi_{ij}}{2} \quad \text{and} \quad c_{ji} = \frac{\psi_{ij} - \xi_{ij}}{2}. \tag{2.54}$$

The corner restriction on the country specific random effect c is imposed by setting the reference level of c_{ij} to zero.

Step 4: $\sigma_\xi^2, \sigma_\psi^2, \sigma_\delta^2, \sigma_\gamma^2 | \Gamma_{-\sigma_\xi^2,\sigma_\psi^2,\sigma_\delta^2,\sigma_\gamma^2}, Y$

Let $\hat{\delta} = u - W_u\theta - W_u\phi - \psi \otimes i_T$ and $\hat{\gamma} = v - W_v\theta - W_v\phi - \xi \otimes i_T$ where $\psi \otimes i_T$

and $\xi \otimes i_T$ are $(n(n-1)/2 * T)$ stacked vectors. Then, given priors (2.28) and (2.29), the variances of the regression error terms δ and γ can be sampled from the following independent inverse-gamma distributions

$$\sigma_\delta^2 \sim IG\left(\frac{mT}{2} + e_{1\delta}, \left[e_{2\delta}^{-1} + \frac{1}{2}\hat{\delta}^T\hat{\delta}\right]^{-1}\right) \tag{2.55}$$

$$\sigma_\gamma^2 \sim IG\left(\frac{mT}{2} + e_{1\gamma}, \left[e_{2\gamma}^{-1} + \frac{1}{2}\hat{\gamma}^T\hat{\gamma}\right]^{-1}\right) \tag{2.56}$$

By analogy, given the independent prior distributions (2.30) and (2.31) we obtain the following posterior densities for the variances of random effects ψ and ξ to draw from

$$\sigma_\psi^2 \sim IG\left(\frac{m}{2} + k_{1\psi}, \left[k_{2\psi}^{-1} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j>i}\psi_{ij}^2\right]^{-1}\right) \tag{2.57}$$

$$\sigma_\xi^2 \sim IG\left(\frac{m}{2} + k_{1\xi}, \left[k_{2\xi}^{-1} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j>i}\xi_{ij}^2\right]^{-1}\right). \tag{2.58}$$

The elements of the original covariance matrices Σ_c and Σ_ϵ can be reconstructed from σ_δ^2, σ_γ^2, σ_ψ^2 and σ_ξ^2 using the following relations

$$\sigma_c^2 = \frac{\sigma_\psi^2 + \sigma_\xi^2}{4} \quad \text{and} \quad \rho_c = \frac{\sigma_\psi^2 - \sigma_\xi^2}{\sigma_\psi^2 + \sigma_\xi^2} \tag{2.59}$$

and

$$\sigma_\epsilon^2 = \frac{\sigma_\delta^2 + \sigma_\gamma^2}{4} \quad \text{and} \quad \rho_\epsilon = \frac{\sigma_\delta^2 - \sigma_\gamma^2}{\sigma_\delta^2 + \sigma_\gamma^2}. \tag{2.60}$$

Note that values of both ρ_c and ρ_ϵ are bound to the interval from -1 to 1 without applying the explicit restrictions which is different from the direct estimation approach suggested in Gill and Swartz (2001, 2007).

2.4 Testing

Similar to Ranjan and Tobias (2007) our primary interest is to study hypotheses concerning the role of the governance indicator variables. In particular, we focus on investigating the significance of the effect of the governance level on the bilateral trade between countries

and the asymmetry of this effect due to the partner's importing or exporting position in the trade process. This task is carried out by comparing model $M0$ to $M1$ and model $M1$ to $M2$, respectively, using the values of Bayes factors associated with corresponding models. There are a variety of methods that allow us to carry out this task, some of which are difficult to employ in more complicated cases. However, since the models of interest are nested, implying a restriction of the form $\omega = 0$ for some subvector of model parameters $\omega \in \Gamma$, the Savage-Dickey density ratio (Verdinelli and Wasserman 1995) is considered to be a useful method. Given that the priors for two models $M1$ and $M2$ satisfy the following condition

$$p(\Gamma|\omega = 0, M2) = p(\Gamma|M2) \tag{2.61}$$

the Theorem 4.1 in Koop (2006) implies that the Bayes factor BF_{12} comparing the restricted model $M1$ to unrestricted $M2$ has the form

$$BF_{12} = \frac{p(\omega = \omega_0|y, M_2)}{p(\omega = \omega_0|M_2)} \tag{2.62}$$

where the expressions in numerator and denominator are the posterior and the prior distributions of the unrestricted model $M2$ both evaluated at $\omega = 0$. Similarly, the Bayes factor comparing models $M0$ and $M1$ can be computed as

$$BF_{01} = \frac{p(\omega = \omega_0|y, M_1)}{p(\omega = \omega_0|M_1)} \tag{2.63}$$

providing $p(\Gamma|\omega = 0, M1) = p(\Gamma|M1)$ is satisfied.

2.5 Data

The trade data used for the study are compiled from the UN Commodity Trade Statistics Database (UN Comtrade). Data are from the Standard International Trade Classification, second revision, (SITC Rev2) and are annual bilateral export and import volumes of trade given in US dollars. We use 1996, 2000 and 2004 annual data series for a sample of 77 countries, for which total agricultural trade volume exceeds 3 million U.S. dollars (see Fig. 2.1). Data on the annual GDP in US dollars and population of trading countries, as well as the

relevant data on governance, are taken from the World Bank's World Development Indicators and World Governance Indicators Database. We use four governance indicators: government effectiveness, regulatory quality, rule of law and the control of corruption. The first two measures describe the government ability to provide effective public policy and private sector regulation, correspondingly. The rule of law reflects the level of law enforcement while the control of corruption suggests the extent the public power is used in the private interests (see Kaufmann, Kraay and Mastruzii (2009) for more details on governance indicators and the procedure of their construction). The data on agricultural land area and length of the exclusive fishery zone are taken from FAO and CIA Factbook, respectively. We use the distance estimates as well as the data on common borders and official languages provided by CEPII.

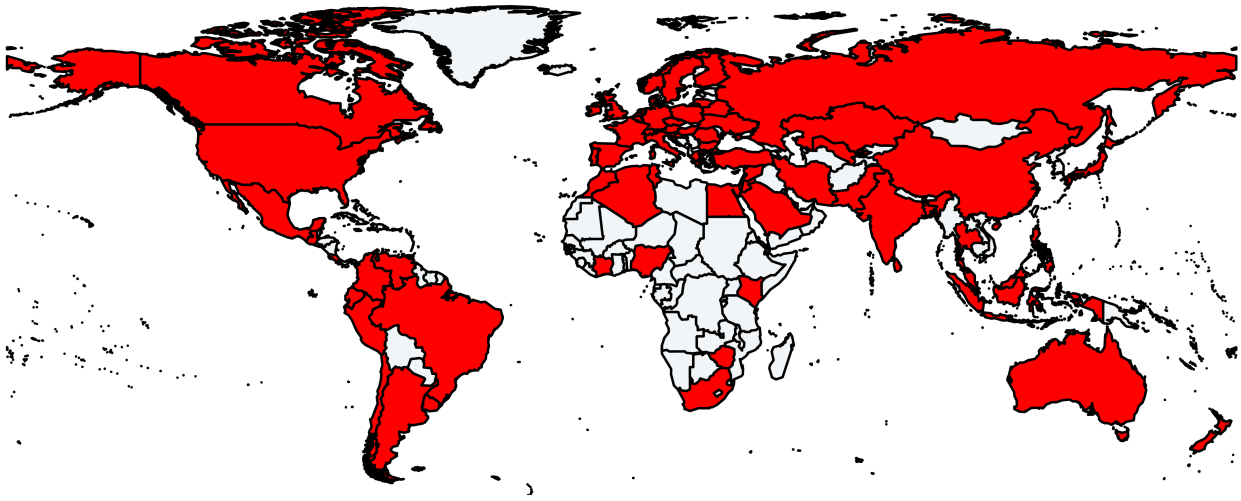


Figure 2.1: Political world map. Red color indicates the countries selected for the study sample

2.6 Empirical Results

We select the prior distributions of the α and β parameters such that $\theta_0 = 0_p$ and $\Sigma_{\theta_0} = 10I_p$, while the prior mean and the prior covariance matrix of country specific effects a_i and b_i in ϕ are taken to be 0_{2n} and I_{2n} by setting $\sigma_a^2 = \sigma_b^2 = 2$ and $\sigma_{ab} = 0$. The hyperparameters

for inverted gamma prior distributions of variance parameters were set to relatively small noninformative values $e1_\delta = e2_\delta = 0.1$, $e1_\gamma = e2_\gamma = 0.1$, $k1_\psi = k2_\psi = 0.1$ and $k1_\xi = k2_\xi = 0.1$. For the threshold parameter we have chosen $\mu_\tau = 10$ and $\sigma_\tau = 500^2$ to reflect the fact that no prior information is available about τ. Both uniform and highly diffuse normal prior distributions for τ worked well, however the latter may be preferred to avoid the problems with possibly improper from of posterior in (2.34). Experiments suggested the value of $s^2 = 0.005^2$ to obtain a 50% acceptance rate for τ'. We run the posterior simulator for 11000 iterations discarding the first 1000 of posterior draws as burn-in. Several chains were used to ensure that convergences of parameters were achieved. The graphical diagnostics of posterior simulator output suggested that chains mixed well.

The posterior means of the parameters and their associated posterior standard deviations (PSD) and numerical errors (NE) are reported in Table 2.1. We present the individual country exporter and importer effects along with their relative allocation in Figures 2.2 – 2.4. Due to the natural uncorrelatedness of the permanent characteristics of trading partners from other factors, the effect of these regressors is robust to the chosen model specifications and it is consistent with the underlying theory of international trade. The estimation results suggest that the larger distance ($DIST$) between the trading countries leads to the lower values of trade flows between them which is the expected result since the physical distance serves as the proxy for the trade costs. At the same time the effect of contiguity ($CONT$) found traditionally to be positive for a variety of reasons, including preferential trade with immediate neighbors and historical ties, is estimated to be insignificant. The possible explanation for that may lay within the covariant dependent threshold specification of the model which uses the distance based threshold parameter accounts implicitly for the effect of contiguity (which is necessary associated with the fact that the neighbor countries are located relatively close to each other) and thus can make the common border variable potentially redundant. This result is consistent with the one obtained by Ranjan and Tobias (2007), who report the absence of a strong and significant border effect in a threshold Tobit setting.

The value of the coefficient of the common language variable ($LANG$) is estimated to be positive which is in line with the theory of gravity models that generally assumes that the language barrier makes it more difficult to trade. On the other hand, we didn't find an evidence of a significant effect of a language of exporter or importer being one of the United Nations official languages on trade, which is potentially due to the operations of the freight forwarders that level conditions for countries with different languages. In general, both trade partners having a common official language can help to reduce the indirect costs of trade and facilitate the negotiation process, however given a contradiction in the language variables results it is reasonable to assume that the coefficient of a common language variable reflects, to a large extent, the effect of historical ties between the trade partners rather than the pure convenience of no language barrier.

The estimation results show the most sensitivity to the introduction the social barriers to trade that can be explained by a strong relation between the quality of governance and socio-economic factors. We used the Savage-Dickey ratio described in Section 2.4 to test the importance of inclusion of the social barriers to trade measured by the absolute difference in governance levels in a symmetric component of the model. The formal test is carried out by examining the associated Bayes factor in (2.63). The logarithm of BF_{01} was calculated to be approximately -150.55 which strongly suggests that the absolute difference in governance levels is indeed a valid factor contributing to the explanation of the changes in trade flows. The data provides strong evidence in favor of the assumption about asymmetry of social costs, as indicated by BF_{21}. In the light of this evidence we focus the following discussion on the results of the complete model $M2$. The effects of the population size and the importer GDP are estimated to be consistent with the conventional results for gravity models. However, the negative coefficient of the exporter GDP deviates from what the theory predicts. Such a difference could be because we only observe the partial effect of gravity due to the substitution between the agricultural and non-agricultural sectors. With a unit increase in wealth the

richer country should be able to import more agricultural commodities shifting its interest towards more advanced industries.

We found no significant effect of the agricultural land endowment. To the contrary, the access to an exclusive fishery zone is beneficial for both exporter and importer. Examining both the symmetric and asymmetric components of the social costs suggests that trade flows are most sensitive to the changes in the governance levels associated with the indicators affecting trade directly, such as the regulatory quality and the rule of law. The ability of the government to formulate effective public policy and control corruption appeared to be not significant in explaining the variation in bilateral trade. In the case of asymmetric social barriers to trade the inference about only the symmetric component of the social costs of trade is misleading since the difference in governance indicators is based directly on their levels. Therefore in order to predict the overall effect of a change in the gap in the regulatory quality of two countries on trade between them one should consider a variety of factors and initial conditions, as well as the possible directions of individual changes.

2.7 Conclusions

The gravity model is one of the most popular empirical tools to study the effects of fundamental characteristics of trading partners on bilateral trade flows between countries. It proved to be a particularly useful method in a variety of applications to both international and the regional trade policy analysis due to the easy estimation and the simplicity of interpretation of the results the gravity models provide. However due to the purely empirical nature that reduced form gravity equations usually have, they suffer from a number of drawbacks often related to the lack of theoretical foundation and to the restrictions a simple univariate log-linear specification of the conventional trade models imposes. In this study we develop a Bayesian social relation gravity model that properly reflects the two-stage decision process and incorporates the social barriers between countries. We generalize the Eaton-Tamura threshold Tobit estimation framework to a special bivariate case that accounts for

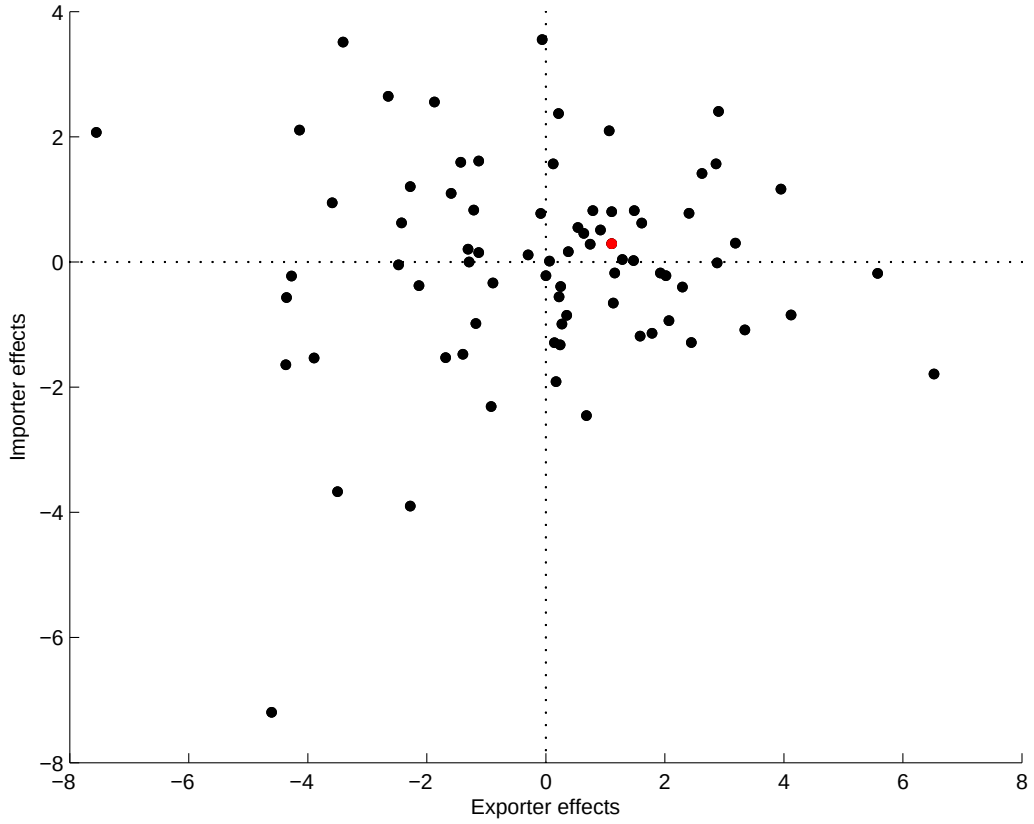


Figure 2.2: Exporter versus importer specific effects. Red dot represents the U.S.

the invariance of the model to labeling the trading partners as implied by the interaction structure of international trade. Our model uses the distance based threshold in order to provide a better flexibility and accuracy of estimation. The proposed methodology was used in empirical application to analyze the effects of social resistance between countries, measured by the difference in the governance levels, on bilateral trade flows. We found the effects of the fundamental variables to be generally consistent with the results reported traditionally in the international trade literature. The data analyzed in this study provided significant evidence in favor of the hypothesis about the importance of the difference in the regulatory quality and the law enforcement the governments of the trading partners provide in predicting the bilateral trade. At the same time no strong effects of the difference in the gov-

ernment effectiveness and corruption levels were found which is probably due to the indirect and not critical influence these characteristics have on trade. Based on the Savage-Dickey density ratio we also conclude that the bilateral trade resistance has asymmetric form, where the position of the trading partner, either in the role of importer or exporter, matters.

Table 2.1: Gravity model estimation results

Variable	Base			Symmetric			Asymmetric		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE	Post. Mean	PSD	NE
τ	0.084	0.006	0.000	0.084	0.006	0.000	0.081	0.006	0.000
α_{1996}	-6.008	1.679	0.017	-6.098	1.689	0.017	-4.118	1.709	0.017
α_{2000}	-4.012	1.679	0.017	-3.770	1.688	0.017	-1.782	1.713	0.017
α_{2004}	-3.921	1.699	0.017	-3.645	1.707	0.017	-0.907	1.735	0.017
GDP_{exp}	0.473	0.093	0.001	0.410	0.094	0.001	-0.955	0.142	0.001
GDP_{imp}	1.217	0.096	0.001	1.146	0.096	0.001	0.251	0.140	0.001
POP_{exp}	0.531	0.131	0.001	0.607	0.132	0.001	2.339	0.185	0.002
POP_{imp}	-0.255	0.133	0.001	-0.178	0.133	0.001	1.026	0.183	0.002
AGR/POP_{exp}	0.096	0.080	0.001	0.085	0.079	0.001	0.079	0.079	0.001
AGR/POP_{imp}	0.050	0.093	0.001	0.034	0.094	0.001	0.098	0.094	0.001
SEA/POP_{exp}	0.807	0.063	0.001	0.791	0.063	0.001	0.595	0.064	0.001
SEA/POP_{imp}	0.352	0.058	0.001	0.328	0.058	0.001	0.184	0.060	0.001
UNL_{exp}	-0.536	0.346	0.004	-0.568	0.345	0.004	-0.495	0.345	0.004
UNL_{imp}	-1.172	0.346	0.004	-1.214	0.345	0.004	-1.113	0.347	0.004
$DIST$	-1.935	0.093	0.002	-1.999	0.090	0.001	-1.986	0.093	0.001
$CONT$	-0.013	0.356	0.006	0.270	0.352	0.006	0.342	0.349	0.006
$LANG$	2.184	0.222	0.004	2.194	0.216	0.003	2.136	0.213	0.003
$\triangle GOVEFF$				0.501	0.169	0.002	0.379	0.167	0.002
$\triangle REG$				-1.117	0.128	0.002	-0.958	0.127	0.002
$\triangle LAW$				1.073	0.168	0.002	1.272	0.170	0.002
$\triangle CORRUPT$				-0.066	0.160	0.002	-0.181	0.159	0.002
$GOVEFF_{exp}$							0.028	0.249	0.003
$GOVEFF_{imp}$							-0.195	0.250	0.003
REG_{exp}							1.040	0.160	0.002
REG_{imp}							1.173	0.158	0.002
LAW_{exp}							1.779	0.277	0.003
LAW_{imp}							1.212	0.274	0.003
$CORRUPT_{exp}$							0.122	0.219	0.002
$CORRUPT_{imp}$							-0.123	0.218	0.002

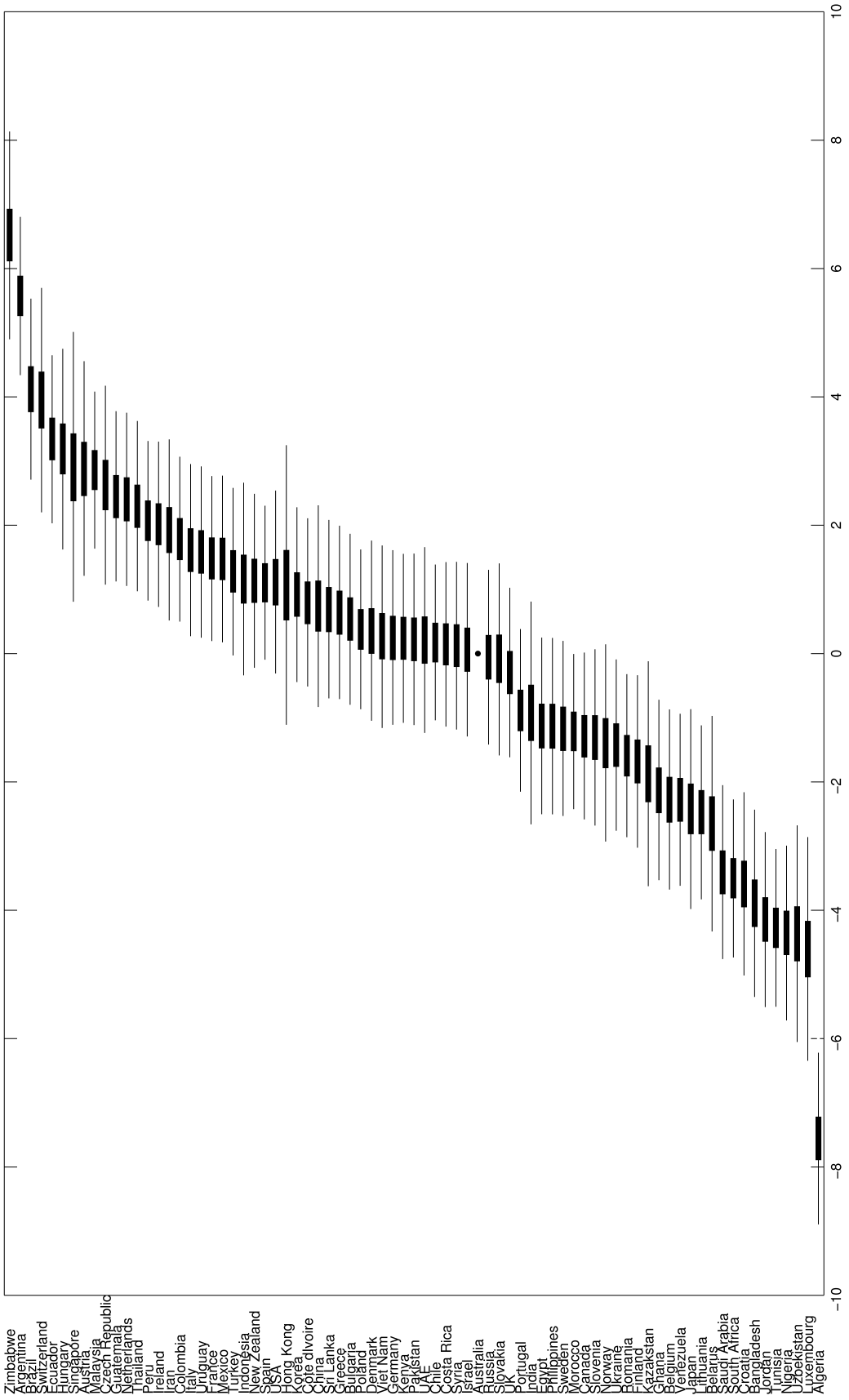


Figure 2.3: Exporter country specific effects

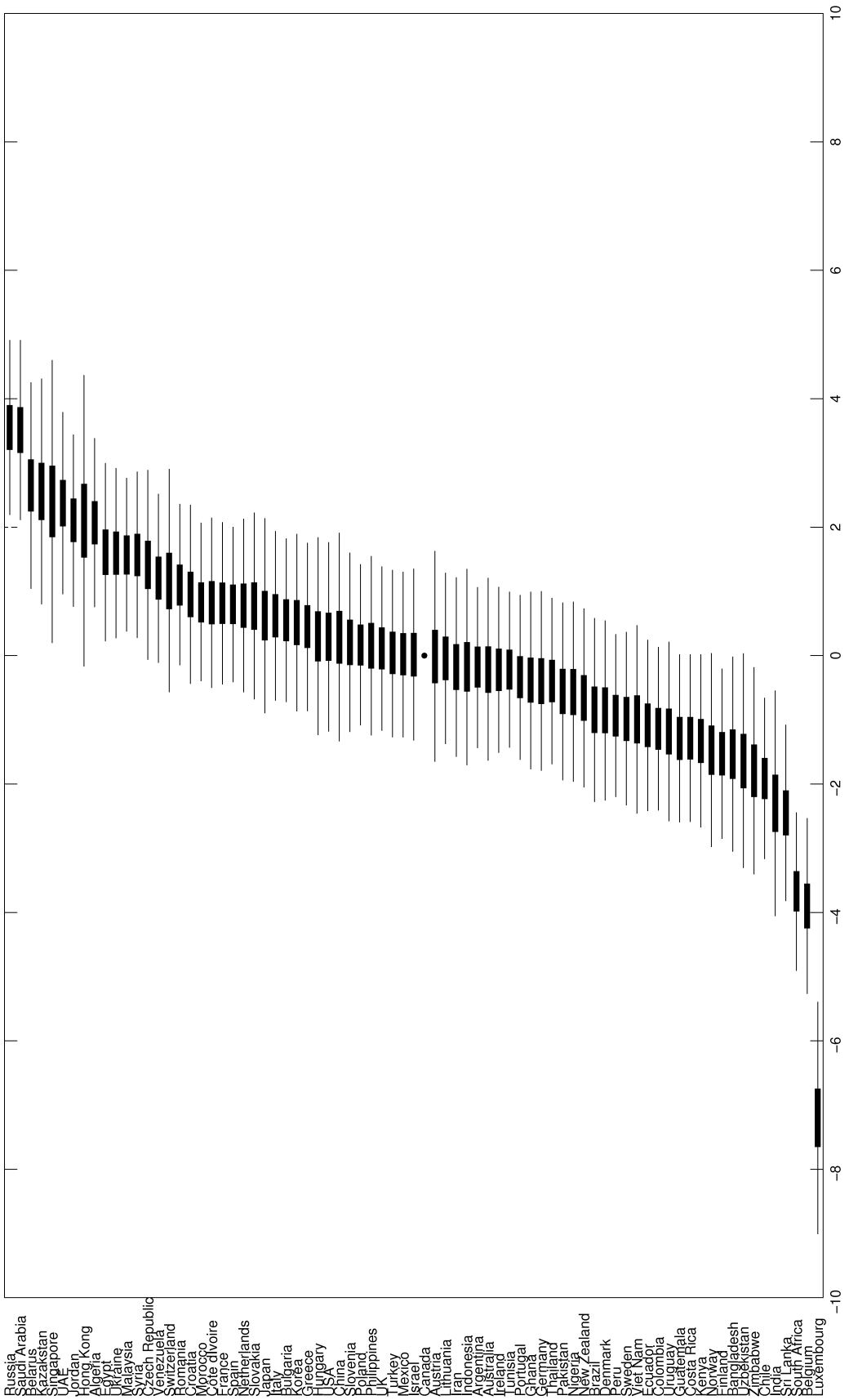


Figure 2.4: Importer country specific effects

Bibliography

[1] Anderson, J.E. and E. van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170 - 192.

[2] Baier, S.L. and Bergstrand. 2007. "Do Free Trade Agreements Increase Members' International Trade?" *Journal of International Economics*, 71(1): 72 - 95.

[3] Bun, M.J.G. and F.J.G.M. Klaassen. 2007. "The Euro Effect on Trade is not as Large as Commonly Thought." *Oxford Bulletin of Economics and Statistics*, 69(4): 473 - 496

[4] Eaton, J. and A. Tamura. 1994. "Bilateralism and Regionalism in Japanese and US Trade and Direct Foreign investment patterns." *Journal of Japanese and International Economies*, 8: 478 - 510.

[5] Gill, P.S. and Swartz, T.B. 2001. "Statistical analysis for round robin interaction data." *The Canadian Journal of Statistics*, 29(2): 321-331.

[6] Gill, P.S. and Swartz, T.B. 2007. "Bayesian analysis of dyadic data." *American Journal of Mathematical and Management Sciences*: Special volume on Modern Advances in Bayesian Theory and Applications, 27(1,2): 73 - 92.

[7] Grant, J.H. and D.M. Lambert. 2008. "Do Regional Trade Agreements Increase Members' Agricultural Trade?" *American Journal of Agricultural Economics*, 90(3) : 765 - 782.

[8] Huang, H.-C. 2001. "Bayesian Analysis of the SUR Tobit Model." *Applied Economics Letters*, 8: 617 - 622.

[9] Kandogan, Y. 2007. "Sensitivity of International Blocs' Trade Effect of Alternative specifications of the Gravity Equation." *Journal of Applied Economics*, 2: 337 - 360

[10] Kaufmann, D., A. Kraay and M. Mastruzii. 2009. "Governance Matters VIII: Aggregate and Individual Governance Indicators, 1996-2008." *World Bank Policy Research Working Paper* No. 4978

[11] Koop, G., D.J. Poirier and J.L. Tobias. 2007. *Bayesian Econometric Methods*. Cambridge University Press. New York, NY.

[12] Ntzoufras, I. 2009. *Bayesian Modeling Using WinBUGS.* John Wiley and Sons, Inc., Hoboken, NJ.

[13] Ranjan, P. and J.L. Tobias. 2007. "Bayesian Inference for the Gravity Model." *Journal of Applied Econometrics*, 22: 817 - 838.

[14] Rose, A.K. 2004. "Do We Really Know That the WTO Increases Trade?" *American Economic Review*, 94(1): 98 - 114.

[15] Santos Silva, J.M.C. and S. Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641 - 658.

[16] Verdinelli, I. and L. Wasserman. 1995. "ComputingBayes Factors Using a Generalization of the Savage-Dickey Density Ratio." *Journal of the American Statistical Association*, 90: 614 - 618.

[17] Ward, M.D. and P.D. Hoff. 2007. "Persistent Patterns of International Commerce." *Journal of Peace Research*, 44(2): 157 - 175.

Chapter 3

Spatio-temporal Model for Agricultural Yield Prediction

3.1 Introduction

Despite the extensive knowledge accumulated over time in the field of modeling crop yield distributions it remains an important area of research due to its role in modern agricultural economics. Accurate information about the behavior of crop yields is a key component of successful policy applications in many areas of agribusiness and finance such as farm decision planning, designing agricultural insurance and government supported policy making. A long-standing result concluded from the empirical studies in the field is the rejection of the assumption of the normality of crop yield distribution in favor of various nonsymmetric alternatives such as the Beta (e.g., Nelson and Preckel 1989, Hennesy, Babcock and Hayes 1997), the Gamma (e.g., Gallagher 1987), the log-normal (e.g. Goodwin, Roberts and Coble 2000) and SU family distributions (e.g. Moss and Shonkwiler 1994, Ramirez 1997). In addition a variety of nonparametric and semi-parametric solutions to the problem have also been offered in the literature (e.g. Goodwin and Ker 1998, Ker and Coble 2003 and, Racine and Ker 2004). In particular, Norwood, Roberts and Lusk (2004) found the method of Goodwin and Ker (1998) to outperform other models in out-of-sample prediction power. However there is still no consensus as to what yield model is superior for empirical work, since the results of normality and performance tests depend significantly on the variety of assumptions and the specifications as well as the data used for each study (see e.g. Ramirez and McDonald 2006 for a comment on the Norwood, Roberts and Lusk 2004 result and Just and Weninger 1999 for a discussion of methodological problems that occur in typical crop yield distribution analyses that can make the validity of results questionable). An interest in the

spatio-temporal component of yield models emerged significantly in recent years (e.g. Wang and Zhang 2003, Ozaki, Ghosh, Goodwin and Shirota 2008, Harri, Erdem, Coble and Knight 2009 and Ozaki and Silva 2009). The spatio-temporal approach to crop yield modeling allows increasing the scale of the studied problems and carrying on the analysis in its full efficiency by avoiding the errors of aggregation through the proper use of spatial information. However the computational complexity of the spatio-temporal methods remains a problem since it often imposes restrictions on one of the components of the analysis – spatial, temporal or distributional.

The objective of this study is to develop a method for modeling crop yield distributions that will allow one to characterize their dynamic behavior by incorporating spatial information to increase the efficiency of analysis and make it available for disaggregate levels of data while retaining the flexibility of the shapes of the crop yield distribution for each spatial unit. The conditional nature of the underlying estimation algorithm assumes that only a small problem will be addressed at a time, providing the advantage of maintaining the efficiency and feasibility of the analysis in large scale models with computational complexity growing linearly in the number of spatial units included.

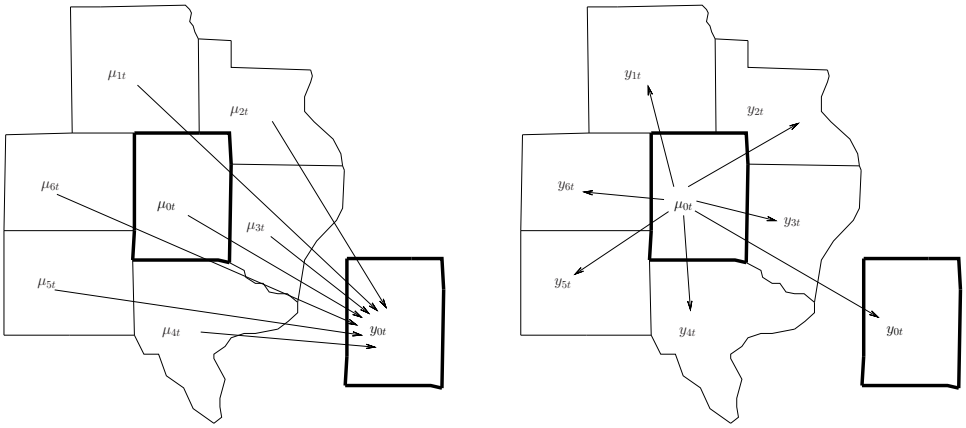
3.2 Model

The proposed approach to agricultural yield prediction models the yield distribution of interest as a spatial mixture of unobserved dynamic processes distributed normally at each period of time such that

$$p(y_{it}) = \sum_{j \in A_i} w_{ij} \phi(\mu_{jt}, \sigma_j^2) \tag{3.1}$$

where $\phi(\mu, \sigma^2)$ denotes the normal density function with mean μ and variance σ^2. A set of spatial unit indices A_i defines the spatial neighborhood for unit i, with $i \in A_i$, describing the primal spatial relation between y_i and μ_j's as shown in Fig 3.1(a). The neighbors' contribution structure is completely described by the set of spatial weights w such that $\sum_{j \in A_i} w_{ij} = 1$ and is assumed to be constant over time. Let B_i denote a set of spatial unit indices j for

which $i \in A_j$, that defines the spatial neighborhood for unit i latent process and describes the dual relation between μ_i and y_j's as shown in the Fig 3.1(b). Let us denote the number of members of A_i (and, correspondingly, the number of mixture components of $p(y_i)$) as k_i and the number of members of B_i as m_i. The nature of the spatial mixture definition of yield distribution (3.1) assumes that only one of k_i latent processes μ_j, $j \in A_i$, will actually contribute to the realization of y_i at any given moment of time $t = 1, \ldots, T$. Therefore, at any given moment of time t only a subset $B_{it} \subseteq B_i$ of the spatial units $j \in B_i$ will be active recipients of the latent signal μ_i. The number of members in B_{it} is thus assumed to vary over time and can be denoted as $m_{it} \leq m_i$, with $m_{it} = 0$ meaning that the spatial unit i is not providing information actively to any distribution, while $m_{it} = m_i$ implying its complete spatial contribution. The stochastic trend specification of the latent spatial



(a) Primal relation (b) Dual relation

Figure 3.1: Spatial relations: a) primal, latent information to observed and b) dual, observed information to latent.

processes estimated in this study is that of a local linear type which is quite general and shown to be well suited for a variety of applications including agricultural yield prediction (see, e.g., Moss and Shonkwiler 1993). It consists of a set of m_{it} measurement equations

$$y_{jt} = \mu_{it} + \epsilon_{jt} \tag{3.2}$$

where $\epsilon_{jt} \sim N(0, \sigma_i^2)$, for all $j \in B_{it}$, and two transition equations that govern the dynamics of the unobserved spatial unit mean μ_{it}

$$\mu_{it} = \mu_{it-1} + \eta_{it-1} + \nu_{it} \tag{3.3}$$

$$\eta_{it} = \eta_{it-1} + \xi_{it} \tag{3.4}$$

where $\nu_{it} \sim N(0, \delta_i\sigma_i^2)$ and $\xi_{it} \sim N(0, \gamma_i\sigma_i^2)$, such that $E(\nu_{it}, \xi_{is}) = 0$ for all t and s. For compactness of notation, let us denote $\alpha_{it} = \{\mu_{it}, \eta_{it}\}$ and $w_{it} = \{\nu_{it}, \xi_{it}\}$ to be stacked (2×1) vectors of state variables and state errors, respectively. Similarly, let $u_{it} = \{y_{jt}\}_{j \in B_{it}}$ and $v_{it} = \{\epsilon_{jt}\}_{j \in B_{it}}$ be $(m_{it} \times 1)$ stacked vectors of observations in (3.2) and measurement errors associated with them. Then using the vector notation introduced above, equations (3.2) - (3.4) form the following state-space model:

$$u_{it} = H_{it}\alpha_{it} + v_{it} \tag{3.5}$$

$$\alpha_{it} = F\alpha_{it-1} + w_{it} \tag{3.6}$$

where $v_{it} \sim N(0, R_i)$ and $w_{it} \sim N(0, Q_i)$, such that $E(v_{it}, w_{is}) = 0$ for all t and s. Matrices H_t and R_i in (3.5) are defined as follows

$$H_{it} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{bmatrix} = [i_{m_{it}} 0_{m_{it}}] \quad \text{and} \quad R_i = \begin{bmatrix} \sigma_i^2 & 0 & \dots & 0 \\ 0 & \sigma_i^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_i^2 \end{bmatrix} = \sigma_i^2 I_{m_{it}} \tag{3.7}$$

and matrices F and Q_i in (3.6) are defined as follows

$$F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad Q_i = \begin{bmatrix} \delta_i\sigma_i^2 & 0 \\ 0 & \gamma_i\sigma_i^2 \end{bmatrix} \tag{3.8}$$

3.3 Estimation

The Bayesian treatment of finite mixture models suggests augmenting the data likelihood (3.1) with the set of mixture component labels, $\{z_{ijt}\}$, where $\{z_{ijt}\} = 1$ indicates that

the observation y_{it} is generated from the jth labeled component of the mixture distribution (see, e.g. Koop, Porier and Tobias 2007 for details). In the context of this study, $\{z_{ijt}\} = 1$ implies that the latent process j contributed to the observed realization of the ith spatial unit yield y_i at time t. The resulting expression for the model parameters likelihood conditional on the values of the latent mixture component indicators is

$$L(\Gamma, \{\alpha_i\}_i^N; \{y_i\}_i^N, \{z_{ijt}\}) = \prod_{t=1}^{T} \prod_{i}^{N} \prod_{j \in A_i} \phi(y_{it}|\alpha_{jt}, \sigma_j^2)^{z_{ijt}} \tag{3.9}$$

where Γ denotes a set of all variance parameters of the model. For computational purposes (3.9) has a more convenient form than the original unconditional likelihood. The model is completed by choosing the set of prior distribution specifications, where two issues have to be considered when working with the mixtures of Normal densities, as discussed in Koop (2006, Sec. 10.3.3). First, the likelihood function for this class of models is unbounded and therefore informative priors are required. Second, the likelihood function is also invariant under relabeling of the mixture components. As the result, any of $k!$ combinations of possible labeling of the k mixture components will yield the same likelihood function. This second property of the mixture of Normals models, called "label switching", is essentially an identification problem and can be irrelevant in cases where the researcher is only interested in analyzing the quantities and functions based explicitly on the likelihood value. However, if we do not have enough prior information to distinguish between the mixture components the invariance of the likelihood to all possible permutations of parameter vectors will lead to a posterior distribution which is also the same for all possible combinations of component labels. One of the conventional ways of dealing with the label switching is to choose prior distribution that will impose labeling restrictions through identifiability constraints on the model parameter space, such as $\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_N^2$, $\mu_1 < \mu_2 < \cdots < \mu_N$ or $w_1 < w_2 < \cdots < w_N$, where only one such constraint is required. In many cases the choice of constraint is naturally suggested by either the underlying economic theory or the type of the mixture used. However, there is no obvious strict ordering of the parameter space, such as the ones discussed above, for the problem studied here. Indeed, there is not enough information

to believe that the variation in yield realization is always higher for one spatial unit than the other and that such a relation exists for all counties and defines the strict ranking of σ_i^2, $i = 1, \ldots, N$. Similarly, identifying the mixture components by restricting the latent process space requires imposing very strong assumptions that lack the formal statistical or economic logic to support them. We suggest that the solution to the identification problem can be found by examining the structure of spatial weights. It still remains difficult to establish the strict ordering of weights within any given spatial neighborhood. At the same time the initial definition of the problem clearly implies that the own effect of the latent spatial process μ_i must dominate the effects of contributions of the rest of spatial neighbors. Formally this condition can be stated as the following inequality restriction, $w_{ii} > w_{ij}$ for all i and $j \in A_i$. Note, that the established inequality will only allow one to differentiate the own mixture component from the rest of the contributors which is generally not enough for identification purposes. However, applying this constraint to each of $i = 1, \ldots, N$ sets of spatial weights w_i provides the required N identifying conditions for each of corresponding N elements of parameter space.

Considering the identification requirements discussed above, we select a prior distribution of the general form

$$p(\Gamma, \{\alpha_i\}_i^N, \{z_i\}_i^N, \{w_i\}_i^N) = \prod_i^N p(\alpha_i|Q_i)p(\sigma_i)p(\delta_i)p(\gamma_i^2)p(z_i|w_i)p(w_i) \tag{3.10}$$

where Q_i is defined in (3.8) and $p(\alpha_i|Q_i) = p(\alpha_{i1}, \ldots, \alpha_{it}, \ldots, \alpha_{iT}|Q_i)$ denotes the prior for each of the state vectors α_i. Rewriting the expression for $p(\alpha_i|Q_i)$ as the product of appropriate conditional densities and applying the Markov property of state space models will yield the following result, which establishes a hierarchical type of construct in prior distribution of α_i

$$p(\alpha_i|Q_i) = p(\alpha_{i1}|Q_i) \ldots p(\alpha_{it}|\alpha_{it-1}, Q_i) \ldots p(\alpha_{iT}|\alpha_{iT-1}, Q_i) \tag{3.11}$$

where the particular form of each component is defined by the state equation (3.6) as multivariate normal distributions (Koop (2006), Sec. 8.3.1)

$$p(\alpha_{it}|\alpha_{it-1}, Q_i) = MVN(\alpha_{it}|F\alpha_{it-1}, Q_i) \quad \text{for } t = 2, T \tag{3.12}$$

and

$$p(\alpha_{i1}|\alpha_{i0}, Q_i) = MVN(\alpha_{i1}|F\alpha_{i0}, Q_i). \tag{3.13}$$

We use the following inverse gamma prior distributions for all three sets of variance parameters in Γ

$$p(\sigma_i^2) = IG(a1_i, a2_i) \tag{3.14}$$

$$p(\delta_i) = IG(b1_i, b2_i) \tag{3.15}$$

$$p(\gamma_i) = IG(c1_i, c2_i). \tag{3.16}$$

The natural choice of prior distribution for the mixture component labels vectors $z_i = \{z_{ij}\}_{j \in A_i}$ is the multinomial distribution for T trials,

$$p(z_i|w_i) = M(T, w_i) \tag{3.17}$$

while component probabilities are assigned the Dirichlet prior

$$p(w_i) = D(\{\alpha_{ij}\}_{j \in A_i})\mathbb{I}(w_{ii} > w_{ij}) \tag{3.18}$$

where the indicator function $\mathbb{I}(w_{ii} > w_{ij})$ imposes the identifiability restriction.

The posterior distribution for model parameters is obtained by combining the likelihood function (3.9) with the priors in (3.12) – (3.18). We fit the model using a Gibbs sampler with data augmentation where the posterior simulations are conducted by iteratively drawing according to Steps 1 – 5 below.

Step 1: $\{z_i\}_i^N|\Gamma, \{\mu_i\}_i^N, y$

Combining the prior information with the augmented likelihood yields the following posterior densities of z_{it}

$$z_{it} \sim M\left(1, \left\{ \frac{w_{ij}\phi(y_{it}|\mu_{it}, \sigma_j^2)}{\sum_{j\in N_i} w_{ij}\phi(y_{it}|\mu_{jt}, \sigma_j^2)} \right\}_{j\in A_i}\right) \tag{3.19}$$

Step 2: $\{\mu_i, \eta_i\}_i^N|\Gamma, \{z_i\}_i^N, y$

Assuming the form of $p(\alpha_i|Q_i)$ in (3.11) – (3.13) the posterior inference about α_i can be carried out using the conventional Bayesian methods for linear models. Note, that the problem of estimating the components of the latent state vectors is essentially the problem of estimating kT time-varying linear regression coefficients (the number of states k equals 2 in our case), k for each time period, leading to a kT-dimensional multivariate normal posterior. Despite the relative ease of this approach, obtaining the draws from the posterior distribution of α_i can become difficult in practice due to a large T and possible high correlation between the coefficients (Koop (2006), Sec. 8.3.1) that result in low numerical stability and inefficiency of the algorithm. The hierarchical structure of the problem allows us to use the Bayesian sequential approach to state vector estimation where the posterior inference about each component of α_i is obtained conditionally on the posterior value of the previous component in the time sequence. The particular results based on the specification of the measurement (3.5) and the state equation (3.6) can be derived according to the Theorem 15.1 in West and Harrison (1989) for general multivariate dynamic linear models as follows. For $t = 2, T$, the prior distribution for α_{it} is implied by the state transition rule as the bivariate normal density

$$(\alpha_{it}|\mathcal{I}_{t-1}) \sim N(d_{t-1}, D_{t-1}) \tag{3.20}$$

with

$$d_{t-1} = Fa_{t-1} \quad \text{and} \quad D_{t-1} = H_{it}P_{t-1}H_{it}^T + Q_i$$

where a_{t-1} and P_{t-1} are the posterior mean and variance of α_{it-1}, respectively, and $\mathcal{I}_{t-1}$ denotes the past information available. Updating the prior information with the observed values of y and corresponding allocation variables z at time t gives the following bivariate normal posterior distribution for α_{it}

$$(\alpha_{it}|\mathcal{I}_t) \sim N(a_t, P_t) \tag{3.21}$$

with $K_{it} = D_{t-1}H_{it}(H_{it}D_{t-1}H_{it}^T + R_i)^{-1}$ such that

$$a_t = d_{t-1} + K_{it}(u_{it} - H_{it}d_{t-1}) \quad \text{and} \quad P_t = D_{t-1} - K_{it}H_{it}^T D_{t-1}$$

where a_t and P_t are the posterior mean and variance of α_{it}, respectively, and $\mathcal{I}_t$ denotes the current information available.

Step 3: $\sigma_j^2|\Gamma_{-\sigma_j^2}, \{z_i\}_i^N, \{\mu_i\}_i^N, y$

The posterior density of σ_j^2 is defined as

$$\sigma_j^2 \sim IG\left(\frac{1}{2}\sum_i n_{ij} + a1_j, \left[a2_j^{-1} + \frac{1}{2}\sum_i\sum_t z_{ijt}(y_{it} - \mu_{jt})^2\right]^{-1}\right) \tag{3.22}$$

where $n_{ij} = \sum_{t=1}^T z_{ijt}$ denotes the number of time periods the latent process j contributed to the observed realization of the ith spatial unit yield y_i. Note that exactly the same expression for posterior of σ_j^2 in (3.22) can be obtained from both finite mixture of normals and Gaussian state space estimation procedures since both models are just two different representations of one spatio-temporal process.

Step 4: $\delta_j, \gamma_j|\Gamma_{-\delta_j,\gamma_j}, \{z_i\}_i^N, \{\mu_i\}_i^N, y$

Assuming the inverse gamma priors for δ_j and γ_j parameters, respectively, the corresponding

posterior densities are derived as

$$\delta_j \sim IG\left(\frac{T}{2} + b1_j, \left[b2_j^{-1} + \frac{1}{2}\sum_t (\mu_{jt} - \mu_{jt-1} - \eta_{jt-1})^2/\sigma_j^2\right]^{-1}\right) \tag{3.23}$$

$$\gamma_j \sim IG\left(\frac{T}{2} + c1_j, \left[c2_j^{-1} + \frac{1}{2}\sum_t (\eta_{jt} - \eta_{jt-1})^2/\sigma_j^2\right]^{-1}\right) \tag{3.24}$$

Note, that the form of prior distributions for δ_j and γ_j defines the degree of smoothness of state variables series and has to be specified by researcher. The choice to favor the higher variation in the stochastic trend will improve the in sample fit of the model; however, it will simultaneously decrease the forecast power. The choice of these priors is crucial to model performance.

Step 5: $w_i|\Gamma_{-w_{ij}}, \{z_i\}_i^N, \{\mu_i\}_i^N, y$

Given the Dirichlet prior for component probabilities the posterior draws of w_i for $i = 1, \ldots, N$ can be obtained from the following conditional densities

$$w_i \sim D(\{n_{ij} + \alpha_{ij}\}_{j \in A_i})\mathbb{I}(w_{ii} > w_{ij}) \tag{3.25}$$

3.4 Prediction

The general one-step ahead prediction can be computed using the one-step forecast result from Theorem 15.1 in West and Harrison (1989) by drawing from the following predictive density of u_{it}

$$(u_{ti}|\mathcal{I}_{t-1}) \sim N(H_{it}d_{t-1}, H_{it}D_{t-1}H_{it}^T + R_i) \tag{3.26}$$

where $H_{it}d_{t-1}$ and $H_{it}D_{t-1}H_{it}^T + R_i$ are the predictive mean and variance of u_{ti}, respectively, and $\mathcal{I}_{t-1}$ denotes the past information available.

3.5 Data

The data used for the study are obtained from the National Agricultural Statistical Service (NASS) and are 1970 - 2009 annual Iowa state county level corn yield data in bushels. The $N \times N$ connectivity matrix C was computed using OpenGeoDa software by applying the first order Queen contiguity criterion (inclusion of all immediate neighbors). The value of the matrix element $C(i, j) = 1$ means that the spatial unit i is the neighbor to the spatial unit j while $C(i, j) = 0$ implies no connectivity between i and j according to the chosen contiguity criterion. Note, that unlike the conventional spatial analysis methods the algorithm proposed in this study requires treating spatial unit i as the neighbor to itself based on the idea of the own latent process contribution. As the result the connectivity matrix we use has the values of its diagonal elements all equal to 1 (see Fig. 3.2). The Iowa state cartographic boundary files was obtained from U.S. Census Bureau Census 2000 County and County Equivalent Areas Cartographic Boundary Files Database at http://www.census.gov/geo/www/cob/co2000.html.

3.6 Results

Based on the general dynamic linear model equations (3.2) - (3.4) we fit two competing models, that are different in the set of spatial weights they use. Model 2 assumes no restrictions on the space of w_{ij} besides the natural nonnegativity and adding up to 1 constraints, thus allowing for spatial correlation between the agricultural yields of any spatial units i and j. In context of the spatial mixture methods discussed in Section 3.2, Model 2 explicitly implies non-normality of the underlying yield distribution y_i if more than one of the spatial weights w_{ij} for $j \in A_i$ is greater than zero. An alternative, Model 1, is a special case of Model 2 that requires $w_{ii} = 1$ for all $i = 1, \ldots, N$. Such a restriction essentially prohibits any spatial correlation by assuming that only own latent process μ_i is going to be an active information contributor to the distribution of yield y_i at any time t thus reducing the Model

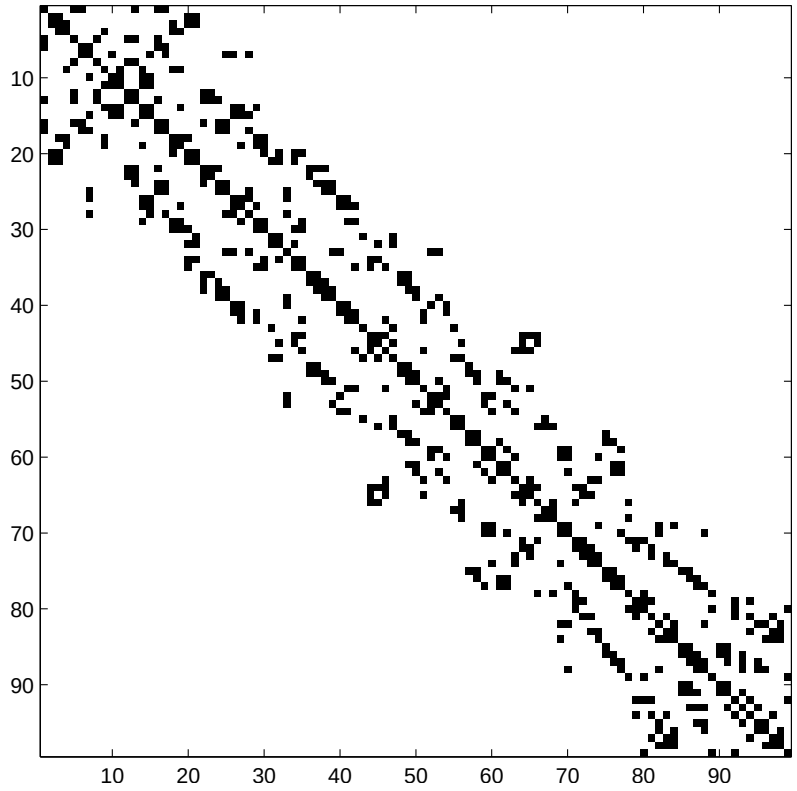


Figure 3.2: Connectivity matrix of Iowa state counties.

2 to a normality case. We have used the first 35 observations covering the period from 1970 to 2004 to fit both models using the algorithm described in Section 3.3 while the last 5 years of data for the period from 2005 to 2009 served as the basis for the model comparison based on the out-of-sample predictive power. To assess the predictive abilities of both models we use the predictive criterion, denoted $D_k(M)$ (M stands for the model index), developed by Gelfand and Ghosh (1998) that incorporates both the goodness-of-fit and the penalty for higher predictive variance estimates, where the value of k defines the tradeoff between these two measures. Assuming the squared error loss function, the expected predictive loss can be computed as

$$D_k(M) = \sum_{i=1}^{N}\sum_{t=1}^{T} V(y_{it}^*|y_{it}, M) + \frac{k}{k+1}\sum_{i=1}^{N}\sum_{t=1}^{T}(E(y_{it}^*|y_{it}, M) - y_{it})^2 \qquad (3.27)$$

where $E(y^*_{it}|y_{it}, M)$ and $V(y^*_{it}|y_{it}, M)$ define the mean and variance of predicted (replicated) observations y^*_{it}. The first term on the right hand side measures the precision of the forecasts while the second measures its quality of fit. Letting $k \to \infty$ yields the limiting case of predictive loss criterion

$$D_\infty(M) = \sum_{i=1}^{N}\sum_{t=1}^{T} V(y^*_{it}|y_{it}, M) + \sum_{i=1}^{N}\sum_{t=1}^{T}(E(y^*_{it}|y_{it}, M) - y_{it})^2 \tag{3.28}$$

To establish the unique basis for comparison the prior distributions for common parameters of both models considered in this study, σ_i^2, δ_i and γ_i, are set to be the same. To ensure the proper form of the resulting posterior distributions and identification of model parameters in the state space equations we choose $p(\sigma_i^2) = IG(200, 10^{-5})$, $p(\delta_i) = IG(500, 100)$ and $p(\gamma_i) = IG(500, 100)$ for all $i = 1, \ldots, N$ to define the prior information about σ_i^2, δ_i and γ_i in both cases. The selected values of prior hyperparameters $a1_i = 200$ and $a2_i = 10^{-5}$ set the prior mean and standard deviation of σ_i^2 to approximately 100 and 50, respectively. This specification makes the prior distribution for σ_i quite informative (having the weight equivalent to 200 observations), but not too restrictive, allowing for enough support of the yield distributions to be explored while moving the model from its degenerate case of $\mu_{it} = y_{it}$ for all i and t. To ensure further model identification and avoid the label switching problem the parameters for Dirichlet prior distribution on w_i were set to $\alpha_{ij} = 1$ for all $i \neq j$ and $\alpha_{ii} = 50$ for all $i = 1, \ldots, N$. We run the posterior simulator for 11000 iterations discarding the first 1000 of posterior draws as burn-in. Several chains were used to ensure that convergence of parameters were achieved. The graphical diagnostics of the posterior simulator output suggested that chains mixed well showing no obvious sign of label switching. Table 3.1 presents the results of comparison of Model 1 and Model 2, based on the assumption of presence and absence of spatial correlation and the associated normality and asymmetry of yield distributions, respectively. In out-of-sample prediction the more flexible Model 2 shows better goodness-of-fit results; however, it is penalized in the precision measure for producing a predictive variance that is about two times higher than the one generated by the normal model. This is to be expected since the approach using the spatial mixtures of distributions

Table 3.1: Predictive loss criterion D_k for competing models of agricultural yield prediction

Model	Model 1	Model 2
Assumption	spatial correlation	no spatial correlation
Goodness-of-fit	$1.1746E+05$	$4.5525E+04$
Penalty	$8.2958E+04$	$1.6771E+05$
D_1	$1.5894E+05$	$1.2938E+05$
D_2	$1.7277E+05$	$1.5733E+05$
D_3	$1.7968E+05$	$1.7131E+05$
D_∞	$2.0042E+05$	$2.1324E+05$

operates technically on a much bigger number of observations by including all neighboring counties (similarly to a panel data methods) and thus is able to capture the variability of yield realizations better than in the case of county by county estimation. At the same time the proposed method, by construction, explores a target distribution that has a wider support (regardless of the probability mass placed on it, however small it is) which leads to a lower precision of the predicted yield values. We suggest that the results can potentially be improved by decreasing the effective range of yield realizations to consider. This can be done by defining the spatial neighborhood using a similarity measure based on a variety of spatial unit characteristics such as the land quality, weather conditions, etc, rather than a simple idea of spatial proximity. Equation (3.27) shows that the predictive loss criterion D_k allows us to compare both models while accounting for a tradeoff between the goodness-of-fit measure and the penalty assigned to the higher variance of prediction. By choosing the value of k the researcher can directly specify the discount placed on the penalty value thus defining the substitution effect between the quality and the precision of predictive model fit in D_k. Given the reasonably low values of k Model 2 has smaller associated predictive loss than its normal alternative, Model 1. Clearly, a low discount on the penalty component of the criterion changes the result in favor of the county by county approach where no spatial cor-

relation is considered; however, such a choice of k may not be reasonable in many practical applications.

Since it is not possible to present the estimation and prediction results for all 99 counties in the study sample, we choose three representative spatial units that have the lowest, median and the highest average yield over the sample time. These counties are Clarke (99.33 bu/ha), Ida (127.78 bu/ha) and Webster (138.53 bu/ha). We show the corresponding smoothed predicted yield densities in Fig. B.1 – B.3. In all three cases the spatial mixtures model exhibits a higher flexibility and thus a better ability to explain the variation in the yields level, especially in-sample, which is important for accurate process estimation. Note, that the proposed model was able to predict or at least correctly suggest the possibility of the extreme events such as unusual decrease of productivity common for all counties in 1983, 1988, 1993 and most of the local unfavorable outcomes (as shown in one step prediction plots in Fig. B.4 – B.6). The out-of-sample results show more similarity in the output from both models due to a large extent to a low variability of yields in the past two decades. Examining Fig. B.1 – B.3 clearly indicates that the yield distributions predicted using the spatial mixtures of distributions are unimodal; however they are rarely close to being symmetric, and usually tend to have a longer left tail. This latter property can be explained by the much higher probability of agricultural yields falling below the spatial neighborhood trend level (including catastrophic events) than to significantly surpass it. Note also, that the proposed approach allows consideration of extreme possibilities without significantly decreasing the quality of the estimation since the higher variation of the predicted outcomes is actually implied by the result of mixing several latent distributions with different (and more compact) support (note, that the estimates of σ_j^2 are lower for Model 2 for every county in the study sample as reported in Table A.1 – A.3 of Appendix A).

3.7 Application to Pricing Crop Insurance Contracts

An accurate description of the lower portion of the yield distributions becomes a crucial feature required for an efficient use of the proposed model in empirical application such as the pricing of the crop insurance contracts since specifically the lower portion of the estimated or predicted density is usually associated with the agricultural risks. The more complete information about the behavior of the yields allows computing premium rates that are close to actuarially fair levels necessary for a proper functioning of the crop insurance programs. The premium rates that are based on the overestimated risk of losses and thus higher than optimal ones, based on the requirement for the agricultural insurance premium to be equal to the expected loss, lead to expensive insurance plans and economic inefficiency. At the same time, the estimates that are too low to reflect the actual probabilities of unfavorable events result in the crop insurance programs costs exceeding their break even threshold level undermining the viability of such programs. The actuarially fair premium rate for a coverage level $0 < \lambda < 1$ can be computed using the following equation given the expected yield y^* (based usually on the actual production history) as

$$\text{Premium Rate} = \frac{\Pr(y < \lambda y^*)(y^* - E(y|y < \lambda y^*))}{\lambda y^*} \tag{3.29}$$

where the numerator of (3.29) calculates the expected loss in terms of yield and requires explicitly the knowledge about the distribution of the yield, while the expression in the denominator λy^* specifies the yield guarantee or the maximum loss covered by the crop insurance contract. Equation (3.29) suggests that the premium rates computed using the empirical predicted yield density to evaluate the expected loss will be close to the actuarially fair level to the extent the empirical model accurately describes the actual distribution of agricultural yield. To demonstrate the performance of the compared yield prediction models in crop insurance application we computed the premium rates for crop insurance contracts using the formulae in (3.29) for Clarke, Ida and Webster counties for 2005 – 2009. The results based on the 10 years rolling actual production history (APH) at the different coverage

levels are presented in Tables 3.2 – 3.4. In almost all the cases the normal county level model clearly underestimates the premium rates compared to the alternative model that assumes non-normality and spatial correlation between counties. This is particularly true for the years and location where the significant decrease in yield levels was observed. For example, the suggested premium rates for Clarke county for the year 2004 based on the information predicted by Model 1 is calculated as 0.000 – 0.010% compared to the Model 2 level of 7.314 – 9.157%. Less difference is observed in the suggested premium rates for the Ida county for the year 2005, reported as 0.000 – 0.005% and 0.039 – 0.254% for two models, correspondingly. In the regular years the magnitude of the discrepancies in the results of the proposed calculations is usually smaller; however, it is still significant. The shape of the yield distribution implied by Model 1 which is characterized by a symmetry and normal tails also leads to another limitation in application to the pricing of the crop insurance contracts: when APH is used as the basis of the yield guaranty the normal county level prediction often do not consider it possible for the yield to fall below the guaranteed level. This situation is almost never the case with Model 2 that assumes a more flexible distribution with a longer left tail. The premium rates proposed by the new approach for such years and locations generally take values in the range of 0.0001 – 0.9000%, exceeding the 1.00% level only for the high risk counties with traditionally low observed yields.

3.8 Conclusions

Accurate information about future yield distributions remains an essential requirement for making successful decisions on agricultural policy and farming. This study proposes a flexible approach of modeling the agricultural yields in a Bayesian hierarchical setting.

In order to overcome the limitations associated with the constraints the conventional modeling imposes on the estimated data generating process, such as the shape of yield density and dynamic specification, we develop a spatio-temporal model for agricultural yield

Table 3.2: Premium rates for Clarke county

Year	Expected Yield (bu/ha)	Coverage Level (%)	Model 1 Premium Rate (%)	Model 2 Premium Rate (%)
2005	114.66	70	n/a	0.98346
		75	0.00024	1.08366
		80	0.00085	1.20285
		85	0.00291	1.35541
		90	0.01398	1.56050
2006	122.79	70	n/a	1.10096
		75	0.00039	1.22098
		80	0.00099	1.35359
		85	0.00277	1.51509
		90	0.01156	1.71834
2007	127.92	70	n/a	1.00549
		75	n/a	1.09354
		80	0.00088	1.19081
		85	0.00415	1.30340
		90	0.02115	1.43674
2008	131.67	70	n/a	7.31418
		75	n/a	7.75851
		80	0.00031	8.20474
		85	0.00231	8.66997
		90	0.01009	9.15690
2009	132.61	70	0.00033	0.35498
		75	0.00504	0.45758
		80	0.01958	0.62495
		85	0.06707	0.87133
		90	0.21366	1.25016

prediction based on the spatial mixture of distributions. Mixing the information from different spatial units allows the proposed model to explore a wide range of possible future yield realizations while increasing the precision of the location level estimation. Resulting predicted yield densities deviate from the normal and have longer tails, in particular the left one, that allow the proposed method to at least suggest the possibility of extreme events, with nontrivial probability to occur, which the competing normal model failed to recognize.

Table 3.3: Premium rates for Ida county

Year	Expected Yield (bu/ha)	Coverage Level (%)	Model 1 Premium Rate (%)	Model 2 Premium Rate (%)
2005	151.22	70	n/a	0.03852
		75	n/a	0.05801
		80	n/a	0.08888
		85	0.00003	0.14524
		90	0.00480	0.25406
2006	155.01	70	n/a	0.00915
		75	n/a	0.01392
		80	n/a	0.02083
		85	0.00160	0.03534
		90	0.01698	0.07088
2007	158.68	70	n/a	0.00232
		75	n/a	0.00432
		80	0.00053	0.00781
		85	0.00189	0.01860
		90	0.01158	0.06032
2008	161.94	70	n/a	0.00122
		75	n/a	0.00457
		80	0.00016	0.01244
		85	0.00222	0.03184
		90	0.01347	0.07846
2009	164.50	70	n/a	n/a
		75	n/a	n/a
		80	n/a	0.00012
		85	0.00191	0.00246
		90	0.01606	0.02058

Table 3.4: Premium rates for Webster county

Year	Expected Yield (bu/ha)	Coverage Level (%)	Model 1 Premium Rate (%)	Model 2 Premium Rate (%)
2005	160.12	70	n/a	0.45652
		75	n/a	0.50802
		80	n/a	0.56497
		85	n/a	0.63029
		90	0.00372	0.70688
2006	164.33	70	n/a	0.37519
		75	n/a	0.41989
		80	n/a	0.46971
		85	n/a	0.54488
		90	0.00386	0.67236
2007	166.08	70	n/a	0.25466
		75	n/a	0.28327
		80	n/a	0.31476
		85	0.00006	0.35671
		90	0.00200	0.42467
2008	169.57	70	n/a	0.54027
		75	n/a	0.60593
		80	n/a	0.69038
		85	0.00078	0.82932
		90	0.00619	1.08317
2009	173.12	70	n/a	0.23423
		75	n/a	0.25767
		80	0.00053	0.29082
		85	0.00438	0.34227
		90	0.02851	0.44118

We applied both prediction methodologies to estimate crop insurance premium rates for a variable level of coverage in order to assess the potential gains offered by the new methodology. The empirical application is carried on for corn yields 2005 - 2009 in Iowa state counties using the 10 years rolling county level APH as the basis for expected yield. We show that the differences between suggested premium rates are significant in particularly for the years and locations where substantial decreases in yield were observed, with the normal model assuming no spatial correlation clearly underestimating the expected losses.

The proposed approach to agricultural yield estimation and prediction fully utilizes the advantages of the hierarchical structure of the spatial mixture dynamic models by presenting the main task as a sequence of more simple estimation problems. The decreased computational complexity associated with the developed method makes it particularly attractive for the modeling of the spatio-temporal problems with large number of units, including farm level analysis.

Bibliography

[1] Gallagher, P. 1987. "U.S. Soybean Yields: Estimation and Forecasting With Nonsymmetric Disturbances." *American Journal of Agricultural Economics*, 69: 796 - 803.

[2] Gelfand, A.E. and S.K. Ghosh. 1998. "Model Choice: A Minimum Posterior Predicitve Loss Approach." *Biometrika*, 85 (1): 1 - 11.

[3] Goodwin, B. and A. Ker. 1998. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts". *American Journal of Agricultural Economics*, 80: 139 - 153.

[4] Goodwin, B.K., M.C. Roberts and K.H. Coble. 2000. "Measurement of Price Risk in Revenue Insurance: Implications of Distributional Assumptions." *Journal of Agricultural and Resource Economics*, 25(1): 195 - 214.

[5] Harri, A.,C. Erdem, K.H. Coble and T.O. Knight. 2009. "Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality". *Review of Agricultural Economics*, 31(1): 163 - 182.

[6] Hennessy, D.A., B.A. Babcock and D.J. Hayes. 1997. "Budgetary and Producer Welfare Effects of Revenue Insurance." *American Journal of Agricultural Economics*, 79: 1024 - 1034.

[7] Just, R. and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *American Journal of Agricultural Economics*, 81: 287 - 304.

[8] Ker, A. and K. Coble. 2003. "Modeling Conditional Yield Densities". *American Journal of Agricultural Economics*, 85(2): 291 - 304.

[9] Moss, C.B. and J.S. Shonkwiler. 1993. "Estimating Yield Distributions Using a Stochastic Trend Model and Nonnormal Errors." *American Journal of Agricultural Economics* 75 (5) : 1056 - 62.

[10] Nelson, C.H. and P.V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics*, 71: 370 - 378.

[11] Norwood, B., M.C. Roberts and J.L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions". *American Journal of Agricultural Economics*, 86(4) : 1032 - 1043.

[12] Ozaki, V.A., S.K. Ghosh, B.K. Goodwin and Ricardo Shirota. 2008. "Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts." *American Journal of Agricultural Economics* 90 (4) : 951 - 961.

[13] Ozaki, V.A., B.K. Goodwin and Ricardo Shirota. 2008. "Parametric and Nonparametric Statistical Modelling of Crop Yield: Implications for Pricing Crop Insurance Contracts." *Applied Economics*, 40 : 1151 - 1164.

[14] Ozaki, V.A. and R.S. Silva. 2009. "Bayesian Ratemaking Procedure of Crop Insurance Contracts with Skewed Distribution." *Journal of Applied Statistics*, 36(4) : 443 - 452.

[15] Racine, J. and A. Ker. 2006. "Rating Crop Insurance Policies with Efficient Nonparametric Estimators that Admit Mixed Data Types". *Journal of Agricultural and Resource Economics*, 31(1): 27 - 39.

[16] Ramirez, O. 1997. "Estimation and Use of a Multivariate Parametric Model for Simulating Heteroscedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean, and Wheat Yields." *American Journal of Agricultural Economics*, 79: 191 - 205.

[17] Ramirez, O. and T. McDonald. 2006. "Ranking Crop Yield Models: A Comment". *American Journal of Agricultural Economics*. 88(4): 1105 - 1110.

[18] Wang, H.H. and H. Zhang. 2003. "On the Possibility of a Private Crop Insurance Market: A Spatial Statistics Approach". *The Journal of Risk and Insurance*, 70(1) : 111 - 124.

[19] West, M. and J. Harrison. 1989. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York Inc, NY.

Chapter 4

Concluding Remarks and Future Research Directions

This dissertation discusses the use of a Bayesian hierarchical approach to analyzing complex economic datasets in two empirical applications. We show that the multilevel modeling of large stochastic systems can be carried out efficiently by introducing an appropriate hierarchical structure which allows us to avoid oversimplification in the underlying model while at the same time maintaining the tractability of results and the feasibility of estimation.

However, while both empirical studies presented here were successful in achieving the goals defined in this dissertation, a number of limitations are still present that should be addressed in future research. Despite the flexibility and robustness the hierarchical methods have to offer, finding the balance between the amount of objective and subjective information to enter the model remains crucial for the validity of the obtained results. The specification of the prior distributions is the weaker point in many cases of Bayesian estimation particularly when there is not enough evidence available. In these situations the use of empirical Bayes methods is advocated as an alternative approach to obtain inference about some of the model parameters (see, e.g. Robert 2007, Chapter 10). Koop (2003, Sec. 8.2) discusses the application of an empirical Bayes approach to estimating the state equation variance in a local linear time series model which is directly related to the problem of defining the δ_i and γ_i variance parameters in the spatio-temporal yield prediction model developed in Chapter 3 of this dissertation. Given little information about δ_i and γ_i *a priori* the choice of these parameters is usually subjective. This leads to selecting prior distributions that are quite restrictive in order to achieve the model identification and the desired level of time series smoothness which is somewhat arbitrary. To improve the objectivity of prediction

more research is required to assess the possibility of using empirical Bayes methods in the estimation framework we propose which represents a more complex class of local linear models.

The results of the study presented in Chapter 2 are more robust to the choice of prior information; however, the model itself despite the improved specification still relies on some of the more strict assumptions characterizing the conventional log-linear gravity equations. To make the analysis more realistic it is important to study the heterogeneity of trade partners beyond the traditional random intercept estimation framework that clearly implies constant marginal effects of all the covariates and can lead to inconsistencies due to the possible aggregation bias. An alternative random coefficients approach offers more flexibility and allows for a number of important features of international trade models such as accounting for small and large country effects that can be modeled naturally through a hierarchical set of prior distributions.

More attention is required for the spatial component of the trade models. Even though the research presented in Chapter 2 is not truly spatial in a sense of geostatistics or spatial analysis on irregular lattice, the gravity model assumes a special spatial relation of the origin-destination type that arise between the trading partners in both the geographic and socio-economic spaces. The potential clustering of countries allows one to research a number of important economic hypothesis related to the growth, development and trade theories, and can be approached using the advanced methods recently developed in the area of statistical mixture modeling, revealing yet one more connection between the two seemingly unrelated empirical studies presented in this dissertation.

Bibliography

[1] Koop, G. 2003. *Bayesian Econometrics*. Wiley & Sons Ltd, Chichester, England.

[2] Robert, C.P. 2007. *The Bayesian Choice*. Springer-Verlag, New York, NY.

Appendix A

Spatial Mixtures Model Estimation Results

Table A.1: Model 2 estimation results for σ_i^2 for Adair – Fayette counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Adair	118.13	8.706	0.087	105.23	8.077	0.100
Adams	120.47	8.855	0.088	105.52	7.913	0.090
Allamakee	108.49	7.679	0.076	101.39	7.320	0.079
Appanoose	124.67	9.215	0.094	106.76	8.504	0.142
Audubon	111.86	8.087	0.079	104.31	7.870	0.090
Benton	113.07	8.079	0.083	104.16	7.757	0.085
Black Hawk	113.93	8.248	0.082	101.95	7.204	0.074
Boone	119.07	8.611	0.087	101.72	7.232	0.072
Bremer	113.24	8.155	0.081	102.26	7.330	0.077
Buchanan	112.45	8.010	0.080	102.47	7.314	0.074
Buena Vista	111.90	8.004	0.079	101.52	7.378	0.078
Butler	109.71	7.805	0.078	103.01	7.587	0.082
Calhoun	112.02	7.995	0.081	103.90	7.562	0.082
Carroll	115.29	8.338	0.084	103.86	7.999	0.100
Cass	115.95	8.368	0.084	104.04	7.720	0.085
Cedar	116.82	8.542	0.090	102.04	7.310	0.079
Cerro Gordo	110.50	7.836	0.078	102.17	7.458	0.083
Cherokee	110.07	7.977	0.079	101.61	7.352	0.079
Chickasaw	113.60	8.151	0.081	104.74	8.114	0.105
Clarke	128.22	9.544	0.101	105.56	8.057	0.098
Clay	114.24	8.203	0.084	102.33	7.387	0.078
Clayton	108.80	7.775	0.078	100.42	7.037	0.072
Clinton	112.64	8.095	0.083	103.44	7.669	0.083
Crawford	112.05	8.045	0.080	104.78	7.749	0.082
Dallas	114.59	8.386	0.085	103.81	7.728	0.085
Davis	130.17	9.834	0.104	105.39	8.121	0.100
Decatur	128.36	9.631	0.100	106.43	8.323	0.110
Delaware	114.84	8.274	0.082	101.83	7.337	0.074
Des Moines	119.31	8.724	0.089	101.63	7.339	0.077
Dickinson	116.15	8.451	0.084	103.31	7.645	0.084
Dubuque	111.09	7.918	0.079	102.18	7.424	0.086
Emmet	117.45	8.530	0.085	103.51	7.516	0.080
Fayette	108.39	7.755	0.077	100.38	7.101	0.073

Table A.2: Model 2 estimation results for σ_i^2 for Floyd – Mitchell counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Floyd	114.50	8.223	0.083	103.44	7.725	0.089
Franklin	111.36	7.942	0.078	102.33	7.309	0.072
Fremont	118.55	8.640	0.091	106.86	8.235	0.113
Greene	118.17	8.632	0.087	104.01	7.517	0.079
Grundy	109.96	7.817	0.078	102.38	7.343	0.079
Guthrie	112.23	7.993	0.079	104.06	7.633	0.081
Hamilton	117.18	8.558	0.087	104.26	7.644	0.081
Hancock	109.49	7.830	0.077	103.02	7.410	0.078
Hardin	112.65	8.055	0.080	102.96	7.494	0.080
Harrison	110.70	8.050	0.081	106.11	7.846	0.084
Henry	122.05	9.036	0.094	102.03	7.283	0.075
Howard	116.81	8.428	0.085	103.24	7.716	0.088
Humboldt	112.29	8.105	0.079	103.50	7.418	0.077
Ida	112.27	8.086	0.082	103.66	7.809	0.095
Iowa	121.41	8.806	0.089	103.40	7.566	0.082
Jackson	112.85	8.161	0.083	104.47	8.002	0.094
Jasper	113.38	8.233	0.083	102.27	7.354	0.082
Jefferson	125.48	9.255	0.095	105.68	8.100	0.109
Johnson	116.67	8.392	0.086	104.22	7.560	0.079
Jones	115.54	8.422	0.086	103.13	7.564	0.080
Keokuk	127.25	9.454	0.097	103.83	7.739	0.088
Kossuth	110.95	7.939	0.079	101.97	7.205	0.073
Lee	133.71	10.039	0.106	104.06	7.562	0.081
Linn	116.71	8.393	0.085	102.12	7.398	0.081
Louisa	120.25	8.825	0.091	101.76	7.243	0.076
Lucas	125.03	9.265	0.097	104.06	7.881	0.086
Lyon	112.92	8.094	0.080	104.82	8.007	0.104
Madison	118.62	8.689	0.089	104.44	7.723	0.084
Mahaska	122.21	8.969	0.092	103.31	7.677	0.085
Marion	120.63	8.785	0.089	103.57	7.961	0.124
Marshall	115.74	8.276	0.082	99.32	6.904	0.070
Mills	116.21	8.405	0.085	105.18	8.130	0.111
Mitchell	114.82	8.207	0.082	103.84	7.759	0.092

Table A.3: Model 2 estimation results for σ_i^2 for Monona – Wright counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Monona	110.95	7.947	0.079	104.93	7.838	0.090
Monroe	123.37	9.157	0.096	104.58	8.075	0.108
Montgomery	114.73	8.157	0.083	103.83	7.639	0.082
Muscatine	115.49	8.400	0.084	102.52	7.383	0.078
O Brien	109.98	7.802	0.078	102.95	7.690	0.099
Osceola	113.34	8.190	0.081	102.32	7.459	0.080
Page	119.85	8.839	0.088	104.91	7.965	0.094
Palo Alto	112.02	7.984	0.082	102.64	7.237	0.074
Plymouth	113.54	8.071	0.080	105.75	8.056	0.097
Pocahontas	115.29	8.315	0.084	103.19	7.377	0.078
Polk	112.18	8.039	0.082	104.87	7.841	0.089
Pottawattamie	116.74	8.448	0.085	104.45	7.722	0.085
Poweshiek	114.41	8.320	0.085	104.05	7.835	0.090
Ringgold	130.00	9.875	0.105	106.64	8.250	0.109
Sac	113.05	8.114	0.081	104.37	7.785	0.088
Scott	114.47	8.419	0.086	102.51	7.348	0.077
Shelby	111.97	8.005	0.080	103.22	7.774	0.094
Sioux	108.82	7.741	0.078	103.27	7.647	0.088
Story	118.72	8.576	0.088	100.74	7.138	0.071
Tama	107.68	7.613	0.075	102.60	7.556	0.088
Taylor	123.84	9.252	0.097	106.08	7.944	0.096
Union	122.93	9.114	0.093	105.84	8.268	0.114
Van Buren	127.51	9.435	0.097	104.80	8.283	0.136
Wapello	122.33	9.084	0.093	104.18	8.219	0.217
Warren	120.93	8.939	0.093	105.90	8.285	0.111
Washington	121.84	8.979	0.095	103.44	7.615	0.085
Wayne	127.58	9.421	0.099	105.37	7.882	0.090
Webster	114.44	8.210	0.083	104.78	7.776	0.089
Winnebago	111.41	8.061	0.083	103.92	7.482	0.076
Winneshiek	109.08	7.808	0.077	101.68	7.277	0.075
Woodbury	111.25	8.092	0.081	106.19	8.008	0.088
Worth	111.71	8.079	0.080	103.26	7.616	0.086
Wright	111.73	7.958	0.080	103.79	7.521	0.077

Table A.4: Model 2 estimation results for δ_i for Adair – Fayette counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Adair	0.0452	0.0130	0.0002	0.0778	0.0435	0.0016
Adams	0.0486	0.0136	0.0002	0.0727	0.0406	0.0014
Allamakee	0.0343	0.0101	0.0001	0.0252	0.0090	0.0002
Appanoose	0.0481	0.0135	0.0002	0.0566	0.0225	0.0006
Audubon	0.0402	0.0116	0.0001	0.1263	0.0728	0.0016
Benton	0.0377	0.0111	0.0001	0.0761	0.0514	0.0017
Black Hawk	0.0376	0.0109	0.0001	0.0322	0.0122	0.0002
Boone	0.0468	0.0132	0.0002	0.0345	0.0114	0.0002
Bremer	0.0375	0.0106	0.0001	0.0319	0.0160	0.0004
Buchanan	0.0364	0.0108	0.0001	0.0327	0.0108	0.0002
Buena Vista	0.0386	0.0112	0.0001	0.0362	0.0125	0.0002
Butler	0.0349	0.0099	0.0001	0.0382	0.0245	0.0009
Calhoun	0.0388	0.0115	0.0001	0.0409	0.0246	0.0008
Carroll	0.0423	0.0120	0.0001	0.1247	0.0963	0.0027
Cass	0.0446	0.0128	0.0002	0.0585	0.0335	0.0012
Cedar	0.0411	0.0117	0.0002	0.0347	0.0126	0.0002
Cerro Gordo	0.0357	0.0102	0.0001	0.0369	0.0275	0.0009
Cherokee	0.0387	0.0109	0.0001	0.0370	0.0140	0.0002
Chickasaw	0.0394	0.0109	0.0001	0.1127	0.0765	0.0024
Clarke	0.0514	0.0147	0.0002	0.0836	0.0449	0.0017
Clay	0.0410	0.0118	0.0001	0.0336	0.0122	0.0002
Clayton	0.0341	0.0100	0.0001	0.0284	0.0096	0.0001
Clinton	0.0386	0.0112	0.0001	0.0458	0.0287	0.0013
Crawford	0.0417	0.0120	0.0001	0.0707	0.0318	0.0006
Dallas	0.0424	0.0123	0.0002	0.0757	0.0533	0.0017
Davis	0.0522	0.0146	0.0002	0.0693	0.0325	0.0010
Decatur	0.0509	0.0148	0.0002	0.0705	0.0362	0.0012
Delaware	0.0378	0.0111	0.0001	0.0327	0.0184	0.0006
Des Moines	0.0430	0.0121	0.0002	0.0348	0.0128	0.0002
Dickinson	0.0429	0.0120	0.0002	0.0544	0.0474	0.0017
Dubuque	0.0357	0.0102	0.0001	0.0380	0.0195	0.0006
Emmet	0.0414	0.0115	0.0001	0.0341	0.0157	0.0004
Fayette	0.0338	0.0098	0.0001	0.0283	0.0092	0.0001

Table A.5: Model 2 estimation results for δ_i for Floyd – Mitchell counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Floyd	0.0403	0.0107	0.0001	0.1213	0.0951	0.0031
Franklin	0.0360	0.0106	0.0001	0.0329	0.0128	0.0003
Fremont	0.0471	0.0132	0.0002	0.0782	0.0449	0.0014
Greene	0.0460	0.0133	0.0002	0.0386	0.0179	0.0005
Grundy	0.0349	0.0101	0.0001	0.0346	0.0222	0.0005
Guthrie	0.0407	0.0117	0.0001	0.0717	0.0523	0.0010
Hamilton	0.0415	0.0121	0.0002	0.0620	0.0656	0.0025
Hancock	0.0353	0.0102	0.0001	0.0327	0.0127	0.0003
Hardin	0.0374	0.0108	0.0001	0.0382	0.0172	0.0005
Harrison	0.0409	0.0116	0.0001	0.0554	0.0195	0.0003
Henry	0.0444	0.0124	0.0002	0.0337	0.0129	0.0003
Howard	0.0440	0.0119	0.0001	0.1468	0.0878	0.0021
Humboldt	0.0373	0.0111	0.0001	0.0359	0.0125	0.0002
Ida	0.0409	0.0116	0.0001	0.0489	0.0239	0.0008
Iowa	0.0449	0.0126	0.0002	0.0521	0.0627	0.0019
Jackson	0.0382	0.0111	0.0001	0.1086	0.0930	0.0021
Jasper	0.0404	0.0118	0.0001	0.0335	0.0129	0.0003
Jefferson	0.0473	0.0136	0.0002	0.0510	0.0211	0.0007
Johnson	0.0402	0.0117	0.0001	0.0404	0.0164	0.0004
Jones	0.0392	0.0112	0.0001	0.0360	0.0148	0.0004
Keokuk	0.0491	0.0140	0.0002	0.0432	0.0253	0.0011
Kossuth	0.0357	0.0106	0.0001	0.0288	0.0090	0.0001
Lee	0.0536	0.0142	0.0002	0.0380	0.0141	0.0003
Linn	0.0401	0.0118	0.0001	0.0347	0.0176	0.0005
Louisa	0.0426	0.0120	0.0001	0.0342	0.0117	0.0002
Lucas	0.0487	0.0139	0.0002	0.0902	0.0393	0.0009
Lyon	0.0449	0.0124	0.0001	0.1019	0.0505	0.0010
Madison	0.0461	0.0133	0.0002	0.0866	0.0501	0.0015
Mahaska	0.0461	0.0133	0.0002	0.0448	0.0379	0.0015
Marion	0.0480	0.0137	0.0002	0.0431	0.0303	0.0016
Marshall	0.0425	0.0119	0.0001	0.0255	0.0083	0.0001
Mills	0.0450	0.0129	0.0002	0.0805	0.0405	0.0019
Mitchell	0.0422	0.0117	0.0001	0.1319	0.0912	0.0031

Table A.6: Model 2 estimation results for δ_i for Monona – Wright counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Monona	0.0425	0.0120	0.0001	0.0527	0.0286	0.0010
Monroe	0.0472	0.0135	0.0002	0.0726	0.0470	0.0017
Montgomery	0.0429	0.0124	0.0001	0.0370	0.0134	0.0002
Muscatine	0.0401	0.0114	0.0001	0.0364	0.0126	0.0002
O Brien	0.0396	0.0112	0.0001	0.0812	0.1037	0.0048
Osceola	0.0429	0.0119	0.0001	0.0442	0.0375	0.0020
Page	0.0481	0.0137	0.0002	0.0717	0.0325	0.0010
Palo Alto	0.0375	0.0107	0.0001	0.0286	0.0092	0.0001
Plymouth	0.0436	0.0119	0.0001	0.0727	0.0302	0.0006
Pocahontas	0.0405	0.0116	0.0001	0.0320	0.0104	0.0002
Polk	0.0400	0.0118	0.0001	0.0606	0.0574	0.0017
Pottawattamie	0.0465	0.0128	0.0002	0.0400	0.0238	0.0011
Poweshiek	0.0392	0.0115	0.0001	0.0476	0.0310	0.0011
Ringgold	0.0557	0.0163	0.0002	0.0719	0.0339	0.0009
Sac	0.0419	0.0120	0.0001	0.0524	0.0345	0.0012
Scott	0.0407	0.0116	0.0001	0.0424	0.0171	0.0003
Shelby	0.0416	0.0120	0.0001	0.0503	0.0233	0.0006
Sioux	0.0403	0.0110	0.0001	0.0495	0.0253	0.0007
Story	0.0460	0.0132	0.0002	0.0304	0.0097	0.0001
Tama	0.0338	0.0099	0.0001	0.0283	0.0130	0.0004
Taylor	0.0516	0.0144	0.0002	0.0934	0.0447	0.0009
Union	0.0495	0.0144	0.0002	0.0744	0.0406	0.0016
Van Buren	0.0495	0.0141	0.0002	0.0516	0.0306	0.0012
Wapello	0.0443	0.0131	0.0002	0.0506	0.0253	0.0015
Warren	0.0488	0.0140	0.0002	0.0920	0.0603	0.0038
Washington	0.0452	0.0128	0.0002	0.0389	0.0150	0.0003
Wayne	0.0493	0.0138	0.0002	0.0549	0.0356	0.0014
Webster	0.0391	0.0115	0.0001	0.0695	0.0660	0.0022
Winnebago	0.0372	0.0106	0.0001	0.0388	0.0261	0.0008
Winneshiek	0.0349	0.0099	0.0001	0.0346	0.0156	0.0004
Woodbury	0.0419	0.0120	0.0001	0.0539	0.0191	0.0003
Worth	0.0371	0.0105	0.0001	0.0415	0.0289	0.0010
Wright	0.0365	0.0108	0.0001	0.0341	0.0112	0.0002

Table A.7: Model 2 estimation results for γ_i for Adair – Fayette counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Adair	0.0022	0.00072	0.00001	0.0028	0.00114	0.00003
Adams	0.0023	0.00077	0.00001	0.0028	0.00110	0.00003
Allamakee	0.0018	0.00062	0.00001	0.0016	0.00059	0.00001
Appanoose	0.0022	0.00073	0.00001	0.0024	0.00091	0.00002
Audubon	0.0021	0.00070	0.00001	0.0037	0.00146	0.00003
Benton	0.0019	0.00066	0.00001	0.0026	0.00114	0.00003
Black Hawk	0.0018	0.00065	0.00001	0.0018	0.00065	0.00001
Boone	0.0022	0.00075	0.00001	0.0019	0.00068	0.00001
Bremer	0.0018	0.00064	0.00001	0.0018	0.00065	0.00001
Buchanan	0.0018	0.00064	0.00001	0.0018	0.00066	0.00001
Buena Vista	0.0020	0.00068	0.00001	0.0020	0.00074	0.00001
Butler	0.0018	0.00063	0.00001	0.0019	0.00075	0.00002
Calhoun	0.0020	0.00071	0.00001	0.0021	0.00086	0.00002
Carroll	0.0021	0.00071	0.00001	0.0035	0.00169	0.00004
Cass	0.0022	0.00073	0.00001	0.0024	0.00098	0.00003
Cedar	0.0021	0.00075	0.00001	0.0021	0.00077	0.00001
Cerro Gordo	0.0018	0.00063	0.00001	0.0019	0.00077	0.00002
Cherokee	0.0021	0.00071	0.00001	0.0022	0.00082	0.00001
Chickasaw	0.0019	0.00066	0.00001	0.0033	0.00151	0.00004
Clarke	0.0023	0.00077	0.00001	0.0030	0.00124	0.00004
Clay	0.0021	0.00072	0.00001	0.0020	0.00071	0.00001
Clayton	0.0018	0.00062	0.00001	0.0017	0.00062	0.00001
Clinton	0.0020	0.00069	0.00001	0.0022	0.00092	0.00003
Crawford	0.0022	0.00073	0.00001	0.0028	0.00104	0.00002
Dallas	0.0021	0.00071	0.00001	0.0027	0.00121	0.00003
Davis	0.0023	0.00076	0.00001	0.0028	0.00108	0.00002
Decatur	0.0022	0.00077	0.00001	0.0027	0.00104	0.00002
Delaware	0.0018	0.00065	0.00001	0.0018	0.00071	0.00001
Des Moines	0.0021	0.00070	0.00001	0.0020	0.00073	0.00001
Dickinson	0.0021	0.00070	0.00001	0.0023	0.00106	0.00003
Dubuque	0.0018	0.00061	0.00001	0.0019	0.00072	0.00001
Emmet	0.0019	0.00064	0.00001	0.0018	0.00069	0.00001
Fayette	0.0018	0.00060	0.00001	0.0017	0.00061	0.00001

Table A.8: Model 2 estimation results for γ_i for Floyd – Mitchell counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Floyd	0.0020	0.00064	0.00001	0.0035	0.00176	0.00005
Franklin	0.0018	0.00064	0.00001	0.0018	0.00066	0.00001
Fremont	0.0023	0.00076	0.00001	0.0028	0.00120	0.00003
Greene	0.0022	0.00076	0.00001	0.0021	0.00078	0.00001
Grundy	0.0018	0.00063	0.00001	0.0018	0.00072	0.00001
Guthrie	0.0021	0.00070	0.00001	0.0027	0.00108	0.00002
Hamilton	0.0020	0.00067	0.00001	0.0024	0.00123	0.00004
Hancock	0.0018	0.00063	0.00001	0.0018	0.00066	0.00001
Hardin	0.0018	0.00064	0.00001	0.0019	0.00071	0.00001
Harrison	0.0022	0.00073	0.00001	0.0026	0.00090	0.00001
Henry	0.0021	0.00070	0.00001	0.0020	0.00075	0.00001
Howard	0.0022	0.00073	0.00001	0.0039	0.00168	0.00004
Humboldt	0.0019	0.00065	0.00001	0.0019	0.00069	0.00001
Ida	0.0022	0.00072	0.00001	0.0025	0.00097	0.00002
Iowa	0.0021	0.00071	0.00001	0.0023	0.00109	0.00003
Jackson	0.0019	0.00066	0.00001	0.0034	0.00156	0.00004
Jasper	0.0020	0.00069	0.00001	0.0019	0.00070	0.00001
Jefferson	0.0021	0.00072	0.00001	0.0023	0.00085	0.00002
Johnson	0.0020	0.00068	0.00001	0.0021	0.00076	0.00001
Jones	0.0019	0.00064	0.00001	0.0019	0.00068	0.00001
Keokuk	0.0022	0.00072	0.00001	0.0021	0.00083	0.00002
Kossuth	0.0018	0.00064	0.00001	0.0016	0.00058	0.00001
Lee	0.0023	0.00074	0.00001	0.0021	0.00076	0.00001
Linn	0.0019	0.00065	0.00001	0.0018	0.00072	0.00001
Louisa	0.0020	0.00069	0.00001	0.0019	0.00070	0.00001
Lucas	0.0022	0.00074	0.00001	0.0030	0.00112	0.00002
Lyon	0.0025	0.00079	0.00001	0.0034	0.00133	0.00002
Madison	0.0022	0.00073	0.00001	0.0030	0.00122	0.00003
Mahaska	0.0021	0.00070	0.00001	0.0021	0.00092	0.00003
Marion	0.0022	0.00074	0.00001	0.0021	0.00086	0.00003
Marshall	0.0020	0.00068	0.00001	0.0017	0.00062	0.00001
Mills	0.0022	0.00075	0.00001	0.0030	0.00124	0.00005
Mitchell	0.0021	0.00071	0.00001	0.0037	0.00176	0.00006

Table A.9: Model 2 estimation results for γ_i for Monona – Wright counties

County	Model 1			Model 2		
	Post. Mean	PSD	NE	Post. Mean	PSD	NE
Monona	0.0024	0.00079	0.00001	0.0025	0.00103	0.00002
Monroe	0.0021	0.00073	0.00001	0.0027	0.00113	0.00003
Montgomery	0.0022	0.00072	0.00001	0.0020	0.00074	0.00001
Muscatine	0.0020	0.00070	0.00001	0.0021	0.00077	0.00001
O Brien	0.0022	0.00073	0.00001	0.0031	0.00175	0.00008
Osceola	0.0022	0.00074	0.00001	0.0022	0.00101	0.00004
Page	0.0023	0.00076	0.00001	0.0030	0.00113	0.00003
Palo Alto	0.0019	0.00065	0.00001	0.0017	0.00063	0.00001
Plymouth	0.0023	0.00075	0.00001	0.0028	0.00105	0.00002
Pocahontas	0.0020	0.00068	0.00001	0.0019	0.00069	0.00001
Polk	0.0020	0.00069	0.00001	0.0024	0.00109	0.00003
Pottawattamie	0.0023	0.00074	0.00001	0.0020	0.00085	0.00003
Poweshiek	0.0019	0.00067	0.00001	0.0022	0.00088	0.00002
Ringgold	0.0024	0.00083	0.00001	0.0028	0.00107	0.00002
Sac	0.0022	0.00073	0.00001	0.0024	0.00101	0.00002
Scott	0.0022	0.00076	0.00001	0.0023	0.00089	0.00001
Shelby	0.0022	0.00072	0.00001	0.0024	0.00088	0.00002
Sioux	0.0024	0.00077	0.00001	0.0025	0.00100	0.00002
Story	0.0021	0.00070	0.00001	0.0018	0.00065	0.00001
Tama	0.0018	0.00062	0.00001	0.0017	0.00066	0.00001
Taylor	0.0024	0.00079	0.00001	0.0033	0.00123	0.00002
Union	0.0023	0.00077	0.00001	0.0028	0.00115	0.00003
Van Buren	0.0022	0.00073	0.00001	0.0024	0.00099	0.00003
Wapello	0.0020	0.00069	0.00001	0.0022	0.00088	0.00003
Warren	0.0022	0.00076	0.00001	0.0030	0.00140	0.00007
Washington	0.0021	0.00072	0.00001	0.0021	0.00076	0.00001
Wayne	0.0022	0.00074	0.00001	0.0024	0.00099	0.00003
Webster	0.0019	0.00067	0.00001	0.0025	0.00128	0.00004
Winnebago	0.0019	0.00065	0.00001	0.0019	0.00076	0.00002
Winneshiek	0.0018	0.00062	0.00001	0.0018	0.00070	0.00001
Woodbury	0.0024	0.00077	0.00001	0.0026	0.00096	0.00002
Worth	0.0019	0.00065	0.00001	0.0019	0.00080	0.00002
Wright	0.0018	0.00064	0.00001	0.0018	0.00064	0.00001

Appendix B

Spatial Mixtures Model Graphical Results

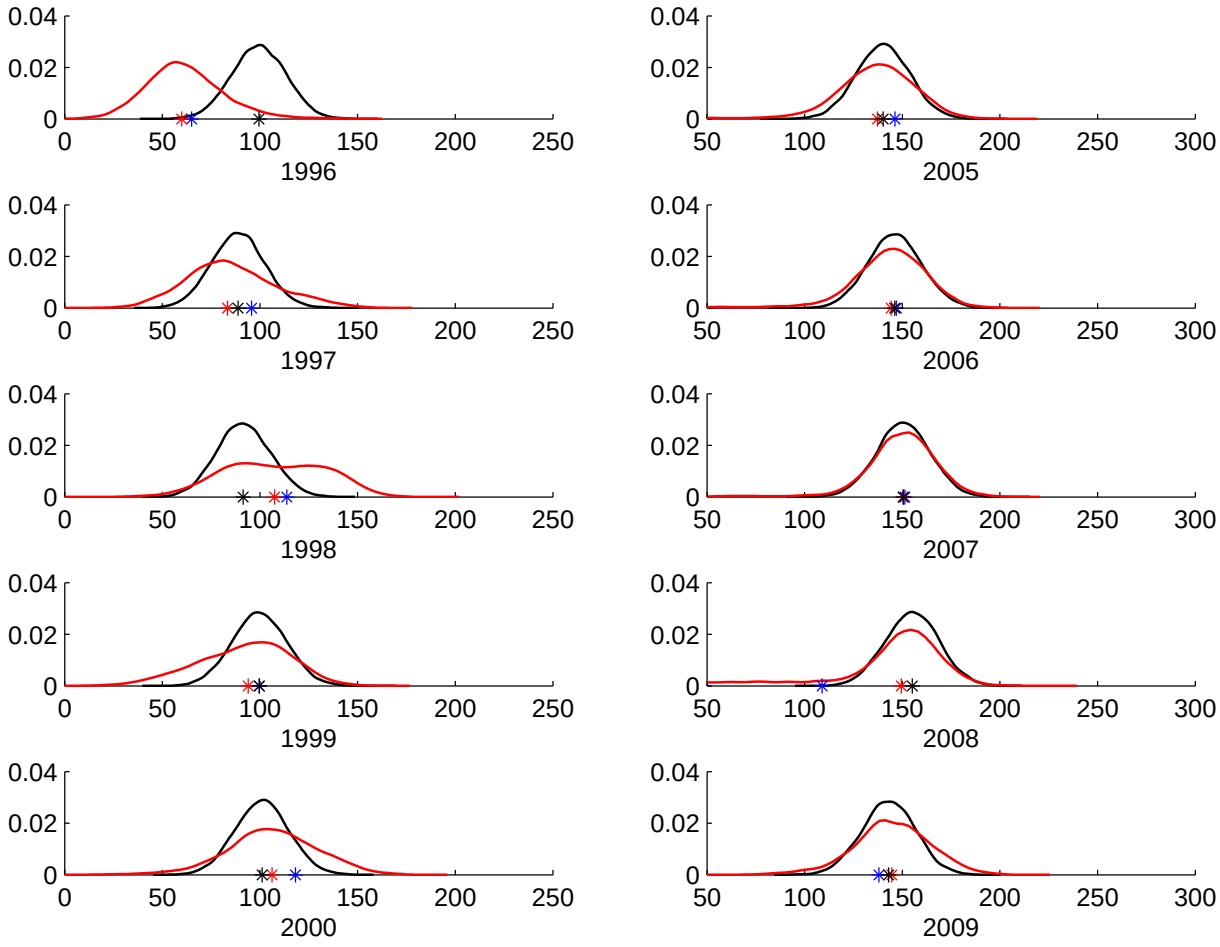


Figure B.1: Smoothed predicted yield densities and observed yields for Clarke county. Left panel: in sample prediction $1996 - 2000$. Right panel: out of sample prediction $2005 - 2009$. Black color: Normal model. Red color: Spatial mixture model. Blue color: Actual observations.

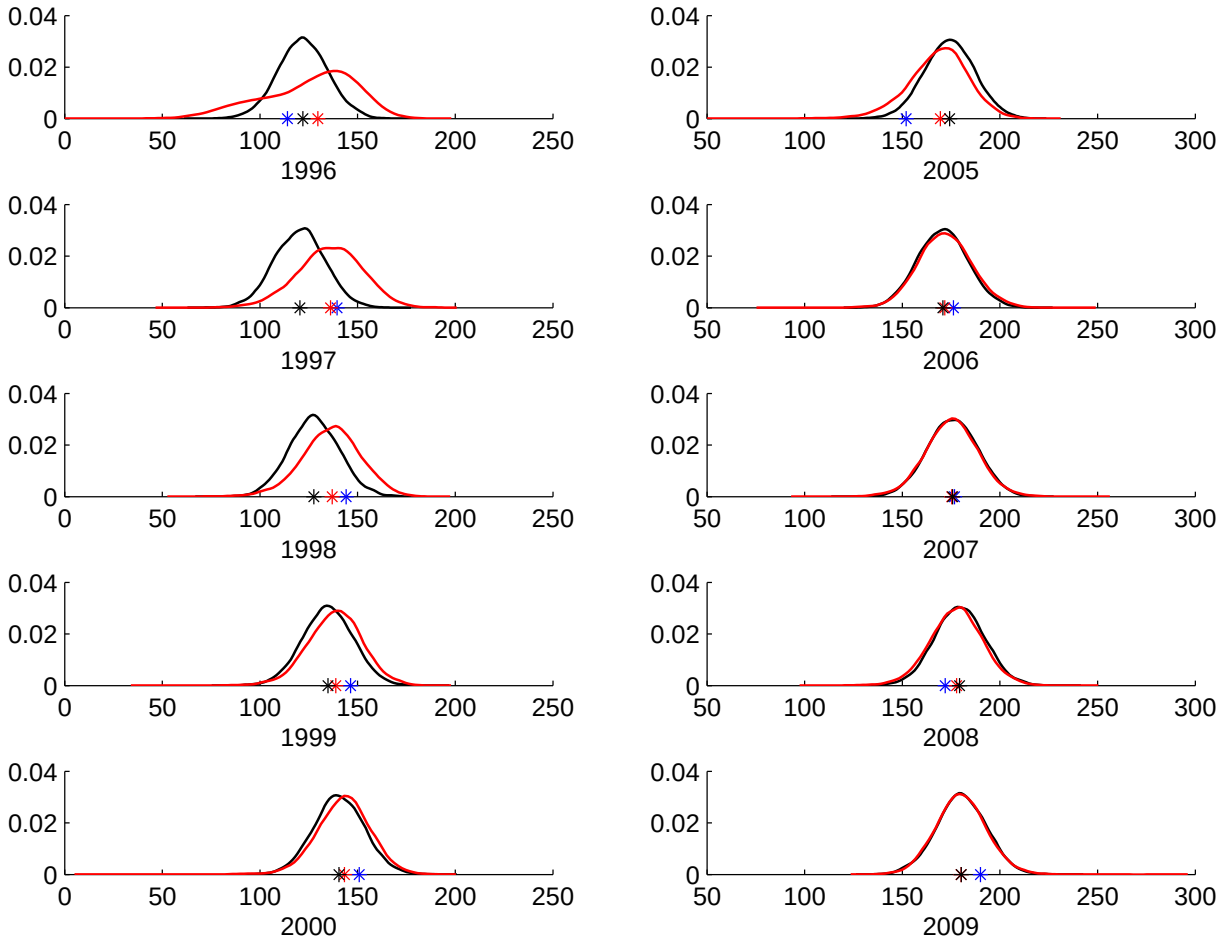


Figure B.2: Smoothed predicted yield densities and observed yields for Ida county. Left panel: in sample prediction $1996 - 2000$. Right panel: out of sample prediction $2005 - 2009$. Black color: Normal model. Red color: Spatial mixture model. Blue color: Actual observations.

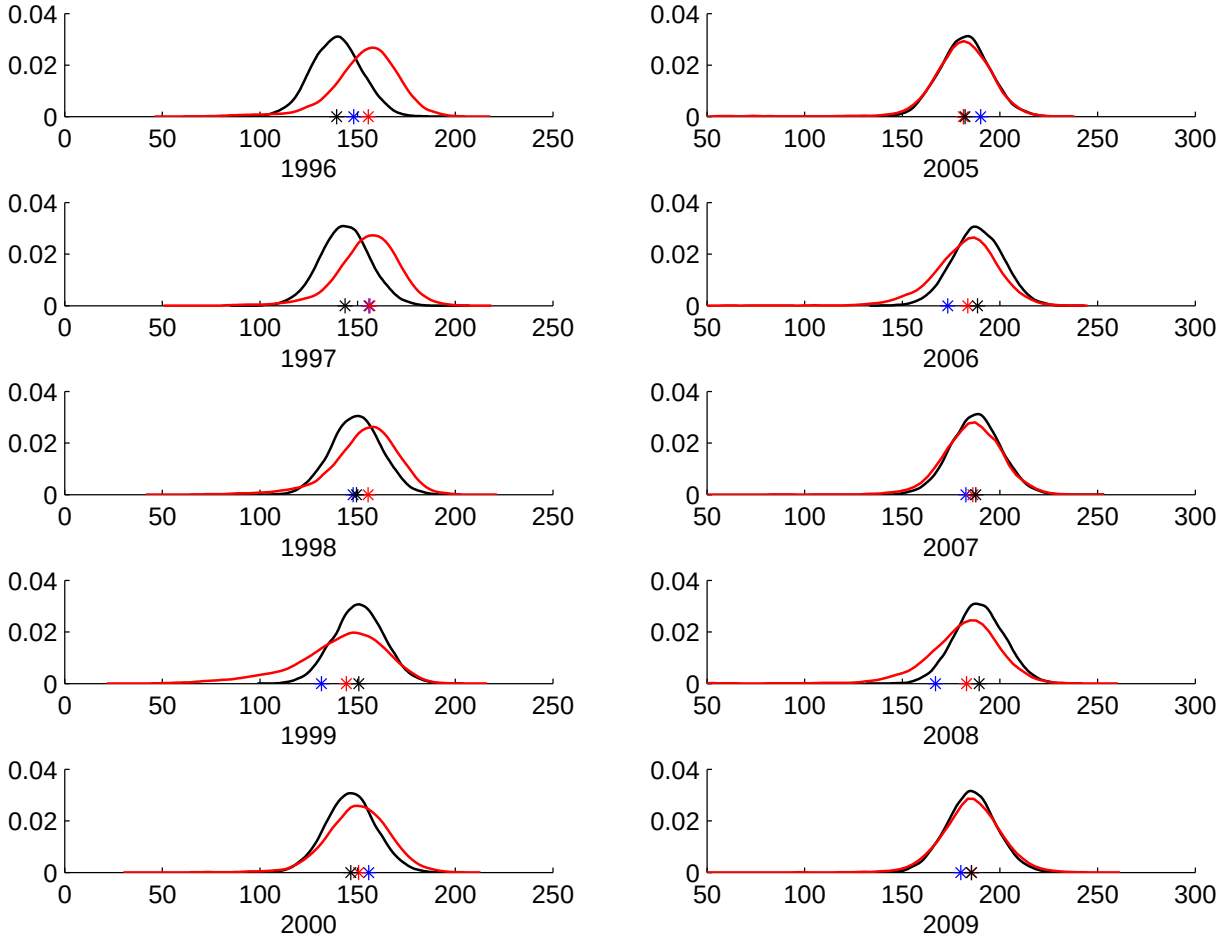


Figure B.3: Smoothed predicted yield densities and observed yields for Webster county. Left panel: in sample prediction $1996 - 2000$. Right panel: out of sample prediction $2005 - 2009$. Black color: Normal model. Red color: Spatial mixture model. Blue color: Actual observations.

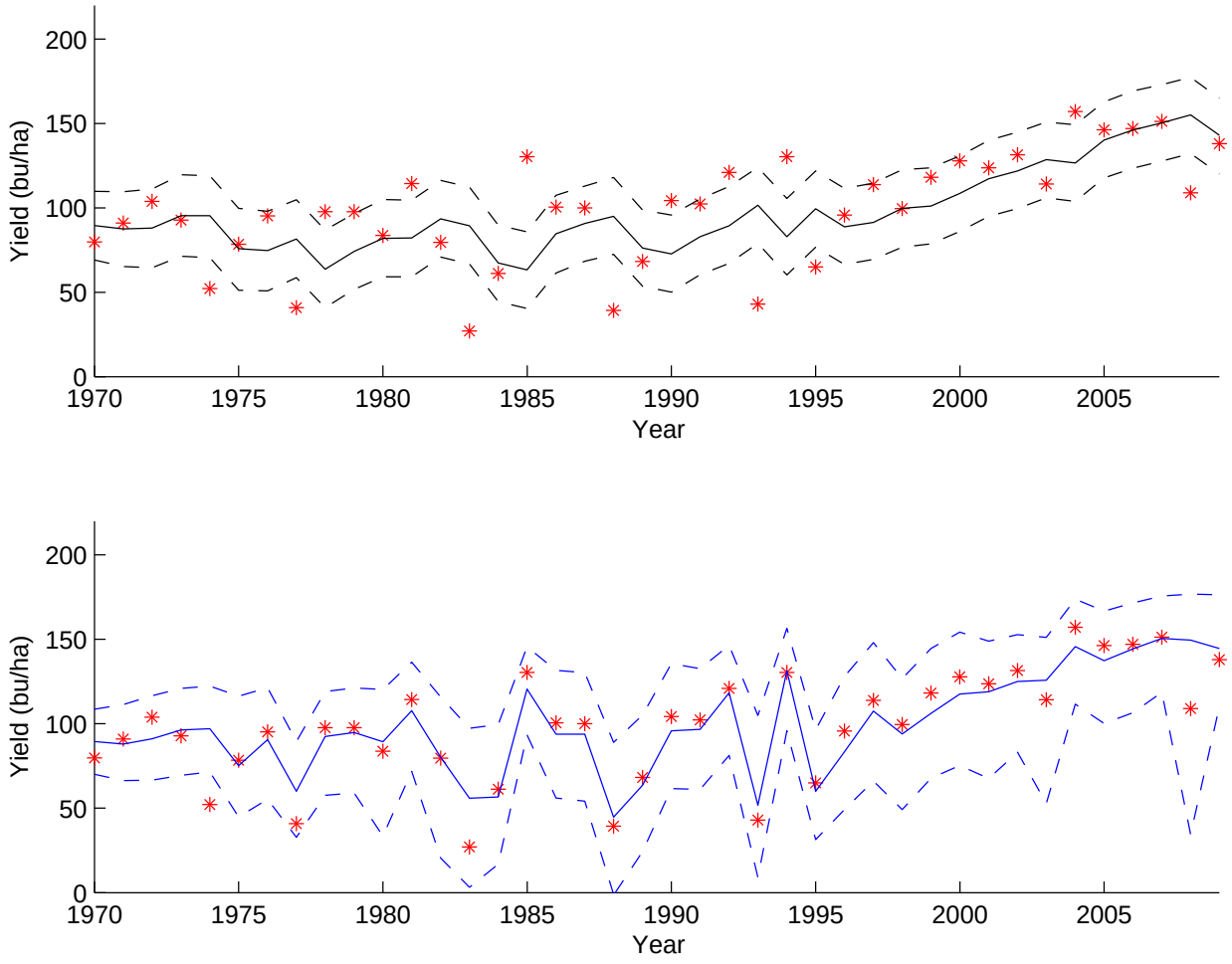


Figure B.4: One step in-sample (1970 – 2004) and out-of-sample (2005 – 2009) yield prediction for Clarke county. Upper panel: Normal model. Lower panel: Spatial mixture model. Solid lines: median predicted yield. Dashed lines: lower and upper bound of 90% credible prediction interval.

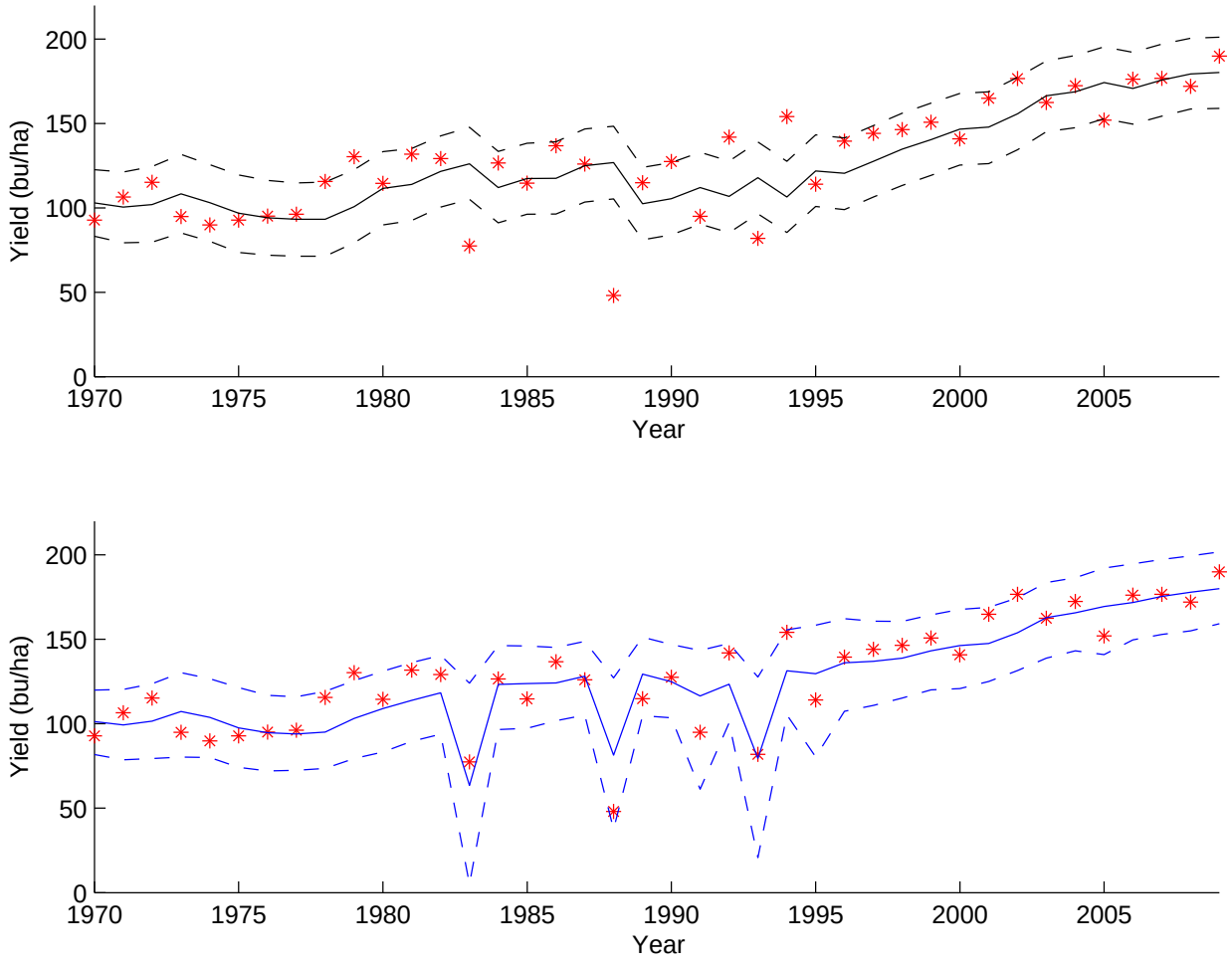


Figure B.5: One step in-sample (1970 – 2004) and out-of-sample (2005 – 2009) yield prediction for Ida county. Upper panel: Normal model. Lower panel: Spatial mixture model. Solid lines: median predicted yield. Dashed lines: lower and upper bound of 90% credible prediction interval.

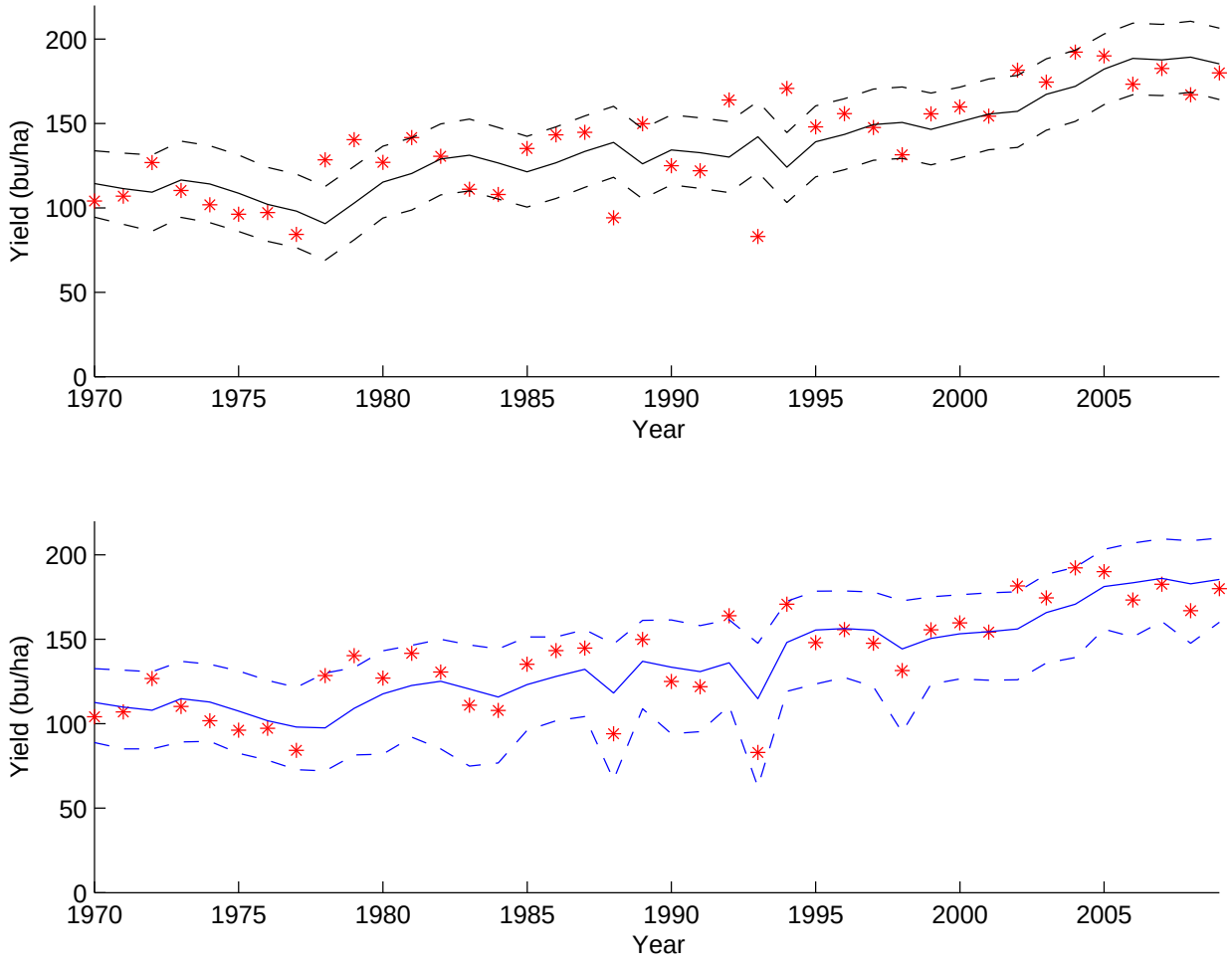


Figure B.6: One step in-sample (1970 – 2004) and out-of-sample (2005 – 2009) yield prediction for Webster county. Upper panel: Normal model. Lower panel: Spatial mixture model. Solid lines: median predicted yield. Dashed lines: lower and upper bound of 90% credible prediction interval.