January 2010

Bayesian Multilevel Analysis of Binary Time-Series Cross-Sectional Data in Political Economy

Xun Pang
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Political Science

Dissertation Examination Committee:
Andrew D. Martin, Chair
Jeff Gill, Chair
Nathan Jensen
Ryan Moore
Edward Greenberg
Nan Lin

Bayesian Multilevel Analysis of Binary Time-Series Cross-Sectional Data in Political Economy

by

Xun Pang

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

March 2010

St. Louis, Missouri

to Wei

ABSTRACT OF THE DISSERTATION

Bayesian Multilevel Analysis of Binary Time-Series Cross-Sectional Data in Political Economy

by

Xun Pang

Doctor of Philosophy in Political Science

Washington University in St. Louis, 2010

Andrew D. Martin and Jeff Gill, Co-Chairs

In this dissertation project, I propose a Bayesian generalized linear multilevel model with pth order autoregressive errors (GLMM-AR(p)) for modeling inter-temporal dependence, con-temporary correlation, and heterogeneity of unbalanced binary Time-Series Cross-Sectional data. The model includes two unnested sources of clustering in the unit- and time-dimensions for analyzing heterogeneities and contemporal correlation which are salient in the era of globalization. Group-level variations are further explained with unit- and time-specific characteristics. For handling dynamics in politics and political economy, I apply the autoregressive error specification to analyze serial correlation which may not be fully captured by the selected covariates.

Two applications on civil war and sovereign default demonstrate how the proposed model controls for multiple potential confounders. It also improves reliability of statistical inferences and helps forecasts by more efficiently using the information in data. The first application focuses on the causal relationship between ethnic minority rule and civil war onset. The GLMM-AR(p) model helps study those background factors which affect the relationship under investigation. The second applied study considers how regime duration affects sovereign default conditional on regime type by putting the national policy-making regarding repaying external debt into the international

context. To model the heterogeneous vulnerability or sensitivity of the developing countries to global shocks, I extend the GLMM-AR(p) model to analyze time-specific unit-varying effects.

ACKNOWLEDGEMENTS

Although only a few names appear on the cover of this dissertation, there are many people to whom I owe my gratitude, because it was they who made this dissertation project possible and turned my graduate experience into one that I will cherish forever.

My deepest gratitude is to my dissertation chairs, Andrew D. Martin and Jeff Gill. I have been amazingly fortunate to have Andrew as my main advisor. As one of the leading scholars in the discipline, he gave me the most important and helpful advice and guidances in the whole process of this dissertation research. I also benefited a lot from his great mentorship in term of professional development, teaching skills, and even presentation style. His support and confidence on me helped me get through many tough times over the past five years. Jeff 's great passion for Bayesian statistics and Markov Chain Monte Carlo is always inspiring and encouraging me. I am grateful to him for carefully reading and commenting on countless revisions of this manuscript. I am also thankful to him for encouraging the use of correct tools for scientific research and consistent notation in my writings. My another advisor, Nathan Jensen, is always there to listen and give advice. He has contributed a lot to the production of this dissertation with his patience and insightful suggestions. I also want to say special thanks to Edward Greenberg who opened the door of Bayesian statistics to me. It was also he who encouraged me to design my first MCMC algorithm and programme it from scratch. A number of other Washington University faculty have been helpful to me along the way, including (but not limited to) Sid Chib, Ryan Moore, Robert Walker, Dawn Brancati, Nan Lin, Sebastian Galiani, James Morley. Thank you also to Christina L. Boyd (Assistant Professor at University at Buffalo, SUNY) and Jong Hee Park (Assistant Professor at The University of Chicago)

I am grateful to my many other graduate school friends and colleagues, including Gordon Arsenoff, Yael Shomer, Dan O'Neill, Noel Pereyra Johnston, Amanda Driscoll, Morgan Hazelton, Hong Min Park, Mariana Medina and Michael Malecki.

I gratefully acknowledge the support for this project by National Science Foundation Dissertation Improvement Grant SES-0918320, the Department of Political Science, the Center for Empirical Research in the Law, and the Center for Applied Statistics at Washington University.

I thank my parents who have always been supportive and wanted the best for me. I am also grateful to my former advisor and mentor, Qingguo Jia, at Peking University, Beijing, China. He encouraged me to go to graduate school in America and is always there to give support.

Finally, I want to thank Dr. Wei Xiong and dedicate this dissertation to him.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Time-series cross-sectional (TSCS) data are very important in political science, especially in the subfields of international relations and comparative politics. TSCS data are rich in structure: multiple units are repeatedly measured over time, and one of the major subjects under scientific investigation is dynamic evolution of those units. At the same time, observations are clustered in the spatial dimension with measurements equally spaced over time, which is not necessarily the case for other types of longitudinal data. In TSCS analysis, the spatial relationship among the units is substantively interesting, especially in the era of globalization (Franzse and Hays, 2007). TSCS data, belonging to the family of longitudinal data, also have spatial structure; therefore, they should be analyzed in a three-dimensional space (Gelman and Hill, 2006; Gill, 2007; Beck and Katz, 2007; Shor et al., 2007). The structure of TSCS data offers a good opportunity for both dynamic analysis and spatial relationship investigation, but it also implies multiple sources of correlation which confounds the causal relationships of research interest and raises several methodological challenges. However, in the literature on longitudinal data analysis, little effort has been made to model inter-temporal dependence, contemporary correlation and heterogeneity at the same time and in the non-linear framework. This dissertation project seeks to remedy this deficit.

Chapter 2, *Bayesian Generalized Linear Multilevel Model with AR(p) Errors* , proposes a very general model for modeling unbalanced binary Time-Series Cross-Sectional (TSCS) data by considering correlation in both the time and spatial dimensions. By controlling for heterogeneities in the two dimensions and modeling the dynamic error process, the proposed model handles the inefficiency and endogeneity problems resulting from the generic TSCS data structure. With the stationarity restriction on the error process, the model can also be used as a residual-based cointegration test on discrete TSCS data. Methodologically, to handle the model estimation difficulties, I develop an efficient Markov Chain Monte Carlo algorithm by orthogonalizing the errors with the Cholesky decomposition and adding an auxiliary variable. I also apply the parameter expansion method to further improve mixing and speed up convergence of the Markov chain. Simulated and empirical examples are used to assess the performance of the model and techniques.

Chapter 3, *Ethnic Minority Rule and Civil War Onset: How Much Background Factors and Dynamics Matter*, revisits the debate about the impact of ethnic minority rule (EMR) on civil war onset. To explain the variation of EMR's effect and to improve statistical prediction of civil war onset, this chapter applies the GLMM-AR(p) model to carefully handles multiple confounders caused by the TSCS structure of the civil war data and by the measures of EMR. The relationship between civil war and EMR is found to be affected by several background factors including regime stability and governance quality. Modeling the stochastic process of the errors dramatically improves forecasting, suggesting that information in the errors, which has often been neglected, is valuable for understanding the dynamics of civil war.

Finally, in Chapter 4, *Sovereign Default: Regime Type, Regime Duration, and Vulnerability to Global Shocks*, I extend the GLMM-AR(p) model with a multifactor specification to exam the regime-specific effect of regime duration on sovereign default by putting this national policy-making into its international context. The empirical findings include that regime duration has different meanings in anocracies and non-anocracies in terms of explaining sovereign default. Empirical evidence also suggests that shocks in the international system strongly affect national decision-making regarding sovereign default in the developing countries, and the impact of globalization varies widely from country to country.

Chapter 2

Bayesian Generalized Linear Multilevel Model with AR(p) Errors

2.1 Motivations

Time-series cross-sectional (TSCS) data are a very important type of data in political science, especially in the subfields of international relations and comparative politics. Political scientists have made a great effort to handle the methodological issues posed by the structural characteristics of such data[1]. TSCS data are rich in structure: multiple units are repeatedly measured over time, and one of the major

[1] There have been numerous great methodological discussions on TSCS data, such as Beck and Katz (1995), Beck and Katz (1996), Beck, Katz and Tucker (1998), Beck et al. (2002), only to name a very small portion of them in the political methodology literature. In 2007, *Political Analysis* published a special issue (volumn 15) on the substantive and methodological questions in TSCS data analysis.

subjects under scientific investigation is dynamic evolution of those units. At the same time, observations are clustered in the spatial dimension with measurements equally spaced over time, which is not necessarily the case for other types of longitudinal data. In TSCS analysis, the spatial relationship among the units is of research interest, especially in the era of globalization (Franzse and Hays, 2007). Hence, TSCS data, belonging to the family of longitudinal data, also have spatial structure; therefore, they should be analyzed in a three-dimensional space (Gelman and Hill, 2006; Gill, 2007; Beck and Katz, 2007; Shor et al., 2007). The structure of TSCS data offers a good opportunity for both dynamic analysis and spatial relationship investigation, but it also implies multiple sources of correlation which confounds the causal relationship of research interest and raises several methodological challenges.

The correlation design of TSCS data is based on both the time and spatial dimensions of hierarchy. Analyzing heterogeneities in both dimensions is easy with multilevel modeling in classical and Bayesian frameworks (Schafer and Yucel, 2002; Renard, Molenberghs and Geys, 2004; Molenberghs and Verbeke, 2005, Chapter 14 and Chapter 22) by using well-developed softwares, such as the `lme4` `R` package, `JAGS/BUGS` and the `SAS` program. However, when serial correlation is considered at the same time, computational complexity dramatically increases, especially for categorical responses and nonlinear model setups. Yet, directly modeling correlated errors is necessary for reliable statistical references and forecasts, and there do not exist easy alternatives. Including lagged values of the response variable (observed or latent) or/and lagged explanatory variables (LDVs/LIDVs) has been often recommended and implemented (Beck, Katz and Tucker, 1998; Beck et al., 2002), but LDVs and LIDVs approaches cannot serve as substitutes for directly modeling serial dependence in the error term (Chib and Jeliazkov, 2006). First, there are multiple sources of intertemporal correlation, including dynamics, unmodeled heterogeneities, omitted variables, measurement error, and so on. Dynamics can be partly captured by LDVs/LIDVs, but this cannot tell whether the errors are still correlated because of other sources of correlation. Therefore, it is necessary to check and correct serial correlation by directly and adequately modeling the dynamic error process even after LDVs or LIDVs are included. Second, unlike in linear models, in generalized linear models the lagged response variable cannot introduce the same covariance matrix as the autoregressive errors do. Hence, including lagged values of the observed response variable (the state-dependence specification as often referred to) does not directly address serially correlated errors, although it can partially control for serial dependence. Moreover, lagged values of the latent response variable should be used with caution: including lagged values into the structural form have to be justified with a "causal" interpretation; if the errors are still correlated, including a lagged latent response variable invites endogeneity; and it often reduces the sample size if the first P values are treated as exogenous. In TSCS data, the loss is $N \times P$ observations which is not a trivial reduction of the sample size (Wilson and Butler, 2007; Skrondal and Rabe-Hesketh, 2008). Finally, with all those pitfalls, estimation of generalized linear models with lagged values of the latent response variable is not necessarily easier than estimating models with autoregressive error specifications.

With binary (or other categorical) TSCS responses and serially correlated errors, the major methodological challenges arise in the following three areas: first, the overall error structure has to be decomposed into three parts, i.e., the unit-level errors, the time-level errors, and the idiosyncratic (individual-level) errors, and an appropriate model specification should be able to analyze the three parts at the same time; second, with the covariance matrix of the idiosyncratic errors not as a diagonal matrix, the likelihood function is intractable and conventional data augmentation methods for sampling the latent responses are inefficient; third, the complication caused by correlated errors is further exacerbated by unbalanced data structures which are common in observational studies. An unbalanced data structure leads to a complicated covariance matrix, and makes it particularly difficult to simulate the time-level errors, because widely-employed approaches fail to orthogonalize the errors in the spatial dimension (Harvey, 1981; Chib, 1993; Mueller and Czado, 2005).

This chapter discusses Bayesian techiniques for analyzing the intertemporal and contemporal correlation of binary TSCS data. The serial dependence consists of an everlasting part of unobserved unit heterogeneity and a time-varying part of serially correlated errors. The spatial dependence results from time-specific common shocks which can be observed or unobserved. In the current research, I do not consider the contemporal correlation caused by interactions among the units, which is often analyzed with spatial dynamic regressions (Franzse and Hays, 2007, 2008b,a). Modeling two-dimensional dynamics at the same time is even more complicated and can be a further extension of the present model and methods. I temporarily leave it to the future research. In this chapter, I specify a hierarchical model with pth-order autoregressive errors and two group-level regressions in the time and spatial dimensions. To estimate the model, I orthognalize the errors by using the Cholesky decomposition and adding an auxiliary parameter. This approach not only solves the problem of constructing conditional distributions of the time-specific random-effect parameters, but also dramatically improves simulation efficiency for the following reasons: first, data augmentation is implemented in one block, unlike in the conventional Geweke method and its modified versions all of which update the augmented data one by one and conditional on one another (Geweke, 1991, 1996; Sandor and Andras, 2004); second, this method simplifies the Bayes Factor computation for model comparison and lag order determination when applying the marginal likelihood approach (Chib, 1995; Chib and Jeliazkov, 2001), because the likelihood ordinate can be computed simply by conducting only one reduced run instead of using those computationally expensive methods such as the GHK method and its variants (Geweke, 1991; Borsch-Supan and Hajivassiliou, 1993; Keane, 1994*a*; Chib and Jeliazkov, 2006) or the auxiliary particle filter (Pitt and Shephard, 1999; Mueller and Czado, 2005; Pang, 2008). Furthermore, because slow MCMC mixing is often a problem of algorithms using Gibbs samplers for hierarchical models with complicated random effects (Carlin, 1996; Olsen and Schafer, 2001), I further improve simulation efficiency and speed up mixing by implementing the parameter expansion method, i.e., the partial group move multigrid Monte Carlo (PGM-MGMC) introduced by Liu and Wu (1999) and Liu and Sabatti (2000).

The model and techniques are espeically useful for investigating dynamics and common shocks by using TSCS data. In political science, path dependence or political inertia is a salient political phenomenon and has strong explanatory power for the evolution of political institutions and events (Thelen, 1999; Pierson and Skocpol, 2002; Peters, Pierre and King, 2005). This dependence or inertia is caused by many unobserved as well as observed factors. Modeling error correlation not only improves reliability of statistical inferences, but also helps explain *sub rosa* political dynamics. Likewise, contemporary correlation is substantively important for political economists who share the consensus regarding great impacts of globalization on almost all significant political economic phenomena. Heterogeneity, if omitted but correlated with the responses, is a cause of serial correlation and heteroskedasticity, but it is also theoretically interesting, especially to comparativists, when the data contain a small or moderate number of identifiable and theoretically interesting units, such as countries, states, or legislators. Appropriately modeling heterogeneity helps the researcher reach generalizable but not over-generalized conclusions. Analyzing heterogeneity across time as well as among units is also essential for correctly understanding dynamics, since ignoring heterogeneity leads to "spurious dynamics" wherein temporal pseudodependence is simply caused by unmodeled differences among units instead of dynamics (Heckman, 1981). This model is also useful for studies with statistical forecasting (building early-warning systems) as one of the major tasks, such as the research on state failure, financial crises, and international investment risk. The proposed model improves statistical predicting by making good use of the information contained in errors which is neglected in conventional models most of the time.

2.2 Model Specification and Assumptions

Suppose the data consist of N units indexed as i, where $i \in \{1, 2, ..., N\}$. Each unit i has T_i observations across the time periods $\{1_i, ..., t_i, ..., T_i\} \subseteq \{1, 2, ..., T\}$. With the contemporary effects modeled, the value of t_i indexes the observation's location in the sequence of $\{1, 2, ..., T\}$, i.e., the time-dimension cluster it belongs to. For unbalanced data structures, it is likely that $t_i \neq t_j$ and $T_i \neq T_j$ for $i \neq j$. Therefore, in the model there are two sources of clustering: an observation belongs to cluster i and t_i at the same time, and the two clusters are not nested. By using the latent variable specification (Albert and Chib, 1993), the generalized linear multi-level model for binary TSCS responses with errors following a pth-order autoregressive process (henceforth, GLMM-AR(p)) can be written as follows:

$$y_{i,t_i} = \mathbb{I}(z_{i,t_i} > 0), \tag{2.1}$$

$$z_{i,t_i} = \mathbf{x}'_{1i,t_i}\boldsymbol{\beta}_1 + \mathbf{w}'_{i,t_i}\boldsymbol{\beta}_{2i} + \mathbf{s}'_{i,t_i}\boldsymbol{\beta}_{3t_i} + \xi_{i,t_i}, \tag{2.2}$$

$$\boldsymbol{\beta}_{2i} = \boldsymbol{A}_i\boldsymbol{\beta}_2 + \boldsymbol{b}_i, \tag{2.3}$$

$$\boldsymbol{\beta}_{3t_i} = \boldsymbol{F}_{t_i}\boldsymbol{\beta}_3 + \boldsymbol{c}_{t_i}, \tag{2.4}$$

$$\xi_{i,t_i} = \rho_1\xi_{i,t_i-1} + ... + \rho_p\xi_{i,t_i-p} + u_{i,t_i}, \tag{2.5}$$

where $\mathbb{I}(\cdot)$ is the indicator function. This specification is an extension of the commonly-applied generalized linear mixed-effect model in longitudinal analysis. In this model, y_{i,t_i} is the observation of unit i at time t_i, and whether it takes value of 0 or 1 is determined by a latent continuous variable z_{i,t_i} and a threshold 0. At the latent level, z_{i,t_i} is assumed to have a linear relationship with the specified covariates (equation (4.7)), and the error ξ_{i,t_i} follows an AR(p) process (equation (4.9)). Although non-Toeplitz errors[2] can be applied, the simple AR(p) error specification should be adequate for analyzing serial correlation of equally spaced TSCS observations. It can also serve as a serial correlation diagnosis following the classical Box-Jenkins procedure: run models with different lag orders and use information-based criteria to choose an appropriate one (Box, Jenskins and Reinsel, 1994). In addition, this error specification does not impose any theoretical or mathematical limitations on the lag order choice or including LDVs/LIDVs. Since LDVs/LIDVs in the model do not lead to any additional methodological issue and can be easily included in any of the design matrices in equation (4.7) as ordinary regressors if we are willing to use them, I do not make them as special terms in the model and will not discuss them as special specifications. In this general specification, there are three groups of covariates: those in $\boldsymbol{x}_{1i,t_i}$ with fixed effects on all the observations, and those in $\boldsymbol{w}_{i,t_i}$ or $\boldsymbol{s}_{i,t_i}$ with unit-specific or time-specific effects (random effects. The model can have time-varing and unit-varing random intercepts, but for identification reasons, a constant should not be included in $\boldsymbol{x}_{1i,t_i}$ at the same time, and when there are two random intercepts, the unit-specific one is centered at 0. At the unit level (equation (2.3)), heterogeneity across units is modeled; another set of covariates $\boldsymbol{a}_i$ in matrix $\boldsymbol{A}_i$, parameters $\boldsymbol{\beta}_2$, and an error term $\boldsymbol{b}_i$, together explain the variation of the random effects $\boldsymbol{\beta}_{2i}$. The same

[2]A Toeplitz matrix is also known as a diagonal-constant matrix, which is a matrix with constant descending diagonals from left to right. The AR(p) and MA(q) covariance matrice are both Toeplitz.

specification applies to the time-dimension clustering in equation (2.4). To estimate the model, I construct its reduced form by setting $\mathbf{x}_{i,t_i} = (\mathbf{x}_{1,it_i}, \mathbf{w}'_{i,t_i}\boldsymbol{A}_i, \mathbf{s}'_{i,t_i}\boldsymbol{F}_{t_i})$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$, which can be written as follows:

$$z_{i,t_i} = \mathbf{x}'_{i,t_i}\boldsymbol{\beta} + \mathbf{w}'_{i,t_i}\boldsymbol{b}_i + \mathbf{s}'_{i,t_i}\boldsymbol{c}_{t_i} + \xi_{i,t_i} \tag{2.6}$$

$$\xi_{i,t_i} = \rho_1\xi_{i,t_i-1} + ... + \rho_p\xi_{i,t_i-p} + u_{i,t_i}. \tag{2.7}$$

The reduced form clearly demonstrates the many possibilities for the error structure: , if assuming $\{\boldsymbol{b}_i\}$ and $\{\boldsymbol{c}_t\}$ are not correlated, the error structure is can be expressed as $\boldsymbol{\Sigma}_{\mathbf{Z}_i} = \boldsymbol{W}'_i\boldsymbol{\Sigma}_{b_i}\boldsymbol{W}_i + T_i\boldsymbol{S}'_i\boldsymbol{\Sigma}_{\boldsymbol{c}_{\boldsymbol{T_i}}}\boldsymbol{S}_i + \boldsymbol{\Sigma}_{\xi_i}$[3]. In this equation, $\boldsymbol{\Sigma}_{Z_i}$, $\boldsymbol{\Sigma}_{b_i}$, $\boldsymbol{\Sigma}_{c_{T_i}}$, and $\boldsymbol{\Sigma}_{\xi_i}$, are the covariance matrices of $\boldsymbol{Z}_i$, $\{\boldsymbol{b}_i\}$, $\{\boldsymbol{c}_{T_i}\}$, and $\boldsymbol{\xi}_i$, respectively. It is important to note that the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}_i}$ not only reflect correlation of the responses, but also contains heteroskedasticity.

Several assumptions are made for identification and estimation reasons: (1) the error term u_{i,t_i} in the autoregressive process is white noise, that is $u_{i,t_i} \sim N(0,1)$; (2) the individual-level errors are not correlated across units; mathematically, $\text{cov}(\xi_{i,t_i}, \xi_{k,t_s}) = 0$, $\forall i \neq k$ and $\forall t_i, t_s \in T$; (3) the error ξ follows a pth-order autoregressive process; (4) if the data structure is unbalanced, it is only for exogenous reasons. In other words, there is no sample selection problem; (5) the covariates in $\boldsymbol{x}_{i,t_i}$ and $\boldsymbol{w}_{i,t_i}$ are sequentially exogenous (compared to the strict exogeneity assumption commonly required in panel data analysis): $(\boldsymbol{w}_{i,t_i} \perp \xi_{i,t_i})|z_{i,t_{s_i}}$ and $(\boldsymbol{x}_{i,t_i} \perp \xi_{i,t_i})|z_{i,t_{s_i}}, \forall t_{s_i} \leq t_i$; (6) the autoregressive process is stationary. The stationarity assumption is restrictive and requires that either the dynamic process under investigation is stationary or a non-stationary process of $\boldsymbol{Z}_i$ is cointegrated with the explanatory variables included

[3]For model specification and estimation, this assumption is not necessary. The assumption used here is only for obtaining in a easily-written mathematical form to illustrate the error structure.

in $\boldsymbol{x_i}$, $\boldsymbol{w_i}$ and $\boldsymbol{s_i}$ (Chib and Greenberg, 1994); otherwise, the model cannot be applied[4]. In fact, this assumption makes the GLMM-AR(p) an informal panel data unit root and cointegration test following the line of the residual-based tests (Engle and Granger, 1987; Kao, 1999; Pedroni, 1999, 2004), which is valuable for avoiding spurious regressions (Hamilton, 1994, pp.557-62). This is particularly important since in the literature conducting formal cointegration tests on discrete panel data is challenging, and this type of statistical tests are rarely done on discrete panel data (or longitudinal data) in practice. Later in this chapter, I will use an empirical example to illustrate how the GLMM-AR(p) model with a stationarity restriction on the error process, serves as a tool to detect likely spurious relationships between the response and explanatory variables; (7) finally, I assign priors on the parameters as follows:

$$\boldsymbol{\beta} \sim N_{K_1}(\boldsymbol{\beta}_0, \mathbf{B}_0), \qquad u_{it} \sim N(0,1), \qquad \{\boldsymbol{b}_i\} \sim N_{K_2}(\mathbf{0}, \boldsymbol{D}), \qquad \boldsymbol{D}^{-1} \sim W_{K_2}(\boldsymbol{\nu}_0, \boldsymbol{D}_0),$$
$$\{\boldsymbol{c}_t\} \sim N_{K_3}(\mathbf{0}, \boldsymbol{E}), \qquad \boldsymbol{E}^{-1} \sim W_{K_3}(\boldsymbol{\eta}_0, \boldsymbol{E}_0), \qquad \boldsymbol{\rho} \sim U_p(\boldsymbol{\rho} : \boldsymbol{\rho} \in S_\rho),$$

where S_ρ is the stationarity space of the autoregressive coefficients[5]. Since u_{it}, $\{\boldsymbol{b}_i\}$, and $\{\boldsymbol{c}_t\}$ are errors at different levels, their prior means are set to be 0. However, centering the prior mean of $\{\boldsymbol{c}_t\}$ at zero is not required for mathematical reasons in the designed MCMC algorithm, which means that the time-level intercept can be omitted in the specification because it can be automatically included in the non-centered posteriors of $\{\boldsymbol{c}_t\}$. All the parameters except $\boldsymbol{\rho}$ have conditional conjugate priors.

[4]The MCMC algorithm will provide information about non-stationarity by not being able to generate or accept legitimate proposals (proposals drawn within the stationary space) for the autoregressive coefficients, and the MCMC simulation process will halt if the chain is stuck with keeping drawing illegitimate proposals in a pre-specified period of time.

[5]An AR process is stationary if all the characteristic roots of the polynomial are outside the unit circle. For different autoregressive processes the stationarity space is different.

Other prior specifications are possible; for instance, the autoregressive coefficients $\boldsymbol{\rho}$ could have a multivariate normal prior distribution truncated in the stationarity space, but the priors of $\boldsymbol{b}_i$ and $\boldsymbol{u}$ by design should be centered at $\mathbf{0}$. I use diffuse but proper priors, and conduct simulations on the prior distributions to ensure they are in reasonable spaces for the substantive questions.

2.3 MCMC Algorithm

2.3.1 Cholesky Decomposition and Auxiliary Variable

Due to the serially correlated errors, the covariance matrix have non-zero off-diagonal elements, which complicates model estimation of the proposed setup. One way to handle this challenge is to use so-called robust standard errors. However, this method is not as convenient as it seems, because three biases[6] have to be overcome when constructing the weight function (Andrews, 1991; Lumley and Heagerty, 1999; Zeileis, 2004), which is difficult for non-linear mixed-effect models. More importantly, using robust standard errors discards valuable information in the error term. Such information could have been used for analyzing dynamics and heterogeneity and for improving forecasts. Other solutions having been developed in the literature include estimating generalized multilevel models with serial correlation by using numerical methods such as Penalized Quasi-Likelihood (PQL), Marginal Quasi-Likelihood (MQL) or EM algorithms (Schafer and Yucel, 2002; Renard, Molenberghs and Geys, 2004; Molenberghs and Verbeke, 2005, Chapter 14 and Chapter 22). However, those methods are complicated with many approximation steps such as linear

[6]The three biases are the bias of the estimator of the variance, bias due to omitted and down-weighted correlations (truncation bias), and bias caused by evaluating the estimator at estimates rather than the true parameters (centering bias).

or high-order Taylor expansion, Gauss-Hermite quadrature, pseudo-data generation, or empirical Bayes estimation. What is even worse is that these procedures often yield estimates and standard errors biased towards zero, especially with the first-order expansion (Ng et al., 2006). The Bayesian approach has been widely used for analyzing multilevel models and enjoys the flexibility in specifying and estimating rich and sophisticated models (Hagenaars, 1990; Singer and Willett, 2003; Yang, Fu and Land, 2004; Skrondal and Rabe-Hesketh, 2004; Molenberghs and Verbeke, 2005), but, even the Markov Chain Monte Carlo methods encounter at least two methodological challenges in estimating the GLMM-AR(p) model.

First, the non-zero off-diagonal elements in the covariance matrix $\mathbf{\Sigma}_{\xi_i}$ complicate constructing the conditional distribution of the time-specific random effect parameter vector $\{\boldsymbol{c}_t\}$. This is because when developing its conditional distribution for constructing the Markov chain, we have to switch to the spatial dimension, and express equation (4.10) as follows:

$$\begin{cases} \boldsymbol{z}_1 = \boldsymbol{x}_1\boldsymbol{\beta} + \boldsymbol{w}_1\boldsymbol{b}_{N_1} + \boldsymbol{s}_1\boldsymbol{c}_1 + \epsilon_1 \\ \boldsymbol{z}_2 = \boldsymbol{x}_2\boldsymbol{\beta} + \boldsymbol{w}_2\boldsymbol{b}_{N_2} + \boldsymbol{s}_2\boldsymbol{c}_2 + \epsilon_2 \\ \qquad \cdots \quad \cdots \quad \cdots \\ \boldsymbol{z}_T = \boldsymbol{x}_T\boldsymbol{\beta} + \boldsymbol{w}_T\boldsymbol{b}_{N_T} + \boldsymbol{s}_T\boldsymbol{c}_T + \epsilon_T \end{cases} \tag{2.8}$$

This is a seemingly unrelated regression (SUR) system, and the errors in the T equations are correlated because of serial dependence. The covariance matrix is $\mathrm{Var}(\boldsymbol{\epsilon}|\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{s}) = \mathbf{\Sigma} \otimes \boldsymbol{I}_T$ for a balanced dataset, and more complicated if $N_t \neq N_k$ for some $t \neq k$. Handling unbalanced structure is not easy in SUR analysis (McDowell, 2004; Schmidt, 1977). In the Bayesian framework, although it is feasible to specify the covariance matrix $\mathrm{Var}(\boldsymbol{\epsilon}|\boldsymbol{X}, \boldsymbol{w}, \boldsymbol{s})$ with unbalanced structure since $\mathbf{\Sigma}_i$ is a Toeplitz matrix, it is computational expensive to compute a $M \times M$ matrix ($M = \sum_{t=1}^{T} N_t$) in each iteration by decomposing Toeplitz matrices and then mapping the elements into a huge covariance matrix.

Second, the Toeplitz error matrix requires sampling from truncated multivariate normal distributions for data augmentation (sampling $\{\boldsymbol{z}_i\}$). In the literature, various approaches have been proposed, such as quasi Monte Carlo, antithetic Monte Carlo, and samples based on orthogonal arrays (Sandor and Andras, 2004). The method, introduced by Geweke (1991, 1996) and widely used in applied works, is to apply the Gibbs sampler to update the latent responses, z_{it_i}, one by one, conditional on other $z_{i,-t_i}$s, which is not efficient and often mixes poorly (Rodriguez-Yam, Davis and Scharf, 2004). In addition to the slow mixing problem, applying this method makes it difficult to compute the Bayes Factor, as the likelihood ordinate in the marginal likelihood contains the latent variable $\boldsymbol{z}$:

$$L = f(y_{it_{1_i}}, ..., y_{it_{J_i}} | \boldsymbol{\theta}) = \int_{a_{i1_i}}^{b_{i1_i}} \int_{a_{i2_i}}^{b_{i2_i}} ... \int_{a_{iT_i}}^{b_{iT_i}} p(z_{it_{1_i}}, ..., z_{it_{J_i}} | \boldsymbol{\theta}) dz_{it_{1_i}} ... dz_{it_{T_i}}, \tag{2.9}$$

where (a_{it_i}, b_{it_i}) is the truncated region determined by y_{it_i}, and $\boldsymbol{\theta}$ represents all the parameters in the model. Note that $\boldsymbol{z}_i$ cannot be averaged out by using the MCMC output, since the samples of $\boldsymbol{z}$ are not drawn from the its marginal distribution (without conditional on $\boldsymbol{y}$) to which this integration is with respect. Instead, the draws are from the distribution conditional on $\boldsymbol{y}$, and cannot be used to average $\boldsymbol{z}$ out. One solution is the GHK simulator (Geweke, 1991; Borsch-Supan and Hajivassiliou, 1993; Keane, 1994*b*; Chib and Jeliazkov, 2006), which is computationally expensive, especially with a large dataset. Alternatively, the likelihood ordinate can be estimated by using the auxiliary particle filter, when the latent response variable is sampled by transforming data with polynomial operator $P(L)z_{it_i}$ to orthogonalize the covariance matrix (Mueller and Czado, 2005; Pang, 2008). However, this SIR-based sampling scheme is not stable for a high order Markov process and often requires a huge number of samples in each iteration to obtain a valid approximation of the likelihood.

I propose an algorithm which orthogonalizes the correlated errors in such a way that z_{it} can be sampled without being conditional on any other zs, and all $\boldsymbol{z}$ updated in one block instead of $\sum_{i=1}^{N} T_i$ blocks, which improves simulation efficiency dramatically. Importantly, this method makes computation of the likelihood ordinate in marginal likelihood calculation as simple as in an ordinary probit model. It also takes care of the off-diagonal elements in the Toeplitz covariance matrix of the errors when constructing the conditional distribution of $\{\boldsymbol{c}_t\}$. The basic idea is as follows: first, we can decompose the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i}$ into two parts $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i} = \boldsymbol{\Omega}_i + \kappa_i \boldsymbol{I}_i$, where $\boldsymbol{\Omega}_i$ is a symmetric positive definite matrix and κ_i is any constant (I choose $\varrho_i/2$ for κ_i, where ϱ_i is the smallest eigenvalue of $\boldsymbol{\Omega}_i$. This choice follows Chib and Jeliazkov (2006) and is to make the algorithm numerically stable); and $\boldsymbol{\Omega}_i$ is further decomposed as $\boldsymbol{V}_i'\boldsymbol{V}_i$, in which $\boldsymbol{V}_i'$ is the lower triangular matrix produced by the Cholesky decomposition. Hence, I can re-express the covariance matrix as $\boldsymbol{\Sigma}_{\boldsymbol{xi}_i} = \boldsymbol{V}_i'\boldsymbol{V}_i + \kappa_i \boldsymbol{I}_T$, and the model in equation (4.10) can be written as

$$\boldsymbol{z}_i = \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{w}_i'\boldsymbol{b}_i + \mathrm{s}_i'\boldsymbol{c}_{T_i} + \boldsymbol{V}_i'\boldsymbol{u}_i + \boldsymbol{\epsilon}_i, \tag{2.10}$$

where the new error term $\boldsymbol{\epsilon}_i \sim N_{T_i}(0, \kappa_i \boldsymbol{I}_{T_i})$ and the auxiliary variable $\boldsymbol{u}_i \sim N_{T_i}(\mathbf{0}, \boldsymbol{I}_{T_i})$, and they are mutually independent. Since in the Bayesian approach we marginalize a parameter with respect to its prior distribution, in equation (2.10) $\boldsymbol{z}_i$ has the exact covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i}$ when integrating $\boldsymbol{u}_i$ out. Conditional on $\boldsymbol{u}_i$ and $\boldsymbol{\rho}$ ($\boldsymbol{V_i}$ is a function of $\boldsymbol{\rho}$) as well as other parameters, the elements in $\boldsymbol{z}_i$ are not correlated and do not need to be updated conditional on one another. Furthermore, there are general formulas to compute the Toeplitz covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i}$ for AR(p) errors, and we even do not need to compute each $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i}$; instead, in each iteration I simply compute $\boldsymbol{\Sigma}_{T\times T}$ (T is the maximum number of time periods) and construct the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i}$ for each i by taking the first T_i rows and columns of $\boldsymbol{\Sigma}_{T\times T}$. Define $(q_1, q_2, ..., q_{T_i}) = \boldsymbol{q}_i \equiv \boldsymbol{V}_i'\boldsymbol{u}_i$ and the algorithm is simplified as follows:

1. $\boldsymbol{\beta}, \{\boldsymbol{b}_i\}, \{\boldsymbol{u}_i\}|\cdot \sim \pi(\{\boldsymbol{u}_i\}|\boldsymbol{\beta}, \{\boldsymbol{b}_i\}, \cdot)\pi(\{\boldsymbol{b}_i|\boldsymbol{\beta}, \cdot)\pi(\boldsymbol{\beta}|\cdot)$[7]
   - $\boldsymbol{\beta}|\cdot \sim N_{K_1}(\bar{\boldsymbol{\beta}}, \boldsymbol{B}_1)$, where $\boldsymbol{B}_1 = (\boldsymbol{B}_0^{-1} + \sum_{i=1}^N \boldsymbol{x}_i'\boldsymbol{H}_i^{-1}\boldsymbol{x}_i)^{-1}$, $\bar{\boldsymbol{\beta}} = \boldsymbol{B}_1\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^N \boldsymbol{x}_i'\boldsymbol{H}_i^{-1}(\boldsymbol{z}_i - \mathbf{s}_i'\boldsymbol{c}_{T_i})\right)$, and $\boldsymbol{H}_i = (\boldsymbol{\Omega}_i + \boldsymbol{w}_i'\boldsymbol{D}\boldsymbol{w}_i)$;
   - $\boldsymbol{b}_i|\boldsymbol{\beta}, \cdot \sim N_{K_2}(\bar{\boldsymbol{b}}_i, \boldsymbol{D}_{1i})$, where $\boldsymbol{D}_{1i} = (\boldsymbol{D}^{-1} + \boldsymbol{w}_i'(\boldsymbol{\Omega}_i)^{-1}\boldsymbol{w}_i)^{-1}$ and $\bar{\boldsymbol{b}}_i = \boldsymbol{D}_{1i}\boldsymbol{w}_i'(\boldsymbol{\Omega}_i)^{-1}(\boldsymbol{z}_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{s}_i'\boldsymbol{c}_{T_i})$;
   - $\boldsymbol{u}_i|\cdot \sim N(\bar{\boldsymbol{u}}_i, \boldsymbol{U}_i)$, where $\boldsymbol{U}_i = (\boldsymbol{I}_T + \boldsymbol{V}_i\boldsymbol{V}_i'/\kappa_i)^{-1}$, and $\bar{\boldsymbol{u}}_i = \boldsymbol{U}_i\boldsymbol{V}_i(\boldsymbol{z}_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{w}_i'\boldsymbol{b}_i - s_i'\boldsymbol{c}_{T_i})/\kappa_i$

2. $\{\boldsymbol{c}_t\}|\cdot \sim N_{K_3}(\bar{\boldsymbol{c}}_t, \boldsymbol{E}_{1i})$, where $\boldsymbol{E}_{1i} = (\boldsymbol{E}^{-1} + \boldsymbol{s}_i'(\boldsymbol{\kappa}_{N_t}\boldsymbol{I}_N)^{-1}\boldsymbol{s}_i)^{-1}$ and $\bar{\boldsymbol{c}}_t = \boldsymbol{E}_{1i}\boldsymbol{s}_i'(\boldsymbol{\kappa}_{N_t}\boldsymbol{I}_N)^{-1}(\boldsymbol{z}_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{s}_t'\boldsymbol{b}_{N_t} - q_t)$, where $\boldsymbol{\kappa}_{N_t}$ is the vector with κ_i's for all the i's observed at time t

3. $z_{it}|\cdot \sim TN(\mathbf{x}_{it}'\boldsymbol{\beta} + \mathbf{w}_{it}'\boldsymbol{b}_i + \mathbf{s}_{it}'\boldsymbol{c}_{t_i} + q_{it}, \kappa_i)$

4. $\boldsymbol{D}^{-1}|\cdot \sim W_{K_2}(\nu_1, \boldsymbol{D}_1)$ and $\boldsymbol{E}^{-1}|\{\boldsymbol{c}_t\} \sim W_{K_3}(\eta_1, \boldsymbol{E}_1)$, where $\nu_1 = \nu_0 + N$, $\boldsymbol{D}_1 = (\boldsymbol{D}_0^{-1} + \sum_{i=1}^N \boldsymbol{b}_i\boldsymbol{b}_i')^{-1}$, $\eta_1 = \eta_0 + T$, and $\boldsymbol{E}_1 = (\boldsymbol{E}_0^{-1} + \sum_{i=1}^T \boldsymbol{c}_t\boldsymbol{c}_t')^{-1}$

5. $\boldsymbol{\rho}|\cdot \sim \Psi(\boldsymbol{\rho}) \times N(\hat{\boldsymbol{\rho}}, \boldsymbol{P})$, following Chib (1993), I use a Metropolis-Hasting algorithm to update $\boldsymbol{\rho}$ by using the tailored kernel $N(\hat{\boldsymbol{\rho}}, \boldsymbol{P})$.

Chib and Jeliazkov (2006) proposed the Cholesky decompostion idea for sampling the nonparametric term in their binary panel model, but they still applied the Geweke method to sample z_{i,t_i} conditional on other zs, which is much less efficient than the algorithm applied here. More important, the algorithm above makes marginal likelihood calculation straightforward which is shown in the next section.

[7]Sampling $\boldsymbol{\beta}, \{\boldsymbol{b}_i\}$ and $\{\boldsymbol{u}_i\}$ in one block improves the efficiency of this algorithm because they are correlated by construction. However, it is not feasible to include $\{\boldsymbol{c}_t\}$ in this block because of the complicated covariance structure caused by marginalization.

2.3.2 Partial Group Move Multigrid Monte Carlo

MCMC algorithms using Gibbs samplers can mix slowly and take a long time for the Markov chain to explore the stationary space (Carlin, 1996; Olsen and Schafer, 2001). With random effects in two dimensions and serially correlated errors, the preliminary simulations I have conducted by implementing the algorithm above justify the concern of slow mixing. To further improve simulation efficiency and speed up mixing, I add a partial group move Multigrid Monte Carlo (PGM-MGMC) updating stage into the algorithm, which dramatically reduces within-chain autocorrelation.

The basic idea of multigrid methods is to use a sequence of auxiliary "coarse-grid" problems in addition to the original "fine-grid" problem so that the information is more efficiently stored and convergence is accelerated (Goodman and Sokal, 1989; Briggs, 1987, Chap.3). This method was first applied in statistical physics and Euclidean quantum physics. Goodman and Sokal (1989) extended the deterministic multigrid method into a multigrid Monte Carlo algorithm by applying partial resampling and fiber construction. Liu and Sabatti (2000) generalized the Gibbs sampler by using the multigrid Monte Carlo method to decompose the sample space into disjoint orbits[8] in order to facilitate information transmission. By generating a transformation group, the Markov chain is moved by a mover from one orbit to another without leaving the sample space of the target distribution (Liu and Wu, 1999). This enables a faster exploration of the sample space and achieves the effects of reparameterization, blocking, or grouping, but has more freedom of decomposing the sample space. They applied this method to state space models and showed that, by choosing a transformation group, the move can be dramatic and autocorrelation reduced to a

[8]In the group theory, define a goup G and a set X, an orbit of $x \in X$ is the set $S \subset X$ to which x can be moved by the elements of G. Disjoint orbits simply means that the orbits $S_1, S_2, \dots$ are disjoint. Refer to Bogopolski (2008) or Aschbacher (2000) for more formally definitions and detailed theories about orbits.

considerable degree. Mueller and Czado (2005) also applied the partial move group multigrid Monte Carlo method to reduce autocorrelation in their autoregressive ordinal probit model, and demonstrated the efficiency of this method by speeding up mixing. The posterior distribution of the GLMM-AR(p) setup facilitates developing a distribution from which a mover can be randomly drawn to transform a subset of m parameters $\boldsymbol{\omega} \equiv (\{\boldsymbol{z}_i\}, \boldsymbol{\beta}, \{\boldsymbol{b}\}, \{\boldsymbol{c}\})$. To apply a partial group move multigrid MC, I choose the scale group $\Gamma = \{\chi > 0 : \chi(x) = \chi x\}$ and calculate the unimodular Haar measure as $L(d\chi) = \chi^{-1} d\chi$, which, together with the posterior distribution, directly implies a standard gamma distribution as the mover distribution:

$$\begin{aligned}
\chi^{m-1}\pi(\chi\boldsymbol{\omega})d\chi \propto & \chi^{m-1}\exp\big(-\frac{1}{2}\sum_{i=1}^{N}(\chi\boldsymbol{z}_i - \mathbf{x}_i\chi\boldsymbol{\beta} - \mathbf{w}_i\chi\boldsymbol{b}_i - \chi\boldsymbol{c})'\boldsymbol{\Omega}_i^{-1}(\chi\boldsymbol{z}_i - \mathbf{x}_i\chi\boldsymbol{\beta} - \mathbf{w}_i\chi\boldsymbol{b}_i - \chi\boldsymbol{c}) \\
& \times \exp\big(-\frac{1}{2}(\chi\boldsymbol{\beta})'\mathbf{B}_0^{-1}(\chi\boldsymbol{\beta})\exp(-\frac{1}{2}\sum_{i=1}^{N}\chi\boldsymbol{b}_i'\boldsymbol{D}^{-1}\chi\boldsymbol{b}_i)\exp\left(-\frac{1}{2}\sum_{t-1}^{T}\chi\boldsymbol{c}_t'\boldsymbol{E}^{-1}\chi\boldsymbol{c}_t\right) \\
\propto & \chi^{m-1}\exp\Big(-\frac{1}{2}\chi^2\big(\underbrace{\sum_{i=1}^{N}(\boldsymbol{z}_i - \mathbf{x}_i\boldsymbol{\beta} - \mathbf{w}_i\boldsymbol{b}_i - \mathbf{s}_i\boldsymbol{c}_{T_i})'\boldsymbol{\Omega}_i^{-1}(\boldsymbol{z}_i - \mathbf{x}_i\boldsymbol{\beta} - \mathbf{w}_i\boldsymbol{b}_i - \mathbf{s}_i\boldsymbol{c}_{T_i})}_{Q_1} \\
& \underbrace{+\boldsymbol{\beta}'\mathbf{B}_0^{-1}\boldsymbol{\beta} + \sum_{i=1}^{N}\boldsymbol{b}_i'\boldsymbol{D}^{-1}\boldsymbol{b}_i + \sum_{t=1}^{T}\boldsymbol{c}_t'\boldsymbol{E}^{-1}\boldsymbol{c}_t)\Big)}_{Q_2} \\
\propto & \chi^{m-1}\exp\Big(-\frac{1}{2}\chi^2\big(Q_1 + Q_2\big)\Big).
\end{aligned} \tag{2.11}$$

This kernel is proportional to a gamma distribution $\Gamma(a, b)$ for χ^2 with parameters $a = (m + 1)/2$ and $b = (Q_1 + Q_2)/2$. I apply this PGM-MGMC as an updating stage in each iteration in the MCMC algorithm: $\{\chi\boldsymbol{z}_i^{(g)}\} \to \{\boldsymbol{z}_i^{(g)}\}$, $\chi\boldsymbol{\beta}^{(g)} \to \boldsymbol{\beta}^{(g)}$, $\{\chi\boldsymbol{b}_i^{(g)}\} \to \{\boldsymbol{b}_i^{(g)}\}$, and $\{\chi\boldsymbol{c}_t^{(g)}\} \to \{\boldsymbol{c}_t^{(g)}\}$.

2.4 Bayesian Model Comparison

Because we always have uncertainty in almost all respects of model specification, implementing information-based criteria for model decision-making is necessary and important (Gill, 2007). In this particular case, among other considerations, we normally have very limited prior information about the order of the autoregressive error process. In the time series literature on linear regressions, lag orders are decided by running models with different lag orders, and the appropriate order is decided by applying information-based criteria. The same procedure has not commonly performed for non-linear models, and lag orders are often decided for convenience (most often the first order). In the Bayesian framework, the Bayes Factor as a criterion of decision making has many advantages, but at the same time, is known as computationally expensive and sometimes numerically unstable, especially for sophisticated models with high dimensionality (Han and Carlin, 2001). With correlated errors and the latent dependent responses as a multivariate normal distribution, approximating the likelihood often requires additional samplers, such as importance samplers (Chib and Jeliazkov, 2006) or recursive importance samplers (Mueller and Czado, 2005). By using the auxiliary parameter approach presented in the previous section, the Bayes Factor can be computed by only using full or reduced MCMC outputs. This section gives the algorithm to estimate the Bayes Factor for model comparison and, especially, lag order determination.

The Bayes Factor is simply defined as the ratio of two marginal likelihoods (Greenberg, 2007, p.34). There are various approaches to approximae this quantity (Han and Carlin, 2001), and I apply the marginal likelihood method (Chib, 1995; Chib and

Jeliazkov, 2001). The Marginal likelihood is the normalizing constant in the Bayesian setup, and can be expressed as follows:

$$m(y) = \frac{f(y|\theta)\pi(\theta)}{\pi(\theta|y)}. \tag{2.12}$$

Adopting the approach developed by Chib (1995) and Chib and Jeliazkov (2001), I fix the values of θ, and the marginal likelihood on the logarithm scale can be computed by using the formula:

$$\ln \hat{m}(y) = \ln \hat{f}(y|\theta^*) + \ln \hat{\pi}(\theta^*) - \ln \hat{\pi}(\theta^*|y). \tag{2.13}$$

By using the Cholesky-decomposition-auxiliary-parameter approach, the likelihood ordinate, $\hat{f}(y|\theta^*)$, is straightforward to compute. Denote $\boldsymbol{\theta}$ as all the parameters except the auxiliary variable $\boldsymbol{u}$, and the likelihood ordinate can be approximated by fixing the values at $\boldsymbol{\theta}^*$ and integrating out $\boldsymbol{u}$ which is a function of q and also related to κ:

$$\hat{f}(\boldsymbol{y}|\boldsymbol{\theta}^*) = \frac{1}{M}\sum_{m=1}^{M}\prod_{i=1}^{N}\prod_{t_i=1}^{T_i}(\Delta_{it_i})^{y_{it_i}}\left(1-\Delta_{it_i}\right)^{1-y_{it_i}}, \tag{2.14}$$

$$\text{where, } \Delta_{it_i} = \Phi\left(\frac{\mathbf{x}'_{it_i}\boldsymbol{\beta}^* + \mathbf{w}'_{it_i}\boldsymbol{b}^*_i + \mathbf{s}'_i\boldsymbol{c}^*_{t_i} + q^{(m)}_{it_i}}{\sqrt{\kappa^{(m)}_i}}\right) \tag{2.15}$$

Because I integrate $\boldsymbol{u}$ out with respect to the conditional distribution of $\pi(\boldsymbol{u}|\boldsymbol{\theta}^*, \boldsymbol{z})$, a reduced run is required, which is $\boldsymbol{u}|\boldsymbol{z}, \boldsymbol{\theta}^*$ and $\boldsymbol{z}|\boldsymbol{u}, \boldsymbol{\theta}^*$.

To approximate the posterior ordinate $\hat{\pi}(\boldsymbol{\theta}^*|y)$, I partition it in the following way:

$$\begin{aligned}\hat{\pi}(\boldsymbol{\beta}^*, \boldsymbol{b}^*, \boldsymbol{D}^*, \boldsymbol{\rho}^*, \boldsymbol{E}^*, \boldsymbol{u}^*|\boldsymbol{y}) = \hat{\pi}(\boldsymbol{\rho}^*|\boldsymbol{y})\hat{\pi}(\boldsymbol{c}^*|, \boldsymbol{\rho}^*, \boldsymbol{y})\hat{\pi}(\boldsymbol{E}^*|\boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y})\hat{\pi}(\boldsymbol{b}^*|\boldsymbol{E}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y}) \\ \times \hat{\pi}(\boldsymbol{D}^*|\boldsymbol{b}^*, \boldsymbol{E}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y})\hat{\pi}(\boldsymbol{\beta}^*|\boldsymbol{D}^*, \boldsymbol{b}^*, \boldsymbol{E}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y}), \qquad (2.16)\end{aligned}$$

and compute each term on the right hand side above in the order from the left to the right:

1. $\hat{\pi}(\boldsymbol{\rho}^*|\boldsymbol{y})$: denote $\boldsymbol{\psi}$ as all parameters except $\boldsymbol{\rho}$ and $\boldsymbol{u}$:

$$\hat{\pi}(\boldsymbol{\rho}^*|\boldsymbol{y}) = \frac{J^{-1}\sum_{i=1}^{N}\Big(\alpha(\boldsymbol{\rho}^{(j)}, \boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\psi}^{(j)}, \boldsymbol{u}^{(j)}, \boldsymbol{z}^{(j)})q(\boldsymbol{\rho}^{(j)}, \boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\psi}^{(j)}, \boldsymbol{u}^{(j)}, \boldsymbol{z}^{(j)})\Big)}{K^{-1}\sum_{k=1}^{K}\Big(\alpha(\boldsymbol{\rho}^*, \boldsymbol{\rho}^{(k)}|\boldsymbol{y}, \boldsymbol{\psi}^{(k)}, \boldsymbol{u}^{(k)}, \boldsymbol{z}^{(k)})\Big)}. \qquad (2.17)$$

The numerator is the sample expectation with respect to $\pi(\boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{z}|\boldsymbol{y})$ and the MCMC output can be directly used to integrate those parameters in the conditional part. The denominator is the sample expectation with respect to the conditional product measure $\pi(\boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{z}|\boldsymbol{y})q(\boldsymbol{\rho}^*, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{z})$. Here, one reduced run is needed: fixed $\boldsymbol{\rho}$ at $\boldsymbol{\rho}^*$, conduct a reduce run to get $\boldsymbol{\psi}$ and $\boldsymbol{\rho}^{(k)}$ in each iteration by using $\boldsymbol{\psi}^{(k)}$, and then plug all those draws of the parameters and augmented data into the denominator, and compute the quantity;

2. $\hat{\pi}(\boldsymbol{c}^*|, \boldsymbol{\rho}^*, \boldsymbol{y})$: directly use the output of the reduced run conducted above;

3. $\hat{\pi}(\boldsymbol{E}^*|\boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{z}, \boldsymbol{y}) = \hat{\pi}(\boldsymbol{E}^*|\boldsymbol{c}^*, \boldsymbol{y})$: no reduced run required;

4. $\hat{\pi}(\boldsymbol{b}^*|\boldsymbol{E}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y}) = \prod_{i=1}^{N}\hat{\pi}(\boldsymbol{b}_i^*|\boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{z}_i, \boldsymbol{y}_i)$: conduct a reduced run by fixing $\boldsymbol{E}, \boldsymbol{c}, \boldsymbol{\rho}$, and keep the output of $\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{z}$ together with the fixed values to compute this quantity;

5. $\hat{\pi}(\boldsymbol{D}^*|\boldsymbol{b}^*, \boldsymbol{E}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y}) = \hat{\pi}(\boldsymbol{D}^*|\boldsymbol{b}^*, \boldsymbol{y})$: no reduced run needed;

6. $\hat{\pi}(\boldsymbol{\beta}^*|\boldsymbol{D}^*, \boldsymbol{b}^*, \boldsymbol{E}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{y}) = \hat{\pi}(\boldsymbol{\beta}^*|\boldsymbol{b}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*, \boldsymbol{z}, \boldsymbol{y})$: conduct a reduced run by fixing $\boldsymbol{D}, \boldsymbol{b}, \boldsymbol{E}, \boldsymbol{c}, \boldsymbol{\rho}$ and keep the output of $\boldsymbol{z}$ together with the fixed values of $\boldsymbol{b}^*, \boldsymbol{c}^*, \boldsymbol{\rho}^*$ to compute this quantity.

2.5 Simulation Study

I exam the performance of the model and estimation techiniques developed in the previous sections with Monte Carlo experiments. The simulations reported in this section are desgined as the following: first, two datasets are generated by roughly the same data-generating process (DGP) but with different autoregressive coefficients ($\rho_1 = 0.7, \rho_2 = 0.2$ for the first dataset, and $\rho_1 = -0.5, \rho_2 = -0.3$ for the second)[9]; second, in both datasets, the number of observations is 2187 and the data structures are unbalanced with the number of observations of a unit varying from 2 to 50 and the number of observations in a time period from 8 to 50; third, both DGPs have a mixed-effect design: 5 covariates have fixed effects (in $\boldsymbol{x}_i$), 5 have unit-specific random effects (in $\boldsymbol{w}_i$), and 2 have time-specific effects (in $\boldsymbol{s}_i$). At the group levels, there are 3 unit-level predictors (in $\boldsymbol{a}_i$) and 2 time-level predictors (in $\boldsymbol{f}_t$).

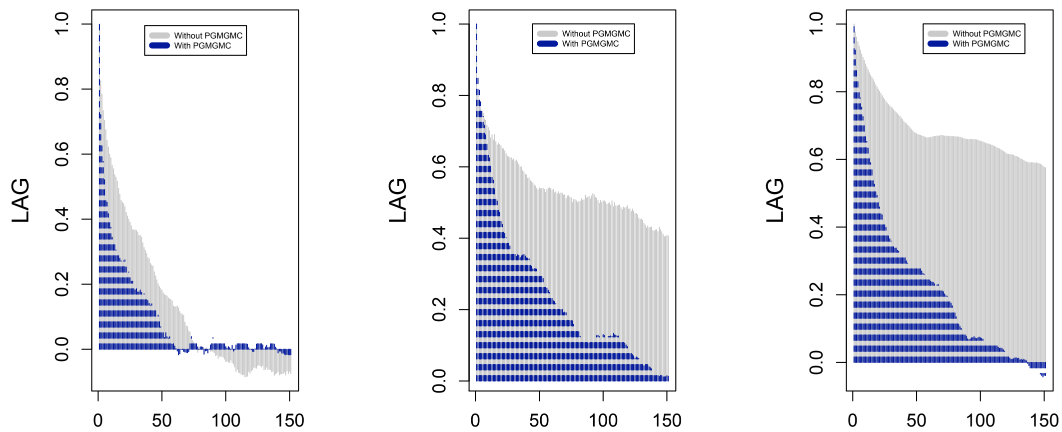
I assign diffuse priors to the parameters. For the auxiliary parameter vector, $\boldsymbol{u}_i$, its prior be $N_{T_i}(\mathbf{0}, \boldsymbol{I}_{T_i})$ by design. The prior choice is also straightforward for the group-level errors $\{\boldsymbol{b}_i\}$ and $\{\boldsymbol{c}_t\}$: as residuals, their distributions are centered at $\mathbf{0}$, and their covariance matrices are treated as hyperparameters. The coefficient parameter vector $\boldsymbol{\beta}$ is assigned with a multivariate normal prior centered at $\mathbf{0}$, having a diagonal

[9] Chib and Jeliazkov (2006) find that in hierarchical models with serially correlated errors, negative autocorrelation seems to be identified better than positive autocorrelation. I use the two comparable datasets to check this point

covariance matrix with all diagonal elements equal to 400. This prior is vague, and the inverse of the prior covariance matrix has all diagonal elements as small as 0.0025, which should not have notable effects on the posteriors. Note that for implementing the PGM-MGMC algorithm, the priors on $\boldsymbol{\beta}$ have to be centered at $\mathbf{0}$. The prior on the autoregressive parameter vector $\boldsymbol{\rho}_p$ is a multidimensional uniform distribution within the stationary space. For the hyperparameter matrices $\boldsymbol{D}$ and $\boldsymbol{E}$, their priors can have dramatic influence on their posteriors for two reasons: they are at the lowest level of the hierarchy, and binary data often have very limited information about them. However, too diffuse priors such as $\boldsymbol{D} = \text{diag}(100)$ are likely to cause trouble in inverting matrices in the Wishart updating and result in numerical instability. My experience from many trials suggests that priors around $\boldsymbol{D} = \text{diag}(20)$ (also for $\boldsymbol{E}$) are good choices for balancing prior influence and numerical stability. The sensitivity of posteriors to the prior choices has been checked by using alternative priors with reasonable changes of locations and scales.

As observed by Carlin (1996) and Olsen and Schafer (2001), the slow MCMC mixing problem for non-linear mixed effect models by using the Gibbs sampler (the MH algorithm in the present algorithm is only for one block) is serious even with data augmentation conducted in one block in the algorithm proposed in the present research. In Figure 2.1, the grey shadow indicates the autocorrelation of a randomly selected fixed-effect parameter chain, a random-effect parameter chain in $\boldsymbol{b}_i$, and a parameter chain in the covariance matrix $\boldsymbol{E}$ (the order is from the left panel to the right). For fixed-effect parameters, the standard Gibbs Markov chains mix not badly at all, but for the random-effect parameters and the ones at low levels in the hierarchy, within-chain autocorrelation decreases very slowly—even after 150 lags, it is still as high as about 0.5, indicating ineffeciency of the algorithm. This slowly-decaying autocorrelations can result in the infeasibility for implementing the methods

Figure 2.1: Comparison of Wihtin-Chain Autocorrelation: Gibbs vs. Gibbs+PGM-MGMC



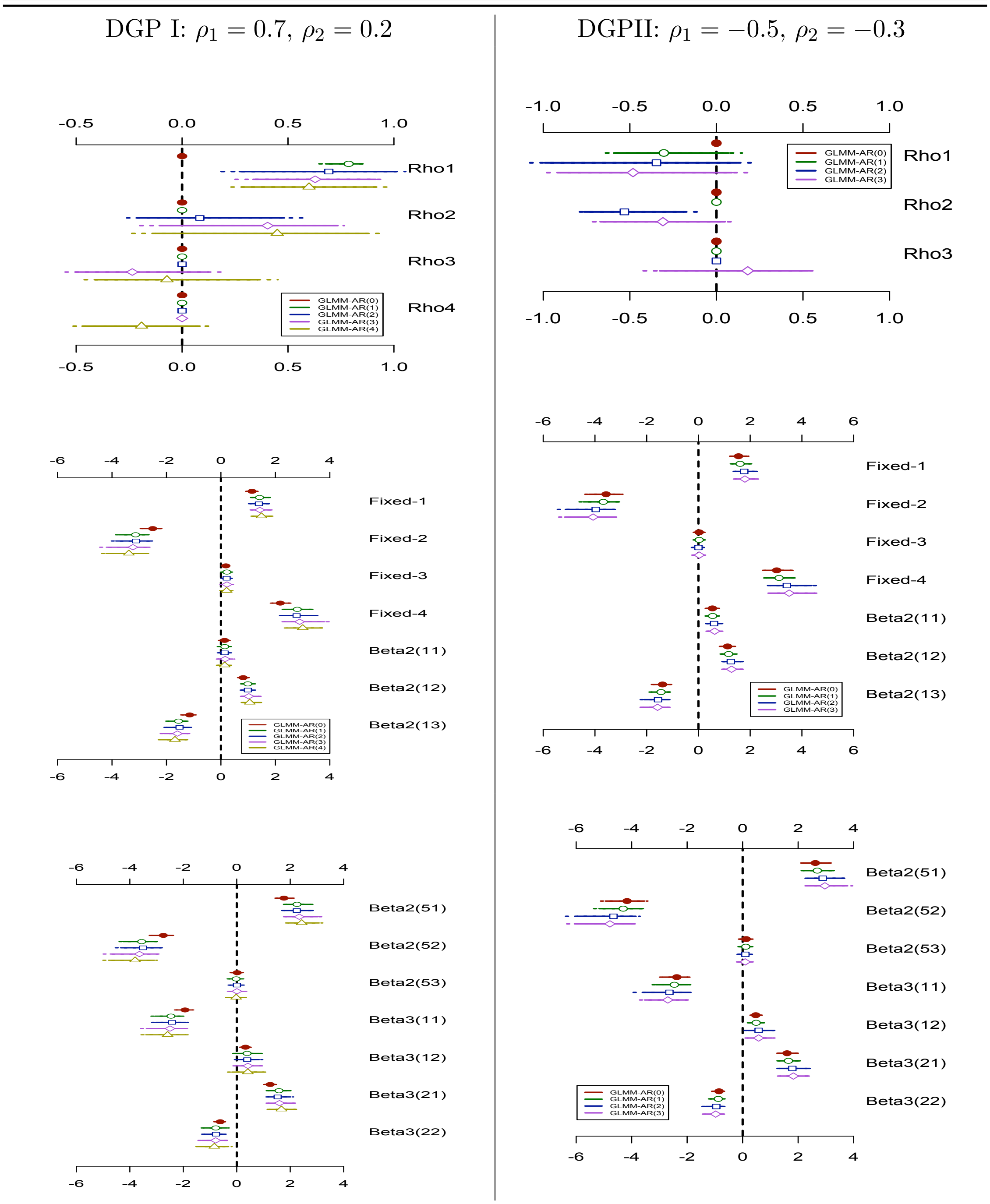
practically to models with more parameters. By using the PGM-MGMC mover, the mixing is dramatically improved, especially for those slowly-mixing parameters (see the blue shadow in Figure 2.1). There is an observable reduction of the mixing time for the fixed-effect parameters, but the improvement is not as dramatic as for the random-effect and low-level parameters. Even with the PGM-MGMC updating, the overall mixing time is much longer than in simple Bayesian models. Considering that this is a sophisticated nonlinear hierarchical model with more than 5,000 parameters, this is not surprising and the MCMC mixing time of the current algorithm can be regarded as satisfactory.

I estimate models with increasing lag orders—if it turns out that a higher order makes the model fit the data better, I increase the lag order further until the marginal likelihood starts to decrease. This procedure illustrates the model with the computed marginal likelihood can serve as a serial correlation diagnosis for binary TSCS data. It is much easier to apply than the score test proposed by Gourieroux, Monfort and

Trongnon (1985), especially for a higher-order autocorrelation diagnosis. Based on the Bayes Factors, for DGP I with positive serial correlation, the GLMM-AR(3) is the best among the five models with the lag order from 0 to 4, while for DGP II (negative serial correlation), the GLMM-AR(2) has the best goodness-of-fit among the competing specifications. In the MH step, because the proposal density is tailored, the acceptance rates are roughly between 75% and 90% in all relevant models. The MCMC outputs analyzed in this section are based on 500,000 iterations after discarding 50,000 burn-in iterations for each model. Multiple convergence diagnostics have been conducted for all parameters except the augmented data and the auxiliary parameters. For the latter two, because there are too many of them ($2178\times$ 2), I randomly drew 100 of each and conducted diagnostics on them.

The posteriors of all fixed-effect parameters and selected random-effect parameters are summarized in Figure 2.2. As discussed in the time series literature, due to falsely assuming serially independent errors, estimators of standard errors are biased (Gourieroux, Monfort and Trognon, 1984; Poirier and Ruud, 1988). In addition, with the probit link and its standard identification assumption (i.e., the unit variance of the latent errors), parameters are estimated with their scales adjusted by a function of the actual standard deviation of the errors. However, the standard deviation is not 1 with the presence of serial correlation, but the parameters will be mistakenly interpreted as if it were. For the two reasons, the GLMM-AR(0) models for both of the DGPs produce estimates with notably smaller scales (moving towards 0) and artificially higher levels of certainty (the error bands are smaller). This biasedness also exists for the random coefficients at the two group levels, as is illustrated in Figure 4.12. There is no clear evidence that negative autocorrelation is easier to identify—they are all estimated correctly in terms of directions, although the scales are not precisely estimated. With a larger number of time periods or units or both, the autoregressive

Figure 2.2: Posteriors of GLMM with Different Lag Order



The true values of the paremeters presented in the graphs are as the following: Fixed-1=2, Fixed-2=-4, Fixed-3=0, Fixed-4=4, $\beta_{2,11} = 0.5$, $\beta_{2,12} = 1$, $\beta_{2,13} = -2$, $\beta_{2,51} = 3$, '$\beta_{2,52} = -4$, $\beta_{2,53} = 0$, $\beta_{3,11} = -3$, $\beta_{3,12} = 0.2$, $\beta_{3,21} = 2$, $\beta_{22} = 1$.

coefficients are estimated better, but there is still no evidence suggesting the difference between positive or negative correlation in terms of autocorrelation estimation. In Figure 2.2, the posteriors after serial correlation correction all cover true values, and are very similar cross the models with $p > 0$. Although the true parameters are exactly the same in the two datasets except the serial correlation, differences in the posteriors based on the two datasets are observed. However, those differences may not be caused by the different directions of serial correlation; there is always randomness in the data generation, and the degree of informativeness of two datasets generated even by the same DGP cannot be the same.

The simulation designs put much heterogeneity in both the time and unit dimensions: the random effects $\boldsymbol{\beta}_{2i}$ and $\boldsymbol{\beta}_{3t}$ are generated from distributions centered at zero with large variances. Figure 4.12 shows that the averages of the random effects in both dimensions are correctly estimated as around 0. For most clusters, the random effects are statistically different from zero (either positive or negative with a 95% credibility level). Without modeling heterogeneity, especially in the time dimension that is ignored most of the time in the literature, inferences based on the estimates are misleading or overgeneralized. The graphs also demonstrate that correcting serial correlation is important for better identifying heterogeneity. Neglecting serial correlation results in smaller error bands of the random-effect parameters which are measuring heterogeneity.

Other parameters such as the covariance matrices of $\boldsymbol{b}_i$ and $\boldsymbol{c}_t$ are not estimated as precisely as the higher-level parameters and more sensitive to prior specifications. However, the estimated correlation parameters are close to the true values with the priors used in the simulations. Furthermore, increasing the number of units helps better estimate $\boldsymbol{D}$ and decreases its sensitivity to prior specification. Similarly, more time periods allows $\boldsymbol{E}$ to be identified more precisely.

Figure 2.3: Heterogeneity in Unit and Time Dimension

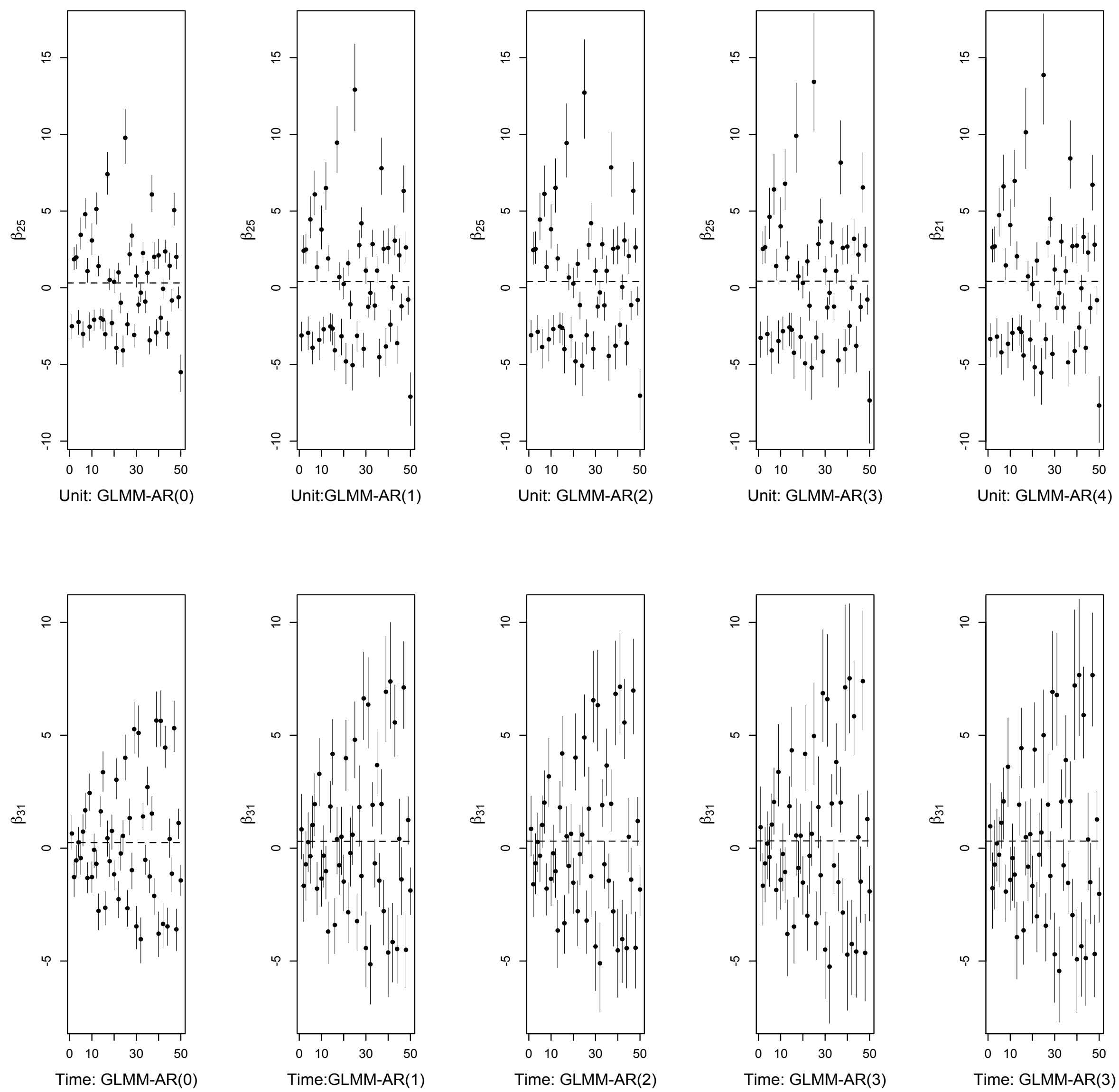


Table 2.1: Marginal Likelihood

DGP	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)
Simulation I ($\rho_1 = 0.7, \rho_2 = 0.2$)	-594.33	-582.44	-574.44	-560.85	-569.47
Simulation II ($\rho_1 = -0.5, \rho_2 = -0.3$)	-570.91	-551.46	-544.25	-558.36	$-$

2.6 Empirical Demonstrations

I use two empirical examples to illustrate how the proposed model and methods can improve reliability of statistical inferences and forecasts in discrete TSCS data analysis. Both examples are about political instability in the sub-Saharan countries: the first studies state failure generally and the second focuses on a particular kind of state failure, civil war. The state failure example shows that by controlling for multiple sources of confounding and making good use of information in the error term, the proposed model fits the data much better than conventional models and dramatically improves within-sample forecasts. The civil war example is used to highlight the danger of ignoring serial correlation diagnosis in TSCS data analysis—spurious regressions when the dynamic process is not stationary or the included regressors fail to cointegrate with the response variable. The GLMM-AR(p) model can serve as a cointegration test because it requires the stationary error process to be stationary, which means that in the model either $\boldsymbol{z}_i$ is stationary or it is cointegrated with the included regressors. In the example, the pooled model finds several "important" explanatory variables, which is unreliable because of the suspected nonstationarity. In the GLMM-AR(p) model, the autoregressive coefficients have the tendency of going out of the stationary space and the MCMC process is terminated when no proposals outside the unit circle can be drawn or accepted.

2.6.1 State Failure in the Sub-Saharan Africa

Since 1994, the Political Instability Task Force (PITF) has been working on detecting important factors which affect state failure risk and building a two-year-ahead early-warning system for the purpose of providing valuable information for policy making. The researchers in the PITF project have built a huge database including all the independent states with population of at least 500,000 since 1955 and collected data on 1,200 variables (see the published four comprehensive resports, *Phase I Findings* to *Phase IV Findings*(Gurr, Harff and Marshall, 2009)). One of their local models focuses on the sub-Saharan countries which are particularly interesting and important because they experienced most state failures occurring in the sample years. King and Zeng (2001*b*) provide a comprehensive critique on the methodology applied by the PITF for drawing causal inferences and conducting forecasts, and they also give suggestions on how to improve state failure data analysis. From a different perspective, I use the sub-Saharan Africa model as an example to illustrate how the GLMM-AR(p) model improves statistical inferences and forecasts.

First, I impute the missing data[10], and, based on the imputed complete dataset, there are 40 sub-Saharan countries observed in the dataset during the time period of 1956 to 1995. Because of the missingness in the response variable, the data structure is not balanced—the minimum number of observations of a country is 3 (Eritrea and Ethiopia), and the maximum number is 40 (Liberia, South Africa, Switzerland). The average number of observations of a country is 30.4 with standard deviation as 8.58. In the spatial dimension, the minimum number of observations in a year is 3 (1956-1959) , and only in two years there is no country missing (1993, 1994). In average, 30.4 countries are observed in a year with the standard deviation as 11.39. There are 1214 country-years in total, and 446 state failures are observed. Because the proportion of events is 36.74%, state failure is not a rare event in the sub-Saharan region. Case-control resampling (Breslow, 1996; King and Zeng, 2001*a*,*c*) does not

[10]Bayesian data augmentation is a better method to handle missingness (Gill, 2007), but due to the very large proportion of missingness in the dataset and more than 2,000 parameters already in the GLMM-AR(p) model in this application, Bayesian data augmentation for missing data will slow down the MCMC process greatly; therefore, I use the multiple imputation method to handle the missingness before the Bayesian MCMC simulation.

Table 2.2: Within-Group Variation of Variables: State Failure Study

Variable	Symbol	Within-Country Variation			Within-Year Variation		
		Min	Mean	Max	Min	Mean	Max
State Failure	`failure`	0.25	0.42	0.58	0.39	0.49	0.58
Democracy	`demo`	0.00	1.51	3.72	2.76	3.38	5.13
Party Fractionalization	`partyfrac`	0.00	0.08	2.26	0.00	0.12	0.86
Party Legitimacy	`partyleg`	0.00	0.77	1.32	0.82	1.09	1.53
Regime Durability	`regundur`	0.00	0.08	0.28	0.06	0.19	0.22
Calories per capita (consumed)	`calory`	0.03	0.17	0.35	0.29	0.35	0.44
GDP per catipa	`gdppc`	0.12	0.57	6.90	0.54	1.68	3.61
Neighbors in Conflict	`neighI`	0.00	0.70	1.78	0.00	0.98	1.58
Neighbors in Civil/Ethnic War	`neighII`	0.00	0.62	1.36	0.00	0.85	1.53
Infant Mortality	`infmort`	0.07	0.29	0.99	0.04	0.53	0.73
Political Terror Scale	`terror`	0.49	0.76	1.15	0.00	0.84	1.23
Political Discrimination	`discrim`	0.00	0.66	0.89	0.00	0.68	0.84
Secondary School Enrollment	`enroll`	0.00	0.01	0.03	0.01	0.02	0.02
Change of Democracy	`cdemo`	0.00	1.14	2.53	0.00	1.08	4.95
Trade Openness	`trade`	0.10	0.34	0.57	0.02	0.48	0.62

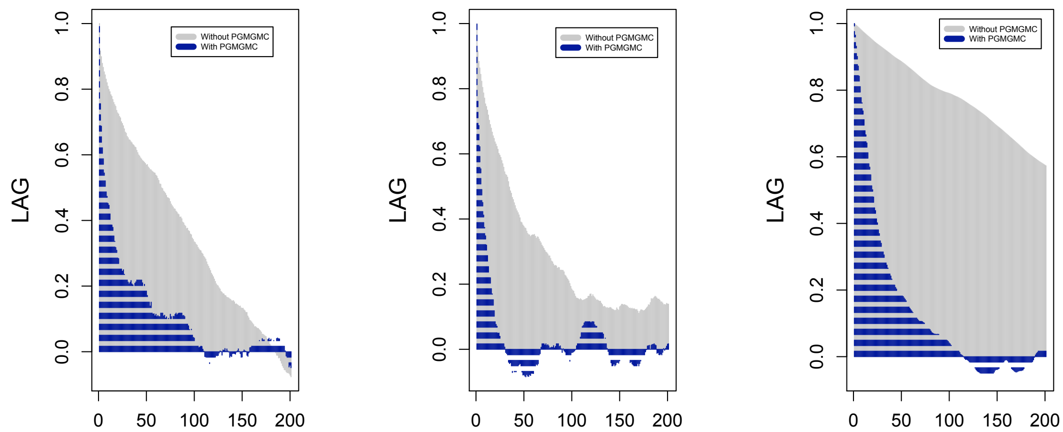
use all the available information and is not needed or preferred in this case, though it is used by both PITF and King and Zeng (2001b). In this chapter, I apply TSCS analysis, instead.

As the PITF and King and Zeng (2001b) do, I lag all the explanatory variables for two years, not only to avoid simultaneity, but also to use the statistical model as an early-warning system. Table 2.2 summarizes the within-country and within-year variances of the variables in the model[11]. For the within-country variations, three variables, *regime durability, party fractionalization and male secondary school enrollment*, are moving slowly; and the within-year variations of the included variables have large variation, suggesting considerable observed heterogeneity among those countries. Because the variable *trade openness* is found important by the PITF and other studies (Beck et al., 2002) but suggested to have an unclear effect by others (King and Zeng (2001b)), I test whether its effect varies across countries by assigning it with a random coefficient. I also include a country-specific and year-specific random intercepts in the model.

The prior assignments in this example are similar to the simulation studies. I run the GLMM-AR(p) models with $p = 0, 1, 2, 3$ and compare them with two other competing models, namely, the completely pooled probit model (PROBIT) and a generalized linear multilevel model only with the country-level random effects (a random coefficient of trade openness and country-specific intercept), which is denoted as GLMM-CL1. The PROBIT, GLMM-CL1 and the GLMM-AR(0) model all assume that the errors are uncorrelated.

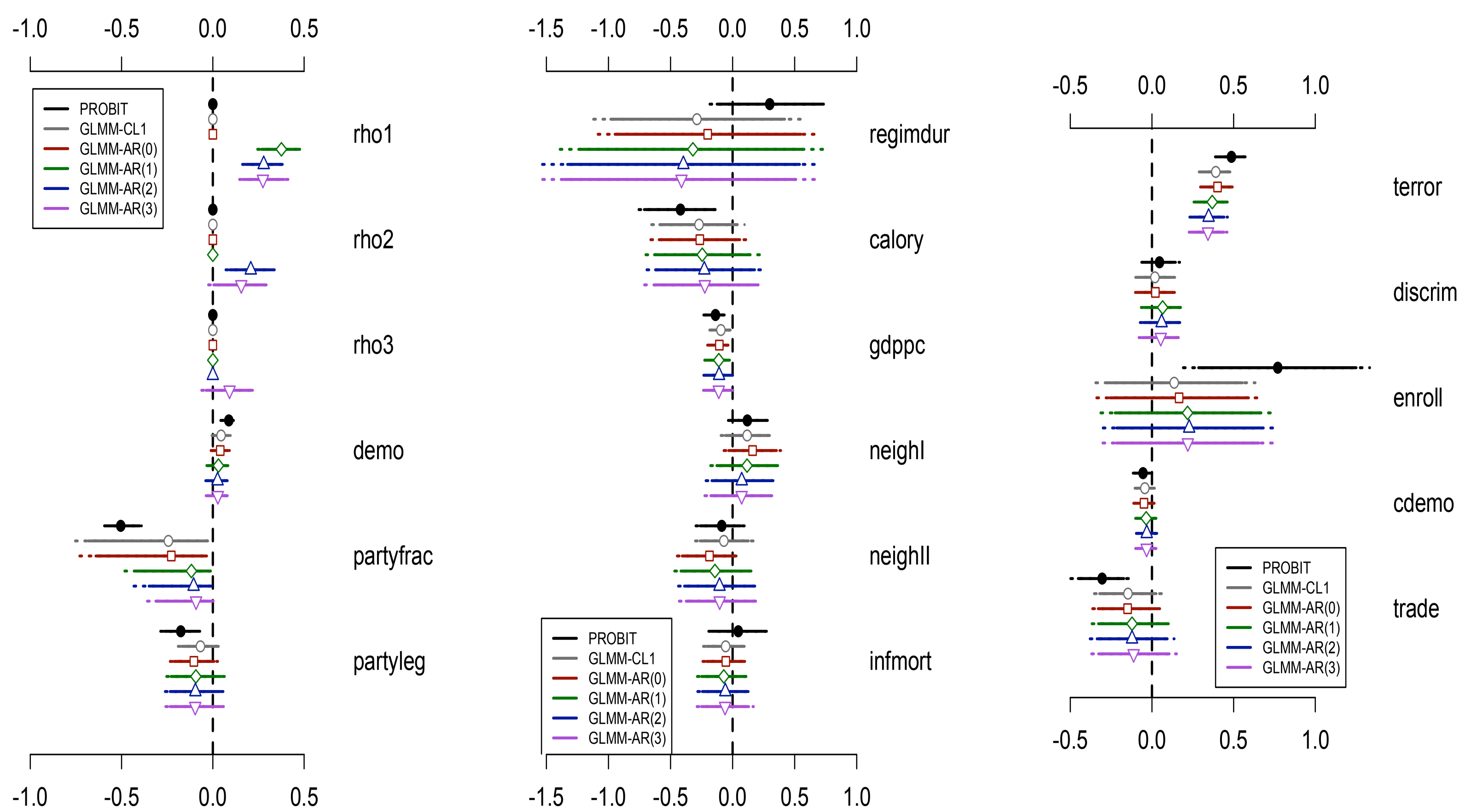
[11]The variables are chosen from the "candidate" covariates in *PITF Phase III Findings*(p. 24) by using stochastic search variable selection method. I include all the variables with posterior probabilities of being included higher than 0.5.

Figure 2.4: Mixing Improvement by the PGM-MGMC Updating: State Failure Study



The PGM-MGMC method reduces within-chain autocorrelations and improve MCMC mixing dramatically, as is shown in Figure 2.4. The posterior summary and marginal likelihood of each model are reported in Figure 2.5. Positive correlation of the errors is found with $\hat{\mathrm{E}}(\rho_1) = 0.37$ based on the GLMM-AR(1) and $\hat{\mathrm{E}}(\rho_1) = 0.27, \hat{\mathrm{E}}(\rho_2) = 0.22$ according to the GLMM-AR(2) model, both of which are with a 95% credible level. The GLMM-AR(3) has a similar posterior of ρ_1 to the GLMM-AR(2), but splits the effect of lag 2 into lag 2 and 3 both of which are with a low credible level. Ignoring serial correlation and heterogeneities, the probit model (black lines in the figures) generates estimates on both the means and standard errors different from those with serial correlation correction and heterogeneity control. It exaggerates the effects of almost all the variables and leads to over-confidence in their effects. It finds 9 out 13 of the explanatory variables are important with high certainty, and strangely suggests that higher male secondary school enrollment leads to higher risk of state failure, which is contradictory to the theories. Controlling for heterogeneity in one or two dimensions makes 6 out of the 9 "important" variables

Figure 2.5: Posterior Summary with 95% Credible Interval (Six Models): State Failure Study

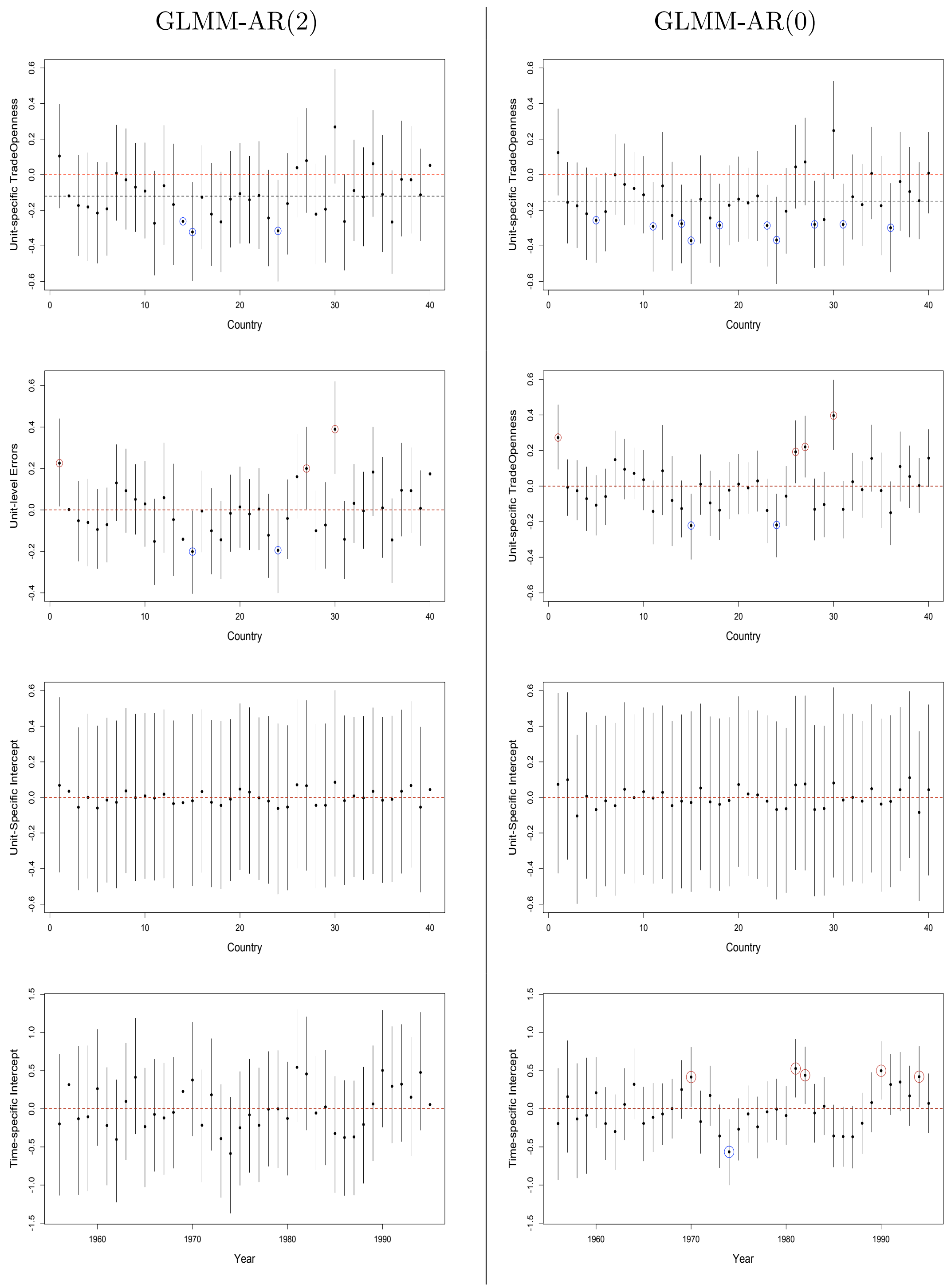


The marginal likelihood of the five PROBIT: −584.794; GLMM-CL1: −511.185; GLMM-AR(0): −503.931; GLMM-AR(1): −470.357; GLMM-AR(2): −468.638.; and GLMM-AR(3): −473.560

based on the simple probit model lose their importance and certainty, which suggests possible edogeneity caused by the correlation between omitted heterogeneites and some of the regressors in the regression. Correcting serial correlation in the errors does not make significant changes in the posteriors compared to the GLMM-CL1 model. Nonetheless, like in the simulation studies, modeling serial correlation leads to larger error bands for all the variables except `partyfrac`, the variable of *party fractionalization*. Somewhat unlike what is observed in the simulated data studies, the posterior means of the coefficients in the GLMM-AR(p) models with $p > 0$ are not always larger than those in the GLMM-AR(0) or GLMM-CL1 model. Note that in the simulation studies, the DGPs do not allow the errors to be directly or indirectly correlated with the regressors, but in this empirical example with all variables lagged for two time periods, it is likely that the error term is correlated with some of the regressors through ϵ_{t-2}. However, the correlation must be weak since the posteriors in the five models are not much different in general. The posteriors in the GLMM-AR(p) models with $p > 0$ are very similar, and the Bayes Factors suggest that the GLMM-AR(2) model is the best one, slightly better than the GLMM-AR(1) model (the Bayes Factor is 0.74), and decisively better than the GLMM-AR(3) model (the Bayes Factor is 2.14). It has much better goodness-of-fit than the multilevel models without modeling serial correlation (the Bayes Factor of the GLMM-AR(2) versus the GLMM-AR(0) is 15.33, and that of the GLMM-AR(2) versus the GLMM-CL1 is 18.48). The PROBIT model has very poor model quality, and the Bayes Factor of it versus the best model is -50.45.

Figure 2.6 compares the random effects estimated by the GLMM-AR(2) and the GLMM-AR(0) and shows that serial correlation correction make much difference in estimating the random-effect coefficients. First, it is easy to see that heterogeneity exits in both the time and spatial dimensions. The effect of *trade openness* varies from

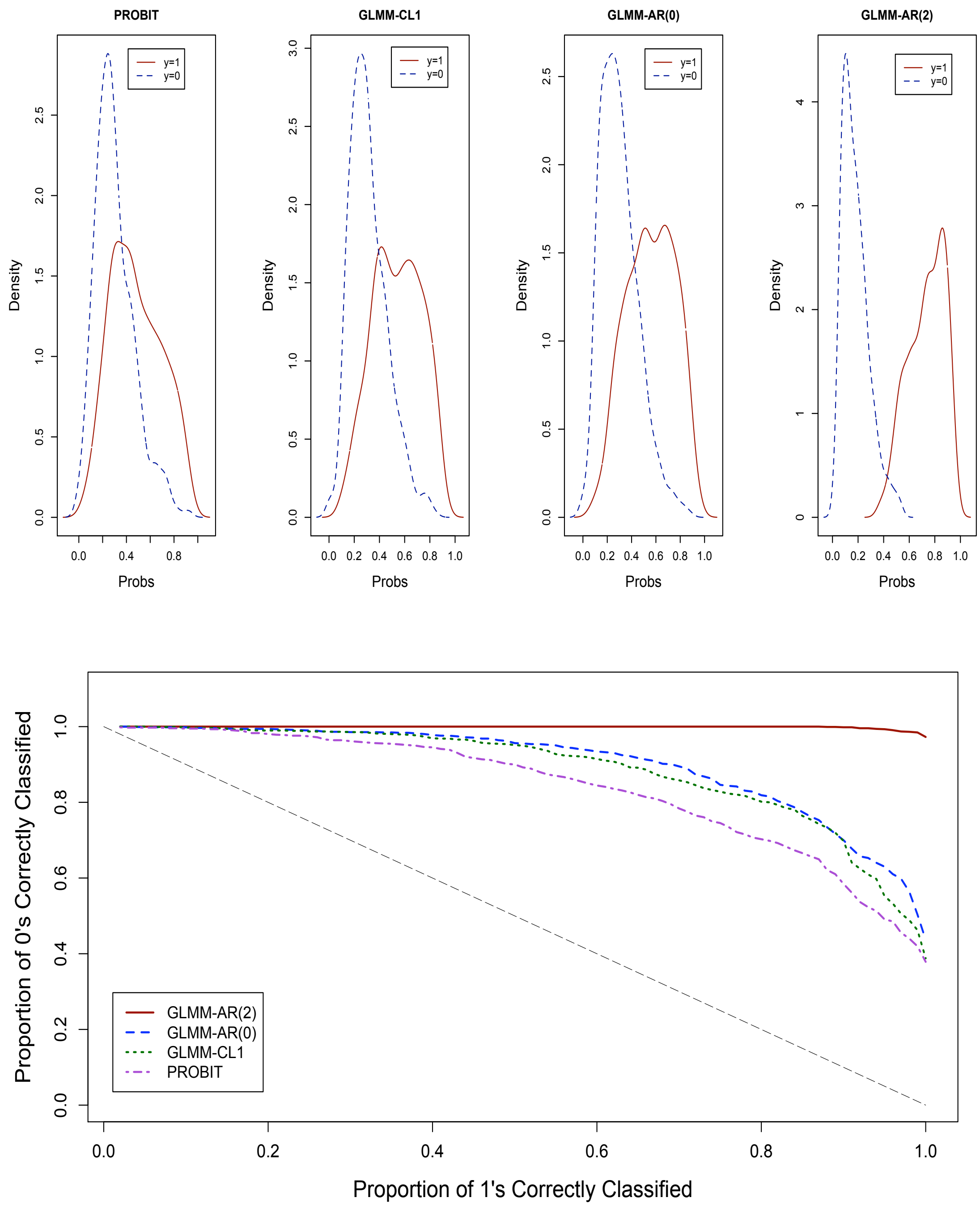
Figure 2.6: Random Intercepts and Random Effects



country to country, which can be substantively attributed to the different structure of international trade in different countries. Trade openness is often suggested to reduce state failure risk because it implies a deeper international engagement and a bigger international influence on domestic politics (*PITF Phase III Findings, 2000*); however, if the major products exported from a given country are primary goods, this trade structure may suggest a country to be more vulnerable to internal instability since primary commodity exports indicate high dependency on natural resources and may be associated with a weak government or more profitable rebellions for more lootable resources (Doyle and Sambanix, 2000; Sambanis, 2001; Collier and Hoeffler, 2004). Working on the global model, King and Zeng (2001b) do not find that trade openness is a statistically reliable variable. Based on the GLMM-AR(2) model of the Sub-Saharan Africa, there are only three countries (Gabon, Gambia, and Malawi) in which trade openness decreases state failure risk at a 90% credible level. For other countries, an open market seems not to have clear effect on state failure. The country-specific random intercept (in the third row in Figure 2.6) shows relative homogeneity, which implies that there is not much unobserved heterogeneity after including the country-level errors (graphs in the second row in Figure 2.6). However, the year-specific intercept (in the bottom row) shows that heterogeneity exists across years. Compared to the random effects estimated by the GLMM-AR(2) model, those based on the GLMM-AR(0) have smaller error bands and consequently are overconfident about the effect of trade openness (they suggest that trade openness has negative effect on state failure risk in 10 instead of 3 countries with a 90% credible interval).

The differences of the posteriors based on the competing models are also reflected by their within-sample forecast performance. Because one of the main goals in the PITF study is to build a forecasting system based on statistical models, and also because prediction is another means of assessing model quality, I further compare

Figure 2.7: Comparison of Within-Sample Predicted Probs.



the models with within-sample predicting. To avoid setting arbitrary or *post hoc* thresholds for classifying failures or non-failures, I simply report the numeric predictive probabilities of state failure in all country-years, and compare the distribution of predictive probabilities of the failure group and that of the non-failure group. As shown in the upper graphs in Figure 3.3, the pooled probit model works poorly to distinguish failures from non-failures, and the density kernels of the two groups have a large overlapping area. The GLMM-CL1 model separates the two densities better than the pooled model, but the two kernels are still not well separated. The GLMM-AR(0) model further considers heterogeneity across time, and slightly reduces the overlapping area compared to the GLMM-CL1 model. The GLMM-AR(2) model, which considers the dynamics in the errors and makes use of the information ignored in the former models, classifies the two groups much more accurately: the two density kernels are well separated from each other, and only a very small part at the tails is connected with each other. The second way to evaluate and compare the predicting performance of the competing models used here is the *Receiver-Operating Characteristic (ROC) curve*, as King and Zeng (2001*b*) do. The idea of using the ROC curve to evaluate models' predicting performance is simple: given a level of correct classification of one group (say, the failure group), the model performs better if it has a higher rate of correct classification of the other group (say, the nonfailure group). Graphically, the curve dominates other curves represents the best model among the competing ones. This approach has the advantage of avoiding assigning any fixed threshold for classification and assessing model performance based on a specific but often arbitrary cutoff value. In Figure 3.3, the diagonal line is just used for reference, indicating the extreme situation that the densities of the two groups are completely overlapped. In the figure, the ROC curve based on the completely pooled probit model is the lowest one. The GLMM-CL1 model improves forecasting by modeling heterogeneity among countries, and its curve globally dominates that of the pooled probit model. The GLMM-AR(0) model has an ROC curve globally but marginally above the one of the GLMM-CL1, but it is dominated everywhere by the curve of the GLMM-AR(2) model. The ROC curve of the GLMM-AR(2) model is almost a horizontal line, indicating that since the densities of the predictive probabilities of the failure and nonfailure groups are well separated, there is barely a trade-off between the two types of classification.

2.6.2 Civil War Duration in the Sub-Saharan Africa

Political scientists have noticed the problem of nonstationarity in dynamic analysis and applied various unit-root and cointegration tests in time series analysis since a long time ago (Beck, 1993; Durr, 1993; Smith, 1993; Williams, 1993; Box-Steffensmeier and Tomlinson, 2000; DeBoef, 2001; Williams, 1993). However, the concern of nonstationary in linear time-series regressions has not been fully extended to TSCS analysis in political science, though in econometrics, unit roots and cointegration testing on panel data is an important on-going research field. Since TSCS data involves dynamic analysis, those problems causing trouble in time series analysis also apply to the TSCS (panel) data. In this example, I use the empirical study on civil war duration to show that the GLMM-AR(p) model can serve as a stationarity test on the TSCS model and helps detect spurious relationships between the response and explanatory variables resulting from nonstationarity and no cointegration. According to the definition of cointegration, the residuals of a model should be stationary whenever the regressors are cointegrated with the response variable (Hamilton, 1994, pp.571-75); otherwise, statistical inferences are based on spurious regressions and, therefore, unreliable. The GLMM-AR(p) model directly analyzes the residual (the error) process by assuming the process is stationary. If this assumption is violated, the model should be re-specified to achieve cointegration or different models should be applied (such as differencing or survival analysis). The MCMC simulation will give valuable information about whether the error process is stationary; if the simulation process shows difficulty (taking an abnormally long time) drawing or accepting legitimate proposals (within the stationary space) for autoregressive coefficients, nonstationary is suspected.

The civil war literature can be categorized into studies on war onset, war duration, and war termination (Sambanis, 2002). Some quantitative research, applying survival analysis, has found that civil war onset and duration are two different processes and require different theories to explain them (Collier, Hoeffler and Soderbom, 2004; Fearon, 2004). However, others, such as Edward Miguel and Sergenti (2004), suggest that the theories applied to civil war onset are also relevant to civil war duration, and their arguments are often based on more limited samples (such as the sub-Saharan countries) and panel data analysis. In this subsection, I do not attempt to solve this debate; instead, I use the GLMM-AR(p) model to show that TSCS analysis on civil war duration without testing cointegration can produce unreliable results, which could shed light on the debate and disagreement mentioned above. I use the same dataset in Edward Miguel and Sergenti (2004); the response variable is civil war duration. As Fearon and Laitin (2003*b*) point out, civil wars tend to be cumulative and last for multiple periods. In the dataset, civil war (coded as 1) is often followed by civil wars in multiple subsequent time periods, sometimes for more than 10 years, demonstrating a very strong path-dependence. The underlying propensity for a country to stay in a war may drift all over the place, but we cannot conduct unit root tests on the observed dichotomous response variable to learn whether this underlying process is stationary or not. On the other hand, many of the explanatory variables included in the model specified in Edward Miguel and Sergenti (2004) and other important models on war duration are slow-moving, as shown in Table 2.3. These data features present the necessity of testing whether the underlying dynamic process of the propensity for a country staying in war is stationary, or whether those covariates are drifting together with, and cointegrated with, the war propensity.

Edward Miguel and Sergenti (2004) use *rainfall* as an instrumental variable for *economic growth* to avoid endogeneity, but this IV is weak and it is difficult to justify that rainfall affects civil war duration only through its effect on economic growth (in other words, it is uncorrelated with the error term) . Also, if economic growth data cannot help achieve cointegration, it is unlikely that using the rainfall data is able to do it. Therefore, I use the lagged *economic growth rate* directly. In their paper, they also test whether economic growth has a varying effect on countries with different conditions for generating grievance and insurgency opportunities. I specify the GLMM-AR(p) model with both country- and year-specific random intercepts and a random coefficient for the variable of *economic growth*. The definitions and within-group variations are summarized in Table 2.3. I do not include two variables often used in civil war onset models—*noncontigeous states* and *new states*—because they either have no overall variation at all or very little variation in the sample country-years, which does not allow us to learn anything about their effects. The variation of *economic growth*'s effect is further explained by the same five explanatory variables—mountainous, oil-exporting country, economic development level, male secondary school enrollment, and democracy— as in Edward Miguel and Sergenti (2004). There are 743 observations (country-years) in total and 182 ongoing wars (24.50%). The data structure is unbalanced—the minimum number of observations of a country is 9 (Namibia), and the maximum number is 19. The average number of observations of a country is 18.1

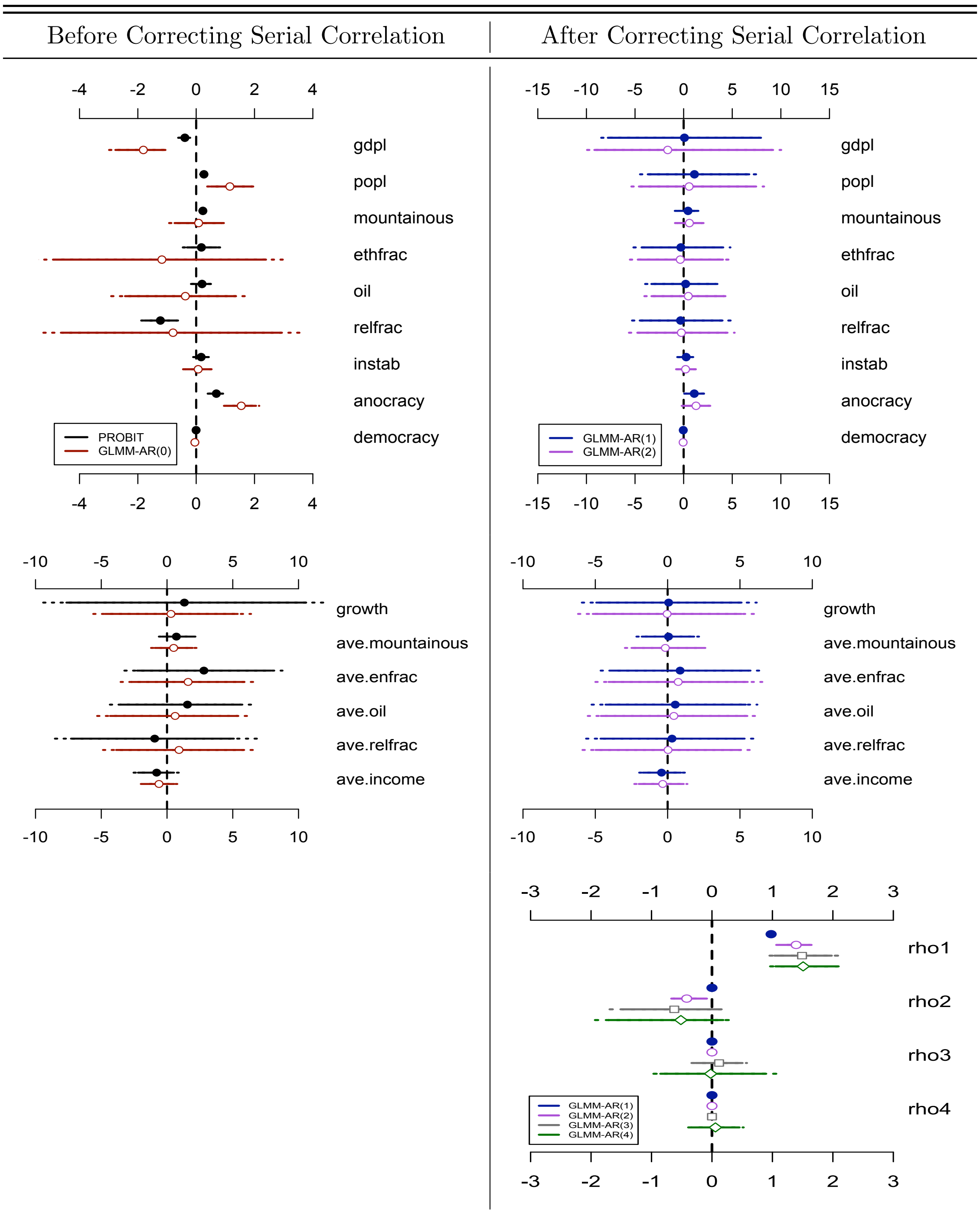
Table 2.3: Within-Group Variation of Variables: Civil War Duration Study

Variable	Symbol	Within-Country Variation			Within-Year Variation		
		Min	Mean	Max	Min	Mean	Max
Ongoing War	`war`	0.00	0.17	0.51	0.36	0.43	0.47
Income per capita (log)$_{t-1}$	`gdpl`	0.02	0.11	0.33	0.55	0.60	0.66
Population (log) $_{t-1}$	`popl`	0.05	0.15	0.21	1.17	1.21	1.23
Mountainous	`mountainous`	0.00	0.00	0.00	1.43	1.45	1.50
Ethno-lingustic fractionalization	`ethfrac`	0.00	0.02	0.51	0.23	0.24	0.25
Religious fractionalization	`relfrac`	0.00	0.02	0.51	0.30	0.33	0.36
Oil-exporting country	`oil`	0.00	0.00	0.00	0.18	0.19	0.20
Political instability	`instab`	0.00	0.35	0.51	0.22	0.38	0.51
Ancocracy $_{t-1}$	`anocracy`	0.00	0.29	0.51	0.22	0.39	0.51
Democracy $_{t-1}$	`democracy`	0.00	3.48	7.21	3.99	5.05	5.96
GDP growth rate $_{t-1}$	`growth`	0.02	0.05	0.20	0.04	0.07	0.13

with standard deviation as 2.56. The minimum number of observations of a year is 36 (1999), and the mean is 39.1 with standard deviation as 1.89.

In Figure 2.8, the graphs in the left column present the posteriors based on the models without considering serial correlation in the errors. The GLMM-AR(0) model, by controlling for heterogeneities in both serial and spatial dimensions, produces different estimates than those produced by the simple probit model, have much larger error bands, and find fewer variables important. Nonetheless, both models suggest that most theories applied to civil war onset are relevant to civil war duration: in terms of their importance and their effect directions, most covariates perform similarly as in civil war onset models such as in Fearon and Laitin (2003*b*), Cederman and Girardin (2007), and Fearon, Kasara and Laitin (2007). Then, I run the GLMM-AR(p) models with $p = 1, 2, 3, 4$, and the autoregressive coefficients in all those models approach unit roots soon after the simulations start. After varying numbers of iterations (from the longest time of about 120,000 iterations for the GLMM-AR(1) model to the shortest of about 70,000 iterations for the GLMM-AR(3) model), those simulations are aborted because they all encounter difficulty drawing or accepting proposals in the stationary spaces. In the right column of Figure 2.8, I summarize the draws based on the GLMM-AR(1) and -AR(2) models before the MCMC simulations are abnormally terminated. Note that the outputs cannot tell us anything about the parameters since they are not justified to be sampled from the ergodic distributions. I use them only to demonstrate how different the empirical results could be when we take seriously the dynamics and cointegration problems. The last graph in the right column reports the samples of the autoregressive coefficients in the four models before the MCMC simulations halt. It is clear that they are very close to the unit circle and have the tendency to go out of the stationary space.

Figure 2.8: Random Intercepts and Random Effects



The marginal likelihood of the five models—PROBIT: −399.514; GLMM-AR(0): −327.632. For GLMM-AR(1) to -AR(4), because the MCMC process halted for not being able to obtain legitimate proposals for autoregressive coefficients, no convergence diagnosis can be done and the marginal likelihood is meaningless since the simulation is terminated abnormally.

The evidence found in the MCMC simulation processes based on the GLMM-AR(p) models suggests that the slow-moving explanatory variables are not cointegrated with civil war duration, and statistical inferences based on the pooled probit or multilevel analysis without considering stationarity of the error process are spurious. The solution might be to add a lagged response variable as a regressor in addition to the autoregressive errors since the lagged on-going war is likely to drift with the current war, or we can try to find other time-varying explanatory variables which can be cointegrated with the propensity of civil war duration. Solving the problem could be a separate research project, and here I only illustrate the importance of considering the dynamic process of the error term in TSCS analysis and show that the GLMM-AR(p) model can be used for the purpose of cointegration testing and avoiding spurious regressions in discrete TSCS analysis.

2.7 Discussion

In TSCS analysis, modeling inter-temporal dependence, contemporary correlation, and heterogeneity at the same time is required by the TSCS data structure and their correlated design. But for categorical responses, this is difficult and complicated to do because of the complex errors structure caused by serially correlated errors and nonlinearity of the model. Moreover, since TSCS analysis investigates dynamic processes, unit roots and cointegration tests are necessary for avoiding spurious regressions; however, those tests are quite challenging for discrete TSCS data and still an active research filed in panel data econometric analysis. This chapter proposes a Bayesian GLML-AR(p) model as a solution, and developes an MCMC algorithm which improves the conventional simulation schemes by orthogonalizing the error term and adding an auxiliary parameter $\boldsymbol{u}$. This method facilitates constructing the conditional distributions of the time-specific random-effect coefficients, and achieves the goal of conducting data augmentation in one block. It also dramatically simplifies and stabilizes estimating the Bayes Factor, which makes it much easier and more reliable to use the Bayes Factor to determine the order of the error autoregressive process. I further improve simulation efficiency by using the PGM-MGMC updating. A possible future extension of this model can be to relax the assumption that time-specific common shocks have the same impact on different units, which may turn out to be a multifactor residual-style model with a hierarchical setup and serial correlated errors.

Chapter 3

Ethnic Minority Rule and Civil War Onset: How Much Background Factors and Dynamics Matter

One of the surprising empirical findings in the quantitative literature on civil war is that ethnicity has an unclear effect on civil war, and the relationship between the two is not statistically robust to different samples or model specifications (see a comprehesive review in Sambanis (2002)), despite strong and direct causalities suggested by nationalist theories (Huntington, 1968; Russett, 1964; Scott, 1976; Muller, 1985; Deutsch, 1953; Anderson, 1983; Horowitz, 1985; Ignatieff, 1993; Huntington, 1996; Wimmer, 2002) and the often highlighted ethnic factors in civil war case studies (Collier and Sambanis, 2005). Two recent APSR papers (Cederman and Girardin, 2007; Fearon, Kasara and Laitin, 2007) revisited this question and focused on the state (group) level. Instead of using the widely-employed measure of ethno-linguistic fractionalization (ELF) and other proxies for nationalist grievance (such as ethnic divisions (Ellingsen, 2000) and polarization (Reynal-Querrol, 2002)), they measured ethnic minority rule (EMR) which is directly related to violence against the state. Their perspective is valuable because the investigation on EMR's effect shortens the causal chain between ethnicity and civil war by clearly asking a more straightforward question: how does ethnic political dominance affect civil war risk (Bates, 1999)? In addition, EMR is an theoretically important and interesting predictor of civil war, and answering the question about the effect of ethnic minority rule, by itself, contributes to understanding civil war.

However, the variable of EMR has serious measurement problems. The measure constructed by Cederman and Girardin (2007, henceforth, CG), focusing on the ethnicities of politicians in important governmental positions, faces aggregation difficulties; and Fearon et al (2007, henceforth, FKL) simply used government leader's ethnicity, which is subject to the problem that the leader's ethnic group is not necessarily the dominant group in politics, such as African-Americans in the United States under the Obama government. Hence, it is unrealistic to expect that EMR, coded with either of the two different criteria, has the same meaning, or generates the same degree of nationalist grievance, or provides the same rebellion opportunities in different countries with different political, economic, and social backgrounds. This poses a challenge for large N studies, because, without considering those variations, the conclusions based on quantitative analyses can be overgeneralized and misleading. As is explicitly suggested by FKL, in order to identify causalities instead of only correlations, we should model the variation of EMR's effect across countries and analyze the background factors which are likely to alter the relationship between EMR and civil war onset. In addition, to find a reliable causal relationship between

EMR and civil war onset, other sources of confounding also need to be controlled for. Because the civil war data are time-series cross-sectional and rich in structure, there are multiple confounders, including serial correlation and unobserved heterogeneity in both the time and spatial dimensions. Without analyzing those factors, endogeneity and inefficiency of estimators will lead to unreliable statistical inferences and poor forecasts.

This chapter focuses on explaining the heterogeneous effect of EMR on civil war onset by handling the challenges resulting from observed and unobserved background factors and the rich structure of the civil war data. The solution proposed in this chaper is to use a generalized linear multilevel model which is able to explain the variation of EMR's effect by using a country-level regression. To control for serial correlation (the dynamic process in the error term) and other unobserved heterogeneity in both the time and spatial dimensions, the model specifies a pth order autoregressive error process and two unnested sources of clustering at the year and country levels. I estimate the model with various lag orders and compare competing models. The empirical results suggest strong positive autocorrelation lasting for multiple time periods and a salient variation of EMR's effect on civil war onset across countries. Political instability (regime stability) is the most important background factor which amplifies EMR's effect on civil war onset. Male secondary school enrollment, which proxies for governance quality and the strength of the government, is likely to reduce the effect of EMR on civil war risk, but its intervening effect is not with high certainty. No evidence is found to support the importance of ethnic diversity for the relationship between EMR and civil war onset. The robustness of those findings is checked with different variable selections and various measures of EMR. The generalized linear multilevel model with an autoregressive error process (henceforth, GLMM-AR(p) model) improves the reliability of statistical inferences and the performance of within-sample forecasts.

3.1 Grievance and Opportunities of Rebellion

Empirical studies on civil war often find that once economic and other material factors are controlled for, the variables proxying for nationalist grievance are not important (Edward Miguel and Sergenti, 2004; Fearon, 2004; Collier and Hoeffler, 2004; Fearon and Laitin, 2003*b*; Collier and Hoeffler, 2002; Sarkees, 2000*b*). This finding has been interpreted with an emphasis on the opportunities for mobilizing and financing rebellion. Political and economic rationales triumph nationalist grievance, and the more important elements in civil war are the expected gain and the opportunity cost on the rebellion side, and the counter-insurgent capabilities on the state side. Particularly, Fearon and Laitin (2003b) argue that civil war risk should be better explained by focusing on a central government's financial, organizational, and political weaknesses and the conditions favoring insurgencies (insurgent relative to counterinsurgent capabilities). Collier and Hoeffler (2004) also stress that understanding civil war should analyze both the motive and opportunity mechanisms: bad economic situations not only generate grievance (which could then become *nationalist* grievance), but also weaken the central government's counterinsurgent capacities, which consequently increases the expected gain from rebellion and decreases the opportunity cost; other material factors, such as lootable resources, can be interpreted not only as favorable conditions for financing rebellion but also mechanisms creating motives for rebellion ("greed"), since looting the resources by itself can be one of the goals or motives of rebellion. Associated with lootable resources, primary commodity exports as a large proportion of GDP has also been used as an indicator of a weaker government, and, accordingly, better chance for the rebels to take over power.

Recently, CG casted doubt on those interpretations which are based on economic and political rational choice theories. CG attributed the empirical null finding on ethnicity's effect to the measures of objective nationalist grievance, especially the widely-used ELF index. They pointed out the gap between a measure of micro-level grievance and an event (civil war) at the macro level. They also proposed an alternative measure, the N^* index, which is attempted to directly measure "control" of the state by different ethnic groups. Since a civil war is a conflict for ownership of the state and occurs at the group level[1], state control by ethnic groups does not have the same difficulty as the ELF index has in explaining how intergroup conflicts escalate to a war against the state. Their research is an effort to overcome the common problem of using micro-level data to analyze a macro-level event in quantitative studies on civil war (Sambanis, 2002). This gap makes empirical findings hard to interpret, and theories are also difficult to test precisely, since the causal chain is long and complex and can be altered by many intervening factors.

However, their study implies only one mechanism through which EMR affects civil war onset: EMR generates nationalist grievance against the government and, consequently, increases civil war risk. This hypothesis is based on the theory that the nation-building is the driving force of civil war (CG, 2007). Their argument ignores the long-existing economic and political rational choice theories, and does not address the question of how fighting for control of the state can be separated from the opportunity problems and the rational cost-benefit calculation. Another difficulty is measuring the variable of ethnic state control (refer to FKL's criticism of

[1] The government has to be involved in all various definitions of civil war (Small and Singer, 1982; Sarkees, 2000*a*; Fearon and Laitin, 2003*b*; Sambanis, 2004).

this measure for details). CG claimed that their coding criterion was to consider the ethnic identities of politicians in senior governmental positions, but there is a serious aggregation problem and an associated heterogeneity concern: the importance of different positions is different in the same country, and the significance of positions with the same or similar name can vary much across different countries. In addition, the N^* index is time-invariant, and only four countries are identified as "minority ethnic group(s) in power (EGIP)" among 90 sample countries[2], which is hard to believe since the sample time period spans 55 years. FKL showed that CG's empirical findings were highly sensitive to reasonable modifications of coding and relied mainly on three of the four cases of minority EGIP. FKL further proposed an alternative measure which focuses on government leader's ethnicity. They also put the research question more straightforwardly: "(a)re countries at greater risk of civil war when the state is controlled by an ethnic minority?" This question is directly related to the nationalist grievance question but does not imply that the effect of EMR on civil war onset is solely through this mechanism. Nonetheless, nationalist grievance is an important channel, which is highlighted by FKL with a quotation from Gellner (Gellner, 1983, p.1):

> there is one particular form of the violation of the nationalist principle to which nationalist sentiment is quite particularly sensitive: if the rulers of the political unit belong to a nation other that of the majority of the ruled, this, for nationalists, constitutes a quite outstandingly intolerable breach of political propriety.

Leader's ethnicity, as a measure of EMR, is straightforward and easy to code. It also has the advantage of being time-varying and being able to include much more

[2]Because of missing data, the number of countries included in their models is 85.

country-years in the analysis (161 countries in 55 years). KFL investigated the relationship between EMR and civil war onset, and their empirical results do not suggest any statistically significant connection between the two. Checking the samples case by case, FKL recognized that using the leader's ethnicity to measure EMR had its own problems. First, in some countries, the ethnic group that the government leader comes from does not necessarily mean that the group is politically dominant. Examples are found in several sample county-years, such as a Scottish prime minister in Britain (Douglas-Home, 1963), the Galician Franco in Spain, the Georgian Stalin in the USSR, Slovak General Secretaries in Czechoslovakia, and rotating presidencies in Switzerland and post-Tito communist Yugoslavia. Recognizing this confounding, they recoded these cases and got similar empirical results. However, there is another important problem which they mentioned in their paper but did not provide a solution. There are observable and unobservable factors that can affect the relationship between leader's ethnicity and civil war since the causal chain is long and indirect. First, as already mentioned, the ethnicity of the leader does not necessarily measure the primary explanatory variable—the dominance of the leader's ethnic group for some (democratic) institutional or (imperial) policy reasons. In this situation the nationalist grievance mechanism is not relevant. Second, both Gellner (1993) and Wimmer (2002) have suggested that minority political dominance may cause nationalist resentment *and* perception of greater opportunities to take over power. Even though leader's ethnicity measures ethnic dominance most of the time as well as FKL claimed, the dichotomous variable does not measure the level of nationalist grievance or rebellion opportunities in different countries. Following the greed-grievance framework of Collier and Hoeffler (2001) and Collier, Hoeffler and Soderbom (2004), motives (mainly grievance) and opportunities are two dimensions for analyzing civil war, and the two are entangled. Rebellion is not solely driven by grievance or anger, but also based on the rational cost-benefit analysis. There are various intervening factors that can alter the mechanism by which EMR affects civil war risk, such as the following factors which theoretically predict the variation of this relationship across countries.

Political Instability

In an unstable regime, a country may be more vulnerable to civil war under ethnic minority rule than under ethnic plurality rule (EPR), at least for two reasons. First, if in a country EMR does generate a higher level of nationalist grievance, political instability can provide greater opportunities to successfully take over the ownership of the state since the government is weak (Collier, Hoeffler and Soderbom, 2004); therefore, increased nationalist grievance under EMR, as increased motives of rebeling, is facilitated with better opportunities provided by political instability, which is consequently more likely to trigger civil war than in politically stable countries. Second, frequent regime transformation by itself can generate more grievance (nationalist or not) and hence higher risk of civil war (Hegre et al., 2001). In this situation, under ethnic minority rule, the ethnic plurality group on the rebellion side can overcome the collective action problem more easily than the minority group(s) because they have a larger pool from which they can mobilize people and recruit soldiers. This is based on the theory that ethnicity provides a tie affiliating the people of the same ethnic group and helps overcome the collective action problem since once an ethnic group is involved in a civil war, participation of the people in that group is hard to avoid (Horowitz, 1985; Rothschild and Foley, 1988). In other words, the threshold of civil war onset can be more easily surpassed by the plurality group (Granovetter, 1978; Sambanis, 2002; Tilly, 2003). Likewise, in a regime which is stable either because the level of grievance is low or because the government is strong and capable, the leader's ethnicity may not make much difference to civil war risk. Therefore, political stability can alter the effect of EMR on civil war propensities and change the difference made by EMR and EPR.

Economy and Governance Capacities

When the economy is good, the opportunity cost of rebellion is high because the expected gain in the labor market is bigger than when the economy is bad. Also, the expected return of rebellion is lower because the government may have more resources to crack down the rebellion and has a better chance to win the war. In a strong economy, whether there is minority rule or majority rule may not make much difference in that the cost-benefit calculation and a strong government predict that rebellion is expected not to be so profitable as in economic bad times (Edward Miguel and Sergenti, 2004). If the economy is bad and the government is weak, both grievance and greed can increase, and then leader's ethnicity may be a more sensitive issue and be blamed for bad policies, and nationalist grievance is generated naturally or artificially; at the same time, the same collective action mechanism stated above also applies to this case. In addition, if EMR is as good as EPR at providing public goods and governance of high quality, leader's ethnicity does not necessarily generate greater nationalist grievance, and civil war risk would not be significantly higher under EMR than uner EPR (Grossman, 1995; Hirschleifer, 1995; Collier and Hoeffler, 2004). Together with the factor of political stability, this suggests that if a government is strong, capable, and willing to provide public goods under a relatively stable regime, leader's ethnicity should make less difference to civil war onset risk.

Ethnicity Diversity

Ethnically more diverse societies may mean a more serious collective action problem (Sambanis, 2002; Collier and Hoeffler, 2000). However, the question on the relationship between EMR and civil war implies that the wars under investigation should occur between the ethnic minority in power and the ethnic plurality group on the rebellion side. If this is true, it can be assumed that the plurality ethnic group is able to recruit soldiers at least from its own ethnic group (Horowitz, 1985; Rothschild and Foley, 1988), and has the advantage of overcoming the collective action problem. According to the measure of ELF, in a less diverse country where the majority has much larger population than the minority group, the politically excluded majority can more easily cross the threshold of civil war on the one hand. On the other, the plurality group will feel more pain at being excluded from the top leadership and hold stronger nationalist grievance. This logic predicts that in less diverse countries or countries with a more polarized ethnicity distribution, EMR may be more likely to cause civil war than in more diverse countries. However, based on the existing literature, ethnic diversity is often suggested to be associated with higher likelihood of civil war because nationalist grievance is more common and inter-ethnic group conflict is more likely. Although the mechanism by which those social conflicts among ethnic groups escalate to a war against the government is unclear, a government headed by an ethnic minority member may be more vulnerable to civil conflicts in an ethnically more diverse country: the population is more equally distributed among several ethnic groups and more groups could be potential challengers to the dominant minority group. Given the high level of nationalist grievance among groups in a more diverse country, EMR provides more opportunities for inter-group grievance to escalate to a war against the government. The collective action approach predicts that *lower* ethnic diversity increases EMR's effect on civil war onset, while the national grievance approach suggests that *higher* ethnic diversity leads to a stronger effect of EMR.

Democracy

Democracy can be another important factor intervening in the relationship between EMR and civil war. Theoretically, a democratically elected leader has his or her legitimacy recognized by the opponents, because one important characteristic of democracy is that the losers in an election admit the legitimacy of the government headed by the winner and will try to take over the office within the democratic institutional framework (in the next election) rather than resorting to rebellion (Dahl, 1989; Gasiorowski, 1995; Acemoglu and Robinson, 2003). Following this logic, it can be expected that, in democracies, EMR does not necessarily incur more civil war than EPR because the leader is elected and has legitimacy, but in non-democracies, the legitimacy of the leader is more questionable and more easily challenged by other groups. Furthermore, there is no "next election" in non-democracies for the ethnic plurality group to lawfully take over power, and they are left with rebellion as the most straightforward, if not the only, avenue to change the ownership of the state. However, this logic is based on the theoretical or ideal type of democracies; in reality, especially in new, immature or unstable democracies, the legitimacy of elected leaders is sometimes not recognized by the challengers who question the fairness and legitimacy of the election. Leader's ethnicity can be easily used as a justification for the plurality group to refuse to recognize the government. At the same time, in immature democracies, repression is reduced and collective action is easier than in authoritarian regimes (Lichbach, 1987; Moore, 1998; Hegre et al., 2001), since the plurality group can use their lawful rights such as gathering, free speech, and protests, for mobilization at early stages of rebellion. Therefore, EMR in those immature democracies may be more vulnerable to civil war than EPR.

Shocks, Historical Memory and other factors

International and historical factors can also alter the degree that EMR affects civil war risk. For example, external shocks could temporally make the country array under the flag, and domestic nationalist grievance is dwarfed by grievance against external entities or by external threats to the survival of the country as a whole. Another example may be historical memory: if a country has experienced conflicts (not necessarily civil war) between the minority group the leader belongs to and the plurality group, or, in other words, if there is unsolved historical hostility between the two groups, it is more likely that EMR will increase civil war risk.

3.2 Methodological Problems

FKL also expressed their concern about the "background" factors which could alter the relationship between EMR and civil war onset. However, the fixed-effect panel model they used is apparently not a solution since the model only controls for unobserved heterogeneity but is not able to capture the coefficient variation of EMR across countries. Another possibility suggested by FKL is to find instrumental variables which are not correlated with the background factors but associated with EMR and civil war. However, it is extremely difficult coming up with such IVs and providing a convincing justification that the those IVs are not uncorrelated with the error term because the background factors are not all known. Not being able to find a satisfactory solution, FKL cautiously interpreted their estimates only as "correlation" instead of "causality".

In fact, to analyze the variation of EMR's impact on civil war across countries, there is a simple solution—a multilevel model with group-level regressions, with which the variation of EMR's effect across countries can be modeled in a group-level regression with its own regressors and error term. Multilevel analysis has been widely used in time-series cross-section data analysis and more generally longitudinal analysis in many disciplines. It has been well developed in both the conventional and Bayesian frameworks (Hagenaars, 1990; Singer and Willett, 2003; Yang, Fu and Land, 2004; Skrondal and Rabe-Hesketh, 2004; Molenberghs and Verbeke, 2005). One particularly important characteristic of multilevel modeling for the current study is that the predictors can enter the model at the correct level, and variations within and between groups can be modeled for both theoretical and methodological reasons (Gill, 2007, p.395). In addition, several slow-moving or time-invariant predictors are important in civil war study, but collinearity is a big concern for the fixed-effect panel model (Shor et al., 2007). In contrast, multilevel modeling avoids this pitfall by placing those variables into the group-level regression (Gelman and Hill, 2006; Gelman, 2006, pp.246-247).

Besides applying a multi-level model to explain the country-level variation of EMR's effect, it is also important to control for other sources of confounding in order to better identify causal relationship rather than only "correlation". The structural characteristics of the civil war data as TSCS data cause correlation from multiple sources. The civil war database used in the present chapter is constructed by Fearon and Laitin (2003*b*); Fearon, Kasara and Laitin (2007). Originally, there are 161 countries in 55 years (from 1945 to 1999) with 6610 observations (country-years). The data structure is not balanced, and some countries start to be observed quite late. Different model specifications contain slightly different sample sizes due to missing data in some variables. Based on the final model specification applied in this chapter which is described in the next section, there are 6210 observations consisting of 155 countries across 55 time periods (years) from 1945 to 1999, among which 104 war onsets are observed[3]. In the unit dimension (countries), the minimum number of

[3]Civil war is a rare event here, but the panel is big and the number of events is not too small. The case-control study is attempted to reduce the expense of collecting enough data for analyzing rare events, and has been used in many empirical studies. King and Zeng (2001*a*) and King and Zeng (2001*c*) introduced it to political science and demonstrated its efficiency and consistency. However, in the present case, the data are given, and a big dataset has already been built for the analysis. From both the conventional and Bayesian point of view, using all information in data is always preferred than resampling from the given dataset and using part of the observations. Resampling will also break the data structure and artificially reduce the TSCS data into cross-sectional data, which results in further information loss. For these reasons, I use TSCS analysis instead of resampling the data and applying the case-control approach.

observations of a country is only 3 and the maximum is 55, and the average number of observations of a unit is 40.1 with a sample standard deviation as large as 15.05. In the spatial dimension, the minimum number of observations in a year is 34, and the maximum is 146. The sample mean of the number of observations in a year is 113 and the standard deviation is 28.15.

Besides the variation of EMR's effect, there is likely to be other unobserved heterogeneity of the sample countries. If this heterpgeneity is correlated with any of the regressors, the estimator is biased. Also, theories of international relations emphasize the importance of the international system. Disturbances in the international system form common observed or unobserved shocks on countries which may affect civil war risk. Again, endogeneity arises when those omitted common shocks are not strictly exogenous. Furthermore, serial correlation is not only a source of inefficiency but also can be a further cause of endogeneity, because several important covariates are conventionally lagged to avoid endogeneity (such as per capita income, population and democracy). In TSCS analysis, lagged values should be used with caution and along with serial correlation correction, because, if the residual in the previous period is correlated with the residual in the current period, i.e., $\mathrm{Cov}(\xi_t, \xi_{t-1}) \neq 0$, and if the some of the covariates are lagged because they are probably correlated with the residual, namely, $\mathrm{Cov}(x_{t-1}, \xi_{t-1}) \neq 0$, we have $\mathrm{Cov}(x_{t-1}, \xi_t) \neq 0$, resulting in endogeneity.

There are multiple ways of controlling for serial correlation sugguested in the political methodological literature (Beck and Katz, 2009, 2004; Beck et al., 2002; Beck, Katz and Tucker, 1998; Beck and Katz, 1995). Fearon and Laitin (2003*b*) report the trouble they encounter when trying to correct serial correlation. They found that conventional methods, such as the lagged dependent variable as a regressor and time splines, did not find serial dependence. Still suspicious of serial independence based on theories and other prior information, they decided to use the variable of lagged *ongoing war* to control for serial correlation. This variable turned out to have considerable explanatory power, and has been adopted by the following studies (Cederman and Girardin, 2007; Fearon, Kasara and Laitin, 2007). Although the lagged ongoing war may partially control for serial correlation as a lagged independent variable approach, there is no direct assurance that serial correlation in the errors is adequately corrected. Also, the sign of the lagged ongoing war in their model is negative, which can be reasonably interpreted as follows: since wars take resources and serve as an outlet of grievance, a ongoing war in the previous time period decreases the propensity of civil war onset in the current period. However, it is not easy to interpret the negative sign as negative serial correlation. In the probit model, serial correlation exists in the latent dynamic process which can be regarded as the dynamic change of the propensity for war onset. Hence, it is difficult to have a theoretical explanation of negative serial correlation of war propensity or risk. In addition, with several explanatory variables lagged (the ongoing war, autocracy, democracy, population, and income), it is important to ensure that the errors approximate a white noise process (uncorrelated), which can only be done by directly modeling the errors. For the purpose of directly and thoroughly correcting serial correlation, I specify a autoregressive process of the error term whose lag order is determined by model comparison based on the Bayes Factor. This method allows the errors to approximate a white noise process and ensures serial independence after correction.

3.3 Model Specification and Data Description

I apply a Bayesian generalized linear multilevel model with pth-order autoregressive errors (GLMM-AR(p)) presented in Chapter refch2 to identify the causal relationship between EMR and civil war onset and to explain the variation of EMR's effect across countries. Based on previous theoretical and methodological discussions and after comparing various preliminary models, I apply the following core model specification to analyze the civil war data and investigate the varying relationship between EMR and civil war risk across countries:

$$\mathtt{Onset}_{i,t_i} = \mathbf{1}(\mathtt{Latent}_{i,t_i} > 0), \tag{3.1}$$

$$\begin{aligned}\mathtt{Latent}_{i,t_i} = {} & \beta_{t_i} + b_{1i} + \beta_{2,i} * \mathtt{EMR} + \beta_{11}\mathtt{ongwar} + \beta_{12} * \mathtt{gdpenl} + \beta_{13} * \mathtt{lpopl} \\ & + \beta_{14} * \mathtt{lmtnest} + \beta_{15} * \mathtt{ncotig} + \beta_{16} * \mathtt{oil} + \beta_{17} * \mathtt{newstate} \\ & + \beta_{18} * \mathtt{instab} + \beta_{18} * \mathtt{polity2} + \beta_{19} * \mathtt{ancol} + \beta_{210} * \mathtt{ethfrac} + \xi_{i,t_i},\end{aligned} \tag{3.2}$$

$$\beta_{2,i} = \beta_{21} + \beta_{22} * \mathtt{minstab} + \beta_{23} * \mathtt{methfrac} + \beta_{24} * \mathtt{second} + b_{2i}, \tag{3.3}$$

$$\beta_{t_i} = \beta_0 + c_{t_i}, \tag{3.4}$$

$$\xi_{i,t_i} = \rho_1 \xi_{i,t_i-1} + ... + \rho_p \xi_{i,t_i-p} + u_{i,t_i}, \tag{3.5}$$

The meanings of the covariate symbols and the within-country and within-year variations of the included variables are summarized in Table 4.2, and the coding criteria of the variables are reported in Fearon and Laitin (2003b,a) and FKL. The individual-level regression, equation (3.2), follows the core specification in Fearon and Laitin (2003), CG, and FKL. Unlike FKL, I include *Democracy* (`polity2`) in the model because CG found that democracy merited inclusion, while FKL did not test it. I use country-specific and year-specific intercepts to control for unobserved heterogeneity in the time and spatial dimensions. Note that this specification cannot be allowed in the fixed-effect panel model because of the slow-moving and time-invariant variables, such as *Mountainous* (`lmtnest`) and *Ethno-linguistic fractionalization* (`ethfrac`). The variable, *Secondary school enrollment*, has no variation across time for almost all countries, and its variation among countries does not change over time, either. This suggests that using its country averages in the group-level regression does not discard information in the data.

In the country-level regression, equation (3.3), the variation of EMR's effect, $\beta_{2,i}$, is explained by both observed (the country-level regressors) and unobserved background factors (the error term, b_{2i}). I do not include *Democracy* as a country-level covariate for two reasons: first, it is a time-varying variable and the average of this variable poorly describes the background; second, preliminary model specifications with democracy averaged within countries have worse model quality than the present one (according to the marginal likelihood), and the posterior distribution of its effect is estimated to be centered at zero and with a small error band. This does not necessarily mean that democracy is not an important factor, and could be attributed to the fact that the average does not capture the background. Therefore, I leave the general regime or institutional factors as unobserved in the error term. *Political instability* is coded by Fearon and Laitin (2003*b*) as "a dummy variable indicating whether the country had a three-or-greater change on the Polity IV regime index in any of the three years prior to the country-year in question", which is labeled by them as "political stability" but actually measures regime stability. The within-unit average of this variable is used in the group level regression. Since whether a regime is stable or not should not be judged by presence or absence of transformation, instead, the frequency of regime change during a certain period time reflects political instability. Therefore, the average is a good measure of instability of a country during the sample years. To measure the background factor of government strength and governance quality, I do not use the average per capita income, even though it is often regarded as a good proxy for the two factors. The major reason is that the within-country variation of per capita income is not small. As is in the case of *democracy*,

Table 3.1: Within-Group Variation of Variables

Variable	Symbol	Within-Country Variation			Within-Year Variation		
		Min	Mean	Max	Min	Mean	Max
War outbreak	`onset`	0.00	0.07	0.35	0.00	0.11	0.24
Ongoing war$_{t-1}$	`ongwar`	0.00	0.15	0.51	0.00	0.32	0.42
Income per capita$_{t-1}$	`gdpenl`	0.02	1.26	17.50	1.81	4.03	6.55
Log(popuplation)$_{t-1}$	`lpopl`	0.00	0.25	0.71	1.36	1.45	1.53
Log(% Mountainous)	`lmtnest`	0.00	0.00	0.00	1.27	1.39	1.45
Noncontiguous state	`ncontig`	0.00	0.00	0.50	0.36	0.38	0.43
Oil producer	`oil`	0.00	0.06	0.50	0.26	0.33	0.39
New state	`nwstate`	0.00	0.17	0.49	0.00	0.13	0.40
Political Instability	`instab`	0.00	0.26	0.53	0.00	0.35	0.45
Anocracy$_{t-1}$	`anocl`	0.00	0.26	0.53	0.31	0.42	0.50
Ethno-linguistic fractionalization	`ethfrac`	0.00	0.00	0.13	0.26	0.28	0.30
Ethnicity Minority Leader	`minldr1`	0.00	0.11	0.51	0.44	0.47	0.49
Democracy$_{t-1}$	`polity2l`	0.00	2.88	9.23	6.73	7.39	7.74
Secondary school Enrollment	`second`	0.00	0.00	0.04	0.11	0.11	0.11

The variations reported in the table is the descriptive summary of the variance of the variables within a country and within a year.

its average is not a good measure of country-specific economic background. Instead, I use *Male secondary school enrollment* which captures the governance quality and the capability of the government to prevent civil war from occurring. *Per capita income* and *male secondary school enrollment* are highly correlation ($\rho \approx 0.8$), and since the latter is almost time-invariant and the former is a time-varying variable, it is better to use *secondary school enrollment* in the group-level regression as a proxy for economic development. Furthermore, secondary school enrollment directly reflects the capability and willingness of the government to provide public goods to the society and measures, more generally, governance quality. The variable of *Ethno-linguistic fractionalization* is a direct measure of the ethnic background of a country, and it is almost time-invariant; the country average of this variable is ready to be used in the group-level regression. Unobserved background factors are included in the group-level error term.

3.4 Empirical Results and Interpretations

I assign diffuse conditional conjugate priors on all parameters except the autoregressive coefficients $\boldsymbol{\rho}$, and use the Gibbs sampler to update those parameters. Since there is no conditional conjugate prior on $\boldsymbol{\rho}$, they are simulated with the Metropolis-Hasting algorithm (see Appendix 2.3). I specify uniform priors to the autoregressive coefficients within the stationary space. Since the sample size is large, the prior specifications will not have a big influence on the posteriors, which is confirmed by sensitivity checks. The GLMM-AR(p) model only handles stationary dynamic processes of the errors, and if the error process is not stationary, the causal relationships found in regressions are spurious (Engle and Granger, 1987; Beck, 1993; Durr, 1993; Smith, 1993; Williams, 1993; Keele and De Boef, 2004). The errors follows a stationary dynamic process either when the process of civil war is stationary, or when the regressors included are co-integrated with the response (Engle and Granger, 1987; Kao, 1999; Pedroni, 1999, 2004). In fact, to study time-series-cross-section data, cointegration testing is necessary to ensure estimates obtained to be reliable. However, there is no good way to conduct cointegration tests for sophisticated nonlinear time-series cross-section models yet. The MCMC simulation designed for the GLMM-AR(p) model will fail to generate legitimate proposals for updating the autoregressive coefficients if the error process is not stationary. I write the `R` code in such a way that if the proposals keep being illegitimate for a pre-specified length of time, the simulation will be halted and ab error message will be returned to suggest further check on stationarity. Therefore, this model can serve as an informal cointegration test, and the following estimates are ensured not to be spurious based on nonstationary panels.

I estimate the GLMM-AR(p) model with increasing lag orders starting with $p = 0$. I compute the marginal likelihood of each model, and stop increasing the lag order when decreasing marginal likelihood is observed. I also compare the more sophisticated GLMM-AR(p) model with two commonly-used alternative specifications, i.e., the completely pooled probit model (denoted as "PROBIT" in Figure 3.1) and the probit multi-level model without year-specific random effects or serial correlation (denoted as "GLMM-CL1" in Figure 3.1[4]). The simple probit model does not model the variation of EMR's effect, and serves as a replication of FKL's model. The GLMM-CL1 model can be regarded as a special case of the GLMM-AR(p) model with two additional assumptions: the time-specific random effects have no variation, and the errors are uncorrelated. I also tried to estimate the GLMM-CL1 model with serially correlated errors, but the error process was likely to be nonstationary—with an

[4]It is equivalent to a fixed effect panel model regarding the country-level regression as interaction terms and merging the group-level error term $b_{2i}$with the random intercept (b_{1i}).

AR(1) and AR(2) process, the first ten thousand draws (without the burnin stage) of the autoregressive coefficients were very close to the unit circle, and then the MCMC simulation halted because no legitimate proposals could be drawn within the stationary space. Therefore, only the AR(0) GLMM-CL1 model could be successfully estimated by ignoring the dynamics. The suspected nonstationary error process makes inferences based on the GLMM-CL1 model unreliable, but for model comparison purposes, I still summarize the MCMC output of the static GLMM-CL1 model in Figure 3.1. However, adding the year-specific effects makes the same specification stationary. This is not surprising, because if time-specific effects are correlated with each other, putting them into the error term will exaggerate serial correlation and create spurious dynamics (Heckman, 1981), or because if time-specific effects are correlated with other regressors, omitted time-dimension heterogeneity changes the relationship between the dynamic process of civil war and the dynamic processes the explanatory variables represent, which affects cointegration.

I compute Bayes Factors for all those models, and they indicate that the model with a second order error autoregressive process is the best one (see Figure 3.1 for the marginal likelihood of each model). I also do an insurance run of the GLMM-AR(4), and find that its quality decreases compared to the AR(2) and AR(3) model. The number of iterations for each model is 500,000 with a burnin stage of 50,000 iterations. The chains of random-effect and low-level parameters mix much more slowly, which slows down the whole process of convergence, since partial convergence should not be regarded as convergence (Gill, 2008). Because there are more than 12,800 chains in the MCMC simulation in each model, it is unrealistic to conduct formal convergence diagnostics on all of them. Instead, I perform multiple convergence diagnostics by using the `coda` package in `R` on all the fixed and random effects parameter posteriors, but for the augmented data and the auxiliary parameter $\boldsymbol{u}$, I randomly select a sample of size 100 from each chain and conduct convergence diagnoses. No evidence of non-convergence is found in those diagnostics. In Figure 3.1, I present all six models with 95% credible intervals and the posteriors of the fixed parameters (at all levels). The random effects (the random intercepts of b_{1i} and β_{t_i}, and the random effects of the ethnic minority leadership ($\beta_{2,i}$) are plotted in Figure 3.2 based on the GLMM-AR(2) model which has the best goodness-of-fit.

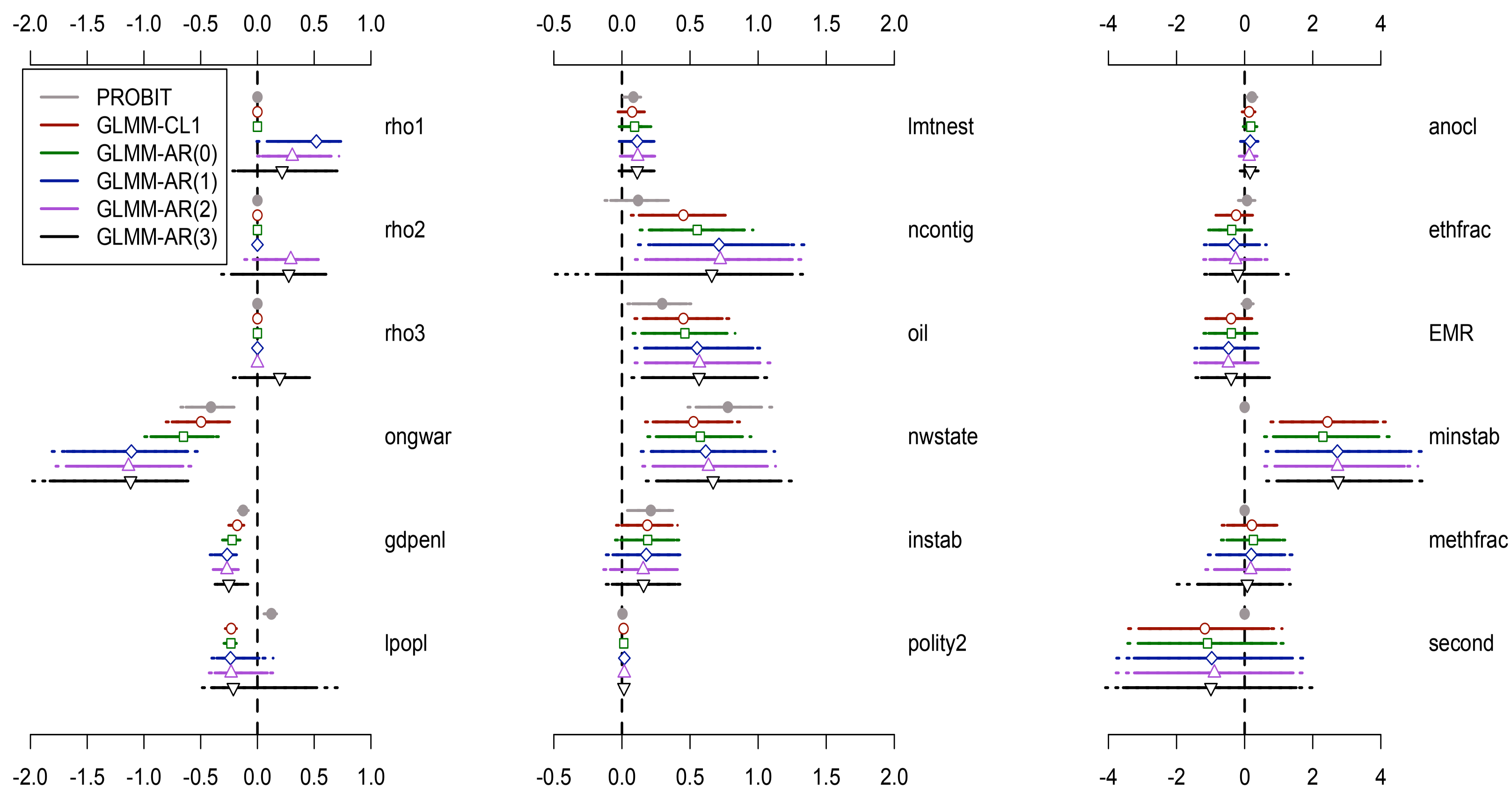
The posteriors based on the probit model are very similar to the results in FKL, though the scales are different because of the different link functions (FKL used the logistic link). With a big sample size, the point estimates of the Bayesian model (posterior mean) are very similar to the maximum likelihood estimates if the same link function were used. Unlike in FKL, I include two additional variables (*democracy* and *ethno-linguistic fractionalization*) in the model. Without modeling serial correlation and heterogeneity among countries, the estimator tends to bias toward 0 and has much smaller error bands compared to posteriors based on the other five models. For some of the variables, it yields very different substantive results. Larger population increases civil war onset propensity with high certainty, but after controlling for heterogeneity, the direction of its effect reverses; modeling serial correlation further, its effect becomes unclear. Similar change has been observed in the existing studies: the effect of population in the fixed-effect panel model is often different from that in completely pooled models (Fearon and Laitin, 2003*b*; Collier, Hoeffler and Soderbom, 2004; Fearon, Kasara and Laitin, 2007). It suggests that population could be correlated with some omitted country-specific characteristics which also affect civil war. Those characteristics might include the territory, economic structure, climate or other geographical features, and so on. It is often argued that more populous country incurs higher civil war risk because large population size enlarges the pool of potential recruits (Fearon and Laitin, 2003*b*), but once confounding factors are controlled for, population does not have a certain effect. The probit model also finds that non-contiguous countries are likely to have higher risk of civil war, but the effect is uncertain, while models controlling for heterogeneity in either time or spatial dimension suggest that it has a positive effect with a high level of certainty. The completely pooled model also exaggerates the effect of *new state status*. For the primary explanatory variable, `EMR`, the results confirm FKL and suggest that EMR may increase the likelihood of civil war onset, but this effect only has a low level of certainty.

After controlling for heterogeneity, the other five models all find that EMR, on average, is more likely to decrease civil war risk, but this statement can only be made with a low level of certainty (only 60%). Among the five multilevel models, the only difference between GLMM-CL1 and GLMM-AR(0) is that the latter considers the heterogeneity across years. The estimated time-dimension hetergeneity is mild. The lower-right graph in Figure 3.2 shows the year-level variation (the year-specific intercept) based on the GLMM-AR(2) model. We can see that there is year-to-year difference but the variation is not large. The Bayes Factor ($\log_{10}$) of the GLMM-AR(0) versus GLMM-CL1 is 2.66, and decisively supports adding the year-level regression (Greenberg, 2008). The improvement on model fit by modeling the year-level heterogeneity is also reflected by the forecasting performance, which is discussed in detail in the next section. Between the two models, the posteriors of the fixed effect parameters are very similar. Nonetheless, for several variables, the GLMM-CL1 model generates posteriors with the tendency of being biased toward 0. The GLMM-AR(p) models with $p > 0$ find positive serial correlation which is not weak (the posterior mean of ρ_1 is 0.5 based on the AR(1) model) and probably lasts for more than one time period (posterior means of ρ_1 and ρ_2 are both about 0.3 based on the AR(2) model). Serial correlation correction and dynamic modeling greatly improve model quality, and the marginal likelihood (natural logarithm) increases by more than 20 (see Figure 3.2). Compared to the probit and GLMM-CL1 models, modeling serial correlation yields three key differences: first, the error bounds are systematically bigger than those of the GLMM model without modeling serial correlation, confirming the statistical theory that ignoring positive serial correlation results in narrower error bands (Gourieroux, Monfort and Trognon, 1984; Poirier and Ruud, 1988). Second, the effect of ongoing war in the previous time period is almost twice as big as in the models without considering serial correlation. Apparently, serial correlation causes biased estimation of this variable's effect: ongoing war at time $t-1$ is very likely to be correlated with ϵ_{t-1} because the omitted factors which affect civil war onset are likely to be correlated with ongoing civil war, too. If ϵ_{t-1} is correlated with ϵ_t, there is endogeneity arising from the correlation between ongoing civil war and the error term. Without modeling serial correlation, the dynamics captured by the variable of ongoing civil war are misleading. Further evidence is that, based on the estimated autoregressive coefficients, the propensity for civil war is positively correlated, which is a different mechanism than that suggested by using the variable of previous ongoing war since their signs are opposite. Third, within-sample prediction is improved dramatically by making good use of the information contained in the dynamic error process, which is analyzed in more detail in the next section.

All of the multilevel models detect salient heterogeneity of EMR's effect on civil war onset across countries. Figure 3.2 shows the heterogeneity estimated by the GLMM-AR(2)[5], which indicates considerable heterogeneity across countries. The effect of EMR is more likely to be negative for about two thirds of the sample countries.

[5] The cross-country heterogeneity estimated by all the multilevel models are very similar to each other, and therefore only the random-effect posteriors based on the GLMM-AR(2) model are presented in the chapter.

Figure 3.1: Posterior Summary with 95% Credibility Interval (Six Models)



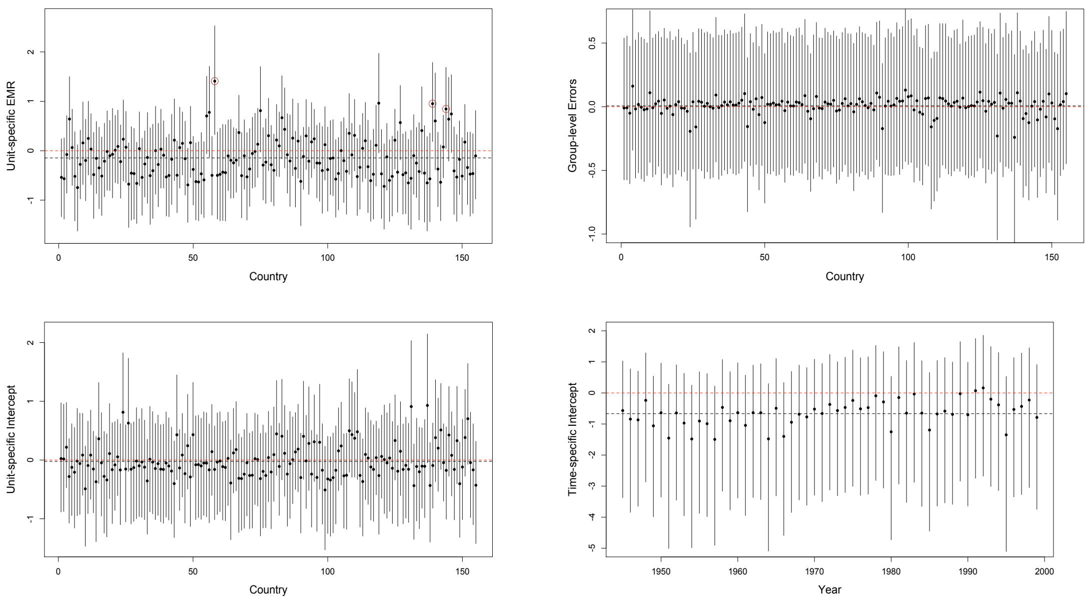
The marginal likelihood of the six models — PROBIT: 531.783; GLMM-CL1: -524.291; GLMM-AR(0): -518.174; GLMM-AR(1): -476.712; GLMM-AR(2): -470.546; and GLMM-AR(3): -494.009

The GLMM-AR(2) model finds that for three countries, i.e., Azerbaijan, Pakistan, and Thailand, ethnic minority rule increases the likelihood of civil war onset with a 95% credible level. However, the effect of EMR is at a low level of credibility for most sample countries, and its directions vary widely across countries. In Figure 3.1, the three posteriors of the group-level regressors partially explain this variation. The effect of political instability is highly robust to various model specifications, and suggests that in a country which experiences frequent regime change, ethnic minority rule is more vulnerable and creates higher risk of civil war than in politically stable countries. Compared with the posteriors of the fixed-effect parameters, the error band of political instability is much larger, which can be explained by the much smaller sample size than at the individual level. The variable of *male secondary school enrollment* is likely to decrease the danger of civil war increased by EMR, but it is not very certain (at a 70% credibility level). *Ethno-linguistic fractionalization* seems not to have an effect on this relationship, which is expected based on the two opposite mechanisms discussed in Section 3.1. The heterogeneity of EMR's effect across countries can be compared with the relatively homogeneous country-level errors presented in the upper-right graph in Figure 3.1. This comparison suggests that the group-level regressors explains a large proportion of the variation of EMR's effect on civil war onset. The lower-left graph displays the country-specific intercept, which shows that, aside from the hetergoneous effect of EMR on civil war, countries still demonstrate differences in other respects and those differences are relevant to civil war onset.

3.5 Prediction

Assessing the goodness-of-fit of a statistical model is necessary to confidently draw causal inferences based on the estimates (Hoeting et al., 1999). The Bayes Factor

Figure 3.2: Random Intercepts and Random Effects



is a comprehensive criterion for this purpose, but within-sample and out-of-sample predictions are not only important alternative ways to assess model quality but also valuable in themselves. Especially for the civil war study, which has important policy implications, better predictions based on a statistical model can serve as a warning system and provide valuable information to policy makers. The GLMM-AR(p) model, better identifying the causal relationships and making better use of the information in the data (the dynamic process of the errors), performs better in forecasting[6].

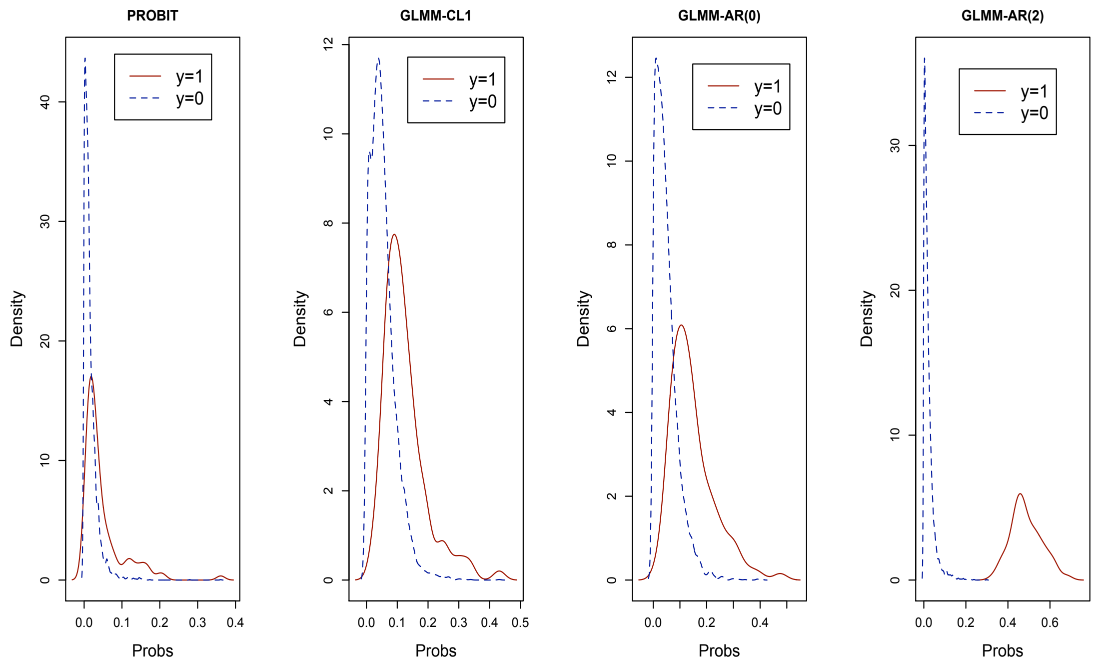
In this section, I compare the forecast performance of the following four models: the probit model, the GLMM-CL1, the GLMM-AR(p) model without serial correlation consideration, and the GLMM-AR(p) with an appropriate lag order (the second order) of the autoregressive errors. The models under comparison use all the observa-

[6]The relationship between better study of causal relationships and better and more stable forecasting is articulated in King and Zeng (2001).

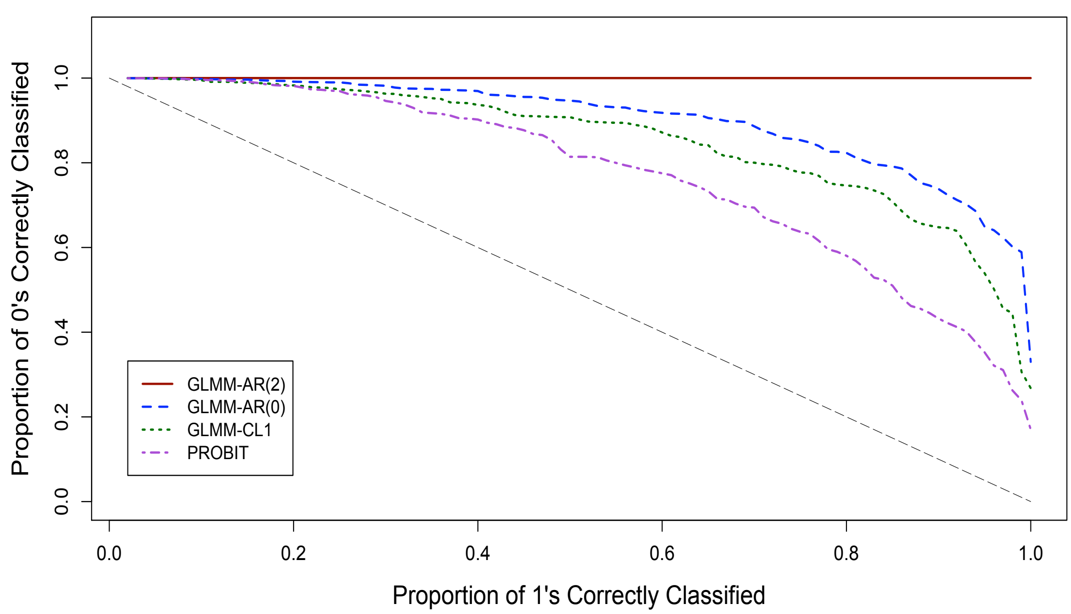
tions, and I only do within-sample predictions. The general within-sample predicting performance of the four models is presented in Figure 3.3. To avoid setting arbitrary or *post hoc* thresholds for classification of civil war or non-civil war, I first only report the numeric predictive probabilities of civil war in all country-years and compare the distributions of predictive probabilities of the civil war group and the non-civil war group. If the densities of the two groups are well separated (the density of the civil war group is plotted on the right side and that of the non-civil war group on the left side in each figure), forecasts based on those predictive probabilities will make few mistakes in both classifications ($y = 0$ and $y = 1$), which suggests that the model performs well in forecasting. As shown in Figure 3.3 on the first row, the pooled probit model does a bad job in differentiating the civil war cases from the non-civil war ones, and the density kernels of the two groups are overlapping most of the time. The GLMM-CL1 model, by modeling the variation of EMR's effect and other country-level heterogeneity, separates the two groups better than the pooled model, but there is still a large overlapping area. The GLMM-AR(0) model further considers heterogeneity in the time periods, and reduces the overlapping compared to the GLMM-CL1 model. It is not a good model, either, though it is better than the GLMM-CL1 model. The GLMM-AR(2) model, which models the dynamics and makes good use of the information contained in the error term, well classifies the two groups: the two density kernels are almost completely separated from each other, and only a very small part at each tail is connected with the other.

The second way of evaluating and comparing the predicting performance of the competing models is to use the *Receiver-Operating Characteristic (ROC) curve*, as King and Zeng (2001*b*) did. The idea of using the ROC curve to evaluate predicting performance of models is simple: given a level of correct classification of one group (say, $y = 1$), the model performs better if the rate of correct classification of the other

Figure 3.3: Comparison of Within-Sample Predicted Probs.



Predictive Prob. of 1's and 0's



ROC Curve

group (say, $y = 0$) is higher based on this model. Graphically, the curve dominating other curves represents the best model among the competing ones. This approach has the advantage of avoiding assigning any fixed threshold for classification and assessing model performance based on any specific cutoff value. In Figure 3.3, the diagonal line is the reference line, indicating the extreme situation that the predictive probabilities of the two groups are completely mixed. The ROC curve based on the completely pooled probit model is the lowest one. The GLMM-CL1 model improves forecasting by modeling heterogeneity among countries, and its curve globally dominates the one of the pooled probit model. The GLMM-AR(0) model has an ROC curve globally above the one of the GLMM-CL1, but is dominated everywhere by the curve of the GLMM-AR(2) model. The ROC curve of the GLMM-AR(2) model is almost a horizontal line, indicating that, since the densities of the predictive probabilities of the 1 and 0 groups are well separated, there is barely any trade-off between the two types of classification[7].

Table 3.2 shows the top 15 best and worst predictions based on the GLMM-AR(2) model and compares them with those based on the GLMM-AR(0) and GLMM-CL1 models. Because the probit model is the worst one, it is omitted here due to space limitation. For the war cases, even the worst predictions of the GLMM-AR(2) are all much better than the other two models, and most of the worst predictions of non-civil war cases are also better. Figure 3.4 presents predictions of 8 selected countries in the sample years. The selected countries experience most to fewest civil war onset events

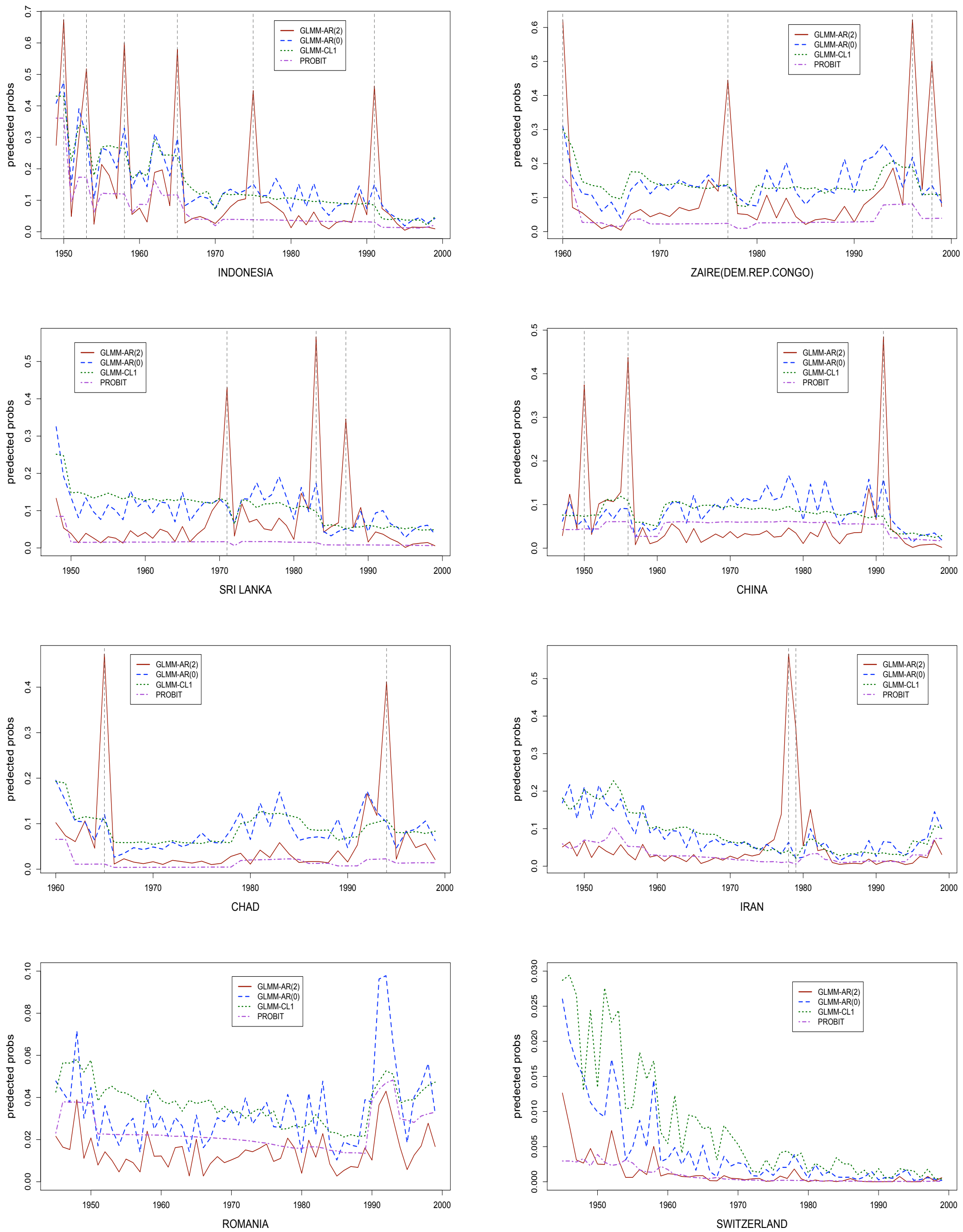
[7] Note that the horizontal ROC curve does not suggest that the GLMM-AR(2) model will correctly classify all the country-years. This is because the threshold of classification should be set *ex ante* based on the risk assessment and pre-specified loss function, rather than being chosen after the pattern of predictive probabilities are observed. For instance, if setting a *post hoc* threshold of 0.32, then the GLMM-AR(2) model makes no misclassification of either civil war or non-civil war; but we may set an *ex ante* threshold of 0.60 if we think misclassification of civil war is much less expensive than mistakenly identifying non-civil war as civil war, then most of the observed civil wars will be misclassified as non-civil wars even based on the GLMM-AR(2) model.

(from 6 episodes to 0). All those figures and tables demonstrate that the GLMM-AR(2) performs much better in statistical forecasting because it takes into account two-dimensional heterogeneity and the dynamic process of the error term. Figure 3.5 shows the autocorrelation and partial autocorrelation of the predicted probabilities for a randomly selected country. For the three models with a static error process, the predicted probabilities demonstrate an AR(2) or AR(1) process, which suggests that their poor forecast performance should be largely attributed to ignoring the dynamics.

3.6 Robustness Checking

To check the robustness of the empirical findings, I have used different model specifications: including some other variables, such as *Democracy*, in the country-group regression and the *Cold war* dummy in the year-level regression; excluding *ethno-linguistic fractionalizationin* from the individual level regression; assigning different priors, and so on (the details are not reported in the present chapter). None of them cause important changes, and the cold war dummy does not have explanatory power. A more important way of checking robustness of the empirical findings is to use alternative measures of EMR, as FKL did. They presented other two alternative measures of EMR: for the first alternative, the previous measure of EMR is modified by examining the data case by case and recoding those country-years as plurality group dominance when leader's ethnicity does not mean the dominance of their minority ethnic groups; the second alternative measure is to further recode white in the same ethnic group as mestizo in Latin America based on the first alternative measure. This changes the proportion of EMR in the sample substantially: the percentage of minority leader country-years decreases from 32.20% to 25.12%. I estimate

Figure 3.4: Random Intercepts and Random Effects



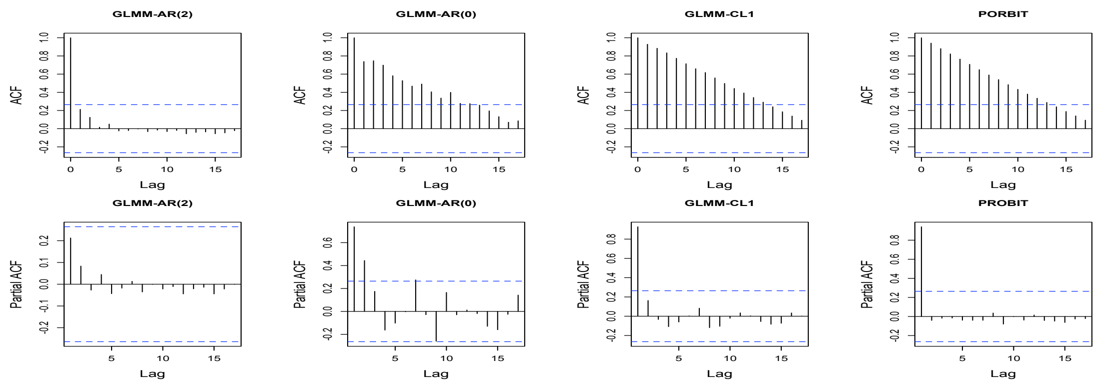
The grey vertical lines in the graphs indicate the episodes of civil war onset.

Table 3.2: Within-Sample Predictions of Civil War Onset

Best Predictions					Worst Predictions									
Probs. (Civil War Observed)					Probs. (Civil War Onbserved)					Probs. (Civil War NOT Observed)				
Country	Year	AR2	AR0	CL1	Country	Year	AR2	AR0	CL1	country	Year	AR2	AR0	CL1
Angola	1975	0.686	0.394	0.338	The U.K.	1969	0.326	0.023	0.028	Indonesia	1952	0.299	0.391	0.338
Indonesia	1950	0.674	0.475	0.431	Sri Lanka	1987	0.346	0.051	0.051	Indonesia	1949	0.274	0.408	0.431
Burma	1948	0.627	0.248	0.199	Argentina	1973	0.365	0.047	0.041	Azerbaijan	1991	0.271	0.343	0.247
Phlippines	1946	0.624	0.303	0.297	Iran	1978	0.366	0.021	0.021	Taijikistan	1991	0.241	0.297	0.209
Congo(DR)	1960	0.623	0.310	0.301	Cyprus	1974	0.367	0.061	0.071	Russia	1998	0.236	0.176	0.127
Congo(DR)	1996	0.623	0.219	0.190	China	1950	0.375	0.067	0.073	Burundi	1991	0.226	0.218	0.132
Azerbaijan	1992	0.612	0.359	0.246	India	1982	0.376	0.067	0.072	Burundi	1992	0.223	0.219	0.129
Rwanda	1962	0.611	0.265	0.257	Colombia	1963	0.377	0.053	0.061	Indonesia	1955	0.214	0.266	0.269
Indonesia	1958	0.601	0.329	0.265	Argentina	1955	0.381	0.075	0.077	Pakistan	1970	0.210	0.230	0.213
Nicaragua	1981	0.600	0.200	0.135	Jordan	1970	0.395	0.053	0.059	Turkey	1983	0.206	0.154	0.078
Pakistan	1971	0.591	0.202	0.199	Bangladesh	1976	0.397	0.083	0.072	Georgia	1991	0.204	0.233	0.149
Tajikistan	1992	0.582	0.283	0.197	Paraguay	1947	0.401	0.058	0.081	Pakistan	1948	0.198	0.390	0.321
Indonesia	1965	0.580	0.297	0.241	South Africa	1983	0.409	0.049	0.026	Indonesia	1963	0.197	0.251	0.243
Haiti	1991	0.577	0.259	0.182	Chad	1994	0.412	0.103	0.109	Rwanda	1989	0.191	0.172	0.099
Colombia	1948	0.570	0.238	0.157	Rwanda	1990	0.414	0.080	0.100	Indonesia	1962	0.189	0.312	0.296

Both “best predictions” and “worst predictions” are sorted based on the predictions of the GLMM-AR(2) model. Also, because the best 20 predictions of those observation of non-civil war based on all the three models are $\Pr(y = 1|\boldsymbol{x}, \boldsymbol{\theta}) = 0.000$, those probabilities are not listed in the table.

Figure 3.5: AutoCorrelation and Partial Autocorrelation of Predictive Prob.



the GLMM-AR(2) model with the two alternative measures and compare them with the results based on the previous measure. The summary of posterior means and 95% credible intervals is presented in Table 3.3. The posteriors are symmetric with their means very close to medians; therefore, I do not report the medians in the table. The posteriors of the parameters in the individual-level regression are very similar among the three models, and the dynamics of the error process is almost the same in the three models. The average effect of EMR demonstrates differences with different measures, but overall, its effect is uncertain in all models. Even though the average effect of EMR is different, the degree that political instability alters EMR's effect on civil war onset is robust and has similar posteriors across models. For the other two explanatory variables, *average ethno-linguistic fractionalization* and *secondary school enrollment*, they seem to be more likely to decrease EMR's effect on civil war onset based on the two alternative measures than the measure adopted in the previous section. Figure 3.6 also shows that the country-specific effect of EMR on civil war onset, as well as its variation across countries, are similar to those in Figure 3.2, only with

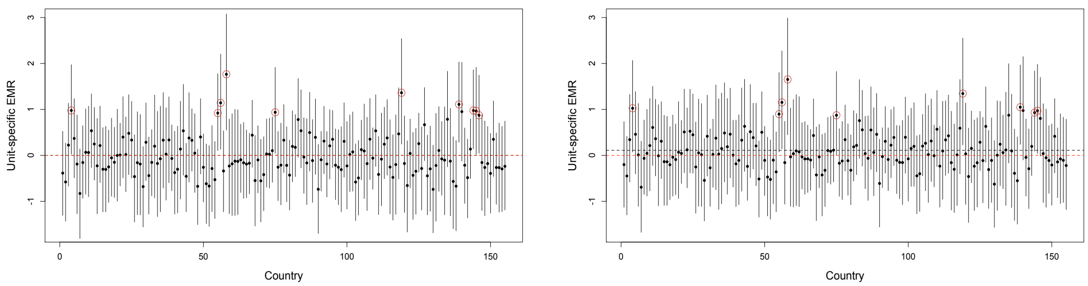
Table 3.3: Robustness Check I: Alternative Measures of EMR

	Measure I		Measure II		Measure III	
Covariate	mean	credible interval	mean	credible interval	mean	credible interval
ρ_1	0.32	(0.00 , 0.72)	0.34	(0.06 , 0.65)	0.32	(0.01 , 0.65)
ρ_2	0.28	(-0.11 , 0.57)	0.32	(-0.07 , 0.62)	0.31	(-0.01 , 0.57)
`ongwar`	-1.15	(-1.78 , -0.58)	-1.25	(-1.97 , -0.66)	-1.23	(-2.26 , -0.60)
`gdpenl`	-0.27	(-0.39 , -0.15)	-0.27	(-0.42 , -0.15)	-0.28	(-0.44 , -0.14)
`lpopl`	-0.20	(-0.43 , 0.19)	-0.22	(-0.64 , 0.32)	-0.21	(-0.70 , 0.47)
`lmtnest`	0.12	(-0.01 , 0.27)	0.12	(-0.02 , 0.28)	0.11	(-0.04 , 0.28)
`ncontig`	0.72	(0.10 , 1.37)	0.78	(0.02 , 1.60)	0.76	(-0.12 , 1.67)
`oil`	0.58	(0.09 , 1.11)	0.58	(0.08 , 1.14)	0.59	(0.09 , 1.18)
`nwstate`	0.64	(0.15 , 1.13)	0.65	(0.15 , 1.17)	0.65	(0.15 , 1.21)
`instab`	0.15	(-0.14 , 0.43)	0.15	(-0.14 , 0.45)	0.15	(-0.13 , 0.43)
`polity2`	0.02	(-0.01 , 0.04)	0.02	(-0.01 , 0.04)	0.02	(-0.01 , 0.04)
`anocl`	0.13	(-0.17 , 0.42)	0.12	(-0.18 , 0.42)	0.13	(-0.20 , 0.43)
`ethrac`	-0.27	(-1.21 , 0.68)	-0.22	(-1.28 , 0.71)	-0.28	(-1.40 , 0.72)
`EMR`	-0.49	(-1.47 , 0.48)	-0.14	(-1.41 , 1.08)	0.09	(-1.25 , 1.47)
`minstab`	2.76	(0.58 , 5.09)	2.92	(0.49 , 5.72)	2.47	(0.01 , 5.32)
`methfrac`	0.17	(-1.15 , 1.43)	-0.08	(-1.47 , 1.35)	-0.17	(-1.75 , 1.27)
`second`	-0.90	(-3.78 , 1.86)	-1.65	(-5.11 , 1.71)	-1.74	(-5.21 , 1.70)
Marg. Lik.	-470.546		-476.891		-481.754	

slightly bigger variations. The marginal likelihoods of the three models are reported in Table 3.6, and they support the original measure over the two alternatives.

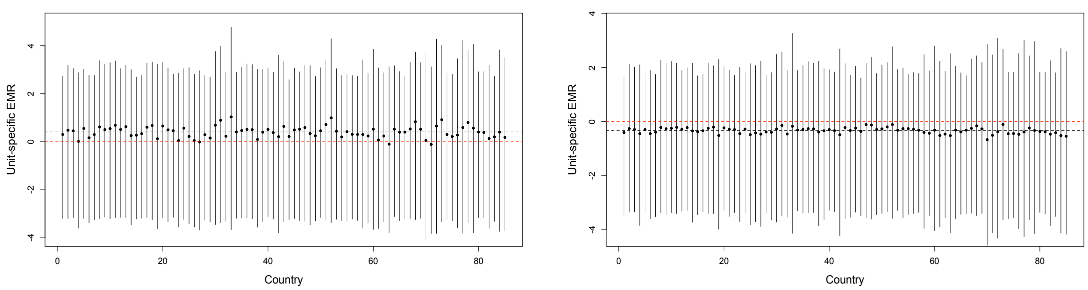
I also use the measure constructed by CG, N^*, and the modified N^* in FKL which "uses dummies for whether the country had a minority EGIP (ethnic group(s) in power) and whether the country had a coalition EGIP as coded by CG". The sample country-years then are limited to Eurasia and North Africa. The sample size is 3327 with 85 countries across 55 years. The posteriors are summarized in Table 3.4. Because some of the posterior distributions are skewed, I report both the means and medians in the table. As in CG, political instability loses its importance in both models, and democracy gains much importance based on the two measures. The dynamic process of the errors is still similar to the one in the other models. With

Figure 3.6: Random Intercepts and Random Effects (Alternative Measures



Measure II of EMR | Measure III of EMR

Figure 3.7: Random Intercepts and Random Effects



Original Measure of N^* | Alternative Measure of N^*

the heterogeneity in the two dimensions and serial correlation being modeled, even the orginal N^* constructed by CG loses its importance on average, and there is not much variation of its effect on civil war across countries (as shown in Figure 3.7). No country-level explanatory variable demonstrates importance since there is not much variation to explain. Because the two models have different response variables from the previous models, they are not comparable. Thus, I omit reporting their marginal likelihoods. This robustness check suggests that CG's argument based on their empirical models is not reliable.

Table 3.4: Robustness Check II: Different Measures of EMR

	Measure I				Measure II			
Covariate	mean	sd	median	credible interval	mean	sd	median	credible interval
ρ_1	0.34	0.16	0.32	(0.07 , 0.71)	0.31	0.17	0.29	(0.03 , 0.75)
ρ_2	0.45	0.19	0.47	(0.04 , 0.77)	0.47	0.20	0.50	(0.01 , 0.77)
`ongwar`	-1.50	0.59	-1.41	(-2.79 , -0.56)	-1.52	0.61	-1.45	(-2.86 , -0.51)
`gdpenl`	-0.29	0.13	-0.31	(-0.50 , -0.08)	-0.30	0.12	-0.30	(-0.51 , -0.02)
`lpopl`	-0.39	1.74	-0.22	(-5.45 , 3.92)	-0.56	1.68	-0.23	(-5.36 , 2.13)
`lmtnest`	0.03	0.36	0.03	(-0.84 , 0.77)	0.04	0.34	0.05	(-0.82 , 0.68)
`ncontig`	0.64	1.43	0.68	(-3.74 , 3.67)	0.74	1.30	0.65	(-2.20 , 3.89)
`oil`	0.96	0.50	0.90	(0.11 , 2.19)	0.94	0.47	0.92	(0.06 , 1.96)
`nwstate`	0.75	0.46	0.72	(-0.09 , 1.75)	0.75	0.46	0.71	(-0.08 , 1.78)
`instab`	-0.06	0.25	-0.05	(-0.58 , 0.42)	-0.09	0.26	-0.09	(-0.62 , 0.39)
`polity2`	0.04	0.02	0.04	(-0.01 , 0.09)	0.04	0.02	0.04	(-0.01 , 0.09)
`anocl`	0.09	0.27	0.09	(-0.44 , 0.61)	0.11	0.27	0.12	(-0.45 , 0.63)
`ethrac`	0.06	1.38	0.36	(-4.13 , 1.76)	0.07	1.37	0.37	(-4.08 , 1.77)
N^*	0.53	2.00	0.70	(-4.61 , 4.01)	-0.31	1.83	-0.14	(-4.87 , 2.91)
`minstab`	1.01	2.83	1.04	(-4.64 , 6.51)	0.35	2.77	0.33	(-5.18 , 5.71)
`methfrac`	-0.46	2.54	-0.48	(-5.39 , 4.64)	-0.46	2.59	-0.47	(-5.51 , 4.65)
`second`	-0.83	2.99	-0.87	(-6.58 , 5.11)	0.20	2.96	0.24	(-5.72 , 5.94)

3.7 Discussion

This chapter revisits the question about the relationship between minority ethnic dominance on civil war onset raised by two recent papers in the civil war quantitative study literature. With the complicated causal chain between the two variables, this relationship is altered by multiple observed and unobserved background factors and varies from country to country. This chapter applies a new model for empirical analyses, which is not only able to directly model the varying degree of EMR's effect on civil war onset, but also carefully controls for multiple sources of confounding caused by the TSCS data structure of the civil war dataset. The major findings include the heterogeneity of EMR's effect on civil war onset across countries, and the countries with an unstable political regime are more prone to civil war when they are under ethnic minority rule than the politically stable countries are. This finding is robust across different model specifications and alternative measures of EMR. In addition, modeling the dynamic process of the errors and differentiating dynamics from the heterogeneity in both year and country dimensions dramatically improve statistical forecasts.

Chapter 4

Sovereign Default: Regime Type, Regime Duration, and Vulnerability to Global Shocks

Is there a "democratic advantage" in sovereign borrowing from the international financial market? The existing studies find mixed and ambiguous evidence about the role of regime type, which may suggest that the relationship is actually nonlinear, conditional on other factors, or confounded by omitted variables. In the decision-making about sovereign default, a government's incentive of to pay back external debt is important but complex (Drazen 1998). Various domestic political institutions, such as regime type, veto power distribution, electoral system, and leadership turnover, have been suggested to shape default incentive. Among those factors, regime type is one of the most interesting but most controversial: despite the strong theories supporting the "democratic advantage", empirical evidence is often found that this advantage does not apply to sovereign default. The null finding can be caused by both theoretical pitfalls and methodological discrepancies. Conventionally, regime type is understood only as a distinct set of domestic institutions, and numerically measured as spot values. However, the default decision involves the problem of time inconsistency. Without analyzing the tradeoff between the present and the future, we cannot really understand the role played by regime type. We should not understand regime type only as a static set of institutional arrangements; instead, regime type may affect how the shadow of the future forms and changes. This means that the relationship between regime type and sovereign default may be more complex than what the extant literature suggests.

Inspired by the theory constructed by Olson (1993) about how time horizons are shaped through different mechanisms in different regimes as time goes by, this paper re-investigates the effect of regime type on sovereign default by focusing on the regime-dependent effect of regime duration. In addition, for both substantive and methodological reasons, sovereign default is explained in both the domestic and international contexts. In the era of globalization, sovereign default crises are usually triggered by macro shocks in the international system, and the variation of default propensity among countries is caused largely by the varying sensitivity and vulnerability of countries to the international system. This has methodological importance, because distinguishing the external sources of variation from the internal sources is necessary for controlling for major confounders in the relationship between regime type and sovereign default. It is also substantively interesting to empirically study in what degree and how differently globalization affects developing countries, which contributes to the IR literature on how the international system affects national decision-making.

In this paper I make the following three arguments: (i) sovereign default is a short-sighted decision which maintains the current consumption level at the expense of future consumption, and the time horizon of a government plays an important role in sovereign default; (ii) regime duration affects of anocracies' time horizons and, accordingly, default likelihood, in a different way from the age of full democracies and stark autocracies; and (iii) sovereign default as a national decision is largely affected by external shocks in the international system which impact all countries but in varying degrees. To empirically test the theoretical hypotheses, I collect time-series cross-sectional data containing 134 developing countries in 14 years, and propose a new method based on the data structure and research goals. By carefully controlling for correlation in both the time and spatial dimensions and measuring the country-specific impacts of common shocks, I find strong evidence supporting the regime-dependent effect of regime duration and salient heterogeneity among countries in terms of their sensitivity and vulnerability to globalization.

4.1 Sovereign Default, Time Horizons, and International System

Sovereign debt is distinct from corporate or individual debt in at leaset two aspects: the government commitment to repay external debt is unbinding in an anarchical international system; and, technically speaking, a country cannot be insolvent, since the government always has avenues, mainly cutting expenditures or raising taxes, to service external debt. Defaulting on sovereign debt is always a "decision" made by politicians based on their cost-benefit calculation, subject to a variety of internal and external constraints (Andritzky, 2006; Drazen, 2002, 1998; Bulow and Rogoff, 1989). In this decision calculus, capability to pay is not the whole story: countries default in bad economic times, but the relationship between economic downturns and sovereign default is surprisingly weak. Sovereign default has occurred when the economic situations are favorable (Tomz and Wright, 2007). Both willingness and capacity matter in sovereign default, and are shaped by many economic and political factors, domestically and internationally.

4.1.1 Why Countries Pay Back? Cost of Sovereign Default

Why does a government keep its commitment to repay external debt in spite of the absence of a coercive authority above to force it to do it? In reality, most countries at most of the time "voluntarily" stick to their debt obligations. The choice between repaying and defaulting should be explained by the cost-benefit calculation of the decision makers. The major benefit of sovereign default is to have breathing space during a certain period of time. It allows the government to maintain or boost current consumption by reducing the debt stock or extending the maturity of its repayments (Paoli and Hoggarth, 2006), but sovereign default hurts economic development in the long run for many reasons.

Sovereign default is bad for the reputation of the defaulting government, and it takes time to rebuild market confidence. The defaulting country can be directly punished by international lenders. Although it remains controversial whether a defaulting country is actually denied future access to the international financial market, empirical evidence confirms that future borrowing cost increases significantly for defaulting countries (Paoli and Hoggarth, 2006). Higher cost of accessing the international capital market has a negative impact on economic growth, especially for developing countries which often have an under-developed domestic financial market and where capital is a scarce resource. In addition, the "sovereign ceiling rule" states that the private sector cannot borrow on better terms than the government in most situations and in most countries. Sovereign default makes it more difficult for the private sector to borrow in the international financial market (Dooley, 2000). The private sector will have to pay higher interest rates in international borrowing and face other worse conditions to access the market (Arteta and Hale, 2005). Furthermore, sovereign default is likely to trigger banking and currency crises; as a consequence, it hurts the entire domestic financial system deeply: the banking system in less developed countries are fragile and vulnerable to debt crises because of the close and complicated connections between the government and banks (Hoelscher and Quintyn, 2003). Sovereign default may also trigger speculative attacks on the currency of the defaulting country, because debt crises send the signal to international speculators that the government is less capable of defending its currency (Cespedes, Chang and Velasco, 2004). Paoli and Hoggarth (2006) find that a twin or triple crisis has a larger effect on the fall in output.

In general, the choice of sovereign default reflects that the government values current consumption rather than long-run economic growth and future taxable capacity (Olson, 1993). Default can be regarded as distortionary taxation: taxing the future for the present (Kim, 2007). If a government values current consumption more than future consumption, it is more likely to choose default in bad economic times. Therefore, the incumbent government's time horizon plays an important role in determining its willingness to repay its external debt. In the IPE literature there are a variety of political institutions suggested to affect time horizons, such as electoral cycles (Alesina, Roubini and Cohen, 1997; Mei and Guo, 2004), political instability (Bussiere and Mulder, 1999; Bordo and Oosterlinck, 2005), regime type and tenure length of chief executives (Clauge, Keefer and Olson, 1996; Manasse, Roubini and Schimmelpfennig, 2003; McGilivray and Smith, 2003; Demir, 2006; Moser, 2006; Chang, 2002). Most of them share a coherent underlying logic that if the incumbent government has a shorter time horizon, it tends to make short-sighted decisions such as sovereign default.

4.1.2 Time Horizons: Regime-Dependent Effect of Regime Duration

Time horizon affects a government's choice between repaying and defaulting on its external debt, since there is a time inconsistency problem. Some address this time horizon issue in sovereign default by focusing on different mechanisms through which citizens punish defaulting political leaders under various regimes (McGilivray and Smith, 2003), and others explain sovereign default as a short-sighted decision of opportunistic politicians with short time horizons (Demir, 2006; Moser, 2006; Bordo and Oosterlinck, 2005; Mei and Guo, 2004; Chang, 2002; Balkan, 1992; Ozler and Tabellini, 1991). However, those studies suffer from difficulties in controlling for various sources of confounding. For example, leadership turnover is suggested to affect the time horizon of the decision makers (McGilivray and Smith, 2003), but there are many different forms of leader change. The Archigos Database categorizes at least four different situations of leadership turnover, namely, leadership turnover in a regular manner, in an irregular manner, through direct removal by another state, and as a result of a natural death (Goemans, Gieditsch and Chiozza, 2009). Leadership turnover does not necessarily mean that the leader at the end of her tenure has a short time horizon regardless of regime type. In democracies, the politician at the end of her tenure may still value the future, because she or her party expects to take office again in the future. A dictator in power transition may still care about the long-run benefits, because she is confident that the successor she chooses will stay in power for a long time or because a strong dominant party forces her to be responsible.

This paper focuses on time horizons to explain sovereign default. It differentiates itself from existing studies by investigating the regime-dependent effect of regime duration. Regime type is one of the most interesting political factors in the literature on sovereign borrowing and other important IPE questions. Most existing studies stress the constraints imposed by regime-specific institutions on sovereign default and analytically reach the conclusion that democracies are less likely to default on their external debt than autocracies and the commitment made by democracies is more credible (Schultz and Weingast, 2003; Tsebelis, 2002; Schultz, 2001; Olson, 2000; Elster, 2000; Smith, 1998; Weingast, 1997, 1995; Fearon, 1994; Firmin-Sellers, 1994; Olson, 1993; Przeworksi and Limongi, 1993; Shepsle, 1991; North, 1989, 1990; North and Weingast, 1989; McGilivray and Smith, 2003; Balkan, 1992; Brewer and Rivoli, 1990; Citron and Nickelsburg, 1987; Abdullah, 1985; Feder and Uy, 1985). However, empirical analyses often fail to support the "democratic advantage" in sovereign borrowing, and the effect of regime type on international capital flow remains controversial (Jensen, 2008; Archer, Biglaiser and DeRouen, 2007; Saiegh, 2005; Jensen, 2003; Li and Resnick, 2003; Sobel, 1999). Null findings are not particular to the studies on politics of international finance; in the broader literature on regime type and economic growth, empirical evidence is also very ambiguous, and a democracy is often found to even have a slightly negative effect (Krieckhaus, 2004; Feng, 2003; Przeworski et al., 2002; Tavares and Wacziarg, 2001; Przeworski et al., 2000; Barro, 1996; Levine and Renelt, 1992), despite widely-accepted theories supporting democracy as an optimal form for economic growth. This has puzzled many researchers. Some argue that it is misleading to treat "democracy" as a static set of institutions and measure it as a spot value. In a country with a long history of democracy, democratic institutions are deeply embedded, and the country will behave differently from a new democracy. Based on this criticism, Gerring et al. (2005) use a cumulative measure of democracy as a stock rather than level, and find empirical support that the effect of regime type on economic growth is conditional on regime duration.

Olson (1982) and Olson (1993) noticed the regime-dependent effect of regime duration on economic development. According to Olson, democracies are subject to "institutional sclerosis" as time goes by, because groups with special interests weaken "encompassing interests" and are becoming better at capturing the state and seeking for rents. In this sense, authoritarian regime enjoys the advantage of more encompassing power. Nonetheless, Olson also argues that this "autocracy advantage" is duration-dependent. Only when the ruler expects the regime to survive in the future, she can have a long time horizon. How long the regime has already been provides important information about the prospective of regime survival in the future. From the perspective of time horizons, Olson's theory predicts that for autocracies with longer regime duration, the ruler is more confident on future regime survival and has longer time horizon, and consequently is less likely to make short-sighted decisions. This is also confirmed by the empirical evidence in Clague et al. (1997). Without much modification, this argument is applied to sovereign default: when an authoritative regime is perceived to last for a long time, the dictator has incentive to ensure future output and consumption, and is less likely to default on external debt.

However, Olson's theory about the effect of regime age for democracies emphasizes dynamic institutional changes instead of time horizons. His theory predicts that with longer regime duration, because of weakened "encompassing interests", a democratic government tends to make short-sighted decisions. However, this does not apply to sovereign default, about which the configuration of special interests of "small groups" is unclear. The dominance of small groups' interests does not necessarily mean that the government is biased towards defaulting on its external debt. If the dominant small groups borrow heavily from the international financial market, it is their interest to ensure an easy and stable access to future financing. The benefit of default is usually diffuse and the cost concentrated to small groups (Celasun and Harms, 2007;

Block and Schrage, 2003), so the dominance of small group interests is more likely to lead to sticking to debt obligations, though the interests of dominant groups can be rather heterogeneous across countries and time. In spite of the ambiguity in the analysis on small group interests in democracies, by focusing on the supply side of economic policies, the time horizon of the government in a stable democracy will be longer than in a new and immature democracy. Parties and individual politicians expect that the rules of the game remain stable in the future, and their record matters when they are trying to win the next election. Therefore, for both full democracies and stark autocracies with a longer regime age, the regime is expected to be more likely to survive in the future, and the government has a longer time horizon, values more the future, and is less likely to default on its external debt.

Besides democracies and autocracies, there is another important type of regime called anocracies. Hegre et al. (2001) define anocracies as "semi-democracies" that are "partly open yet somewhat repressive," distinguished from "institutionally consistent democracies and stark autocracies." Fearon and Laitin (2003) also describe anocracies as a mixed regime of a democracy with autocratic features. Vreeland (2008) characterizes this regime as "a mix of institutional characteristics, some democratic and others distinctively authoritarian." Since anocracies are an intermediate state between full democracies and stark autocracies (Baliga, Lucca and Sjöström, 2009), by definition and by nature, anocracies are an unstable regime in the middle (Rost, Schneider and Kleibl, 2009), and are at a transitory stage (Goetze and Guzina, 2008). "[A] key feature of anocracies is that they are relatively unstable and that instability is often associated with a shift to autocracy or democracy... [Anocracies] denote a transitional and potentially unstable stage on the way to more stable democratic governance (or state failure)" (Elgie and McMenamin, 2008). Anocracies' characteristic of being highly transitory is also reflected by empirical evidence. For example, Marshall and Gurr (2003) find, among all ancroacies, "over fifty percent experiencing a major regime change within five years and over seventy percent within ten years."

Because anocracies are transitory, regime duration of annocracies has different implications to sovereign default. Annocracies are expected to last for a short period of time; as time goes by, regime change is expected more likely to occur in either direction. No matter whether it changes to be more democratic or more authoritarian, the expected change causes high political uncertainty and shortens the time horizon of the incumbent government. It is highly uncertain whether the incumbent government is still in power after a major regime change, and the incumbent government will value less future income and consumption, *ceteris paribus.* From the perspective of time horizons, the age of an anocracy will increase the likelihood that the incumbent government chooses sovereign default rather than repaying external debt by increasing taxes or cuting expenditures in order to avoid output loss caused by sovereign default.

In all, because sovereign default is a short-sighted decision to maintain the current consumption level at the expense of future consumption, incumbent governments with longer time horizons will be less likely to default on their external debts. In full democracies and stark autocracies, longer regime duration stabilizes the expectation of regime survival and regime stability in the future, and the government values more future consumption and is less likely to make short-sighted decisions such as defaulting on external debt; but for anocracies which are intrinsically transitory and unstable, with longer regime duration, a major regime change is expected to be more likely, and the government will have a shorter time horizon and tend to make opportunistic decisions.

4.1.3 External Shocks and Heterogeneous Sensitivity to International System

There is no need to argue that sovereign default is largely triggered by external shocks in the international system[1]. However, the IPE literature on sovereign borrowing focuses on domestic economic and political institutions; besides time dummies used for controlling for unobserved common shocks, only several observed common shocks in the international system are included in empirical studies as control variables, such as global GDP growth, US interest rate, default proportion (Kim, 2008; Kraay and Nehru, 2006; Saiegh, 2004). More importantly, both observed and unobserved global shocks are assumed to have unit-invariant impacts despite the salient heterogeneities among sample countries. This assumption is unrealistic because the pressure of external events is filtered or cushioned by domestic political, economic, and social characteristics, and will affect domestic policy-making in different ways. Modeling unit-varying effects of common shocks is not only methodologically desirable (I will discuss this in more detail in the next section), but also has important theoretical meaning for understanding sovereign default by taking account of countries' vulnerability to financial globalization. For less developed but open economies, the effects of global shocks can be considerably large and vary widely across countries. Without analyzing the international factors together with domestic institutional variables, both inferences and forecasting would be misleading.

In the era of globalization, "the world is now really a single economy in the macroeconomic sense (Glyn and Sutcliffe, 1992)." Complex interdependence shrinks national control of resources and erodes the autonomy of national governments (Berger, 2000).

[1]Here international system is not formally defined as in the IR literature (primarily the definition given by Waltz (1979)); instead, it is informally referred to and broadly defined as a group of interacting, interrelated, or interdependent countries forming a complex whole

We can no longer understand national policy-making without putting it in the international context (Keohane and Milner, 1996). Common shocks in the international system, such as global macroeconomic changes, establishment of new economic and political rules, technological development, and so on, create new opportunities but also imposes new constraints on nation-states (Frieden and Rogowski, 1996). More than three decades ago, Gourevitch (1978) invented the term, “the second image reversed”, and raised a series of research questions about how international factors affect domestic economy and politics and whether the shocks in the system have the same impacts everywhere. Gourevitch (1986) showed, by using case studies, large variation across countries of domestic responses reflected by economic policies under the pressure of external shocks. In the paradigm of “the second image reversed” (Keohane and Milner, 1996), there has been a huge IPE literature on how globalization affects national decision making (Pierson, 1996; Williamson, 1997; Kapstein, 2000; Rudra, 2002; Schmukler, 2004; Wibbels and Arce, 2003; Wibbels, 2006, to name only a few). Most of them use qualitative research since empirically measuring how the effects of both observed and unobserved common shocks vary is methodologically challenging.

The international financial market is integrated in a complex way compared with other aspects of globalization. Financial globalization is “the integration of countries into international financial markets,” posing complex problems and challenges to all countries, especially developing countries (Torre, Yeyati and Schmukler, 2002). The financial connections among countries are vulnerable to shocks within and outside the financial sector. Many scholars are interested in studying how financial globalization affects national decision making differently given different domestic institutions. A good example is the classical Mundell-Fleming model, which establishes a framework to analyze how international capital flow affects domestic policy making (monetary policy) conditional on other domestic institutions (exchange rate regime). Focusing more on less-developed countries, Torre, Yeyati and Schmukler (2002) emphasize three domestic factors, international currency, flexible exchange rate, and sound contractual and regulatory environment, which together are called the "blessed trinity". They use this "blessed trinity" to explain why some countries benefit from financial globalization, but others handle volatility in the international financial market less successfully. The large literature on spillover and contagion of financial crises is essentially about how external shocks affect countries in globalization through real links, financial links, and imperfect information and sentiments in the international capital market (Cuadra and Sapriza, 2008; Demir, 2006; Dungey and Martin, 2005; Hedge and Paliwal, 2005; Calvo, 2005; Lizarazo, 2005; Abreu, 2003; Giannetti, 2003; Gelos and Wei, 2002; Change and Majnoni, 2001; Kawai, Newfarmer and Schmukler, 2001; Bordo and Murshid, 2000; Calvo and Mendoza, 1998; Calvo and Reinhart, 1996; Banerjee, 1992; Bikhchandani, Hirshleifer and Welch, 1992). There is very little controversy that the risks and rewards of globalization, especially for developing countries, vary widely from country to country, according to their sensitivity and vulnerability to the international system (Kenen, 2007; Kose et al., 2006).

The financial market is subject to various shocks, some of which are tangible, such as macroeconomic shocks (global economic situations), financial shocks (*e.g.*, changes of liquidity), political shocks (*e.g.*, wars), institutional shocks (*e.g.*, establishment or demise of international organizations), and natural shocks (*e.g.*, natural disasters); others, though, are intangible shocks, such as a sentiment shift of international investors, herding behavior caused by market and information imperfection, and changes of international pressure on servicing external debt (the peer pressure). External shocks affect decisions about sovereign default, and the relationship is altered by domestic factors and through various channels. If an empirical analysis had been able to include all the relevant mediate domestic factors, we would not be concerned about the direct effects of external shocks, but the channels through which the external pressure transforms into domestic policy making are too complicated to be identified or included. It is also unrealistic to include all common shocks in our theoretical framework and empirical models. Therefore, when analyzing sovereign default, we always face the omitted variable problem. Furthermore, as mentioned before, the effect of financial globalization is likely to vary from country to country, and it is not easy to model the varying level of vulnerability or sensitivity of countries' decision making regarding sovereign default to common shocks.

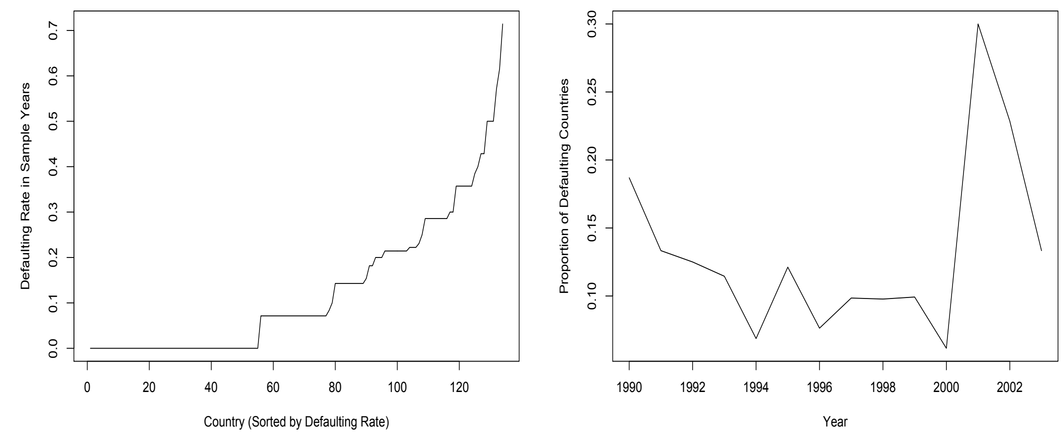
4.2 Variables and Data

To test the regime-specific effect of regime age and to analyze the unit-varying impact of the international system on national decision of sovereign default, I collect data of 134 developing countries from 1990 to 2003. In this section, I describe the definition, measurement, and descriptive characteristics of the variables.

4.2.1 Dependent Variable: Sovereign Default

In the literature, there are several different ways to determine and measure sovereign default. Moody's Investors Service (1999) defines sovereign default as "(1) any missed or delayed payment of interest and/or principal, or (2) any exchange where the debtor offers the creditor a new contract that amounts to a diminished financial obligation, or (3) where the exchange has the apparent purpose of helping the borrower avoid default." Similarly, in academia, sovereign default usually refers to rescheduling or restructuring debt, including arrears on principal or interests (Reinhart, Rogoff and Savastano, 2003; van Rijicheghem and Weder, 2004; Kraay and Nehru, 2006; Kohlscheen, 2006; Tomz and Wright, 2007; Saiegh, 2008). In the present paper, I

Figure 4.1: Default Rate (Country and Year)



follow a widely-applied definition and measurement: sovereign default is an event in which a sovereign fails to make pre-scheduled principal or interest payment or restructures its external debt. As in Kim (2008), the two criteria used to measure an episode of sovereign default are the accumulation of debt payment arrears and a rescheduling arrangement. The variable, `Default`, is measured as the same in Kim (2008): a country-year is coded as 1 "if the increase in the stock of total arrears exceeds 2% of total debt from private creditors or if the total amount of debt rescheduled exceeds 2.5% of total debt from private creditors unless the stock of total arrears decreases by more than the amount of debt rescheduled in the same year, and 0 otherwise." The data are from Global Development Finance (GDF 2005).

The sample year starts from 1990 due to the large amount of missing data before. There are 134 developing countries in the dataset from 1990 to 2003. The data structure is unbalanced: the minimum number of observations of a country is 3 (Bosnia and Herzegovina), and the maximum number is 14 (97 countries); the average number of observations of a country is 12.70 (the median is 14) with standard deviation 2.52.

In the cross-section dimension, the minimum number of observations in a year is 102, and the maximum is 133. In average 123 countries are observed in a year with standard deviation 11.00. There are 1718 country-years in the dataset, and 217 default episodes are observed, a proportion of 12.60% [2]. For more information about the sample countries and years, see Appendix 4.6.4. Figure 4.1 presents the default rate of the sample countries and years. The left figure shows that there are 53 countries which do not default even once in their sample time period, and the country that defaulted most frequently (in about 70% of its sample years) is Jordan. The right graph illustrates the proportion of defaulting countries in a given year. The default rate roughly decreases between 1990 and 2000. There is a surge in 2001, when about 30% of the sample countries default. The default rate remains high in 2002 (about 23%), and then back to around the average level in 2003.

4.2.2 Regime Type, Regime Duration, and External Shocks

There are three different regime types under investigation in this paper, namely, democracy, anocracy, and autocracy. In the Polity IV database, the variable `polity2` measures the degree of democracy as 21-scale scores ranging from -10 (strongly autocratic) to $+10$ (strongly democratic) (Marshall and Jaggers, 2007). This measure is used here as the variable `Democracy` in the model to estimate the effect of the level of democracy. To test the regime-specific effect of regime duration, I use a discrete measure for the three types and, following the most used coding criterion in the literature such as Gurr et al (Polity V), Fearon and Laintin (2003) and many others:

[2] I tried to use alternative coding criteria as in Kim (2008), but based on the current data set the thresholds of 4% or 6% for arrears and 5% or 7.5% for rescheduled debt generate sovereign default as a rare event (6.35% and 4.19%, respectively). Since rare events requires special analytical methods, I do not use those alternative measures in this paper.

based on the `polity2` score in Polity IV, I code the regime authority years with a polity score between -5 and $+5$ as "anocracies"(coded as 1), and those with scores falling out of this range as "non-anocracies"(coded as 0), including full democracies (between $+6$ and $+10$) and stark autocracies (between -10 and -6). Alternative coding criteria are also used, though less frequently, in the literature: since "anocracies" are understood as a regime "in the middle", some studies use equal-lengthed subintervals to define autocracies, anocracies, and democracies, and coded regime authority years with polity scores between -3 and 3 as anocracies (Relter and Meek, 1999; Baliga, Lucca and Sjöström, 2009). I use this alternative coding criterion as a robustness check. Using the coding criterion of $[-5, 5]$, there are 643 country-years of anocracies, a proportion of 37.4%. For the anocracy group, regime duration is 8.3 years on average, while countries in the group of full democracy and autocracy last for 17.0 years on average before a major regime change occurs. The t-test rejects the null hypothesis that the two groups have the same mean regime duration. Alternatively, if using the subinterval of $[-3, 3]$, 361 country-years are anocracies, a proportion of 18%. The anocracy group has average regime duration as 7.93 years, and the group of full democracies and autocracies as 15.3 years. The t-test also rejects the null hypothesis. In both cases, we can see that anocracies are transitory and last for a much shorter time period than both democracies and autocracies.

The measure of regime duration is based on the variable `durable` in Polity IV, which has been used to measure regime/political stability in many empirical studies in political science (Gartzke, 2001; Fearon and Laitin, 2003*b*; Fearon, Kasara and Laitin, 2007; Calderon and Chong, 2007; Vreeland, 2008; Chang and Golden, 2009, for instance). It is defined in the Polity IV User's Manual (p.13) as following:

> Regime Durability: the number of years since the most recent regime change (defined by a three- point change in the POLITY score over a period of three years or less) or the end of transition period defined by the lack of stable political institutions (denoted by a standardized authority score). In calculating the DURABLE value, the first year during which a new (post-change) polity is established is coded as the baseline "year zero" (value = 0) and each subsequent year adds one to the value of the DURABLE variable consecutively until a new regime change or transition period occurs.

From the definition we can see that this variable directly measures the age of a regime before a major regime change occurs and is the actual regime duration. There is another variable, regime "durability", which is often used to measure the time horizon associated with regime survival. One way to measure regime "durability" is to use a regression model (survival analysis, for example) to estimate the survival probability of a regime at each time period (Wright, 2008). I do not use this variable or apply the estimated regime survival probability for two major reasons. First, the regime durability is a related but different concept from regime duration or regime age. Although regime duration is closely associated with the expected regime survival probability, it is essentially a cumulative measure and based on the whole history of a certain regime, but the estimated survival probability is a spot value in each time period, responding to the changes of other relevant situations and probably being affected by the dynamics (what happened in the previous one or two time periods). The second reason is that the survival probability is an estimates based on a particular regression model, and the measure's quality completely depends on correctly specifying the model and how well the model predicts regime changes. The better the prediction, the better the measure. This means that the model is likely to be over-fitted and include most of the variables to be selected in the sovereign default model. This will confound the relationships of major interest. In addition, to include the uncertainty of the predicted probability is difficult, but ignoring the uncertainty and arbitrarily picking up point values make the measure misleading.

4.2.3 Macroeconomic Situations and Global Shocks

There are several macroeconomic variables widely regarded to have effects on the capability of a government to repay external debt. The ratio of external debt to output measures how deep a country is in debt and its capability to repay debt. The higher the ratio, the higher the expected default likelihood. GDP per capita reflects the economic development level of a country. Richer countries should be more capable of sticking to debt obligations, but for developing countries, those that are more deeply involved in the international economy are likely to be richer than those that are not, but more sensitive and vulnerable to the changes in the system. Therefore, GDP per capita may have an ambiguous effect. The rate of output growth can also be an important predictor of sovereign default, because a decline of GDP growth causes a long-term insolvency problem. Trade openness has been suggested to reduce the probability of sovereign default, because the more open the economy, the more costly sovereign default, since international creditors can impose punishment on international trade of the defaulting country. However, the creditors in the commercial market are not necessarily those in the financial market, and the former may not have incentive to punish a defaulting country. The effect of trade openness could be negative or uncertain. A large proportion of short-term debt in a country's debt stock, a low level of foreign exchange reserve relative to external debt stock, and a

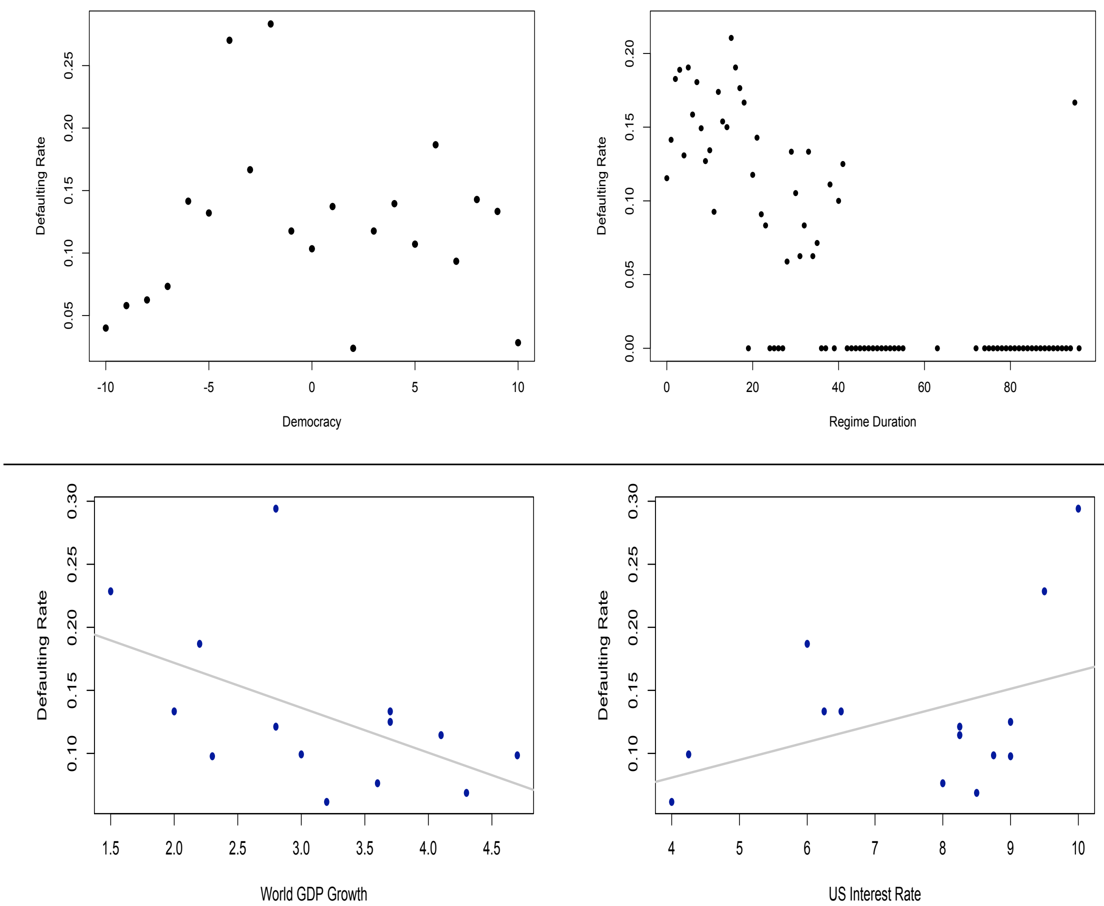
Table 4.1: Within-Group Variation of Variables)

Symbol	Individual		Within-Country		Within-Year	
	Mean	SD	Mean	SD	Mean	SD
Default	0.126	0.332	0.232	0.207	0.328	0.062
Duration	13.749	17.380	8.430	7.280	17.400	1.440
Democracy	1.796	6.296	3.580	1.920	6.250	0.204
Annocracy	0.153	0.360	0.187	0.204	0.358	0.022
Total Debt/GDP $(_{t-1}, log)$	11.406	1.919	0.244	0.194	1.930	0.081
GDP per capita $(_{t-1}, log)$	6.855	1.170	0.151	0.203	1.170	0.025
GDP Growth $(_{t-1})$	1.517	6.281	4.810	4.090	6.000	1.680
Trade Openness $(_{t-1}, log)$	3.342	0.658	0.238	0.196	0.655	0.052
Short-term Debt/Total Debt $(_{t-1}, log)$	2.003	1.325	0.652	0.665	1.310	0.183
Reserves/Total Debt $(_{t-1}, log)$	2.742	1.414	0.631	0.403	1.400	0.161
Current Account $(_{t-1})$	−0.325	4.139	1.480	2.890	3.720	1.600
World GDP Growth	3.192	0.889				
US Interest Rate	7.569	1.816				

high ratio of current account deficit to GDP are all dangerous signs, because the country is more prone to liquidity debt crises (Min et al., 2003; Rodrik and Velasco, 1999; Jeanne, 1997; Sachs, 1984). The economic data are mainly from the IMF and Work Bank data bases, and are lagged for one time period to avoid the simultaneity problem. All variables and their variation at both the individual and group levels are summarized in Table 4.2.

For observed common shocks, I simply include the two most used variables: US interest rate and global GDP growth. Global GDP growth represents the global macroeconomic situation in a given year and the output shocks in the international system. Changes in the US prime interest rate reflect the fluctuation of international liquidity and the general cost of international borrowing. Figure 4.2 shows a simple univariate analysis of the relationship between default rate and regime type, regime duration, global GDP growth, and US interest. Without controlling for other variables, being more democratic seems not to have a linear relationship with higher default rate, but regime duration is negatively correlated with sovereign default rate.

Figure 4.2: Default Rate and Primary Explanatory Variables



While World GDP growth rate may reduce sovereign default, increase of US interest is associated with higher default rate.

4.3 Methodology

The response variable, sovereign default, is dichotomous and time-series cross-sectional. The data characteristics mean that to investigate the relationships of theoretical interest, several methodological issues have to be handled with caution. First, the 135 sample countries, though all less developed economies, are saliently different in many important aspects. Country-specific characteristics may affect sovereign default and be correlated with the covariates at the same time. Omitted heterogeneities are likely to lead to inconsistent as well as inefficient estimators. Second, each sample country is repeatedly measured over time, and serial correlation is a natural concern. There are multiple sources of inter-temporal correlations, including dynamics in the financial market, path dependence of states and investors' behavior, policy inertia, measurement error, and omitted variables, etc. Unmodeled or uncorrected serial correlation will cause biased or inconsistent estimates of standard errors. Third, to investigate the country-specific effect of globalization is methodologically challenging: for observed common shocks we can use country-specific coefficients, but most common shocks are not measured. How, then, can we estimate the unit-varying effects of the unobserved common shocks?

4.3.1 Methodological Issues

Observations of the same country are likely to be correlated because of country-specific characteristics and dynamics. Those of the same year may also be related: changes in the international system have impacts on all the countries albeit at different levels; and interactions among countries cause spillover of debt crises across borders. Hence, observations of the 134 countries in the 14 years are clustered in both the time and spatial dimensions, which suggests multilevel analysis to handle heterogeneity and different sources and levels of variation (Gelman and Hill, 2006; Gill, 2007; Beck and Katz, 2007; Shor et al., 2007). For the serial correlation issue, the lagged variable approach (including lagged values of response variables (observed or latent) and explanatory variables (LDVs/LIDVs)) have been often recommended and implemented (Beck, Katz and Tucker, 1998; Beck et al., 2002), but those methods cannot substitute for direct error serial dependence correction. Because there are multiple sources of intertemporal correlation, serial dependence may be partially corrected by LDVs or LIDVs, but they cannot speak to whether the errors are still correlated. In addition, for generalized linear models, including lagged values of observed response variable does not introduce the same error covariance matrix as a same order autoregressive errors, but lagged values of the latent response variable raise several potential theoretical and methodological problems (Wilson and Butler, 2007; Skrondal and Rabe-Hesketh, 2008), and are as complex to estimate as a model with a dynamic error process.

In this paper, modeling the impacts of common shocks is not only substantively interesting but also has methodological importance. Longitudinal analysis usually focuses on studying dynamic relationships and heterogeneity among units (Molenberghs and Verbeke, 2005; Frees, 2004; Verbeke and Molenberghs, 2000). For the sovereign default data analyzed in the current paper, concerns about heterogeneity and correlation are naturally extended to the cross-section dimension, since the observations of all countries are sampled at equally spaced time periods (Gelman, 2006). Cross-sectional correlation can be caused by many sources, such as omitted common shocks, spillover or contagion (spatial effects), or interactions among units (network effects) (Pesaran, 2006). Spatial correlation, just like serial dependence, yields misleading inferences if standard panel models are applied. When omitted common effects are correlated with the regressors, we will have an endogeneity problem and, as a consequence, inconsistent estimators (Philips and Sul, 2003).

How to solve this problem depends on the data structure and the substantive question to answer. If the unit dimension is much smaller than the time dimenstion, i.e., $N << T$, TSCS data can be treated as a seemingly unrelated system, and generalized least squares is applied. This is apparently not the case for the sovereign default data in this paper: with the cross-section dimension much larger than the time dimension, SUR analysis cannot apply. Two major alternative specifications have been used to model spatial correlation for such a data structure ($N >> T$). One approach includes spatial regressions, such as spatial autoregressive, moving average, and spatial error component models. Spatial regressions are based on the assumption that correlation among units is determined by a pre-specified metric based on geographical, economic, social, political or other kinds of distance among units (Franzse and Hays, 2008*b*, 2007; Pesaran, Schuermann and Weiner, 2004; Conley and Topa, 2002; Conley, 1999). They are often applied to the substantive and methodological issues caused by spillover or interactions. However, because of complex interdpendence in the era of globalization, we do not have the prior information required for specifying the metric, since geographical distance has lost its conventional meaning, and the distance in other senses, such as economic distance or social distance, is not clearly defined and remains controversial.

An alternative approach is the multifactor residual model, focusing on spatial correlation caused by common shocks in varying degrees for different units. Unlike ordinary mixed-effect models, the multifactor model admits that, due to heterogeneity among units, the impacts of common factors vary from unit to unit. The residual term in the model includes two parts: the linear combination of unobserved common factors with their particular impacts on each unit, and an idiosyncratic error term. Because of the substantive interest of this paper and the purpose of keeping the statistical model from being too complicated, I do not consider spillover effects at the same time.

Conventionally, the multifactor residual model is estimated with a full maximum likelihood procedure (Robertson and Symons, 2000) or principal component analysis (Coakley, Fuertes and Smith, 2002). For those estimators to be consistent, an aditional assumption is required: regressors are uncorrelated with the unobserved common factors. Since this assumption is too restrictive in empirical studies, there are several methods in the literature used to allow the time-specific disturbances to be correlated with the covariates, such as linear panel models with instrumental variables (Holtz-Eakin, Newey and Rosen, 1988), generalized methods of moments (Ahn, Lee. and Schmidt, 2001), and least squares with auxiliary regressions, also referred to as common correlation effects estimators (Pesaran, 2006; Pesaran and Tosetti, 2007). With a multilevel specification, those frequentist and likelihoodist approaches suffer from complications, especially when developing asymptotic properties of the estimators. In this paper, I take advantage of a Markov Chain Monte Carlo method and develop a more straightfoward Bayesian estimator.

4.3.2 GLMM-TSUV.AR(p): Specification and Estimation

This subsection specifies the generalized linear multilevel model with time-specific unit-varying effects and an autoregressive idiosyncratic error term (henceforth, GLMM-TSUV.AR(p)), and addresses the Bayesian strategy of model estimation and comparison.

Suppose the data consist of N units each of which is indexed as $i \in \{1, 2, ..., N\}$. Unit i has T_i observations in time periods $\{1_i, ..., t_i, ..., T_i\} \subseteq \{1, 2, ..., T\}$. With the contemporary effects modeled, index $t_i = t$ means that y_{i,t_i} is observed in year t. Since the data structure is unbalanced, the observation y_{i,t_i} is not necessarily the tth element in $\{1_i, ..., t_i, ..., T_i\}$ and it is likely that $t_i \neq t_j$ and $T_i \neq T_j$ for $i \neq j$. By using the latent variable approach (Albert and Chib, 1993), the general specification of GLMM-TSUV.AR(p) is as follows:

$$y_{i,t_i} = \mathbf{1}(z_{i,t_i} > 0), \tag{4.1}$$

$$z_{i,t_i} = \mathbf{x}'_{1i,t_i}\boldsymbol{\beta}_1 + \mathbf{w}'_{i,t_i}\boldsymbol{\beta}_{2i} + \boldsymbol{v}'_{i,t_i}\boldsymbol{\beta}_{3t_i} + \zeta_{i,t_i} + \xi_{i,t_i}, \tag{4.2}$$

$$\boldsymbol{\beta}_{2i} = \mathbf{A}_{1i}\boldsymbol{\beta}_2 + \mathbf{b}_{1i}, \qquad \boldsymbol{\beta}_{3t_i} = \mathbf{U}_{t_i}\boldsymbol{\beta}_3 + \boldsymbol{\eta}_{t_i}, \qquad \zeta_{i,t_i} = \mathbf{s}'_{t_i}\boldsymbol{\beta}_{4i} + \mathbf{f}'_{t_i}\boldsymbol{\gamma}_i, \tag{4.3}$$

$$\boldsymbol{\beta}_{4i} = \mathbf{A}_{2i}\boldsymbol{\beta}_4 + \mathbf{b}_{2i}, \qquad \mathbf{f}_{t_i} = \mathbf{f} + \mathbf{c}_{t_i}, \qquad \boldsymbol{\gamma}_i = \boldsymbol{\gamma} + \boldsymbol{\zeta}_i, \tag{4.4}$$

$$\xi_{i,t_i} = \rho_1\xi_{i,t_i-1} + ... + \rho_p\xi_{i,t_i-p} + \epsilon_{i,t_i}. \tag{4.5}$$

Unlike ordinary generalized linear multi-level models, this model has a parameter ζ_{i,t_i} measuring the country-specific effect of the international system. This effect is further decomposed into two parts in the third equation of line (4.3): the unit-varying impacts of the observed common shocks $\mathbf{s}$ (a q_1-dimensional vector) and the unobserved ones $\mathbf{f}$ (a q_2-dimensional vector). Both the observed and unobserved common shocks are called factors in this model, and $\mathbf{F}_{t_i} \equiv (\mathbf{s}_{t_i}, \mathbf{f}_{t_i})$ is a factor vector, and $\mathbf{\Gamma} \equiv (\boldsymbol{\beta}_{4i}, \boldsymbol{\gamma}_i)$ is the vector of factor loadings which measure the unit-specific effects of the common shocks. In the model the unobserved factors are treated as latent variables distributed as $\mathbf{f}_{t_i} \sim N(\mathbf{f}, \mathbf{\Omega}_{\mathbf{f}_{t_i}})$ and their factor loadings $\boldsymbol{\gamma}_i \sim N(\boldsymbol{\gamma}, \mathbf{\Sigma}_{\boldsymbol{\gamma}_i})$ (as in the last two equations of line (4.4)). However, these general distribution assumptions make the effects of the factor loadings unidentifiable. If all of the overall, time-specific, and unit-specific intercepts are included, then, with essentially no loss of generality, we can instead assume $\mathbf{f}_{t_i} \sim N(\mathbf{0}, \mathbf{\Omega}_{\mathbf{f}_{t_i}})$ and $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{\gamma}_i})$. The identifiable model can be written as follows:

$$y_{i,t_i} = \mathbf{1}(z_{i,t_i} > 0), \tag{4.6}$$

$$z_{i,t_i} = \mathbf{x}'_{1i,t_i}\boldsymbol{\beta}_1 + \mathbf{w}'_{i,t_i}\boldsymbol{\beta}_{2i} + \boldsymbol{v}'_{i,t_i}\boldsymbol{\beta}_{3t_i} + \mathbf{s}'_{t_i}\boldsymbol{\beta}_{4i} + \mathbf{f}'_{t_i}\boldsymbol{\gamma}_i + \xi_{i,t_i}, \tag{4.7}$$

$$\boldsymbol{\beta}_{2i} = \mathbf{A}_{1i}\boldsymbol{\beta}_2 + \mathbf{b}_{1i}, \qquad \boldsymbol{\beta}_{3t_i} = \mathbf{U}_{t_i}\boldsymbol{\beta}_3 + \boldsymbol{\eta}_{t_i}, \qquad \boldsymbol{\beta}_{4i} = \mathbf{A}_{2i}\boldsymbol{\beta}_4 + \mathbf{b}_{2i}, \tag{4.8}$$

$$\xi_{i,t_i} = \rho_1\xi_{i,t_i-1} + \ldots + \rho_p\xi_{i,t_i-p} + \epsilon_{i,t_i}. \tag{4.9}$$

As in the general multilevel model for TSCS data (Pang, 2009), there are three different types of individual-level covariates: $\boldsymbol{x}$ with fixed effect coefficients, $\boldsymbol{w}$ with unit-specific effects, and $\boldsymbol{v}$ with time-specific effects, all of which contain a constant term. The variations of unit- and time-specific effects may be further explained by observed unit- and time-specific characteristics and unobserved random effects as in the first two equations of line (4.8). I relax the unrealistic assumption that the common shocks have unit-invariant effects and allow each unit to have its own coefficients $\boldsymbol{\beta}_{4i}$ for the observed global shocks and $\boldsymbol{\gamma}_i$ for the unobserved ones. The variation of $\boldsymbol{\beta}_{4i}$ may be future explained with a matrix of predictors, $\mathbf{A}_2$, which can be the same as $\mathbf{A}_1$ or its subset or contain a different set of unit-level predictors. With the assumption that $\mathbf{f}_{t_i}$ and $\boldsymbol{\gamma}_i$ are distributed with mean $\mathbf{0}$, the term $\mathbf{f}'_{t_i}\boldsymbol{\gamma}$ is a component of the overall residual term in equation (4.7). The idiosyncratic error ξ_{i,t_i} follows a pth order autoregressive process to correct serial correlation. The reduced form of this model is just a standard multifactor residual model (m factor loadings) with autoregressive errors. To see it more clearly, set $\mathbf{X}_{i,t_i} = (\mathbf{x}_{1,it_i}, \mathbf{w}'_{i,t_i}\mathbf{A}_{1i}, \boldsymbol{v}'_{t_i}\mathbf{U}_{t_i}, \mathbf{s}_{t_i}\mathbf{A}_{2i})$,

$\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)$ (K_1 dimensions), $\mathbf{W}_{i,t_i} = (\mathbf{w}_{i,t_i}, \mathbf{s}_{t_i})$, and $\mathbf{b}_i = (\mathbf{b}_{1i}, \mathbf{b}_{2i})$($K_2$ dimensions), and

$$z_{i,t_i} = \mathbf{X}'_{i,t_i}\boldsymbol{\beta} + \mathbf{W}'_{i,t_i}\mathbf{b}_i + \boldsymbol{v}'_{i,t_i}\boldsymbol{\eta}_{t_i} + \mathbf{f}'_{t_i}\gamma_i + \xi_{i,t_i} \tag{4.10}$$

$$\xi_{i,t_i} = \rho_1\xi_{i,t_i-1} + ... + \rho_p\xi_{i,t_i-p} + \epsilon_{i,t_i}. \tag{4.11}$$

Normally the factors $\mathbf{f}$ have no direct theoretical interpretation, and the decision of how many factors to be included is not made based on substantive consideration (Lord and Novick, 1968; Mulaik, 1988*a*,*b*; Skrondal and Laake, 2001). The so-called Kaiser-Guttmann criterion is often applied for estimation reasons, which requires the number of factors be equal to the number of eigenvalues of the correlation matrix that are larger than one. With latent response variables and complex covariance matrices, this criteria is difficult to apply, and the choice of the number of factors is often *ad hoc* (Skrondal and Rabe-Hesketh, 2004, pp.63-71). In empirical models we can always increase the dimension of $\mathbf{f}$ and investigate how the results differ from a lower dimensional $\mathbf{f}$, and further apply information-based criteria to determine the number of factors which makes the model fit the data better. This model is also a two-way mixed effects model, but no assumption of independence between random coefficients is needed; this assumption is normally required in standard two-way mixed effects models (Weerahandi, 2004, p.107-113). Finally, we need another assumption to identify the parameters of $\boldsymbol{\gamma}$. Two identification strategies are often adopted in the literature of multifactor residual models, namely, "anchoring" (set $\boldsymbol{\gamma}_1 = \mathbf{1}$) and "factor standardization" (set $\sum_{n=1}^{N} \gamma_i^2 = 1$). The "anchoring" method is preferred because it achieves "factorial invariance" (Skrondal and Rabe-Hesketh, 2004, pp.67), which is adopted in this paper.

Although the model is very general, it is difficult to estimate because of multiple sources of correlation in different dimensions in this model. Philips and Sul (2003) apply a GLS-SURE method for an autoregressive linear model with random coefficients, but their approach can only apply to unidimensional factor specifications and the asymptotic properties of their estimator are not well developed. A similar model with an addition assumption of homogeneous effects of common shocks can be estimated by using the hybrid Markov Chain Monte Carlo algorithm developed in Pang (2009). The big advantages of using the Bayesian approach lie in its modularity of estimating algorithms and its flexibility of accommodating unbalanced data structures. By treating unobserved factors and factor loadings as parameters, the Bayesian approach allows them to be arbitrarily correlated with regressors. The algorithm applied to estimate the model is based on Pang (2009): it uses the Cholesky decomposition to diagonalize the variance-covariance matrix and adds an auxiliary variable to orthogonalize the idiosyncratic error term. The algorithm includes five Gibbs steps for augmenting the latent data and updating most parameters, and an MH chain to update the autoregressive coefficients. A partial group move multigrid Monte Carlo method is used to speed up MCMC mixing, as is applied in Pang (2009). The detailed prior assignments, MCMC algorithm, and Bayes factor computational scheme are reported in the Appendices 4.6.1, 4.6.2, and 4.6.3.

4.4 Empirical Results and Interpretation

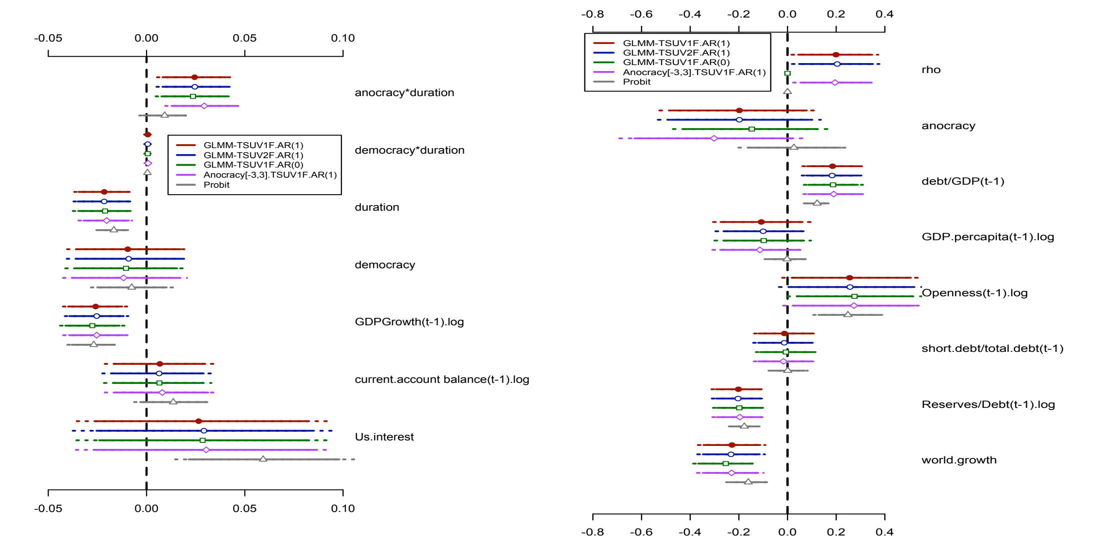
I apply the GLMM-TSUV.AR(p) model to empirically test the regime-dependent effect of regime duration by using an interaction term,
`anocracy*duration`, and investigate the impacts of observed and unobserved common shocks on sovereign default by estimating the time-specific unit-varying effects

of common shocks. Another interaction, `democracy*duration`, is included to test whether the effect of regime duration varies with democracy level or whether the effect of the level of democracy is altered by regime duration. To estimate the country-varying effects of common shocks, I use country-specific coefficients for both observed (`US.Interest` and `world.growth`) and unobserved ($\mathbf{f}$) common shocks. A country-specific random intercept is included in the model to control for unobserved heterogeneity. I estimate the model with different lag orders and different numbers of unobserved factors. The Bayes factors support the model with an AR(1) error process and one unobserved factor. The empirical results based on models with more factors and high lag orders are all very similar to GLMM-TSUV1F.AR(1). I only report two of them, i.e., GLMM-TSUV2F.AR(1) and GLMM-TSUV1F.AR(2). I also estimate the same specification (GLMM-TSUV1F.AR(1)) with an alternative measure of the variable `anocracy` by using the subinterval $[-3, 3]$. Furthermore, just to show how much difference this more realistic model makes on the fixed-effect coefficients, I compare the model with a simple probit model and a model with the assumption that common shocks have the same effects on all countries. Interestingly, for the model with constant effects of common shocks, the MCMC draws demonstrates very strong serial correlation, which casts the doubt on stationarity of the error process and cointegration relations. With more than 5,000,000 iterations the Markov chain still demonstrates strong evidence of nonconverge, as shown in Appendix 4.6.5. Although the Markov chain fails to converge, I graphically summarize the MCMC output in Appendix 4.6.5, but do not compare the output with other models, since the draws are likely not from the stationary distribution.

In Figure 4.3, I summarize the posterior distributions of the fixed-effect coefficients based on five models and also present their marginal likelihood. The posterior distributions based on various GLMM-TSUV models are very similar, and using one or two unobserved factors do not make any important difference. Because serial correlation is estimated to be positive but weak (0.2 at the mean level), correcting serial correlation improves model quality (the Bayes factor of GLMM-TSUV.AR(1) vs. GLMM-TSUV.AR(0) is 1.140), but does not make notable difference in posteriors of the fixed-effect coefficients. Coding `anocracy` as polity score between -3 and $+3$ changes the magnitude of the posterior mean of `anocracy` and its interaction with `duration`, but the directions are unchanged and error bands roughly the same. It also leaves the coefficient parameters associated with other variables almost unaffected. The posteriors of the completely pooled probit model are saliently different from other models: the error bands are much smaller and biased towards 0 most of the time. The model suggests that US interest rate has a positive effect at a 95% level of credibility, but based on the four other models, this positive effects is only at a 70% level of credibility; as well, the interaction term of `anocracy` and `duration` has a much smaller and more uncertain positive effect on sovereign default based on the probit model than other models. The Bayes factors least prefer the probit model and decisively support the GLMM-TSUV1F.AR(1) model over all four other models. In the rest of this paper, I analyze the empirical results mainly based on the GLMM-TSUV1F.AR(1) model.

Based on GLMM-TSUV1F.AR(1), the posterior of the interaction term, `anocracy*duration`, confirms the theory that, regime duration of anocracies affects sovereign default differently from that of non-anocracies. As shown in the left graph in Figure 4.4, under nonanocratic regime the effect of regime duration is negative with high certainty. The substantive implication is that as regime age increases, the likelihood of sovereign default decreases in nonanocracies, *ceteris paribus*, but anocracy drives the effect to an opposite direction: more than half of the posterior draws of regime duration lie on the right hand side of zero, and at the 70% credibility level we can conclude that as

Figure 4.3: Posterior Summary: Fixed-Effect Coefficients in Five Competing Models

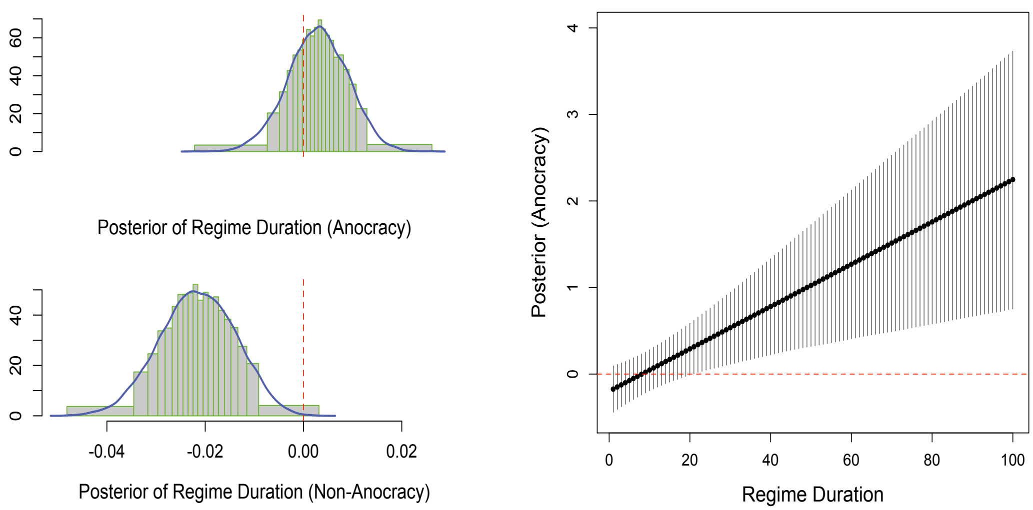


Marginal Likelihood GLMM-TSUV1F.AR(1): -598.775, GLMM-TSUV2F.AR(1): -604.273, GLMM-TSUV1F.AR(0): -601.409, Anocracy$[-3, 3]$: -601.076, Probit: -654.583

regime age increases, an anocratic government is more likely to default on its external debt, holding other things constant. In Figure 4.4, the right graph presents the duration-dependent effect of anocracy: anocracies tend to be more likely to default when regime duration is longer. In the first couple of years of anocracies, regime type does not have a clear effect on sovereign default, but upon staying in anocracy longer, major regime changes are more likely. A shorter time horizon makes an anocratic government value future benefits less and more prefer sovereign default to sacrificing current consumption.

Figure 4.5 shows estimated heterogeneity among countries. In all three graphs, the sample countries are sorted by the posterior mean of the random effect. The left graph shows the posteriors of the country-specific intercepts, which capture unobserved heterogeneity across countries. With the selected covariates and unit-varying effects of common shocks, the posteriors do not show much heterogeneity. The middle and

Figure 4.4: Regime-Dependent Effect of Regime Duration and Duration-Dependent Effect of Anocracy



right graphs in Figure 4.5 illustrate the *variations* of the effects of global GDP growth and the US interest rate. In the model, the variations are captured by the group-level residual term (b_{2i} and b_{3i}). The grand (average) effects of the two observed common shocks are presented in Figure 4.3. Countries demonstrate different levels of sensitivity or vulnerability to both of the observed common shocks in terms of sovereign default: the range of the posterior mean of b_{2i} for global GDP growth is from -0.278 to 0.379 and of b_{3i} for the US interest rate -0.128 to 0.241. Only ten countries which are not sensitive (countries in the middle range in the graph) to global GDP growth are also insensitive to US interest rate change. There are 12 countries benefiting from both macroeconomic and liquidity shocks (the left third of countries in both graphs), and13 countries vulnerable to both types of common shocks (the right third of countries in both graphs). More than 70% of the sample countries respond to the macro-economic shock differently from the global liquidity shock.

Figure 4.5: Unobserved Heterogeneities and Varying Effects of Observed Common Shocks

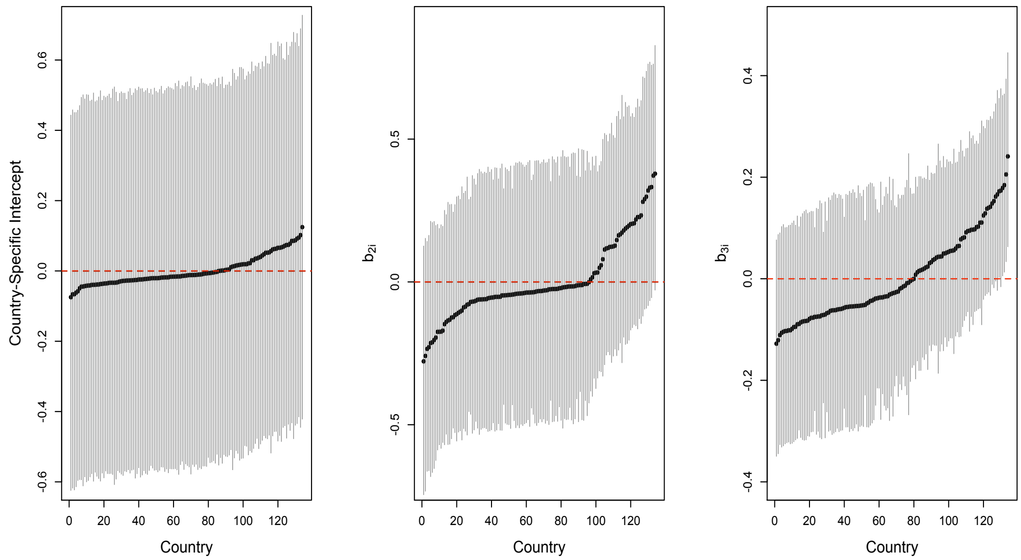


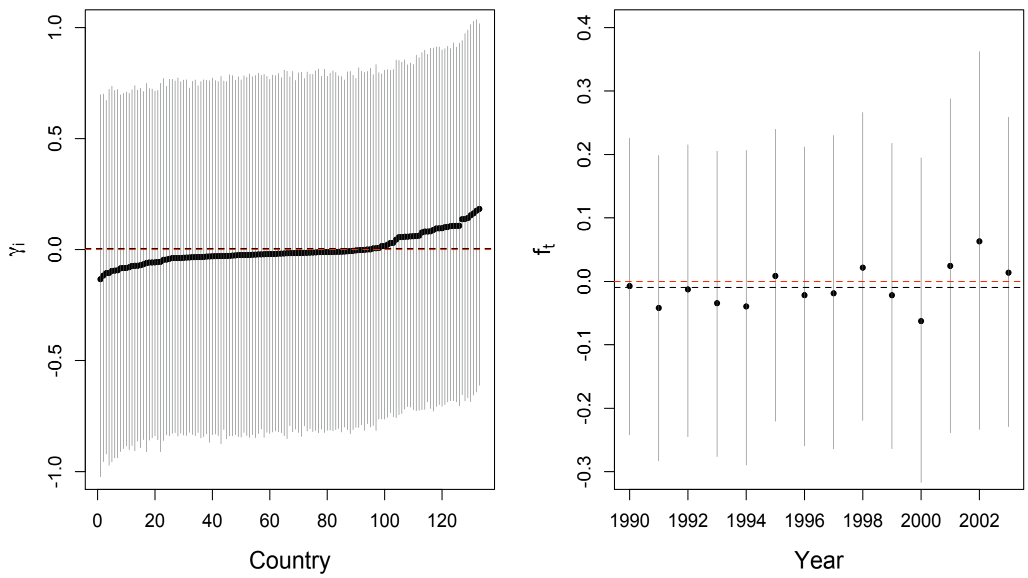
Figure 4.6 shows the country-specific effects ($\boldsymbol{\gamma}_i$) of unobserved common shocks ($\mathbf{f}_t$), the augmented unobserved common shocks themselves, and the time-specific intercepts which capture time heterogeneity. The left graph shows that the sensitivity of the sample countries to unobserved common shocks varies at the posterior mean level from -0.1333 to 0.183. With large error bands (a 95% credible intervals), the variation is not very salient but still demonstrates that countries respond to unobserved common shocks differently even after the observed common shocks are controlled for. Interestingly, countries follow a very similar pattern in terms of responding to unobserved common shocks as to global GDP growth: for 88.6% of the countries (39 out of 44) which benefit from global GDP growth (the right 1/3 countries in Figure 4.5), unobserved common shocks are also most likely to reduce their likelihood of sovereign default (the right 1/3 countries in Figure 4.6); 93.2% of the most vulnerable countries (41 out of 44) to global GDP growth are also the countries

most vulnerable to unobserved common shocks (the left 1/3 countries); and 82.6% of countries insensitive to global GDP growth are barely affected by the unobserved common shocks (the middle 1/3 countries). However, the pattern of country-specific response to unobserved shocks is quite different from to that of the US interest rate, and the percentages are 28.9%, 19.6% and 61.4% respectively. This may suggest that the omitted or unobserved shocks are mostly macro-economic shocks besides global GDP growth.

The middle graph in Figure 4.6 summarizes the augmented overall unobserved shocks which have country-specific effects, and the right graph shows the unobserved common shocks with the same effect on all countries, i.e., the time-specific intercepts. To compare the two different types of shocks or the two component effects of unobserved shocks, I put the two graphs on the same scale. The mean time trends of the two types of shocks are different from each other, and the error bands are much smaller for the country-invariant unobserved shock than for the country-varying ones. Both types of shocks demonstrate that in 2001 the unobserved shocks in the international system are strongest and in the direction of increasing sovereign default probability. Methodologically, relaxing the assumption of constant effects of common shocks avoids the non-cointegration problem: within-chain serial correlation reduces greatly, and the posterior mean of the autoregressive coefficient is only around 0.2. The chain converges much faster, and within-chain mixing is greatly improved (see Appendix 4.6.5).

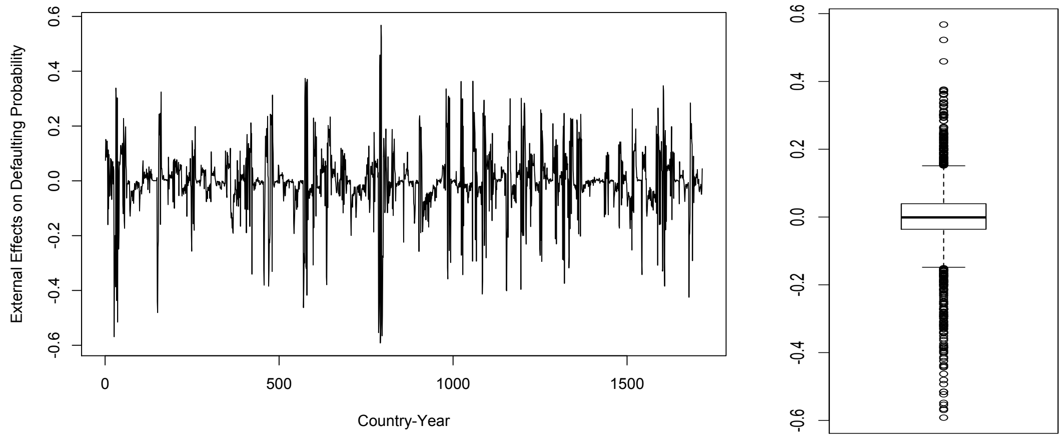
It is more intuitive to see how the common shocks in the system affect sovereign default by looking at their effects on default probability instead of on the absolute value at the latent level. Country-specific impacts of common shocks are calculated by using all the MCMC posterior draws, conducting within-sample prediction, and calculating the difference between predicted probabilities with and without taking

Figure 4.6: Unobserved Factor and Factor Loading



account of the effects of common shocks. In Figure 4.7, I plot the mean levels of the impacts of common shocks on default probability of all the sample country-years. The left graph, with the sample country-years not sorted, shows the volatility of the impact of the international system on default probability, and the right graph shows the distribution of this impact. Both graphs demonstrate that the international system plays an important role in explaining countries' decision making regarding sovereign default: in some country-years, common shocks in the system can increase default probability by as large as 56.80% or decrease it by 59.20%. For half of the country-years, external shocks change default probability by more than 3.80%, holding other variables constant. The impact on default probability also varies greatly, and its standard deviation is 0.118. Figure 4.8 shows the mean-level impact of external shocks on each individual country over their sample years. On average, 44% of the sample countries (59 countries) generally benefit from the international system in terms of repaying their external debt, but system shocks increase default probability for other

Figure 4.7: Unit-Varying Impact of Time-Specific Shocks I



75 countries. Common shocks reduce the default likelihood most dramatically for Jordan by 46.80%, other things equal; but the system increases default probability most for Mauritania by 19.7% on average. There are 15 countries which are particularly vulnerable to the international system. It increases their default risk by more than 10%. And there are another 3 countries besides Jordan (i.e., Angola, Gabon, Democratic Republic of the Congo) whose default probabilities are lowered by the international system by more than 10% at the mean level. Figure 4.9 shows the impacts of common shocks in each year. In each year the response of countries to common shocks varies greatly because of their own characteristics; and in different years the volatility of the effect of common shocks are also different. In some years, such as 1999 and 2000, the international system is more tranquil, but in 1992, 2001 and 2002, the overall impact of external shocks is much bigger than in the rest of the sample years.

Finally, as for the control variables, the posteriors based on the GLMM-TSUV1F.AR(1) model are summarized in Figure 4.3. Domestic GDP growth is important for countries to service their external debt. Slow economic growth predicts higher likelihood

Figure 4.8: The Impact of Unobserved Unit-Varying Time-Specific Shocks II

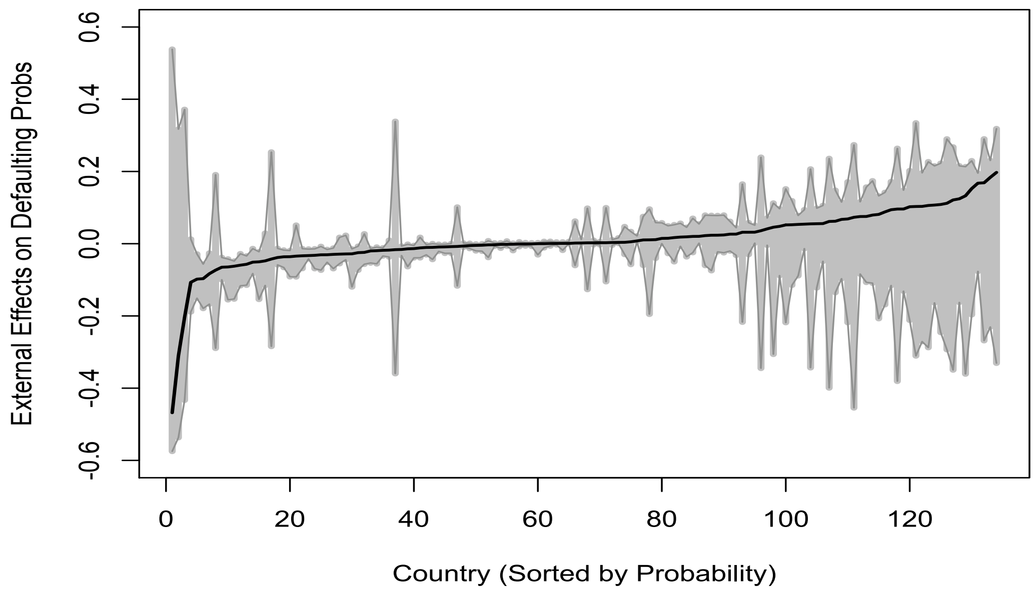
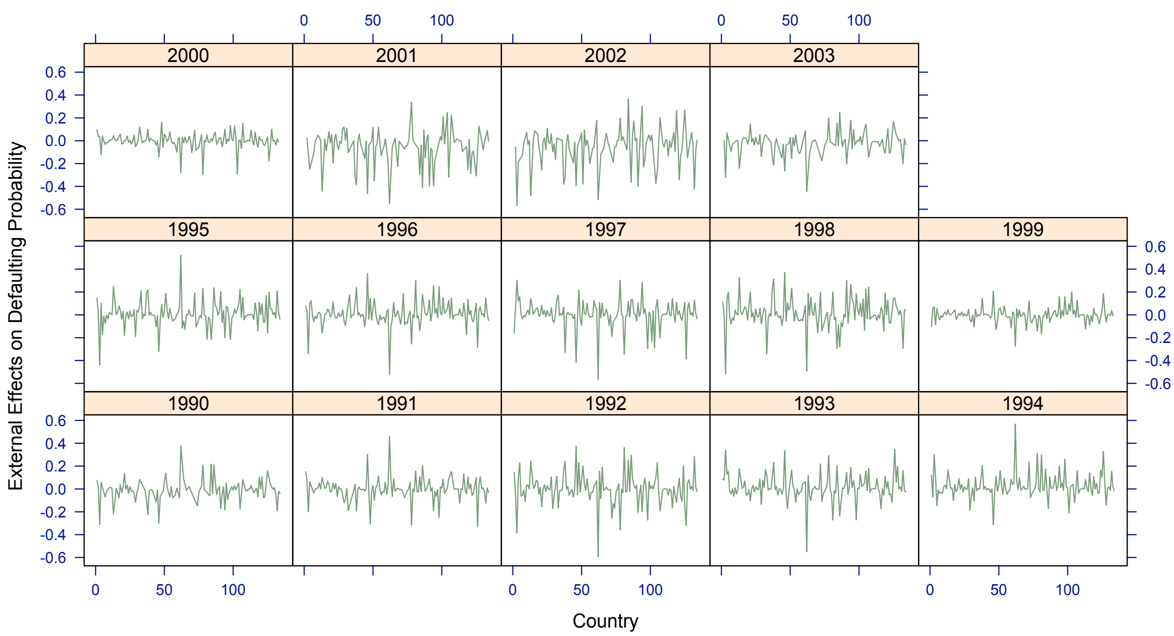


Figure 4.9: The Impact of Unobserved Unit-Varying Time-Specific Shocks III



of sovereign default. While current account balance does not have a clear effect on sovereign default, a high ratio of external debt to output is a dangerous sign that a country may be insolvent and more likely to default. It seems that the configuration of external debt does not matter when explaining sovereign default, and short-term debt is not necessarily associated with high default probability, but having a big foreign exchange reserve relative to output in a country is important to avoid sovereign default. Interestingly, trade openness is likely to be associated with higher default probability, which does not confirm the conventional theory. This may be attributed to the fact that for developing countries, an open economy often has a high degree of sensitivity and vulnerability to external shocks. Even if trade openness can increase a government's incentive to repay debt, it may imply the capability to repay the external debt of an open economy can be volatile because of international shocks.

4.5 Discussion

There are many studies in the literature trying to find a stable and reliable relationship between regime type and sovereign default, but empirical evidence is often mixed when the relationship is treated as linear, unconditional, and outside the international context of sovereign default. This paper investigates this question from a historical and international perspective by emphasizing the regime-dependent effect of regime duration and controlling for countries' sensitivity or vulnerability to the international system. I proposed a Bayesian multifactor generalized linear multilevel model motivated by the characteristics of the empirical data and the major research goals of the current research. I applied the model to analyze time-series cross-sectional data with 134 developing countries from 1990 to 2003, and the empirical findings suggest that regime duration of anocracies has a very different meaning from non-annocracies when explaining sovereign default. Anocratic governments have shorter time horizons with increasing regime age, but governments of nonanocracies expect higher likelihood of regime survival with a longer regime history. This finding is also based on carefully controlling for an important but often neglected source of variation, i.e., the country-specific effect of the external environment. Empirical evidence suggests that shocks in the international system strongly affect the national decision-making regarding sovereign default in developing countries, and the impacts of globalization vary widely from country to country. This finding not only suggests that models without controlling for the international source of variation would generate misleading inferences, but also sheds light on questions about the importance of globalization and its different meanings to different countries.

4.6 Appendix

4.6.1 Priors

The prior assignment for a GLMM-TSUV.AR(p) model with m unobserved factors is as follows

$$\boldsymbol{\beta} \sim N_{K_1}(\boldsymbol{\beta}_0, \mathcal{B}_0), \qquad \epsilon_{i,t_i}, \sim N(0,1), \qquad \{\mathbf{b}_i, \boldsymbol{\gamma}_i\} \sim N_{K_2+m}(\boldsymbol{\varsigma}, \mathbf{D}),$$

where the vector $\boldsymbol{\varsigma}$ is specified as: $\boldsymbol{\varsigma} = (\underbrace{0,...,0}_{K_2}, \underbrace{a_1,...,a_m}_{m})$

$$\mathbf{D}^{-1} \sim W_{K_2+m}(\boldsymbol{d}_0, \mathbf{D}_0), \qquad \{\boldsymbol{\eta}_t, \mathbf{f}_t\} \sim N_m(\mathbf{0}, \mathbf{E}), \qquad \mathbf{E}^{-1} \sim W_{K_3+m}(\boldsymbol{e}_0, \mathbf{E}_0),$$

$\boldsymbol{\rho} \sim U_p(\boldsymbol{\rho} : \boldsymbol{\rho} \in S_\rho)$,

4.6.2 MCMC Algorithm

Following Pang (2009), I use the Cholesky decomposition to diagonalize the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i}$: $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i} = \boldsymbol{\Omega}_i + \kappa \boldsymbol{I}_i$ (where κ is any constant). The symmetric positive definite matrix $\boldsymbol{\Omega}_i$ is further decomposed as $\boldsymbol{V}_i'\boldsymbol{V}_i$, in which $\boldsymbol{V}_i'$ is the lower triangular matrix produced by the Cholesky decomposition. The covariance matrix is then expressed as $\boldsymbol{\Sigma}_{\boldsymbol{xi}_i} = \boldsymbol{V}_i'\boldsymbol{V}_i + \kappa \boldsymbol{I}_T$. By adding an auxiliary variable $\boldsymbol{u}_i \sim N(\mathbf{0}, \boldsymbol{I})$, the error term $\boldsymbol{\xi}_i$ can be written as $\boldsymbol{V}_i'\boldsymbol{u}_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i \sim N(0, \kappa_i \boldsymbol{I}_T)$. Then, the MCMC algorithm is simplified as follows:

1. $\mathbf{D}^{-1}|\{\mathbf{b}_i\}, \{\gamma_i\} \sim W_{K_2+m}(\boldsymbol{d}_1, \mathbf{D}_1)$ and $\mathbf{E}^{-1}|\{\boldsymbol{\eta}_t, \mathbf{f}_t\} \sim W_m(\boldsymbol{e}_1, \mathbf{E}_1)$, where $\boldsymbol{d}_1 = \boldsymbol{d}_0 + N$, $\mathbf{D}_1 = (\mathbf{D}_0^{-1} + \sum_{i=1}^N (\mathbf{b}_i, \boldsymbol{\gamma}_i)(\mathbf{b}_i, \boldsymbol{\gamma}_i)')^{-1}$, $\boldsymbol{e}_1 = \boldsymbol{e}_0 + T$, and $\mathbf{E}_1 = (\boldsymbol{E}_0^{-1} + \sum_{i=1}^T \{\boldsymbol{\eta}_t, \mathbf{f}_t\}\{\boldsymbol{\eta}_t', \mathbf{f}_t'\})^{-1}$

2. $\boldsymbol{u}_i|\cdot \sim N(\bar{\boldsymbol{u}}_i, \mathbf{U}_i)$, where $\mathbf{U}_i = (\boldsymbol{I}_T + \boldsymbol{V}_i\boldsymbol{V}_i'/\kappa)^{-1}$, and $\bar{\boldsymbol{u}}_i = \mathbf{U}_i\boldsymbol{V}_i(\boldsymbol{z}_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \mathbf{W}_{i,t_i}'\boldsymbol{\mathcal{B}}_i - \boldsymbol{v}_{i,t_i}'\boldsymbol{\eta}_{t_i} - \mathbf{f}_{t_i}\gamma_i)/\kappa_i$

3. $z_{it}|\cdot \sim TN(\mathbf{X}_{i,t_i}'\boldsymbol{\beta} + \mathbf{W}_{i,t_i}'\mathbf{b}_i + \boldsymbol{v}_{i,t_i}'\boldsymbol{\eta}_{t_i} + \mathbf{f}_{t_i}'\gamma_i + q_{it}, \kappa_i)$

4. $\boldsymbol{\beta}|\cdot \sim N_{K_1}(\bar{\boldsymbol{\beta}}, \mathbf{B}_1)$, where $\mathbf{B}_1 = (\mathbf{B}_0^{-1} + \sum_{i=1}^N \boldsymbol{x}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{x}_i)^{-1}$,
$\bar{\boldsymbol{\beta}} = \mathbf{B}_1 \left(\mathbf{B}_0\boldsymbol{\beta}_0 + \sum_{i=1}^N \mathbf{x}_i'\boldsymbol{\Omega}_i^{-1}(\mathbf{z}_i - \mathbf{W}_{i,t_i}'\mathbf{b}_i - \boldsymbol{v}_{i,t_i}'\boldsymbol{\eta}_{t_i} - \mathbf{f}_{t_i}\gamma_i)\right)$;

5. $\{\mathbf{b}_i\}, \{\gamma_i\}|\cdot \sim N_{K_2+m}(\bar{\gamma}_1, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma} = (\boldsymbol{D}^{-1} + (\boldsymbol{W}_i, \mathbf{f}_{t_i})'\boldsymbol{\Omega}^{-1}(\boldsymbol{W}_i, \mathbf{f}_{t_i}))^{-1}$, and
$\bar{\gamma}_1 = \boldsymbol{\Gamma}\left((\boldsymbol{W}_i, \mathbf{f}_{t_i})'\boldsymbol{\Omega}^{-1}(\mathbf{z}_i - \mathbf{x}_i\boldsymbol{\beta} - \boldsymbol{v}_{i,t_i}'\boldsymbol{\eta}_{t_i}) + \mathbf{D}^{-1}\boldsymbol{\varsigma}\right)$.

6. $\{\boldsymbol{\eta}_t, \mathbf{f}_t\}|\cdot \sim N_m(\bar{\mathbf{f}}_t, \mathbf{F}_1)$, where $\mathbf{F}_1 = (\mathbf{E}^{-1} + (\boldsymbol{v}_{it}, \boldsymbol{\gamma}_i)'(\kappa_{N_t}\boldsymbol{I}_N)^{-1}(\boldsymbol{v}_{it}, \boldsymbol{\gamma}_i))^{-1}$
and $\bar{\mathbf{f}}_t = \boldsymbol{F}_1(\mathbf{v}_{it}, \boldsymbol{\gamma}_i)'(\kappa_{N_t}\boldsymbol{I}_N)^{-1}(\mathbf{z}_i - \mathbf{x}_i'\boldsymbol{\beta} - \mathbf{W}_t'\mathbf{b}_{N_t} - q_t)$;

7. $\boldsymbol{\rho}|\cdot \sim \Psi(\boldsymbol{\rho}) \times N(\hat{\boldsymbol{\rho}}, \boldsymbol{P})$, following Chib (1993), I use a Metropolis-Hasting algorithm to update $\boldsymbol{\rho}$ by using the tailored kernel $N(\hat{\boldsymbol{\rho}}, \boldsymbol{P})$.

4.6.3 Bayes Factor Computational Scheme

I use the marginal likelihood approach of Chib (1995) and Chib and Jeliazkov (2001) to compute the Bayes Factor. For the likelihood ordinate, denote by $\boldsymbol{\theta}$ all the parameters, except the auxiliary parameter $\boldsymbol{u}$, and apply the following formula:

$$\hat{f}(\boldsymbol{y}|\boldsymbol{\theta}^*) = \frac{1}{M}\sum_{m=1}^{M}\prod_{i=1}^{N}\prod_{t_i=1}^{T_i}\Delta^{y_{it_i}}\left(1-\Delta\right)^{1-y_{it_i}},$$
$$\text{where, } \Delta = \Phi\left(\frac{= \mathbf{X}'_{i,t_i}\boldsymbol{\beta} + \mathbf{W}'_{i,t_i}\mathbf{b}_i + \boldsymbol{v}'_{i,t_i}\boldsymbol{\eta}_{t_i} + \mathbf{f}'_{t_i}\gamma_i + q_{it_i}^{(m)}}{\sqrt{\kappa_i^{(m)}}}\right) \tag{4.12}$$

The approach to approximate the posterior ordinate is as follows:

$$\hat{\pi}(\boldsymbol{\beta}^*, \mathbf{b}^*, \gamma^*, \mathbf{D}^*, \boldsymbol{\rho}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \mathbf{E}^*|\mathrm{y}) = \hat{\pi}(\boldsymbol{\rho}^*|\boldsymbol{y})\hat{\pi}(\boldsymbol{\eta}^*, \mathbf{f}^*|\boldsymbol{\rho}^*, \mathrm{y})\hat{\pi}(\mathbf{E}^*|\boldsymbol{\rho}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{y})\hat{\pi}(\mathbf{b}^*, \gamma^*|\boldsymbol{E}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \boldsymbol{y})$$
$$\times\, \hat{\pi}(\boldsymbol{D}^*|\mathbf{b}^*, \gamma^*, \mathbf{E}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathrm{y})\hat{\pi}(\boldsymbol{\beta}^*|\mathbf{D}^*, \mathbf{b}^*, \gamma^*, \mathbf{E}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathrm{y})$$

1. For the ordinate $\hat{\pi}(\boldsymbol{\rho}^*|\boldsymbol{y})$, denote all parameters except $\boldsymbol{\rho}$ and $\boldsymbol{u}$ as $\boldsymbol{\psi}$:

$$\hat{\pi}(\boldsymbol{\rho}^*|\boldsymbol{y}) = \frac{J^{-1}\sum_{i=1}^{N}\Big(\alpha(\boldsymbol{\rho}^{(j)}, \boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\psi}^{(j)}, \boldsymbol{u}^{(j)}, \boldsymbol{z}^{(j)})q(\boldsymbol{\rho}^{(j)}, \boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\psi}^{(j)}, \boldsymbol{u}^{(j)}, \boldsymbol{z}^{(j)})\Big)}{K^{-1}\sum_{k=1}^{K}\Big(\alpha(\boldsymbol{\rho}^*, \boldsymbol{\rho}^{(k)}|\boldsymbol{y}, \boldsymbol{\psi}^{(k)}, \boldsymbol{u}^{(k)}, \boldsymbol{z}^{(k)})\Big)}. \tag{4.13}$$

   The numerator is the sample expectation with respect to $\pi(\boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{z}|\boldsymbol{y})$, and the MCMC output can be directly used to integrate those parameters in the conditional part. The denominator is the sample expectation with respect to the conditional product measure $\pi(\boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{z}|\boldsymbol{y})q(\boldsymbol{\rho}^*, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\psi}, \boldsymbol{u}, \boldsymbol{z})$. Here, one reduced run is needed: fix $\boldsymbol{\rho}$ at $\boldsymbol{\rho}^*$, run a reduce run to get $\boldsymbol{\psi}$ and $\boldsymbol{\rho}^{(k)}$ in each iteration by using $\boldsymbol{\psi}^{(k)}$, and then plug all the draws of the parameters and augmented data into the denominator, and compute the quantity.

2. $\hat{\pi}(\boldsymbol{\eta}^*, \mathbf{f}^*|\boldsymbol{\rho}^*, \mathrm{y})$: directly use the reduced run conducted above;

3. $\hat{\pi}(\mathbf{E}^*|\boldsymbol{\rho}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{y}) = \hat{\pi}(\mathbf{E}^*|\boldsymbol{\eta}^*, \mathbf{f}^*)$: no reduced run is required;

4. $\hat{\pi}(\mathbf{b}^*, \gamma^* | \mathbf{E}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathbf{y}) = \prod_{i=1}^{N} \pi(\mathbf{b}^*, \gamma^* | \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathbf{z}_i)$: Conduct a reduced run by fixing $\mathbf{E}, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}$, and keep the output of $\boldsymbol{\beta}, \mathbf{D}, \mathbf{z}$ together with the fixed values to compute this quantity;

5. $\hat{\pi}(\boldsymbol{D}^* | \mathbf{b}^*, \gamma^*, \mathbf{E}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathbf{y}) = \hat{\pi}(\boldsymbol{D}^* | \mathbf{b}^*, \gamma^*)$: no reduced run is required here;

6. $\hat{\pi}(\boldsymbol{\beta}^* | \mathbf{D}^*, \mathbf{b}^*, \gamma^*, \mathbf{E}^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathbf{y}) = \prod_{i=1}^{N} \pi(\boldsymbol{\beta}^* | \mathbf{b}^*, \gamma^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*, \mathbf{z})$: conduct a reduced run by fixing $\mathbf{D}, \mathbf{b}, \gamma, \mathbf{E}, \boldsymbol{\eta}, \mathbf{f}, \boldsymbol{\rho}$ and keep the output of $\mathbf{z}$ together with the fixed values of $\mathbf{b}^*, \gamma^*, \boldsymbol{\eta}^*, \mathbf{f}^*, \boldsymbol{\rho}^*$ to compute this quantity.

4.6.4 Sample Countries and Years

4.6.5 Nonstationarity of Error Process with Unit-Identical Effects of Common Shocks

Figure 4.10: Parameter "Posteriors" Based on GLMM-AR(1)

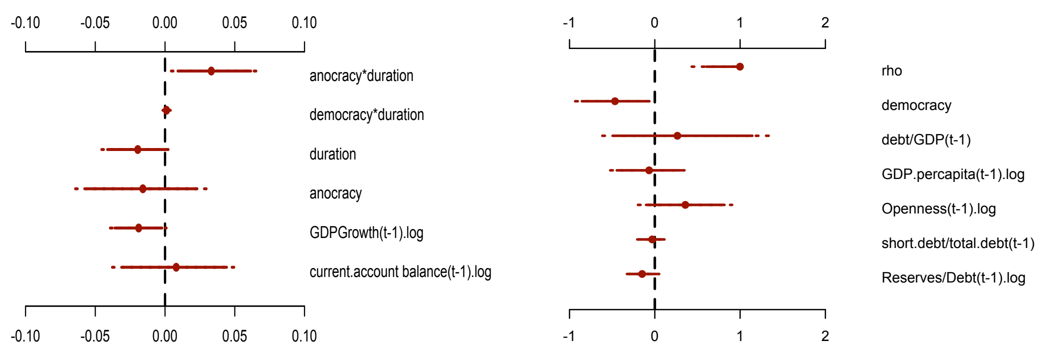
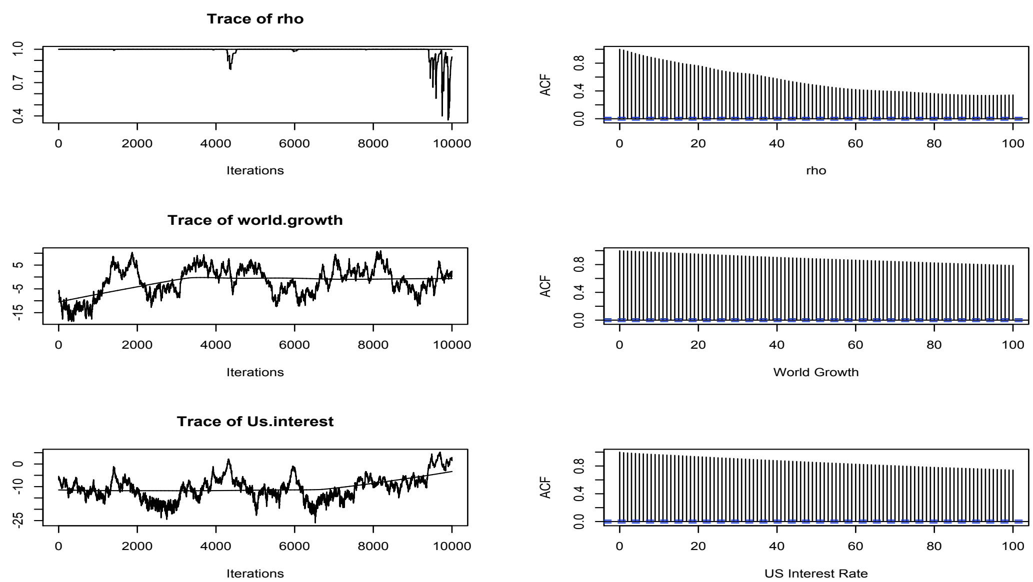


Table 4.2: Sample Countries and Years

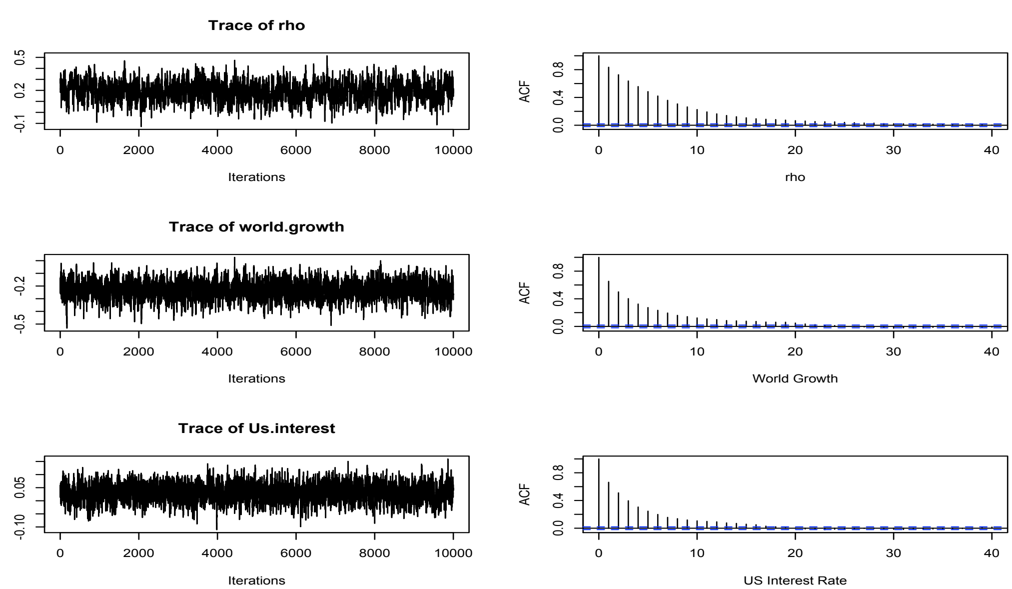
Country	Years	Defaulting Episoda	Country	Years	Defaulting Episoda
Albania	1993-2003	2000-2001	Lebanon	1990-2003	1990
Algeria	1990-2003	1995, 1997	Lesotho	1990-2003	
Angola	1991-2003	1991-1995, 1998-2001	Liberia	1998-2003	
Argentina	1990-2003	1990-1991, 2002-2003	Lithuania	1994-2003	
Armenia	1995-2003	1997-1998	Macedonia	1995-2003	1995, 2000
Azerbaijan	1995-2003		Madagascar	1990-2003	
Bangladesh	1990-2003		Malawi	1990-2003	
Barbados	1990-2003		Malaysia	1990-2003	
Belarus	1995-2003		Maldives	1990-2003	
Belize	1990-2003	1992	Mali	1991-2003	1998, 2002
Benin	1990-2003	1996	Mauritania	1990-2003	1994-1995, 1999-2003
Bhutan	1990-2003		Mauritius	1990-2003	
Bolivia	1990-2003	1990-1993, 1999	Mexico	1990-2003	1990
Bosnia and Herzegovina	2001-2003		Moldova	1994-2003	1996, 1998, 2000, 2002
Botswana	1990-2003		Mongolia	1994-2003	
Brazil	1990-2003	1990, 1993	Morocco	1990-2003	
Bulgaria	1993-2003	1993, 1997	Mozambique	1990-2003	1990, 1998-1999, 2001
Burkina Faso	1990-2003	1996	Nepal	1990-2003	
Burundi	1990-2003		Nicaragua	1990-2003	1990-1991, 1996-1997, 2001
Cambodia	1990-2003	1991	Niger	1990-2003	1996, 2001-2002
Cameroon	1990-2003	1990-1991, 1995-1996, 1998	Nigeria	1990-2003	1990, 1992-1993, 2001
Cape Verde	1990-2003		Oman	1990-2002	
Central African Republic	1990-2003	1999	Pakistan	1990-2003	2000-2001
Chad	1990-2003		Panama	1990-2003	1990-1993, 1995
China	1990-2003		Papua New Guinea	1990-2003	
Chile	1990-2003	1990	Paraguay	1990-2003	1990
Colombia	1990-2003		Peru	1990-2003	1990, 1993-1995, 1998
Comoros	1990-2003	1993-1994	The Philippines	1990-2003	1990-1992
Congo, Dem. Rep.	1990-2003	1990	Poland	1991-2003	1991, 1993-1994
Congo, Rep.	1990-2003	1991, 1996, 1998	Romania	1991-2003	
Costa Rica	1990-2003	1991	Russia	1994-2003	1994-1996, 1998, 2000
Cote d'Ivoire	1990-2003	1990-1991, 1994, 1997, 2002	Rwanda	1990-2003	1998
Croatia	1995-2003	1995, 2001	Samoa	1990-2003	
Czech Republic	1995-2003		Sao Tome & Principe	1990-2003	1997, 2000
Djibouti	1996-2003		Senegal	1990-2003	1994, 1998
Dominica	1990-2003		Serbia and Montenegro	2000-2003	2000, 2003
Dominican Republic	1990-2003	1990-1993	Seychelles	1990-2003	1991, 2001
Ecuador	1990-2003	1990-1994, 2000	Sierra Leone	1990-2003	1990-1991, 1993, 1997
Egypt	1990-2003		Slovak Republic	1995-2003	
El Salvador	1990-2003	1993	Solomon Islands	1990-2003	1992, 1995, 2000
Equatorial Guinea	1990-2001		South Africa	1996-2003	
Eritrea	1996-2003		Sri Lanka	1990-2003	
Estonia	1994-2003		St. Kitts & Nevis	1990-2003	
Ethiopia	1990-2003		St. Lucia	1990-2003	
Fiji	1990-2003		St. Vincent & the Grenadines	1990-2003	2003
Gabon	1990-2003	1990-1993, 1997-1998, 2000, 2003	Sudan	1990-2003	1993-1994, 2001
Gambia	1990-2003		Swaziland	1990-2003	
Georgia	1994-2003	1994, 1996, 2002	Syrian Arab Republic	1990-2003	
Ghana	1990-2003	2001	Tajikistan	1994-2003	31995-1996, 2002
Grenada	1990-2003	2000	Tanzania	1990-2003	1990
Guatemala	1990-2003	1990-1992, 1994	Thailand	1990-2003	
Guinea	1990-2003	1990	Togo	1990-2003	1990, 1992, 1995
Guinea-Bissau	1990-2003	1994	Tonga	1990-2003	
Guyana	1990-2003	1993, 1996, 1999	Trinidad and Tobago	1990-2003	1990-1991
Haiti	1990-2003	1993	Tunisia	1990-2003	
Honduras	1990-2003		Turkey	1990-2003	
Hungary	1990-2003		Turkmenistan	1995-1998	1996
India	1990-2003		Uganda	1990-2003	1990, 1995, 1998
Indonesia	1990-2003	1998-1999, 2002-2003	Ukraine	1994-2003	1994-1995, 1999-2000,
Iran	1990-2003	1994	Uruguay	1990-2003	1991, 2003
					2003
Jamaica	1990-2003	1990, 1994	Uzbekistan	1994-2003	1997, 2002
Jordan	1990-2003	1990-1992, 1995-1996, 2000-2003	Vanuatu	1990-2003	
Kazakhstan	1994-2003	1998-1999	Venezuela	1990-2003	1990
Kenya	1990-2003	1993	Vietnam	1990-2003	
Kyrgyzstan Republic	1994-2003	1996, 1998	Yemen	1991-2003	1991-1994, 2001
Laos	1990-2003	1991	Zambia	1990-2003	1991, 1993, 1996
Latvia	1994-2003		Zimbabwe	1990-2001	2001

Figure 4.11: Convergence and Fixing (GLMM-AR(1) Model)



The trace plot and the acf graphs are based on the last 10,000 iterations of the thinned chain (50,000 iterations with thinning interval 100)

Figure 4.12: Convergence and Fixing (GLMM-TSUV.AR(1) Model)



Chapter 5

Conclusion

In this dissertation, I propose a Bayesian generalized linear multilevel model with pth-order autoregressive errors for modeling inter-temporal dependence, contemporary correlation, and heterogeneity of unbalanced binary. I also extended this model with a multifactor specification to analyze time-specific unit-varying effects. The models are applied to several political economy studies including civil war, state failure, and sovereign default.

This dissertation makes several contributions to the literature. The proposed models extend the existing methods in longitudinal analysis to allow very rich dynamics and carefully distinguishes political dynamics and heterogeneity from one another. This facilitates political scientists to better understand political dynamics and conduct cross country comparative studies. Modeling heterogeneity also avoids drawing over-generalized conclusions in Time-Series Cross-Section data analyses. I develop an efficient algorithm for posterior estimation by innovatively orthogonalizing errors (the Cholesky decomposition and auxiliary parameter approach) and applying the parameter expansion method (the partial group move multigrid Monte Carlo updating). In addition, I provide an algorithm for Bayes factor computation, which is used for serial correlation diagnostics, lag order determination, variable selection, and forecasts. Finally, the model and methods are applied to an empirical study to test important theories in political economy. It also contributes to answering the practical questions on sustainable development of the developing countries and global financial stability.

Bibliography

Abdullah, Fuad. A. 1985. “Development of an Advanced Warning Indicator of External Debt Servicing Vulnerability.” *Journal of International Business Studies* 16:135–141.

Abreu, Margarida. 2003. “Contagion Phenomena in Financial Crises: Evidence from the Portuguese and Spanish Exchange Rate Crises in the Early 19909s.” *International Finance* 6(2):201–225.

Acemoglu, Daron and James A. Robinson. 2003. *Political Origins of Dictatorship and Democracy.* Massachusetts Institute of Technology.

Ahn, Seung Chan, Yong Hoon Lee. and Peter Schmidt. 2001. “GMM estimation of linear panel data models with time-varying individual effects.” *Journal of Econometrics* 101(2):219–255.

Albert, James A. and Siddhartha Chib. 1993. “Bayesian Analysis of Binary and Polychotomous Response Data.” *Journal of the American Statistical Association* 88(422):669–679.

Alesina, Ablerto, Nouriel Roubini and Gerhard Cohen. 1997. *Political Cycles and the Macroeconomy.* MIT Press.

Anderson, Benedict. 1983. *Imagined Communities.* London: Verso.

Andrews, Donald W. K. 1991. “Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation.” *Econometrica* 59(3):817 –858.

Andritzky, Jochen. 2006. *Sovereign Default Risk Evaluation: Implication of Debt Cirses and Bond Restructurings*. New York: SprinerSpringer.

Archer, Candace C., Glen Biglaiser and Karl DeRouen. 2007. "Sovereign Bonds and the "Democratic Advantage": Does Regime Type Affect Credit Rating Agency Ratings in the Developing World?" *International Organization* 61:341–365.

Arteta, Carlos and Galina Hale. 2005. "Are Private Borrowers Hurt by Sovereign Debt Rescheduling." Working Paper.

Aschbacher, Michael. 2000. *Finite Group Theory*. Cambridge, New York: Cambridge University Press.

Baliga, Sandeep, David O. Lucca and Tomas Sjöström. 2009. "Domestic Political Survival and International Conflict: Is Democracy Good For Peace?" Unpublished Manuscript.

Balkan, Erol M. 1992. "Political Instability, Coutry Risk and Probability of Default." *Applied Economics* 24:999–1008.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *American Economic Review* 88:724–748.

Barro, Robert. 1996. "Democracy and Growth." *Journal of Economic Growth* 1:1–27.

Bates, Robert H. 1999. "Ethnicity, Capital Formation, and Conflict." CID Working Paper, No.7.

Beck, Nathaniel. 1993. "The Methodology of Cointegration." *Political Analysis* 4:237–248.

Beck, Nathaniel, David Epstein, Simon Jackman and Sharyn O'Halloran. 2002. "Alternative Models of Dynamics in Binary Time-Series–Cross-Section Models: The Example of State Failure." Working Paper.

Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time-Series Cross-Section Data." *American Political Science Review* 89:634–47.

Beck, Nathaniel and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6:1–36.

Beck, Nathaniel and Jonathan N. Katz. 2004. "Time-Series–Cross-Section Issues: Dynamics, 2004." Working Paper.

Beck, Nathaniel and Jonathan N. Katz. 2007. "Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments." *Political Analysis* 15:182–195.

Beck, Nathaniel and Jonathan N. Katz. 2009. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." Working Paper.

Beck, Nathaniel, Jonathan N. Katz and Richard Tucker. 1998. "Taking Time Seriously: Time-Series–Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42(4):1260–88.

Berger, Suzanne. 2000. "Globalization and Politics." *Annual Review of Political Science* 3:43–62.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Changes as Information Cascades." *Journal of Political Economy* 100:992–1026.

Block, Steven and Burkhard N. Schrage. 2003. "Democratization's Risk Premium: Partisan and Opportunistic Political Business Cycle Effects on Sovereign Ratings in Developing Countries.".

Bogopolski, Oleg. 2008. *Introduction to Group Theory.* Zürish, Switzerland: European Mathematical Society.

Bordo, Michael D. and Antu Panini Murshid. 2000. "Are Financial Crises Becoming Increasingly More Contagious? What is the Historical Evidence on Contagion." NBER Working Papers Series.

Bordo, Michael D. and Kim Oosterlinck. 2005. ""Do Political Changes Trigger Debt Default? And Do Defaults Lead to Political Changes?"." Working paper.

Borsch-Supan, A. and V. Hajivassiliou. 1993. "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Modelsn." *Journal of Econometrics* 58:347–368.

Box, George E. P., Gwilym M. Jenskins and Gregory C. Reinsel. 1994. *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliffs, N.J. : Prentice Hall.

Box-Steffensmeier, Janet M. and Andrew R. Tomlinson. 2000. "Fractional Integration Methods in Political Science." *Electoral Studies* 19:63–76.

Breslow, Norman E. 1996. "Statistics in Epidemiology:The Case-Control Study." *Journal of the American Statistical Association* 91:14–28.

Brewer, Thomas. L. and Pietra Rivoli. 1990. "Politics and Perceived Country Creditworthiness in International Banking." *Journal of Money, Credit and Banking* 22:357–369.

Briggs, William L. 1987. *A Multigrid Tutorial*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Bulow, Jeremy and Kenneth Rogoff. 1989. "A Constant Recontracting Model of Sovereign Debt." *Journal of Political Economy* 97:155–178.

Bussiere, Matthieu and Christian B. Mulder. 1999. "Political Instability and Economic Vulnerability." IMF Working Paper No.99/46.

Calderon, Cesar and Alberto Chong. 2007. "Rent Seeking and Democracy: Empirical Evidence for Uruguay." *Economic Inquiry* 45:592–601.

Calvo, Guillermo A. 2005. *Contagion in Emerging Markets: When Wall Street is a Carrier*. Cambridge, Massachusetts: The MIT Press.

Calvo, Guillermo and Enrique Mendoza. 1998. "Rational Herd Behavior and the Globalization of Securities Market." mineo, Univeristy of Maryland.

Calvo, Sarah and Carmen Reinhart. 1996. Capital Flows to Latin America: Is there Evidence of Contagion Effect? In *Private Captital Flows to Emerging Markets After the Mexican Crisis*, ed. G.A. Calvo, M. Goldstein and E. Hochreiter. Institute for International Economics, Washington D.C.

Carlin, Bradley P. 1996. Hierarchical Longitudinal Modeling. In *Markov Chain Monte Carlo in Practice*, ed. S. Richardson W.R Gilks and D. J. Spiegelharlter. London: Chapman & Hall pp. 303–319.

Cederman, Lars-Erik and Luc Girardin. 2007. "Beyond Fractionalization: Mapping Ethnicity Onto Nationalist Insurgencies." *American Political Science Review* 101(1):173–185.

Celasun, Oya and Philipp Harms. 2007. "How does Private Foreign Borrowing Affect the Risk of Sovereign Default in Developing Countries." Working Paper.

Cespedes, Luis, Roberto Chang and Andrea Velasco. 2004. "Balance Sheets and Exchange Rate Policy." *American Economic Review* 94:mising.

Chang, Eric and Miriam A. Golden. 2009. "Sources of Corruption in Authoritarian Regimes." Working Paper.

Chang, Roberto. 2002. "Financial Crises and Political Crises." Working Paper.

Change, Roberto and Giovanni Majnoni. 2001. International Contagion: Implications for Policy. In *International Financial Contagion*, ed. Stijn Claessens and Kristin J. Forbes. Boston: Kluwer Academic Publishers.

Chib, Siddhartha. 1993. "Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach." *Journal of Econometrics* 58(3):275–294.

Chib, Siddhartha. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90:1313–1321.

Chib, Siddhartha and Edward Greenberg. 1994. "Bayesian Inference in Regression Models with ARMA (p, q) Errors." *Journal of Econometrics* 64:183–206.

Chib, Siddhartha and Ivan Jeliazkov. 2001. "Marginal Likelihood from the Metropolis-Hastings Output." *Journal of the American Statistical Association* 96:270–281.

Chib, Siddhartha and Ivan Jeliazkov. 2006. "Inference in Semiparametric Dynamic Models for Binary Longitudinal Data." *Journal of the American Statistical Association* 101(474):685–700.

Citron, Joel-Tomas and Gerald Nickelsburg. 1987. "Country Risk and Political Instability." *Journal of Development Economics* 25:385–392.

Clague, Christopher, Philip Keefer, Stephen Knack and Mancur Olson. 1997. Democracy, Autocracy, and the Institutions Supportive of Economic Growth. In *Institutions and Economic Development: Growth and Governance in Less-Developed and Post Socialist Countries*, ed. Christopher Clague. Baltimore: Johns Hopkins University Press pp. 91–120.

Clauge, Christopher, Philip Keefer and Mancur Olson. 1996. "Property and Contract Rights under Democracy and Distatorship." *The Journal of Economic Growth* 1:243–276.

Coakley, Jerry, Ana-Maria Fuertes and Ron Smith. 2002. "A Principal Components Approach to Cross-Section Dependence in Panels." Birkbeck College Discussion Paper.

Collier, Paul and Anke Hoeffler. 2000. "Greed and Grievance in Civil War." World Bank Policy Research Paper.

Collier, Paul and Anke Hoeffler. 2001. "Greed and Grievance in Civil War." Policy Research Working Paper 2355.

Collier, Paul and Anke Hoeffler. 2002. "On the Incidence of Civil War in Africa." *Journal of Conflict Resolution* 46:13–28.

Collier, Paul and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic papers* 56:563–595.

Collier, Paul, Anke Hoeffler and Mans Soderbom. 2004. "On the Duration of Civil War." *Journal of Peace Research* 41:253–273.

Collier, Paul and Nicholas Sambanis. 2005. *Understanding Civil War: Evidence and Analysis.* Vol. 1, 2 Washington, DC: World Bank.

Conley, Timothy G. 1999. "GMM Estimation with Cross Sectional Dependence." *Journal of Econometrics* 92:1–45.

Conley, Timothy G. and Giorgio Topa. 2002. "Socio-economic Distance and Spatial Patterns in Unemployment." *Journal of Applied Econometrics* 17:303–327.

Cuadra, Gabriel and Horacio Sapriza. 2008. "Sovereign Default, Interest Rates and Political Uncertainty in Emerging Markets." *Journal of International Economics* 76:78–88.

Dahl, Robert A. 1989. *Democracy and Its Critics.* New Haven: Yale University Press.

DeBoef, Suzanna. 2001. "Modeling Equilibrium Relationships: Error Correction Models with Strongly Autoregressive Data." *Political Analysis* 9:78–94.

Demir, Firat. 2006. "Volatility of Short Term Capital Flows and Socio-Political instability in Argentina, Mexico and Turkey." MPRA Paper No. 1943.

Deutsch, Karl W. 1953. *Nationalism and Social Communication.* Cambridge, MA: MIT Press.

Dooley, Michael P. 2000. "Can Output Losses Following International Financial Crises be Avoided?" NBER Working Paper No. 7531.

Doyle, Michael W. and Nicholas Sambanix. 2000. "International Peacebuilding: A Theoretical and Quantitative Analysis." *American Political Science Review* 94:779–802.

Drazen, Allen. 1998. Towards a Political Economic Theory of Domestic Debt. In *The Debt Durden and Its Consequences for Monetary Policy*, ed. Guillermo A. Calvo and Mervyn A. King. London: Macmillan.

Drazen, Allen. 2002. *Political Economy in Macroeconomics.* Princeton: Princeton University Press.

Dungey, Mardi and Vance L. Martin. 2005. "Unravelling Financial Market Linkages During Crises." Working Paper.

Durr, Robert. 1993. "An Essay on Cointegration and Error Correction Models." *Political Analysis* 4:185–228.

Edward Miguel, Shanker Satyanath and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 112(4):725753.

Elgie, Robert and Iain McMenamin. 2008. "Semi-presidentialism and Democratic Performance." *Japanese Journal of Political Science* 9:323–340.

Ellingsen, Tanja. 2000. "Ethnic Witches Brew: Linguistic, Religious and Ethnic Fragmentation. 1945-94." *Journal of Conflict Resolution* 44:228–249.

Elster, Jon. 2000. *Ulysses Unbound.* New York: Cambridge University Press.

Engle, Robert F. and Clive W. J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55:251–276.

Fearon, James D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88:577–592.

Fearon, James D. 2004. "Why Do Some Civil Wars Last So Much Longer Than Others." *Journal of Peace Research* 41:275–301.

Fearon, James D. and David D. Laitin. 2003*a*. "Additional Tables for "Ethnicity, Insurgency, and Civil War." Unpublished Manuscript.

Fearon, James D. and David D. Laitin. 2003*b*. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* 97(1):75–90.

Fearon, James D., Kmuli Kasara and David D. Laitin. 2007. "Ethnic Minority Rule and Civil War Onset." *American Political Science Review* 101(1):187–193.

Feder, Gershon and Lily V. Uy. 1985. "The Determinants of International Creditworthiness and their Implications." *Journal of Policy Modeling* 7:133–156.

Feng, Yi. 2003. *Democracy,Governance,and Economic Performance:Theory and Evidence*. Cambridge: MIT Press.

Firmin-Sellers, K. 1994. "The Politics of Property Rights." *American Political Science Review* 89:867–888.

Franzse, Robert J. and J. C. Hays. 2008*a*. Empirical Modeling of Spatial Interdependence in Time-Series Cross-Sections. In *Methods of Comparative Political and Social Science: New Developments & Applications*, ed. S. Pickel, G. Pickel, H-J. Lauth and D. Jahn. Wiesbaden: Westdeutscher Verlag.

Franzse, Robert J. and Jude C. Hays. 2007. "Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data." *Political Analysis* 15:140–164.

Franzse, Robert J. and Jude C. Hays. 2008*b*. Empirical Models of Spatial Interdependence. In *Oxford Handbook of Political Methodology*, ed. J. Box-Steffensmeier, H. Brady and D. Dollier. Oxford: Oxford University Press pp. 570–604.

Frees, Edward W. 2004. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. New York: Cambridge University Press.

Frieden, Jeffery. A. and Robert R. Rogowski. 1996. The Impact of the International Economy on National Policies: An Analytical Overview. In *Internationalization and Domestic Politics*, ed. R. Robert and H. Milner. New York: Cambridge University Press pp. 25–47.

Gartzke, Erik. 2001. "Democracy and the Preparation for War: Does Regime Type Affect States' Anticipation of Casualties." *International Studies Quarterly* 45:467–484.

Gasiorowski, Mark J. 1995. "Economic Crisis and Political Regime Change: An Event History Analysis." *The American Political Science Review* 89:882–897.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca, NY: Cornell University Press.

Gelman, Andrew. 2006. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do." *Technometrics* 48.

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York : Cambridge University Press.

Gelos, R. Gastpn and Shang-Jin Wei. 2002. "Transparency and International Investor Behavior." NBER Working Paper Series No.9260.

Gerring, John, Philip Bond, William T. Barndt and Carola Moreno. 2005. "Democracy and Economic Growth: A Historical Perspective." *World Politics* 57(3):323–364.

Geweke, John. 1991. Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constaints. In *Computing Science and Statistics: Pro-*

ceedings of the Twenty Third Symposium on the Interface, ed. E. M. Keramidas. Fairfax, VA: Interface Foundation of North America pp. 571–578.

Geweke, John. 1996. Bayesian Inference for Linear Models Subject to Linear Inequality Constraints. In *Modeling and Prediction: Honouring Seymour Geisser*, ed. W. O. Johnson, J. C. Lee and A. Zellner. New York: Springer.

Giannetti, Mariassunta. 2003. "Bank-Firm Relationships and Contagious Banking Crises." *Journal of Money, Credit and Banking* 35(April):2.

Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach.* 2nd ed. Boca Raton, Fla: Chapman & HallCRC.

Gill, Jeff. 2008. "Is Partial-Dimension Convergence a Problem for Inferences from MCMC Algorithms?" *Political Analysis* 16(2):153–178.

Glyn, Andrew and Bob Sutcliffe. 1992. Global But Leaderless: The New Capitalist Order. In *The Socialist Register 1992*, ed. R. Miliband. London: Merlin pp. 76–95.

Goemans, Henk E., Kristian Skrede Gieditsch and Giacomo Chiozza. 2009. "Introducing Archigos: A Dataset of Political Leaders." *Journal of Peace Research* 46(2):269–283.

Goetze, Catherine and Dejan Guzina. 2008. "Peacebuilding, Statebuilding, Nationbuilding–Turtles All the Way Down." *Civil Wars* 10(4):319–347.

Goodman, Jonathan and Alan D. Sokal. 1989. "Multigrid Monte Carlo Method: Conceptual Foundations." *Physical review. D* 40:2035–72.

Gourevitch, Peter. 1978. "The Second Image Reversed." *International Organization* 32:881–912.

Gourevitch, Peter. 1986. *Politics in Hard Times.* Ithaca, NY: Cornell University Press.

Gourieroux, Christian, A. Monfort and A. Trognon. 1984. Estimation and Test in Probit Models with Serial Correlation. In *Alternative Approaches to Time Series Analysis*, ed. J. P. Florens, M. Mouchart, J. P. Raoult and L. Simar. Brussels: Publications des Facultes Universitaires Saint-Louis.

Gourieroux, Christian, A. Monfort and A. Trongnon. 1985. "A General Approach to Serial Correlation." *Econometric Theory* 1(3):315–340.

Granovetter, Mark. 1978. "Threshold Models of Collective Action." *The American Journal of Sociology* 83:1420–1443.

Greenberg, Edward. 2007. *Introduction to Bayesian Econometrics.* Cambridge ; New York : Cambridge University Press.

Greenberg, Edward. 2008. *Introduction to Bayesian Econometrics.* New York: Cambridge University Press.

Grossman, Herschel I. 1995. Insurrections. In *Handbook of Defense Economics*, ed. K. Hartley and T. Sandler. Vol. 1 Elsevier, Amsterdam pp. 191–212.

Gurr, Ted Robert, Barbara Harff and Monty G. Marshall. 2009. "State Failure Task Force Report: Phase I-IV Findings.".

Hagenaars, Jacques A. 1990. *Categorical Longitudinal Data: Log-Linear Analysis of Panel, Trend and Cohort Data.* London: Sage.

Hamilton, James Douglas. 1994. *Time Series Analysis.* Princeton , N.J.: Princeton University Press.

Han, Cong and Bradley Carlin. 2001. "Markov Chain Monte Carlo Methods for Computing Bayes Factors: A Comprehensive Review." *Journal of the American Statistical Association* 96(455):1122–1132.

Harvey, Andrew C. 1981. *The Econometric Analysis of Time Series*. Oxford, UK: Philip Allen.

Heckman, James. 1981. Heterogeneity and State Dependence. In *Labor Markets*, ed. S. Rosen. Chicago: University of Chicago Press pp. 91–131.

Hedge, Shantaram O. and Rupendra Paliwal. 2005. "Financial Contagion and Market Liquidity: Evidence from the Asian Crisis." Working Paper.

Hegre, Havard, Tanja Ellingsen, Scott Gates and Nils P.Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1861-1992." *American Political Science Review* 95:33–48.

Hirschleifer, Jack. 1995. Theorizing about Conflict. In *Handbook of Defense Economics*, ed. K. Hartley and T. Sandler. Vol. 1 Elsevier, Amsterdam pp. 165–192.

Hoelscher, David and Marc Quintyn. 2003. "Managing Systemic Banking Crises." IMF Occasional Paper.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery and Chris T. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial (with discussion)." *Statistical Science* 14:382–401.

Holtz-Eakin, Douglas, Whitney K. Newey and Harvey Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometica* 56:1371–1395.

Horowitz, Donald L. 1985. *Ethinic Groups in Conflict*. Berkeley: University of California Press.

Huntington, Samuel P. 1968. *Social Order in Changing Societies*. New Haven CT: Yale University Press.

Huntington, Samuel P. 1996. *The Clash of Civilizations and the Remaking of World Order*. New York: Simon and Shuster.

Ignatieff, Michael. 1993. *Blood and Belonging.* London: Moonday Press.

Jeanne, Olivier. 1997. "Are Currency Crises Self-Fulfilling? A Test." *Journal of International Economics* 43:263–286.

Jensen, Nathan. 2008. "Political Risk, Democratic Institutions, and Foreign Direct Investment." *The Journal of Politics* 70(4):1040–1052.

Jensen, Nathan M. 2003. "Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment." *International Organization* 57(3):587–616.

Kao, Chihwa. 1999. "Spurious Regression and Residual-Based Tests for Cointegration in Panel Data." *Journal of Econometrics* 90:1–44.

Kapstein, Ethan B. 2000. "Review: Winners and Losers in the Global Economy." *International Organization* 54(2):359–384.

Kawai, Masahiro, Richard Newfarmer and Sergio Schmukler. 2001. "Crisis and Contagion In East Asia: Nine Lessions." Working Paper.

Keane, Michael P. 1994*a*. "A Computational Practical Simulation Estimator for Panel Data." *Econometrica* 62:95–116.

Keane, Michael P. 1994*b*. "A Computational Practical Simulation Estimator for Panel Data." *Econometrica* 62:95–116.

Keele, Luke and Suzanna De Boef. 2004. "Not Just for Cointegration: Error Correction Models with Stationary Data." Working Paper.

Kenen, Peter B. 2007. "The Benefits and Risks of Financial Globalization." *Cato Journal* 27:179–183.

Keohane, Robert and Helen Milner. 1996. *The Political Economy and Economic Regionalism.* NY: Cambridge University Press.

Kim, Yong Kyun. 2007. "Democracy and Sovereign Default Risk in International Credit Market: Empirical Implications from Modeling Strategies Selection." Working Paper.

Kim, Yong Kyun. 2008. "Inequality and Sovereign Default under Democracy: Empirical Implications from Modeling Strategic Selection." Unpublished Manuscript.

King, Gary and Langche Zeng. 2001*a*. "Explaining Rare Events in International Relations." *International Organization* 55:693–715.

King, Gary and Langche Zeng. 2001*b*. "Improving Forecasts of State Failure." *World Politics* 53:623–658.

King, Gary and Langche Zeng. 2001*c*. "Logistic Regression in Rare Events Data." *Political Analysis* 9:137–163.

Kohlscheen, Emanuel. 2006. "Sovereign Risk: Consitutions Rule." Warwick Economic Research Papers, No. 731.

Kose, M. A., Eswar Prasad, Kenneth S. Rogoff and Shang-Jin Wei. 2006. "Financial Globalization: A Reappraisal." IMF Working Paper 06/179.

Kraay, Aart and Vikram Nehru. 2006. "When is External Debt Sustainable." *The World Bank Economic Review* 20:341–365.

Krieckhaus, Jonathan. 2004. "The Regime Debtate Revisted: A Sensitivity Analysis of Democracy's Economic Effect." *British Journal of Political Science* 34:635–655.

Levine, Ross and David Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82:942–963.

Li, Quan and Adam Resnick. 2003. "Reversal of Fortunes: Democratic Institutions and Foreign Direct Inflows to Developing Countries." *International Organization* 57(1):175–211.

Lichbach, M. I. 1987. "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent." *Journal of Conflict Resolution* 31:266–297.

Liu, Jun S. and Chiara Sabatti. 2000. "Generalised Gibbs Sampler and Multigrid Monte Carlo for Bayesian Computation." *Biometrica* 87(2):353–369.

Liu, Jun S. and Ying Nian Wu. 1999. "Parameter Expansion for Data Augmentation." *Journal of American Statistical Association* 94:1264–1274.

Lizarazo, Sandra Valentina. 2005. "Contagion of Financial Crises in Sovereign Debt Markets." Working Paper.

Lord, F.M. and M. R. Novick. 1968. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison Wesley.

Lumley, Thomas and Patrick Heagerty. 1999. "Weighted Empirical Adaptive Variance Estimators for Correlated Data Regression." *Journal of the Royal Statistical Society: Series B* 61:459–477.

Manasse, Paolo, Nouriel Roubini and Axel Schimmelpfennig. 2003. "Predicting Sovereign Debt Crises." IMF Working Paper No.03/221.

Marshall, Monty G. and Keith Jaggers. 2007. "Polity IV Project: Political Regime Charateristics and Transitions, 1800-2008: Dataset Users' Manual." Center for Systemic Peace.

Marshall, Monty G. and Ted Robert Gurr. 2003. *Peace and Conflict 2003: A Global Survey of Armed Conflicts, Self-Determination Movements, and Democracy*. College Park, MD: Center for International Development and Conflict Management, University of Maryland.

McDowell, A. 2004. "From the Help Desk: Seemingly Unrelated Regression with Unbalanced Equations." *The Stata Journal* 4:442–448.

McGilivray, Fiona and Alastair Smith. 2003. "Who can be trusted? Sovereign Debt and the Impact of Leadership Change." Working Paper.

Mei, Jianping and Limin Guo. 2004. "Political Uncertainty, Financial Crisis, and Market Volatility." *European Financial Management* 10(4):639–657.

Min, Hong-Ghi, Duk-Hee Lee, Changi Nam, Myeong-Cheol Park and Sang-Ho Nam. 2003. "Determinants of Emerging-Market Bond Spreads: Cross-Country Evidence." *Global Finance Journal* 14:271–286.

Molenberghs, Geert and Geert Verbeke. 2005. *Models for Discrete Longitudinal Data.* New York: Spriner.

Moore, Will H. 1998. "Repression and Dissernt: Substitution, Context, and Timing." *American journal of Political Science* 41:851–873.

Moser, Christoph. 2006. "The Impact of Political Risk on Sovereign Bond Spreads – Evidence from Latin America." Working Paper.

Mueller, Gernot and Claudia Czado. 2005. "An Autoregressive Ordered Probit Model with Application to High-Frequency Financial Data." *Journal of Computational & Graphical Statistics* 14:320–338.

Mulaik, Stanley A. 1988*a*. "A Brief History of the Philosophical Foundations of Explanatory Factor Analysis." *Multivariate Behavioral Research* 23:267–305.

Mulaik, Stanley A. 1988*b*. Confirmatory Factor Analysis. In *Handbook of Multivariate Experimental Psychology*, ed. J. H. Nesselroade and R. B. Cattell. Number 259-288 New York: Plenum.

Muller, Edward N. 1985. "Income Inequality, Regime Repressiveness, and Political Violence." *American Sociology Review* 50:47–61.

Ng, Edmond SW, James R Carpenter, Harvey Goldstein and Jon Rasbash. 2006. "Estimation in Generalized Linear Mixed Models with Binary Outcomes by Simulated Maximum Likelihood." *Statistical Modelling* 6:23–42.

North, Douglass C. 1989. "Institutions and Economic Growth: An Historical Introduction." *World Developement* 17:1319–1332.

North, Douglass C. and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutional Governing Choice in Seventeenth-Century England." *The Journal of Economic History* 49(4):803–832.

North, Douglass D. 1990. *Institutions, Institutional Change and Economic Performance.* New York: Cambridge University Press.

Olsen, Karen K and Joseph L. Schafer. 2001. "A Two-Part Random-Effects Model for Semicontinuous Longitudinal Data." *Journal of the American Statistical Association* 96:730–745.

Olson, Mancur. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and social Rigidities.* New Haven: Yale University Press.

Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *American Political Science Review* 87(3):567–576.

Olson, Mancur. 2000. *Power and Prosperity.* New York: Basic Books.

Ozler, Sule and Guido Tabellini. 1991. "External Debt and Political Instability." Working Paper.

Pang, Xun. 2008. "Binary Time Series with AR(p) Errors: Bayes Factor for Lag Order Determination and Model Selection." Unpublished Manuscript.

Paoli, Bianca De and Glenn Hoggarth. 2006. "Costs of sovereign default." SSRN.

Pedroni, Peter. 1999. "Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors." *Oxford Bulletin of Economics and Statistics* 61:653–670.

Pedroni, Peter. 2004. "Panel Cointegration: Asymptotic and Finite Sample Properties of Pooled Time Series Tests with an Application to the PPP Hypothesis." *Econometric Theory* 3:579–625.

Pesaran, Hashem M. 2006. "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure." *Econometrica* 74:967–1012.

Pesaran, Hashem M. and Elisa Tosetti. 2007. "Large Panels with Common Factors and Spatial Correlations." Unpublished Manuscript.

Pesaran, Mashem H., Til Schuermann and Scott M. Weiner. 2004. "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model." *Journal of Business & Economic Statistics* 22:129–162.

Peters, B. Guy, Jon Pierre and Desmond S. King. 2005. "The Politics of Path Dependency: Political Conflict in Historical Institutionalism." *The Journal of Politics* 67(4):1275–1300.

Philips, Peter C.B. and Donggyu Sul. 2003. "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence." *The Economic Journal* 6:217–259.

Pierson, Paul. 1996. "The New Politics of the Welfare State." *World Politics* 48(2):143–179.

Pierson, Paul and Theda Skocpol. 2002. Historical Insitutionalism in Contemporary Political Science. In *Political Science: State of the Discipline*, ed. Ira Katznelson and Helen V. Milner. New York: W.W. Norton pp. 692–721.

Pitt, Michael K. and Neil Shephard. 1999. "Filtering via Simulation: Auxilary Particle Filters." *Journal of American Statistical Association* 94(446):590–599.

Poirier, Dale J. and Paul A. Ruud. 1988. "Probit with Dependent Observations." *The Review of Economic Studies* 55(4):593–614.

Przeworksi, Adam and Fernando Limongi. 1993. "Political Regimes and Economic Growth." *Journal of Economic Perspectives* 7:51–69.

Przeworski, Adam M., Michael E. Alvarez, Jose A. Cheibub and Femando Limongi. 2002. *Democracy and Development: Political Regimes and Economic Well-being in the World, 1950-1990*. Cambridge: Cambridge University Press.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Material Well-being in the World, 1950-1990*. New York: Cambridge University Press.

Reinhart, Carmen M., Kenneth S. Rogoff and Miguel A. Savastano. 2003. "Debt Intolerance." NBER Working Paper 9908.

Relter, Dan and Curtis Meek. 1999. "Determinants of Military Strategy, 1903-1994: A Quantitative Empirical Test." *International Studies Quarterly* 43:363–387.

Renard, Didier, Geert Molenberghs and Helena Geys. 2004. "A Pairwise Likelihood Approach to Estimation in Multilevel Probit Models." *Computational Statistics & Data Analysis* 44:649–667.

Reynal-Querrol, Marta. 2002. "Ethnicity, Political Systems and Civil Wars." *Journal of Conflict Resolution* 46:29–54.

Robertson, Donald and James Symons. 2000. "Factor Residuals in SUR Regressions: Estimating Panel Allowing for Cross Sectional Correlation." Unpublished Manuscript.

Rodriguez-Yam, Gabriel, Richard A. Davis and Louis L. Scharf. 2004. "Efficient Gibbs Sampling of Truncated Multivariate Normal with Application to Constrained Linear Regression." Unpublished Manuscript.

Rodrik, Dani and Andres Velasco. 1999. "Short-Term Capital Flows." World Bank Working Paper.

Rost, Nicolas, Gerald Schneider and Johannes Kleibl. 2009. "A Globle Risk Assessment Model for Civil Wars." *Social Science Research* 38:921–933.

Rothschild, Donald and Michael Foley. 1988. African States and the Politics of Inclusisve Coalitions. In *The Precarious Balance: State and Society in Africa*, ed. D. Rothschild and Chazan. N. Boulder, CO: Westview Press.

Rudra, Nita. 2002. "Globalization and the Decline of the Welfare State in Less-Developed Countries." *International Organization* 56(2):411–445.

Russett, Bruce M. 1964. "Inequality and Instability." *World Politics* 16:442–54.

Sachs, Jeffery. 1984. "Theoretical Issues in International Borrowing." Princeton Studies in International Finance 54.

Saiegh, Sebastian M. 2004. "Income Redistribution, Political Competition, and Sovereign Debt Repudiation." *Comparative Political Studies* . Working Paper.

Saiegh, Sebastian M. 2005. "Do Countries Have a "Democratic Advantage"?: Political Institutions, Multilaterial Agencies and So." *Comparative Political Studies* 38:366–387.

Saiegh, Sebastian M. 2008. "Coalitions Governments and Sovereign Debt Crises." Working Paper.

Sambanis, Nicholas. 2001. "Do Ethnic and Nonethnic Civil Wars Have the Same Causes?:A Theoretical and Empirical Inquiry (Part I)." *The Journal of Conflict Resolution* 45(3):259–282.

Sambanis, Nicholas. 2002. "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War." *Defence and Peace Economics* 13:215–243.

Sambanis, Nicholas. 2004. "What is Civil War?: Concepts and Empirical Complexities of an Operational Definition." *Journal of Conflict Resolution* pp. 814–858.

Sandor, Zsolt and Peter Andras. 2004. "Alternative Sampling Methods for Estimating Multivariate Normal Probabilities." *Journal of Econometrics* 120:207–234.

Sarkees, Meredith Reid. 2000*a*. "The Correlation of War Data on War: An Updata to 1997." *Conflict Management and Peace Science* 18:123–144.

Sarkees, Meredith Reid. 2000*b*. "The Correlations of War Data on War: An Updata to 1997." *Conflict Management and Peace Science* 18:123–144.

Schafer, Joseph L. and Recai M. Yucel. 2002. "Computational Strategies for Multivariate Linear Mixed-Effects Models with Missing Values." *Journal of Computational & Graphical Statistics* 11(2):437–457.

Schmidt, Peter. 1977. "Estimation of Seemingly Unrelated Regression with Unequal Numbers of Observations." *Journal of Ecobometrics* 5:365–377.

Schmukler, Sergio L. 2004. "Financial Globalization: Gain and Pain for Developing Countries." *Economic Review (Federal Reserve Bank of Atlanta)* 2:39–66.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy.* New York: Cambridge University Press.

Schultz, Kenneth A. and Barry R. Weingast. 2003. "The Democratic Advantage: Institutional Foundation of Financial Power in International Competition." *International Organization* 57(1):3–42.

Scott, James C. 1976. *The Moral Economy of the Peasant.* New Haven CT: Yale University Press.

Shepsle, K. 1991. Discretion, Institutions and the Problem of Government Commitment. In *Social Theory for a Changing Society*, ed. Pierre Bourdieu and J. Coleman. Boulder: Westview Press pp. 245–260.

Shor, Boris, Joseph Bafumi, Luke Keele and David Park. 2007. "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data." *Political Analysis* 15:165–181.

Singer, Judith D. and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modelling Change and Event Occurrence*. New York: Oxford University Press.

Skrondal, Anders and Peter Laake. 2001. "Regression Among Factor Scores." *Psychometrika* 66:563–576.

Skrondal, Anders and Sophia Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. New York: Chapman & Hall/CRC.

Skrondal, Anders and Sophia Rabe-Hesketh. 2008. Multilevel and Related Models for Longitudinal Data. In *Handbook of Multilevel Analysis*, ed. Jan de Leeuw and Erik Meijer. New York : Springer pp. 275–300.

Small, Melvin and J. David Singer. 1982. *Resort to Arms: International and Civil War, 1816-1980*. Beverly Hills, CA: Sage.

Smith, Alastair. 1998. "International Crises and Domestic Politics." *American Political Science Review* 92:623–638.

Smith, Robert. 1993. "Error Correction, Attractions, and Cointegration." *Political Analysis* 4:249–254.

Sobel, Andrew C. 1999. *State Institutions, Rpivate Incentives, Global Capital*. Ann Arbor: The University of Michigan Press.

Tavares, Jose and Romain Wacziarg. 2001. "How Democracy Affects Growth." *European Econonic Review* 45:1341–1378.

Thelen, Kathleen. 1999. "Historical Institutionalism in Comparative Politics." *Annual Review of Political Science* 2:369–404.

Tilly, Charles. 2003. *The Politics of Collective Violence.* Cambridge ; New York : Cambridge University Press.

Tomz, Michael and Mark L. J. Wright. 2007. "Do Countries Default in Bad Times?" *Journal of the European Economic Association* 5(2-3):1–9.

Torre, Augusto De La, Eduardo Levy Yeyati and Sergio L. Schmukler. 2002. "Financial Globalization: Unequal Blessings." *International Finance* 5(3):335–357.

Tsebelis, George. 2002. *Veto Players: How Political Institutions Work.* Princeton: Princeton University Press.

van Rijicheghem, Caroline and Beatrice Weder. 2004. *The Politics of Debt Crises.* London: Center for Economic Policy Research.

Verbeke, Geert and Geert Molenberghs. 2000. *Linear Mixed Models for Longitudinal Data.* New York: Springer.

Vreeland, James Raymond. 2008. "The Effect of Political Regime on Civil War: Unpacking Anocracy." *Journal of Conflict Resolution* 52:401–425.

Waltz, Kenneth. 1979. *The Theory of International Politics.* Reading, MA: Addison Wesley.

Weerahandi, Samaradasa. 2004. *Generalized Inference in Repeated Measures: Exact Methods in Manova and Mixed Models.* Hoboken, N.J. : Wiley & Sons, Inc.

Weingast, Barry R. 1995. "The Economic Role of Political Insitutions: Market Preserving Federalism and Economic Development." *Journal of Law, Economics and Organization* 11:1–31.

Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91:245–263.

Wibbels, Erik. 2006. "Dependency Revisited: international Markets, Business Cycles, and Social Spending in the Developing World." *International Organization* 60:433–468.

Wibbels, Erik and Moises Arce. 2003. "Globalization, Taxation, and Burden-Shifting in Latin America." *International Organization* 57(1):111–136.

Williams, John. 1993. "What Goes Around, Comes Around: Unit Root Tests and Cointegration." *Political Analysis* 4:229–236.

Williamson, Jeffrey A. 1997. "Globalization and Inequality, Past and Present." *World Bank Research Observer* 12:117–135.

Wilson, Sven E. and Daniel M. Butler. 2007. "A Lot More to Do: The Sensitivity of Time-Series- Cross-Section Analyses to Simple Alternative Specifications." *Political Analysis* 15:101–123.

Wimmer, Andreas. 2002. *Nationalist Exclusion and Ethnic Conflict: Shadow of Modernity.* Cambridge: Cambridge University Press.

Wright, Joseph. 2008. "To Invest or Insure? How Authoritarian Time Horizons Impact Foreign Aid Effectiveness." *Comparative Political Studies* 41:971–1000.

Yang, Yang, Wenjiang Fu and Kenneth C. Land. 2004. "A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models." *Sociological Methodology* 34:75–110.

Zeileis, Achim. 2004. "Econometric Computing with HC and HAC Covariance Matix Estimators." *Journal of Statistical Software* 11:1–17.