Political Science Research and Methods Vol 4, No. 3, 595–620 September 2016

doi:10.1017/psrm.2015.81

A Bayesian Approach to Dynamic Panel Models with Endogenous Rarely Changing Variables*

TSUNG-HAN TSAI

Whether democratic and non-democratic regimes perform differently in social provision policy is an important issue to social scientists and policy makers. As political regimes are rarely changing, their long-term or dynamic effects on the outcome are of concern to researchers when they evaluate how political regimes affect social policy. However, estimating the dynamic effects of rarely changing variables in the analysis of time-series cross-sectional data by conventional estimators may be problematic when the unit effects are included in the model specification. This article proposes a model to account for and estimate the correlation between the unit effects and explanatory variables. Applying the proposed model to 18 Latin American countries, this article finds evidence that democracy has a positive effect on social spending both in the short and long term.

In national-level studies of public social spending, many efforts have been made to explain why some countries consistently pursue more welfare-enhancing policies than others. From the perspective of institutional scholars, democratic institutions enable politicians to make credible policy commitments to the electorate through competitive elections, so citizens who prefer redistributive policies are able to influence policy decisions by electing their representatives, compared with non-democratic ones (Meltzer and Richard 1981; Persson and Tabellini 2000). The realization of redistributive policies in democracies, however, may only be observed in the long term (Huber et al. 2008; Keefer and Vlaicu 2008). This raises one empirical question: Does democracy matter for social welfare programs in the short term, in the long term, or both?

This article addresses this question by focusing on the estimation issues of dynamics of rarely changing variables such as political regimes in time-series cross-sectional (TSCS).[1] In the analysis of TSCS data or panel data, which have both intertemporal and cross-sectional variations, researchers are able to control for unobserved heterogeneity across units to eliminate omitted variable bias in estimation by including unit-specific effects.[2] However, it is problematic to estimate a panel model with a lagged dependent variable (LDV), time-invariant and/or rarely changing variables, and unit effects by two conventional approaches: fixed-effects (FE) models and random-effects (RE) models (Wooldridge 2002; Hsiao 2003; Baltagi 2005).

* Tsung-han Tsai is an Assistant Professor in the Department of Political Science, National Chengchi University, Taipei 11605, Taiwan (thtsai@nccu.edu.tw). An earlier version of this manuscript was presented at the 2011 Annual Meeting of the American Political Science Association, Seattle, WA. The author thanks Jeff Gill, Guillermo Rosas, Jacob Montgomery, Chia-yi Lee, Jamie Monogan, Keith Schnakenberg, and Justin Grimmer for helpful comments and suggestions at different stages of this manuscript. The author also thanks two anonymous reviewers and the PSRM editor for helping improve the manuscript. Any remaining errors are the author's responsibility. To view supplementary material for this article, please visit http://dx.doi.org/10.1017/psrm.2015.81

[1] Generally speaking, a rarely/slowly changing variable refers to a variable that does not change often or has very low within variation (Plümper and Troeger 2007), e.g., the level of democracy in the empirical example.

[2] TSCS data have a small number of units over a reasonable-sized time period while panel data have a large number of units for a short time period (Beck and Katz 1996). Although the discussion in this article focuses mainly on TSCS data, the proposed model can be applied to analyzing panel data as well.

Building on the structure of simultaneous equation modeling and error-component formulations, a Bayesian simultaneous equation model (hereafter, BSEM) is developed here with hierarchical features that accommodate the correlation between the unit effects and explanatory variables, for which I focus on time-varying and rarely changing variables. Unlike conventional approaches, the BSEM provides additional information on the degree of the correlation. The complexity of this specification requires estimation with Markov chain Monte Carlo (MCMC) methods. A Bayesian approach offers flexibility for complex model specifications and resolves the inferential problems that arise in non-Bayesian multilevel models (MLMs) (Carlin and Louis 2000; Gill 2008a).

To assess the performance of the BSEM presented in this article, I employ a Monte Carlo study, in which I compare the proposed model with alternative estimators in estimating the coefficients of correlated time-varying and rarely changing variables. The simulation results show that the proposed model not only performs as well as, or better than alternative estimators in terms of unbiasedness and efficiency, but also can estimate the correlation between covariates and the unit effects. The proposed model is applied to analyze the effects of political regimes on social spending in Latin America, where countries have different democratic experiences and different social welfare systems. This article finds evidence that democracy has a positive effect on social spending both in the short and long term.

This article has two primary contributions. First, methodologically, the proposed model in this article can deal with the problem of the correlation between the unit effects and covariates under the framework of multilevel modeling. Doing so, the proposed model is more efficient than the FE models and less biased than the RE models. Second, substantively, this article contributes to our understanding of the effects of political regimes on social spending in general. The evidence of positive democratic effects on social spending in Latin American countries support existing theories of comparative political economy that predict more redistributive policies in democracies.

The remainder of this article proceeds as follows. The A Model for Endogenous Rarely Changing Variables section discusses modeling dynamics of a endogenous, rarely changing variable, and a Bayesian approach to the model specification, followed by a Monte Carlo study in the Monte Carlo Simulations section. The Application: Social Spending in Latin America section presents the application of the proposed model to social security and welfare spending (SSW) in Latin America, and the Concluding Remarks section concludes.

A MODEL FOR ENDOGENOUS RARELY CHANGING VARIABLES

Suppose that there exist TSCS data with a continuous outcome variable, y_{jt}, and a rarely changing variable, w_{jt}, for unit $j = 1, \ldots, J$ measured at time $t = 1, \ldots, T$, which indicates a balanced TSCS data structure. A major advantage of TSCS data is that researchers are able to control for unobserved heterogeneity across units and/or through time to avoid omitted variable bias by including unit-specific and/or time-specific effects (Wooldridge 2002; Hsiao 2003; Baltagi 2005). For simplicity, I consider only unit heterogeneity in the model specification. Furthermore, to model dynamics, one standard approach is to include an LDV, $y_{j(t-1)}$, into the model. Thus, the model is specified as

$$y_{jt} = \mu + \phi y_{j(t-1)} + \beta w_{jt} + \delta_j + \varepsilon_{jt}, \qquad (1)$$

where μ denotes the intercept, ϕ the autoregressive coefficient, β the coefficient parameter, δ_j the unit-specific effects, and ε_{jt} the error term. It is assumed that the error term is independently, identically distributed with mean 0 and finite variance σ^2_ε.

When the inclusion of the LDV appropriately captures the dynamics of explanatory variables, autocorrelation can be eliminated (Beck and Katz 1996, 2011; Keele and Kelly 2006). This works particularly well when covariates are rarely changing and their effects decay over time. To represent the decaying effects of covariates, it is assumed that the process $\{y_{jt}\}$ is stationary for individual units, i.e., formally, $|\phi| < 1$, which means that there is no unit roots. With the stationary assumption, the dynamic effects of a rarely changing variable on the outcome can be obtained by the dynamic multiplier, also referred to as the impulse response function in time-series analysis (see e.g., Hamilton 1994; Enders 2004).[3]

When researchers are interested in the dynamics of slowly changing variables and specify a dynamic panel model, e.g., Equation 1, as my motivation stated in the first section, they face a quandary between FE and RE models. For those who prefer FE models, FE models not only provide inefficient estimates for rarely changing variables (Plümper and Troeger 2007), but also give biased and inconsistent coefficient estimates for finite samples as the inclusion of the LDV violates the assumption of strict exogeneity of the regressors (Nickell 1981). For those who prefer RE models, which can be considered as a special case of MLMs, the consistency of parameter estimates of RE models depends not only on the assumption of the independence between the unit effects and covariates, but also on the initial conditions of the outcome variable (Anderson and Hsiao 1981; Anderson and Hsiao 1982; Hsiao 2003).

It has been shown that the RE model for dynamic panel models is consistent, assuming that the initial values of outcomes (i.e., y_{j0}) are treated as fixed constants (Anderson and Hsiao 1982; Sevestre and Trognon 1985). This property holds only if explanatory variables are not correlated with the unit effects or, at least, the correlation is explicitly modeled. A common approach to deal with the correlation between the unit effects and covariates is to include the within-group means of the covariates as group-level predictors suggested by Mundlak (1978). However, there is little discussion on finite sample properties of this approach for dynamic panel models in estimating correlated rarely changing variables although some literature implies that this approach works fine asymptotically (Anderson and Hsiao 1982; Hsiao 2003).

Building on the structure of simultaneous equation models and error-component formulations, an RE-based model is proposed here to accommodate the dependence between the unit effects and rarely changing variables. This model not only explicitly deals with the problem of endogenous explanatory variables, but also maintains the advantages of multilevel modeling, e.g., the extension to models with random coefficients. For illustration, consider the model presented in Equation 1, which contains an LDV, a rarely changing variable, and the unit effects. The resulting analysis also holds for correlated time-varying variables and can be generalized to both static and dynamic panel models containing both exogenous and endogenous explanatory variables.

I start with the assumption that a rarely changing variable can be decomposed into three components: a mean effect, differences between units, and differences within units. Differences between units are effects specific to individual units, which do not vary over time, and differences within units are changes over time within a specific individual unit. Thus, the slowly changing variable can be expressed as

$$w_{jt} = \zeta_0 + \eta_j + \xi_{jt}, \tag{2}$$

[3] The derivation of the dynamic multiplier and, thus, the dynamic effects of covariates can be found in textbooks about panel models such as Baltagi (2005), Hsiao (2003), and Wooldridge (2002) and in political science research such as Beck (2001), Beck and Katz (1996), De Boef and Keele (2008), Keele and Kelly (2006).

where ζ_0 denotes the mean effect, η_j between differences, which can be treated as the "unit effects" for w_{jt}, and ξ_{jt} within differences.[4] Combining Equation 2 with Equation 1, we have a simultaneous equation model as follows:

$$y_{jt} = \mu + \phi y_{j(t-1)} + \beta w_{jt} + \delta_j + \varepsilon_{jt}, \tag{3}$$

$$w_{jt} = \zeta_0 + \eta_j + \xi_{jt}. \tag{4}$$

Equation 3 is an ordinary panel model for a continuous outcome variable and it is assumed that

$$\varepsilon_{jt} \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2). \tag{5}$$

Note that ξ_{jt} in Equation 4 denotes differences within units for a rarely changing variable. Although the distribution of ξ_{jt} may be limited to certain values empirically, which depends on the operationalization of variables, ξ_{jt} can be any positive or negative values theoretically. In other words, ξ_{jt} ranges from negative infinity to positive infinity in general. Moreover, as the value of w_{jt} rarely changes for a given unit j, ξ_{jt} is highly likely to be 0. Based on these two general points, I assume that ξ_{jt} follows a normal distribution with mean 0 and finite variance σ_ξ^2, i.e.,

$$\xi_{jt} \overset{\text{iid}}{\sim} N(0, \sigma_\xi^2). \tag{6}$$

Next, to model the correlation across equations, i.e., the dependence between δ_j and w_{jt}, I assume that there exists a time-invariant, unobserved common factor related to some features of the individual units influencing both y_{jt} and w_{jt} through δ_j and η_j, respectively.[5] That is to say, the dependence between δ_j and w_{jt} results from the correlation between δ_j and η_j. Following the conventional approach in which the unit effects are assumed to be normally distributed with mean 0 and finite variance (Mundlak 1978; Gelman and Hill 2007), it is assumed that the joint distribution for the vector $\psi_j = (\eta_j, \delta_j)$ is a bivariate normal distribution:

$$(\eta_j, \delta_j) \sim N_2(\mathbf{0}, \Omega), \quad \text{for} \quad j = 1, \cdots, J, \tag{7}$$

where the variance–covariance matrix is of the form

$$\Omega = \begin{pmatrix} \sigma_\eta^2 & \sigma_{\delta\eta} \\ \sigma_{\delta\eta} & \sigma_\delta^2 \end{pmatrix}. \tag{8}$$

[4] One may argue that we can simply decompose w_{jt} into three components by the fact that $w_{jt} = \overline{\overline{w}} + (\overline{w}_j - \overline{\overline{w}}) + (w_{jt} - \overline{w}_j)$. By this transformation, we do not need additional distributional assumptions for these three components. However, when we substitute this transformation into Equation 1 and obtain $y_{jt} = \alpha_j + \beta\overline{\overline{w}} + \beta(\overline{w}_j - \overline{\overline{w}}) + \beta(w_{jt} - \overline{w}_j) + \varepsilon_{jt}$, we can see that the overall mean effect, within effect, and between effect of w on y are equal, which is an unrealistic assumption.

[5] Although the proposed model does not explicitly model the correlation between the LDV and the unit effects, it provides consistent model estimates by assuming that y_{j0} is fixed just like dynamic RE models (Anderson and Hsiao 1982; Sevestre and Trognon 1985; Hsiao 2003).

Consequently, the dependence between δ_j and w_{jt} is represented by the covariance $\sigma_{\delta\eta}$. As our interest is the correlation between δ_j and w_{jt}, given $\boldsymbol{\Omega}$ and σ_ξ^2, we can simply derive this value by the definition of correlation given by

$$\begin{aligned} \text{corr}(\delta_j, w_{jt}) &= \frac{\text{Cov}(\delta_j, w_{jt})}{\sqrt{\text{Var}(\delta_j)\text{Var}(w_{jt})}}, \\ &= \frac{\text{Cov}(\delta_j, \eta_j)}{\sqrt{\text{Var}(\delta_j)\text{Var}(w_{jt})}}, \\ &= \frac{\sigma_{\delta\eta}}{\sqrt{\sigma_\delta^2(\sigma_\xi^2 + \sigma_\eta^2)}}. \end{aligned} \tag{9}$$

The second line of Equation 9 is derived because the correlation between w_{jt} and δ_j in fact results from the correlation between η_j and δ_j.

The main distinction between Mundlak's (1978) approach and the BSEM presented in this section is that the former provides no further information about the degree of the correlation between covariates and the unit effects while the latter does, given that both approaches explicitly model the correlation. The additional information gained from the BSEM over Mundlak's approach is important for at least two reasons. First, the estimated correlation shows the extent of the dependence with regard to a particular covariate. With this information, we know which covariates suffer from the problem of endogeneity and how serious this problem is. Second, in practice, the information on the degree of the correlation shows us which variables are correlated to the features of the individual units that do not vary over time. When a covariate is highly correlated to the unit effects, it means that the differences in this covariate are mainly from the differences in the features of individual units. This information sometimes has theoretical meanings.

I estimate this model by a Bayesian approach, so I complete the model specification by defining the prior distribution. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\varepsilon, \zeta_0, \sigma_\xi, \boldsymbol{\Omega})$ denote the model parameters where $\boldsymbol{\beta} = (\phi, \mu, \beta)'$. Following the conventional approach (Carlin and Louis 2000; Robert 2001; Gelman et al. 2004), I use conjugate prior distributions for these parameters as follows:

$$\sigma_\varepsilon^2 \sim \text{Inv}-\text{Gamma}(v_0/2, d_0/2), \tag{10}$$

$$\beta \,|\, \sigma_\varepsilon^2 \sim N_3(\boldsymbol{b}_0, \sigma_\varepsilon^2 \boldsymbol{B}_0), \tag{11}$$

$$\sigma_\xi^2 \sim \text{Inv}-\text{Gamma}(g_0\,/\,2, h_0\,/\,2), \tag{12}$$

$$\zeta_0 \,|\, \sigma_\xi^2 \sim N(m_0, \sigma_\xi^2\,/\,a_0), \tag{13}$$

$$\boldsymbol{\Omega} \sim \text{Inv}-\text{Wishart}(\nu_0, \boldsymbol{\Lambda}_0^{-1}), \tag{14}$$

where v_0, d_0, $\boldsymbol{b}_0$, $\boldsymbol{B}_0$, g_0, h_0, m_0, a_0, ν_0, and $\boldsymbol{\Lambda}_0$ are hyperparameters, which can be assigned values to reflect prior information about the corresponding parameters or to give diffuse forms.[6] Furthermore, I assume priori independence of $(\beta, \sigma_\varepsilon^2), (\zeta_0, \sigma_\xi^2)$, and $\mathbf{\Omega}$ and, therefore, our prior distribution of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}) = IG\left(\sigma_\varepsilon^2 \mid \frac{v_0}{2}, \frac{d_0}{2}\right) N_3\left(\beta \mid \sigma_\varepsilon^2, \boldsymbol{b}_0, \boldsymbol{B}_0\right)$$
$$\times IG\left(\sigma_\xi^2 \mid \frac{g_0}{2}, \frac{h_0}{2}\right) N\left(\zeta_0 \mid \sigma_\xi^2, m_0, a_0\right) IW\left(\mathbf{\Omega} \mid \nu_0, \mathbf{\Lambda}_0^{-1}\right). \tag{15}$$

Let $\boldsymbol{y}_j = (y_{j1}, \; y_{j2}, \ldots, \; y_{jT})'$, $\boldsymbol{y}_{j,-1} = (y_{j0}, \; y_{j1}, \ldots, \; y_{j(T-1)})'$, $\boldsymbol{w}_j = (w_{j1}, \; w_{j2}, \ldots, w_{jT})'$, $\boldsymbol{\xi}_{\mathrm{j}} = (\xi_{j1}, \xi_{j2}, \cdots, \xi_{jt})'$, and $\varepsilon_j = (\varepsilon_{j1}, \varepsilon_{j2}, \cdots, \varepsilon_{jt})'$ be $T \times 1$ vectors of observations for unit $j = 1, 2, \ldots, J$. The simultaneous equation model has the form

$$\begin{pmatrix} \mathbf{w}_j \\ \mathbf{y}_j \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \beta \mathbf{I}_T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{w}_j \\ \mathbf{y}_j \end{pmatrix} + \begin{pmatrix} \zeta_0 \mathbf{I}_T & \mathbf{0} \\ \mu \mathbf{I}_T & \phi \mathrm{I}_T \end{pmatrix} \begin{pmatrix} \boldsymbol{e} \\ \boldsymbol{y}_{j,-1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{e} & \tilde{0} \\ \tilde{0} & \boldsymbol{e} \end{pmatrix} \begin{pmatrix} \eta_j \\ \delta_j \end{pmatrix} + \begin{pmatrix} \boldsymbol{\xi}_j \\ \boldsymbol{\varepsilon}_j \end{pmatrix}, \tag{16}$$

where $\mathbf{0}$ is a $T \times T$ matrix, $e = (1, \ldots, 1)'$ is a $T \times 1$ vector, $\tilde{0} = (0, \cdots, 0)'$ is a $T \times 1$ vector. The multivariate regression representation of the structural form is given by

$$\boldsymbol{Y}_j = \boldsymbol{B}\boldsymbol{Y}_j + \Gamma \boldsymbol{X}_j + \boldsymbol{Z}\boldsymbol{\psi}_j + \boldsymbol{U}_j, \tag{17}$$

where $\boldsymbol{\psi}_j = (\eta_j, \delta_j)'$, $\boldsymbol{U}_j \sim N(\mathbf{0}, \mathbf{\Sigma})$, and

$$\Sigma = \begin{pmatrix} \sigma_\xi^2 \mathbf{I}_T & 0 \\ 0 & \sigma_\varepsilon^2 \mathbf{I}_T \end{pmatrix}. \tag{18}$$

Also, let $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_J)$, $\boldsymbol{w} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_J)$, and $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_J)$. The likelihood function for the simultaneous equation model denoted by $p(\boldsymbol{w}, \boldsymbol{y} \mid \boldsymbol{\psi}, \boldsymbol{\theta})$ has the form

$$p(\boldsymbol{w}, \boldsymbol{y} \mid \boldsymbol{\psi}, \boldsymbol{\theta}) = \prod_{j=1}^{J} p(\boldsymbol{w}_j, \boldsymbol{y}_j \mid \boldsymbol{\psi}, \boldsymbol{\theta}), \tag{19}$$

where $p(\boldsymbol{w}_j, \boldsymbol{y}_j \mid \boldsymbol{\psi}, \boldsymbol{\theta}) = N_{2T}(\boldsymbol{B}\boldsymbol{Y}_j + \Gamma \boldsymbol{X}_j + \boldsymbol{Z}\boldsymbol{\psi}_j, \sum) N_2(\boldsymbol{\psi}_j \mid \mathbf{0}, \Omega)$.

From the Bayes theorem, the joint posterior distribution of interest, $\pi(\theta \mid \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{\psi})$, is as follows:

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{\psi}) \propto p(\boldsymbol{y}, \boldsymbol{w} \mid \boldsymbol{\psi}, \boldsymbol{\theta}) p(\boldsymbol{\psi} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}). \tag{20}$$

This joint distribution is of a type that can be efficiently processed by MCMC methods (see Gelfand and Smith 1990; Casella and George 1992; Chib and Greenberg 1995), which can be implemented in computer programs such as WinBUGS (Lunn et al. 2000) and JAGS (Plummer 2003).

[6] In the Bayesian approach, proper prior distributions can help in the identification of the sampling model (Lindley 1972). In this case, the simultaneous equation is identified as ζ_0 is assumed to be normally distributed with mean known.

MONTE CARLO SIMULATIONS

In this section, I employ Monte Carlo simulations to assess the finite sample properties of the BSEM, compared with alternative estimators that have been proposed or suggested to estimate correlated explanatory variables.

Simulation Design

The data generation process (DGP) for the simulations is as follows:

$$y_{jt} = \mu + \phi y_{j(t-1)} + \alpha_1 x_{1jt} + \alpha_2 x_{2jt} + \alpha_3 x_{3jt} + \beta_1 w_{1jt} + \beta_2 w_{2jt} + \beta_3 w_{3jt} + \delta_j + \varepsilon_{jt}, \tag{21}$$

where x_1, x_2, and x_3 are time-varying variables and w_1, w_2, and w_3 are rarely changing variables, δ_j denotes the unit effects, and ε_{jt} the error term. The two groups of variables except w_3 are drawn from a normal distribution with different means and variances; δ_j is normally distributed with mean 0 and variance 1; the error term ε_{jt} is assumed to be white noise and drawn from a standard normal distribution. Among the covariates, x_3 and w_3 are correlated with the unit effects δ_j and, thus, x_3, w_3, δ, and ε vary for each replication while x_1, x_2, w_1, and w_2 are fixed across all experiments. In the experiments, the correlation between x_3 and the unit effects, corr (x_3, δ_j), and the correlation between w_3 and the unit effects, corr(w_3, δ_j), are varied. In specific, corr(x_3, δ_j) = {0, 0.1, 0.2, … , 0.9} and corr(x_3, δ_j) = {0, 0.1, 0.2, … , 0.9}.[7]

The true values of the coefficients are held constant throughout all experiments as follows:

$$\mu = 1,\ \phi = 0.8,\ \alpha_1 = 0.5,\ \alpha_2 = 2,\ \alpha_3 = -1.5,\ \beta_1 = -2.5,\ \beta_2 = 1.8,\ \beta_3 = 3.$$

I simulate the data with different numbers of time periods T = {15, 30, 60} and fixed number of units $J = 20$. The number of replications is 500 for each experiment. Moreover, another experimental design is conducted in which the data are simulated with different numbers of units J = {40, 60} and fixed number of time periods $T = 30$. For this design, the number of replications is 100.

The DGP presented in Equation 21 follows the simulation design in Plümper and Troeger (2007) but differs in three fundamental respects. First, the unit effects vary across replications for each experiment, which account for random variation in the unit effects (Breusch et al. 2011). Second, the rarely changing variables I consider are persistent for the rest of periods under analysis once they change at a certain (randomly assigned) time point. Finally, the dynamics of explanatory variables are taken into account by the inclusion of the LDV in the DGP.

The BSEM employed here contains two endogenous covariates, x_3 and w_3, which is an extension of the one discussed in the A Model for Endogenous Rarely Changing Variables section. This Bayesian model with vague priors is estimated with MCMC techniques and is implemented in JAGS 3.1.0 called from R version 3.1.0 (Su and Yajima 2012).[8] The estimation was performed with three parallel chains of 10,000 iterations each. The first half of the iterations

[7] The correlated covariates are generated via the following procedure: first, after the samples of δ_j are drawn, the unit-level covariates (x_{3j} and w_{3j}) are generated with the assigned values of correlation; second, for $j = 1, 2, \cdots, J$, samples of x_{3jt} are drawn from a normal distribution with means x_{3j} and variances $1 - [\text{corr}(x_{3j}, \delta_j)]^2$; third, samples of w_{3jt} are generated to be $w_{3jt} = w_{3j}$ for $t = 1, 2, \ldots, k$, where k is a random number from $\{1, 2, \ldots, T\}$. For $t = k+1, k+2, \ \ldots, T$, $w_{3jt} = w_{3j} + \tau_j$, where $\tau_j \sim N(1, 0.25)$. By this procedure, both δ_j and x_{3jt} have a univariate normal distribution while w_{3jt} does not. Moreover, δ_j and x_{3jt} jointly have a bivariate normal distribution while δ_j and w_{3jt} do not. Therefore, the distributional assumptions made in the simulation do not exactly match the model assumptions.

[8] The prior values used for the simulation are presented in Appendix A.

were discarded as a burn-in and five as thinning, and thus 3000 samples in total were generated. The convergence of Markov chains was tested by standard diagnostic tools such as Geweke, Gelman-Rubin, Raftery-Louis, and Heidelberger-Welch (Gill 2008b) and was conducted by an easy to use R function superdiag that integrates all of the standard empirical diagnostics (Tsai and Gill 2012). The results show no evidence of non-convergence.

The alternative estimators considered include the pooled ordinary least squares (OLS) estimator, the FE estimator, the fixed-effects vector decomposition (FEVD) estimator (Plümper and Troeger 2007), the RE model, and a MLM with the within-group means of the correlated variables (Mundlak 1978; Gelman 2006). This procedure is implemented in R: the FE and RE models are estimated through the plm package (Croissant and Millo 2008); the MLM is estimated by the restricted maximum likelihood estimation through the lme4 package (Bates, Maechler and Ben 2015).[9]

To compare competing estimators, I follow the literature in reporting the average bias and the root mean squared error (RMSE) of estimates. To avoid the bias to be canceled out, I calculate the absolute value of bias (hereafter, AVB). The RMSE captures both the bias and efficiency of the estimators calculated based on the formula $\sqrt{\frac{\sum_{s=1}^{R}\left(\hat{\beta}^{(s)}-\beta_{\text{true}}\right)^2}{R}}$, where $\hat{\beta}^{(s)}$ denotes the estimate, s in the superscript denotes the sth replicate within R replications, and β_{true} is the true value.

Moreover, the uncertainty associated with simulation results is reported along with the estimates of AVB and RMSE (Koehler, Brown and Haneuse 2009). Let θ denote some target quantities of interest, i.e., AVB and RMSE, and $\hat{\theta}_R$ the estimate of θ from a Monte Carlo simulation with R replications. The variability between simulations, also called the Monte Carlo error (MCE), is defined as the standard deviation of the Monte Carlo estimator:

$$\text{MCE}(\hat{\theta}_R) = \sqrt{\text{Var}(\hat{\theta}_R)}. \tag{22}$$

That is to say, MCE(AVB_R) and MCE($RMSE_R$) of a number of coefficient parameters are presented in the simulation results to show the variation of the Monte Carlo sampling distribution.

Results of Simulation

In Table 1, I report the AVB and RMSE over ten experiments for three different designs: $T = \{15, 30, 60\}$ with $J = 20$. The results of experiments in which the correlation between the time-varying variable x_3 and the unit effects δ is varied are displayed from the second to the seventh columns; those in which the correlation between the rarely changing variable w_3 and the unit effects δ is varied are displayed in the eighth to the 13th columns. Notice that in the former corr(w_3, δ_j) is fixed at the value of 0.3 while in the latter corr(x_3, δ_j) is fixed at the value of 0.3.

The results in Table 1 are summarized as follows. First, on average, the pooled OLS produces the poorest estimates in terms of AVB and RMSE. As we know and the results confirm, the correlation between the LDV and the unit effects would seriously bias the pooled

[9] I am especially interested in the performance of the MLM and FEVD in a dynamic setting as they are increasingly used in empirical research, but there is few discussion on their performance. I do not consider generalized method of moments (GMM) estimators here and leave them for future research because GMM estimators explicitly account for the correlation between the LDV and unit effects while the estimators compared here do not and because GMM estimators are not suited for data that do not have many units, which is what I consider here.

TABLE 1 *Absolute Value of Bias (AVB) and Root Mean Squared Error (RMSE) over 500 Replications Times Ten Experiments with T Varied*

	corr(x_3, δ) = {0, …, 0.9}						corr(w_3, δ) = {0, …, 0.9}					
	AVB			RMSE			AVB			RMSE		
	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3
OLS	0.019	0.240	0.104	0.022	0.277	0.129	0.020	0.141	0.182	0.022	0.160	0.213
	(0.003)	(0.025)	(0.026)	(0.003)	(0.027)	(0.031)	(0.003)	(0.023)	(0.031)	(0.004)	(0.024)	(0.033)
FE	0.006	0.063	0.123	0.007	0.080	0.151	0.006	0.050	0.124	0.007	0.061	0.152
	(0.001)	(0.016)	(0.030)	(0.002)	(0.021)	(0.035)	(0.001)	(0.012)	(0.031)	(0.002)	(0.014)	(0.037)
FEVD	0.006	0.063	0.312	0.007	0.080	0.351	0.006	0.050	0.355	0.007	0.061	0.398
	(0.001)	(0.016)	(0.057)	(0.002)	(0.021)	(0.059)	(0.001)	(0.012)	(0.048)	(0.002)	(0.014)	(0.050)
RE	0.005	0.073	0.108	0.006	0.095	0.133	0.005	0.049	0.129	0.006	0.060	0.157
	(0.001)	(0.017)	(0.025)	(0.001)	(0.023)	(0.030)	(0.001)	(0.012)	(0.030)	(0.001)	(0.014)	(0.034)
MLM	0.005	0.063	0.120	0.006	0.079	0.148	0.005	0.050	0.120	0.006	0.061	0.147
	(0.001)	(0.016)	(0.028)	(0.001)	(0.021)	(0.034)	(0.001)	(0.012)	(0.031)	(0.001)	(0.014)	(0.036)
BSEM	0.005	0.062	0.114	0.006	0.077	0.140	0.005	0.049	0.115	0.006	0.060	0.141
	(0.001)	(0.016)	(0.027)	(0.001)	(0.020)	(0.032)	(0.001)	(0.012)	(0.029)	(0.001)	(0.014)	(0.035)
	corr(w_3, δ) = 0.3, J = 20, T = 15						corr(x_3, δ) = 0.3, J = 20, T = 15					
OLS	0.017	0.248	0.116	0.020	0.283	0.145	0.018	0.146	0.184	0.021	0.164	0.216
	(0.003)	(0.023)	(0.028)	(0.003)	(0.025)	(0.035)	(0.003)	(0.021)	(0.033)	(0.003)	(0.022)	(0.036)
FE	0.004	0.044	0.087	0.005	0.055	0.107	0.004	0.035	0.086	0.005	0.043	0.107
	(0.001)	(0.010)	(0.021)	(0.001)	(0.014)	(0.024)	(0.001)	(0.008)	(0.021)	(0.001)	(0.009)	(0.026)
FEVD	0.004	0.044	0.281	0.005	0.055	0.312	0.004	0.035	0.330	0.005	0.043	0.370
	(0.001)	(0.010)	(0.048)	(0.001)	(0.014)	(0.049)	(0.001)	(0.008)	(0.040)	(0.001)	(0.009)	(0.042)
RE	0.004	0.054	0.084	0.004	0.072	0.103	0.003	0.034	0.099	0.004	0.042	0.120
	(0.001)	(0.012)	(0.020)	(0.001)	(0.018)	(0.023)	(0.001)	(0.008)	(0.021)	(0.001)	(0.009)	(0.024)
MLM	0.003	0.044	0.085	0.004	0.055	0.104	0.003	0.035	0.085	0.004	0.042	0.104
	(0.001)	(0.011)	(0.020)	(0.001)	(0.014)	(0.024)	(0.001)	(0.008)	(0.021)	(0.001)	(0.009)	(0.025)
BSEM	0.003	0.043	0.083	0.004	0.055	0.102	0.003	0.034	0.083	0.004	0.042	0.103
	(0.001)	(0.010)	(0.020)	(0.001)	(0.014)	(0.023)	(0.001)	(0.008)	(0.020)	(0.001)	(0.009)	(0.024)
	corr(w_3, δ) = 0.3, J = 20, T = 30						corr(x_3, δ) = 0.3, J = 20, T = 30					

TABLE 1 *(Continued)*

	corr(x_3, δ) = {0, … , 0.9}						corr(w_3, δ) = {0, … , 0.9}					
	AVB			RMSE			AVB			RMSE		
	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3
OLS	0.020	0.260	0.166	0.023	0.292	0.209	0.020	0.156	0.219	0.023	0.173	0.260
	(0.003)	(0.023)	(0.039)	(0.003)	(0.026)	(0.048)	(0.003)	(0.019)	(0.040)	(0.003)	(0.020)	(0.048)
FE	0.003	0.030	0.071	0.004	0.038	0.088	0.003	0.024	0.073	0.004	0.030	0.089
	(0.001)	(0.008)	(0.017)	(0.001)	(0.011)	(0.021)	(0.001)	(0.006)	(0.017)	(0.001)	(0.007)	(0.020)
FEVD	0.003	0.030	0.286	0.004	0.038	0.311	0.003	0.024	0.327	0.004	0.030	0.363
	(0.001)	(0.008)	(0.040)	(0.001)	(0.011)	(0.040)	(0.001)	(0.006)	(0.038)	(0.001)	(0.007)	(0.039)
RE	0.003	0.036	0.072	0.004	0.049	0.088	0.003	0.024	0.082	0.004	0.029	0.101
	(0.001)	(0.009)	(0.017)	(0.001)	(0.014)	(0.020)	(0.001)	(0.006)	(0.019)	(0.001)	(0.007)	(0.022)
MLM	0.003	0.030	0.069	0.004	0.038	0.085	0.003	0.024	0.070	0.003	0.030	0.086
	(0.001)	(0.008)	(0.017)	(0.001)	(0.011)	(0.020)	(0.001)	(0.006)	(0.017)	(0.001)	(0.007)	(0.019)
BSEM	0.003	0.030	0.068	0.004	0.038	0.084	0.003	0.024	0.070	0.003	0.029	0.086
	(0.001)	(0.008)	(0.017)	(0.001)	(0.011)	(0.020)	(0.001)	(0.006)	(0.016)	(0.001)	(0.007)	(0.019)
	corr(w_3, δ) = 0.3, J = 20, T = 60						corr(x_3, δ) = 0.3, J = 20, T = 60					

Note: Monte Carlo errors are presented in parentheses.
OLS = ordinary least squares; FE = fixed effects; FEVD = fixed-effects vector decomposition; RE = random effects; MLM = multilevel models; BSEM = Bayesian simultaneous equation model.

OLS estimator. Furthermore, the correlation between covariates and the unit effects exacerbates the problem. The magnitude of AVB and RMSE does not decrease as T increases.

Second, among these six estimators, the FEVD estimator produces estimates with the largest AVB and RMSE for the rarely changing variable in all cases while its estimates for the LDV and the time-varying variable are exactly the same with those under the FE model. The results indicate that the FEVD does not perform well for estimating rarely changing variables in dynamic panel data models.

Third, the RE is more biased than the FE in estimating correlated covariates x_3 and w_3 while the RE is more efficient than the FE in estimating slightly correlated covariates, i.e., corr (x_3, δ_j) = 0.3 or corr(w_3, δ_j) = 0.3. As the correlation is modeled under the MLM, the MLM and the FE perform equally well in estimating coefficients for the LDV and correlated covariates in terms of AVB and RMSE. In other words, when the correlation between the unit effects and covariates is modeled by Mundlak's (1978) approach, the MLM is less biased at the expense of efficiency.

Last but not least, the BSEM performs as well as, or better than the FE, RE, and MLM in terms of AVB and RMSE. We can see that, in almost all of the cases, the BSEM is the least biased and the most efficient estimator. These properties can be more clearly observed when T is small.

In Table 2, I report the AVB and RMSE over ten experiments for three different designs: $J = \{10, 40, 60\}$ with $T = 30$. Combining the results in Table 2 and the results in the middle block of Table 1 in which $J = 20$ and $T = 30$, we can see how the performances of these estimators change as the number of units increases. We find that the results are similar to those summarized above: the pooled OLS produces the poorest estimates and does not perform better as J increases; the FEVD does not perform well for estimating rarely changing variables in dynamic panel data models; the RE is more biased than the FE, MLM, and BSEM; the MLM and BSEM performs equally well with the FE in terms of AVB but are more efficient than the FE; the BSEM performs better than the MLM especially when J is small, e.g., $J = 10$.

I then show the effects of the correlation on the AVB and RMSE of the estimates for the LDV, time-varying variable x_3, and rarely changing variable w_3. To save space, I only present the results from the design where $J = 20$ and $T = 30$, but the results remain the same for $T = \{15, 60\}$ with J fixed and $J = \{40, 60\}$ with T fixed. Figure 1 presents the AVB and RMSE of the six estimators for the LDV, x_3, and w_3 when corr(x_3, δ_j) is varied and corr(w_3, δ_j) is fixed at the value of 0.3. The three panels on the left-hand side [(a), (c), and (e)] show the comparison in terms of AVB and, as can be seen, the pooled OLS produces estimates of the LDV and x_3 with the largest AVB. The RE performs worse for strongly correlated covariate x_3. The FEVD estimator has the poorest estimates of rarely changing variables and the FE, MLM, and BSEM perform more or less equally well in estimating all of these three variables no matter what the size of the correlation is.

Furthermore, the same pattern appears in the RMSE of these estimators represented in the three panels on the right-hand side of Figures 1(b), 1(d), and 1(f). Simply put, the pooled OLS has the largest RMSE for the LDV and time-varying variable x_3 while the FEVD estimator performs poorly for the estimate of rarely changing variable w_3. The RE has poor performance in RMSE for the covariate strongly correlated to the unit effects. The BSEM performs equally well with the FE and MLM.

Figure 2 presents the AVB and RMSE of the six estimators for the LDV, time-varying variable x_3, and rarely changing variable w_3 when corr(x_3, δ_j) is fixed at the value of 0.3 and corr(w_3, δ_j) is varied. It shows that, in general, the pooled OLS has the largest AVB and

TABLE 2 *An Absolute Value of Bias (AVB) and Root Mean Squared Error (RMSE) over 100 Replications Times Ten Experiments with J Varied*

	corr(x_3, δ) = {0, … , 0.9}						corr(w_3, δ) = {0, … , 0.9}					
	AVB			RMSE			AVB			RMSE		
	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3
OLS	0.020	0.214	0.217	0.025	0.258	0.271	0.019	0.127	0.252	0.024	0.155	0.307
	(0.005)	(0.036)	(0.062)	(0.006)	(0.044)	(0.074)	(0.004)	(0.029)	(0.058)	(0.006)	(0.033)	(0.071)
FE	0.008	0.060	0.160	0.010	0.075	0.198	0.007	0.051	0.157	0.009	0.062	0.196
	(0.002)	(0.016)	(0.045)	(0.002)	(0.020)	(0.059)	(0.002)	(0.011)	(0.042)	(0.002)	(0.012)	(0.049)
FEVD	0.008	0.060	0.382	0.010	0.075	0.443	0.007	0.051	0.405	0.009	0.062	0.461
	(0.002)	(0.016)	(0.075)	(0.002)	(0.020)	(0.079)	(0.002)	(0.011)	(0.068)	(0.002)	(0.012)	(0.074)
RE	0.007	0.064	0.147	0.009	0.083	0.179	0.007	0.050	0.156	0.008	0.061	0.191
	(0.002)	(0.016)	(0.040)	(0.002)	(0.022)	(0.049)	(0.002)	(0.010)	(0.041)	(0.002)	(0.012)	(0.047)
MLM	0.007	0.059	0.152	0.009	0.075	0.190	0.007	0.050	0.152	0.008	0.062	0.190
	(0.002)	(0.016)	(0.043)	(0.002)	(0.020)	(0.055)	(0.002)	(0.010)	(0.039)	(0.002)	(0.012)	(0.047)
BSEM	0.007	0.058	0.143	0.008	0.073	0.178	0.007	0.050	0.145	0.008	0.061	0.181
	(0.002)	(0.016)	(0.040)	(0.002)	(0.020)	(0.051)	(0.002)	(0.010)	(0.038)	(0.002)	(0.012)	(0.044)
	corr(w_3, δ) = 0.3, J = 10, T = 30						corr(x_3, δ) = 0.3, J = 10, T = 30					
OLS	0.021	0.272	0.109	0.023	0.299	0.136	0.022	0.166	0.174	0.023	0.177	0.203
	(0.002)	(0.017)	(0.024)	(0.002)	(0.019)	(0.029)	(0.002)	(0.014)	(0.026)	(0.002)	(0.014)	(0.028)
FE	0.003	0.031	0.058	0.004	0.040	0.071	0.003	0.024	0.057	0.003	0.029	0.070
	(0.001)	(0.008)	(0.015)	(0.001)	(0.012)	(0.017)	(0.001)	(0.006)	(0.013)	(0.001)	(0.007)	(0.016)
FEVD	0.003	0.031	0.313	0.004	0.040	0.325	0.003	0.024	0.349	0.003	0.029	0.377
	(0.001)	(0.008)	(0.030)	(0.001)	(0.012)	(0.029)	(0.001)	(0.006)	(0.028)	(0.001)	(0.007)	(0.028)
RE	0.003	0.048	0.063	0.003	0.068	0.075	0.002	0.024	0.083	0.003	0.029	0.102
	(0.001)	(0.012)	(0.016)	(0.001)	(0.019)	(0.018)	(0.001)	(0.006)	(0.017)	(0.001)	(0.006)	(0.020)
MLM	0.003	0.031	0.057	0.003	0.040	0.069	0.002	0.024	0.057	0.003	0.029	0.070
	(0.001)	(0.008)	(0.014)	(0.001)	(0.012)	(0.016)	(0.001)	(0.006)	(0.013)	(0.001)	(0.007)	(0.015)
BSEM	0.003	0.031	0.056	0.003	0.039	0.069	0.002	0.024	0.057	0.003	0.029	0.069
	(0.001)	(0.008)	(0.014)	(0.001)	(0.012)	(0.016)	(0.001)	(0.006)	(0.012)	(0.001)	(0.007)	(0.015)
	corr(w_3, δ) = 0.3, J = 40, T = 30						corr(x_3, δ) = 0.3, J = 40, T = 30					

OLS	0.022	0.280	0.127	0.025	0.306	0.154	0.023	0.173	0.187	0.024	0.184	0.217
	(0.002)	(0.012)	(0.024)	(0.002)	(0.014)	(0.028)	(0.002)	(0.012)	(0.024)	(0.002)	(0.011)	(0.026)
FE	0.003	0.025	0.050	0.003	0.031	0.060	0.002	0.020	0.048	0.003	0.025	0.060
	(0.001)	(0.006)	(0.014)	(0.001)	(0.008)	(0.017)	(0.001)	(0.005)	(0.011)	(0.001)	(0.006)	(0.014)
FEVD	0.003	0.025	0.321	0.003	0.031	0.328	0.002	0.020	0.354	0.003	0.025	0.379
	(0.001)	(0.006)	(0.023)	(0.001)	(0.008)	(0.023)	(0.001)	(0.005)	(0.026)	(0.001)	(0.006)	(0.023)
RE	0.002	0.047	0.058	0.003	0.069	0.070	0.002	0.020	0.083	0.003	0.025	0.101
	(0.001)	(0.008)	(0.014)	(0.001)	(0.013)	(0.016)	(0.001)	(0.005)	(0.014)	(0.001)	(0.006)	(0.016)
MLM	0.002	0.025	0.047	0.003	0.031	0.058	0.002	0.020	0.045	0.003	0.025	0.056
	(0.001)	(0.006)	(0.013)	(0.001)	(0.008)	(0.016)	(0.001)	(0.005)	(0.011)	(0.001)	(0.006)	(0.013)
BSEM	0.002	0.025	0.047	0.003	0.031	0.057	0.002	0.020	0.045	0.003	0.025	0.056
	(0.001)	(0.006)	(0.013)	(0.001)	(0.008)	(0.016)	(0.001)	(0.005)	(0.011)	(0.001)	(0.006)	(0.013)
	corr(w_3, δ) = 0.3, J = 60, T = 30						corr(x_3, δ) = 0.3, J = 60, T = 30					

Note: Monte Carlo errors are presented in parentheses.
OLS = ordinary least squares; FE = fixed effects; FEVD = fixed-effects vector decomposition; RE = random effects; MLM = multilevel models; BSEM = Bayesian simultaneous equation model.

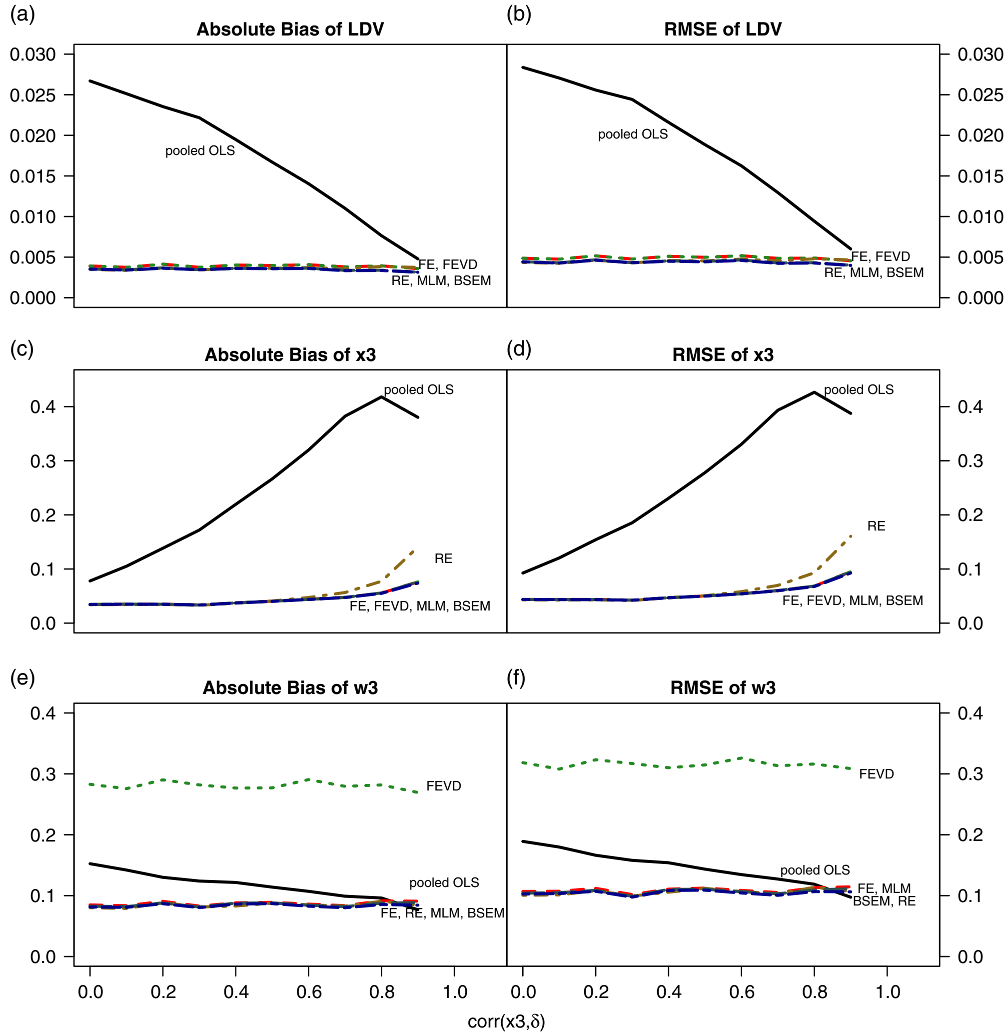

Fig. 1. Change in the absolute value of bias and root mean squared error (RMSE) over variations in the correlation between the unit effects and the time-varying variable x_3; corr(w_3, δ_j) = 0.3, J = 20, and T = 30
Note: LDV = lagged dependent variable; OLS = ordinary least squares; FE = fixed effects; FEVD = fixed-effects vector decomposition; RE = random effects; MLM = multilevel models; BSEM = Bayesian simultaneous equation model.

inefficient estimates of the LDV and x_3; the RE has biased and inefficient estimates for the LDV and w_3; the FEVD performs poorly in terms of the AVB and RMSE in estimating the coefficient of w_3. The FE, MLM, and BSEM perform equally well in estimating the LDV, x_3, and w_3.

Finally, I present the estimates of the correlation between the unit effects and the time-varying variable and rarely changing variable from the proposed model. To save space, I only show the results from the design where $T = 30$ and $J = 20$. In Figure 3, the two panels on the top are from the design where corr(x_3, δ_j) is varied while corr(w_3, δ_j) is fixed at the value of 0.3; the two panels at the bottom are from the design where corr(x_3, δ_j) is fixed at the value of 0.3

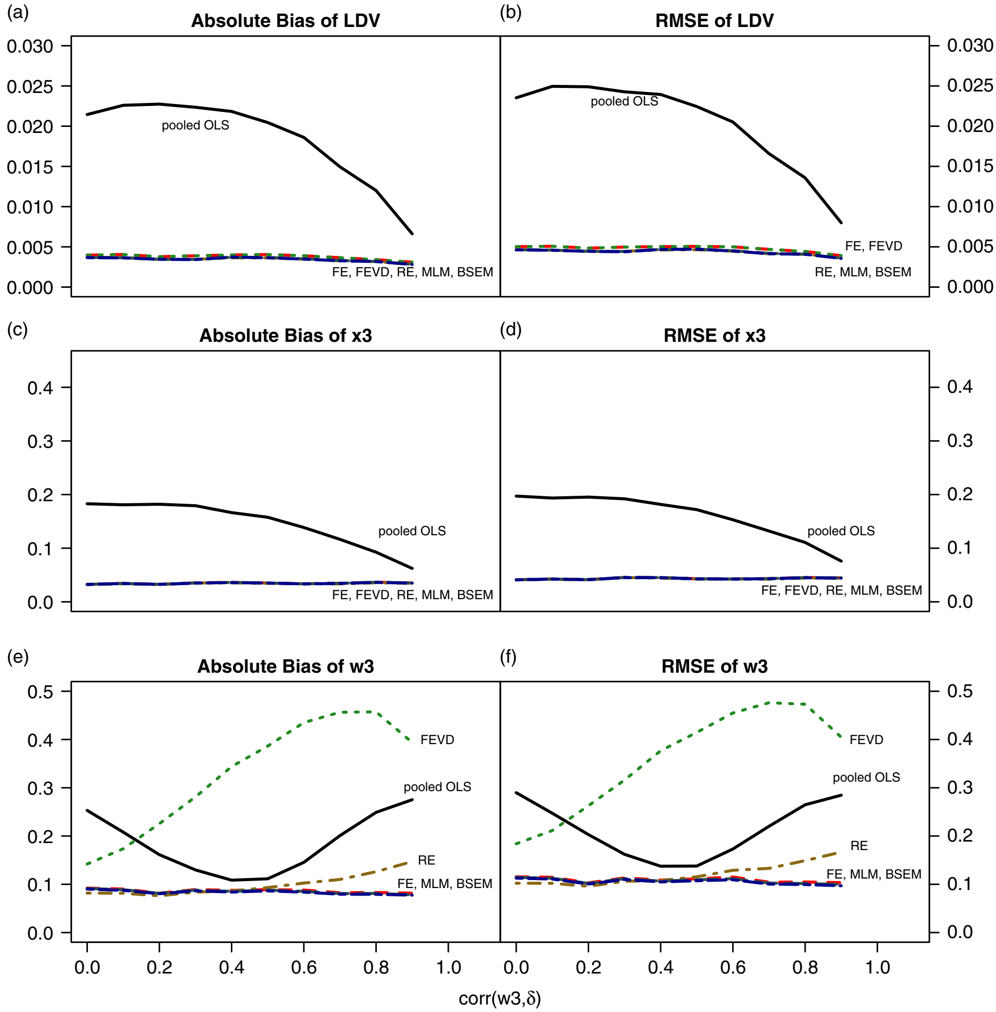

Fig. 2. Change in the absolute value of bias and root mean squared error (RMSE) over variations in the correlation between the unit effects and the rarely changing variable w_3; corr(x_3, δ_j) = 0.3, J = 20, and T = 30
Note: LDV = lagged dependent variable; OLS = ordinary least squares; FE = fixed effects; FEVD = fixed-effects vector decomposition; RE = random effects; MLM = multilevel models; BSEM = Bayesian simultaneous equation model.

while corr(w_3, δ_j) is varied. Moreover, the solid lines represent the 90 percent highest posterior density (HPD) intervals of the correlation between the unit effects and covariates from generated data while the dotted lines represent the 90 percent HPD intervals of the estimated correlation parameters. As can be seen in Figure 3, the estimates of correlation between the unit effects and the time-varying variable x_3 and those between the unit effects and rarely changing variable w_3 are approximately close to those in the generated data.[10] As the dotted lines represent

[10] The results show that the BSEM underestimates the correlation, especially when the correlation is high. In real data, however, the correlation is rarely >0.7 as shown in the application. When the correlation is low or medium, the BSEM actually estimates the correlation well.

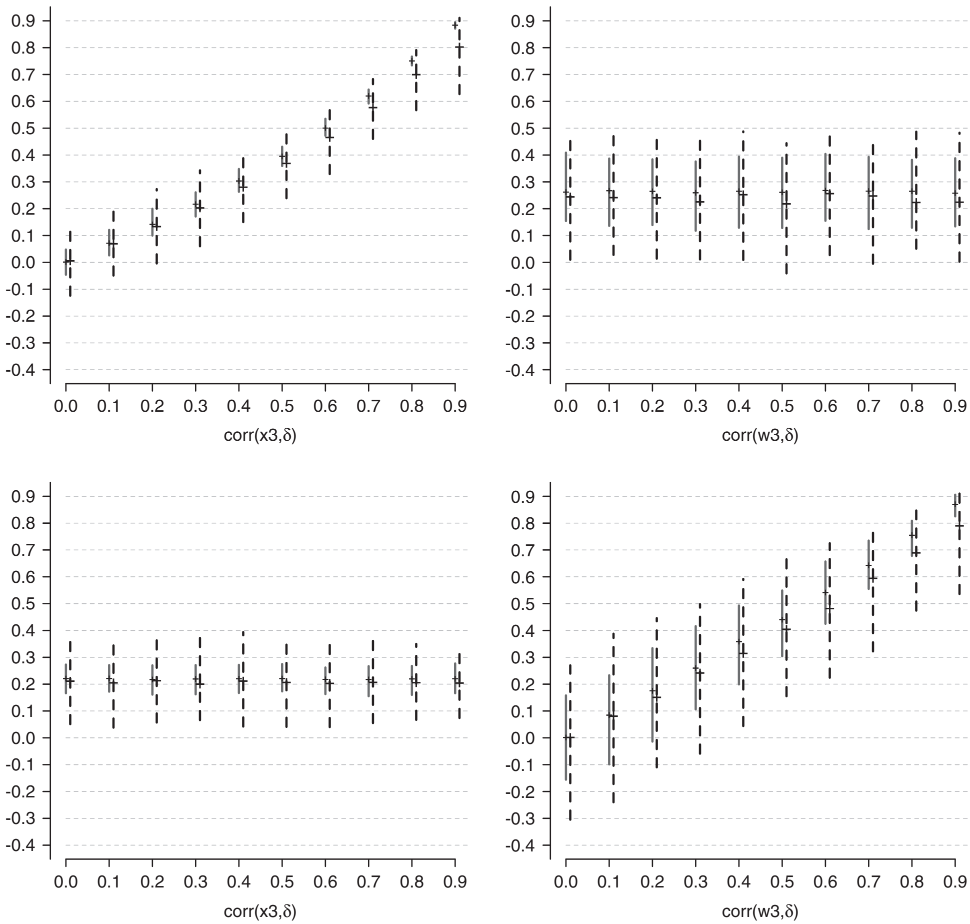

Fig. 3. The 90 percent highest posterior density intervals of the simulated and estimated correlations between the unit effects and the covariates x_3 and w_3 for T = 30 and J = 20

Note: The plus signs denote the median. The solid lines present the correlation from the simulated data while the dotted lines present the estimates of correlation.

the estimates, they have larger uncertainty than the simulated correlation coefficients. But the wide range of uncertainty decreases as J or T increases, which is not shown here to save space.

In short, the proposed model not only performs as well as, or better than classical estimators in estimating coefficients for correlated covariates in terms of bias and efficiency especially when J or T is small, but also provides estimates for the correlation between the unit effects and covariates. Therefore, when the sample size is small or the degree of the correlation between the unit effects and covariates is of interest, the proposed model is preferred to the conventional estimators such as the FE models which eliminate the unit effects, and the RE models which assume no correlation.

Robustness Check for Distributional Assumptions

One concern about the simulation results is that the setup of a bivariate normal distribution for δ_j and x_{3j} matches the model assumption. In this subsection, I conduct a robustness analysis to check whether the data are simulated in favor of the proposed model and against the alternative estimators. Instead of drawing samples from a standard normal distribution, I draw the samples of δ_j from a uniform distribution in the interval between −2 and 2. Then, the unit-level covariates x_{3j} and w_{3j} are generated based on the assigned values of correlation and the individual-level samples x_{3jt} and w_{3jt} are generated in turn, which is the same procedure as that is stated in the Simulation Design section (and in footnote 7 for details). By drawing samples of δ_j from a uniform distribution, the joint distribution of δ_j and x_{3jt} is a bimodal distribution rather than a bivariate normal distribution. The comparison between alternative estimators is conducted under the setting that $J = 20$, $T = 15$, and $R = 100$.

The results are shown in Table 3, which indicate that the performances of these estimators are consistent to what we find in the experiment with $J = 20$ and $T = 15$ presented in the top block of Table 1. The results remain the same when J or T is changed. Simply put, the results in Table 3 suggest that the changes in the distributional setup for the joint distribution of (δ_j, x_{3jt}) and that of (δ_j, w_{3jt}) do not influence the estimates. These results indicate that the BSEM still performs well even if the distributional assumptions of the BSEM are not matched by the simulation design, which implies that the simulation results are quite robust.

Sensitivity Analysis for the Choice of Priors

As discussed in the A Model for Endogenous Rarely Changing Variables section, one of the advantages of the proposed model against alternative estimators is that the BSEM can estimate the correlation between covariates and the unit effects. The major component for the estimation of correlation lies in the variance–covariance matrix $\mathbf{\Omega}$, which is assumed to be an inverse Wishart distribution with the degrees of freedom ν_0 and the scale matrix $\mathbf{\Lambda}_0$. In both the simulation and the application, these hyperparameters of the inverse Wishart distribution are assigned values to be an uninformative prior for the correlation. In specific, suppose that $\mathbf{\Omega}$ is a $p \times p$ matrix. The degree of freedom ν_0 is set to be $p+1$ and the scale matrix $\mathbf{\Lambda}_0$ is an identity matrix $\mathbf{I}_p$, which implies a uniform distribution for the correlation.

To check the sensitivity to the choice of this parameterization, I use three different parameterizations for the degrees of freedom while the scale matrix is fixed to an identity matrix $\mathbf{I}_p$: $\nu_0 = \{p, p+3, p+7\}$, where $p = 3$ in the simulation. The sensitivity analysis is conducted under the setting that $J = 20$, $T = 15$, and $R = 100$. To save space, I only present the 90 percent HPD intervals of varied correlations for the designs where $\nu_0 = p$ and $p+7$ displayed in Figure 4.[11] As can be seen, panels (a) and (c) in Figure 4 show the same pattern as the upper-left panel in Figure 3, which suggests that the estimated correlations are close to the simulated ones although the BSEM underestimates the correlation when the correlation is high. By the same token, panels (b) and (d) in Figure 4 show the same pattern as the bottom-right panel in Figure 3. In sum, these results show that the posterior distribution of correlation parameters from the BSEM is not affected by the choice of prior values.

APPLICATION: SOCIAL SPENDING IN LATIN AMERICA

Previous studies on democracy and social welfare policy focus on Latin American countries because, first, there is great variation in social welfare systems among Latin American countries

[11] p and $p+7$ are selected to show two extremely distinct distributions for the correlation. When $\nu_0 = p$, the prior for the correlation is a bimodal distribution which has two peaks around −1 and 1. When $\nu_0 = p+7$, the prior for the correlation is a unimodal distribution centered on 0.

TABLE 3 *Absolute Value of Bias (AVB) and Root Mean Squared Error (RMSE) over 100 Replications Times Ten Experiments with Joint Distributions Changed*

	$\text{corr}(x_3, \delta) = \{0, \ldots, 0.9\}$						$\text{corr}(w_3, \delta) = \{0, \ldots, 0.9\}$					
	AVB			RMSE			AVB			RMSE		
	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3	y_{t-1}	x_3	w_3
OLS	0.019	0.244	0.108	0.022	0.281	0.132	0.019	0.138	0.176	0.022	0.157	0.208
	(0.003)	(0.024)	(0.027)	(0.003)	(0.025)	(0.033)	(0.003)	(0.021)	(0.028)	(0.003)	(0.022)	(0.030)
FE	0.006	0.063	0.120	0.007	0.080	0.149	0.006	0.050	0.122	0.007	0.061	0.150
	(0.001)	(0.013)	(0.029)	(0.002)	(0.018)	(0.035)	(0.001)	(0.011)	(0.032)	(0.002)	(0.014)	(0.038)
FEVD	0.006	0.063	0.312	0.007	0.080	0.356	0.006	0.050	0.359	0.007	0.061	0.403
	(0.001)	(0.013)	(0.051)	(0.002)	(0.018)	(0.052)	(0.001)	(0.011)	(0.046)	(0.002)	(0.014)	(0.049)
RE	0.005	0.072	0.104	0.006	0.093	0.130	0.005	0.048	0.128	0.006	0.060	0.155
	(0.001)	(0.016)	(0.024)	(0.001)	(0.022)	(0.029)	(0.001)	(0.011)	(0.027)	(0.002)	(0.013)	(0.030)
MLM	0.005	0.063	0.117	0.006	0.079	0.146	0.005	0.050	0.118	0.006	0.061	0.145
	(0.001)	(0.014)	(0.027)	(0.001)	(0.018)	(0.033)	(0.001)	(0.011)	(0.032)	(0.002)	(0.013)	(0.037)
BSEM	0.005	0.062	0.111	0.006	0.077	0.139	0.005	0.049	0.113	0.006	0.060	0.140
	(0.001)	(0.013)	(0.025)	(0.002)	(0.016)	(0.031)	(0.001)	(0.011)	(0.029)	(0.002)	(0.013)	(0.033)
	$\text{corr}(w_3, \delta) = 0.3, J = 20, T = 15$						$\text{corr}(x_3, \delta) = 0.3, J = 20, T = 15$					

Note: Monte Carlo errors are presented in parentheses.
OLS = ordinary least squares; FE = fixed effects; FEVD = fixed-effects vector decomposition; RE = random effects; MLM = multilevel models; BSEM = Bayesian simultaneous equation model.

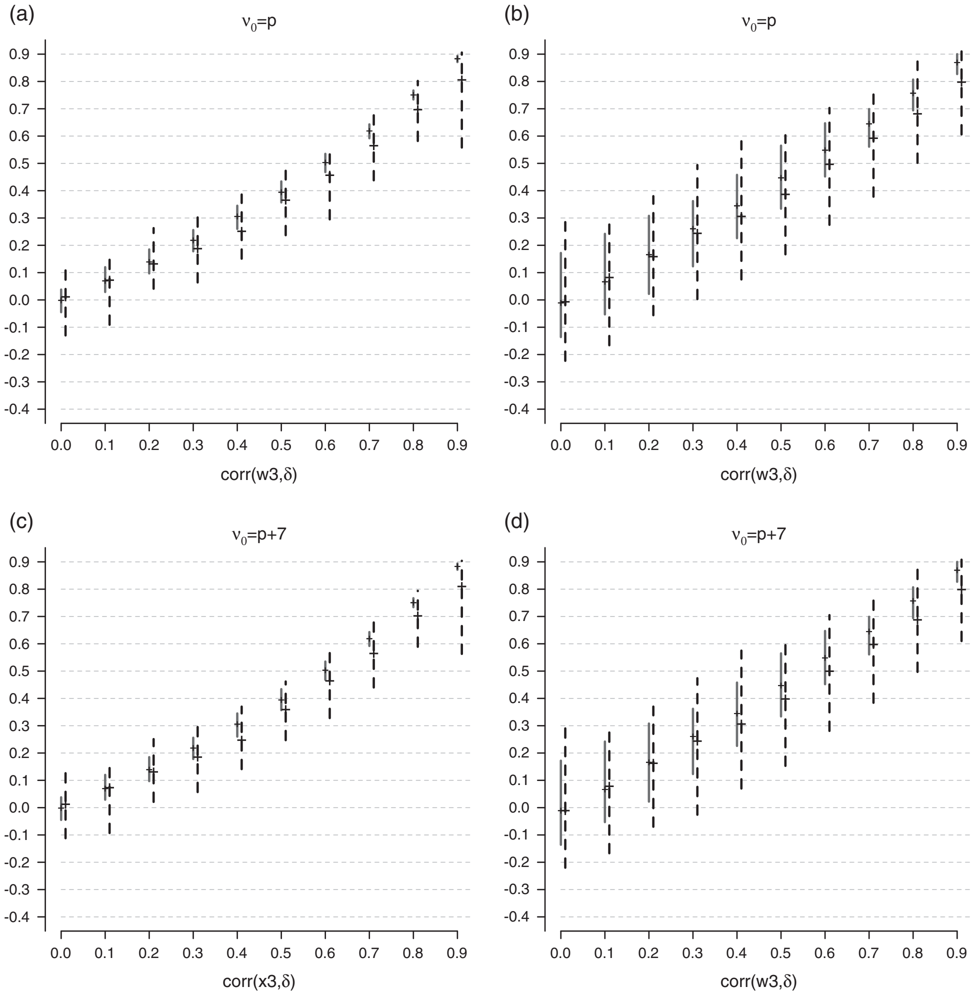

Fig. 4. The 90 percent highest posterior density intervals of the simulated and estimated correlation coefficients between the unit effects and the covariates for T = 30 and J = 20
Note: The plus signs denote the median. The solid lines present the correlation from the simulated data while the dotted lines present the estimates of correlation.

and, second, most Latin American countries experienced regime change between democracy and authoritarianism. However, the empirical results are not consistent. For example, while some studies show that democratic regimes tend to spend more on overall social programs than authoritarian regimes (Brown and Hunter 1999; Avelino, Brown and Hunter 2005), others find no robust evidence that democracy has an impact on aggregate social expenditures (Kaufman and Segura-Ubiergo 2001). Moreover, for disaggregate spending, some find that democracy has a positive effect on social security spending in the long term (Huber et al. 2008), while others do not find strong evidence in either the short term or the long term (Kaufman and Segura-Ubiergo 2001; Avelino, Brown and Hunter 2005).

The inconsistent results in previous studies may result from at least two methodological issues. First, the slowly changing property of political regimes are not explicitly illustrated and different model strategies are used. Some of these studies include an LDV to correct for serial correlation (Kaufman and Segura-Ubiergo 2001; Avelino, Brown and Hunter 2005). Some other studies consider the long-term properties of political regimes, but inappropriately model the long-term effects of democracy by cumulating yearly values of democracy variable (Huber et al. 2008).

Second, the absence of country-specific effects in statistical models is one of the problems in past empirical research (Ross 2006). One of the reasons to exclude unit effects is that these unit effects influence the coefficients of factors that vary mainly between countries (Plümper, Troeger and Manow 2005). As discussed before, excluding the unit effects may be at the risk of having omitted variable bias, which leads to invalid inferences. Furthermore, as discussed in the A Model for Endogenous Rarely Changing Variables section, it is better to include an LDV and unit effects into model specifications to evaluate the effects of democracy on social welfare spending because it is believed that the effects of democracy is distributed across several extended time periods.

Data and Measurements

I apply the BSEM presented in the A Model for Endogenous Rarely Changing Variables section to analyze social spending in Latin America, which handles the above issues. The data set I use was collected by Huber et al. (2008).[12] This data set covers a number of political and economic variables in 18 Latin American countries from 1970 to 2000. In this application, the outcome variable is SSW, and the main explanatory variable is the level of democracy.

Before discussing the results of analysis, two important points need to be observed. First, SSW is measured as a percentage of gross domestic product (GDP). As it is bounded between 0 and 100 percent, we can argue that the proportion of GDP spent on social security is stationary (Beck and Katz 2011). Second, the distribution of social security spending is right-skewed, so I use the logarithm transformation, which is better described by a normal distribution.[13]

The level of democracy (DEM), which is rarely changing, is measured by the Polity index ranging from −10 to 10 with 10 as the highest degree of democracy (Marshall, Jaggers and Gurr 2010). I also include several control variables that may influence social spending discussed in the literature. These control variables are gross domestic production per capita (GDPPC) (million US dollars) adjusted for purchasing power parities, the percentage of population that lives in the urban area (UBNPOP), the percentage of aged population (POP65), export and import as a percentage of GDP as a measurement of trade openness (TRADE), and foreign direct investment (FDI) as a percentage of GDP.[14]

[12] The data on SSW were collected by Huber et al. (2008) from the International Monetary Fund (IMF) and those on education and health spending were from the Economic Commission for Latin American and the Caribbean, Cominetti’s (1996) data set, and the IMF. The data set can be downloaded from the website http://www.unc.edu/ jdsteph/common/data-common.html

[13] Six observations (1981–1986 in Peru) have values of 0 in the measurement of social security/welfare spending, which makes logarithm transformation produce negative infinity. For these six observations, I treat them as missing rather than replace them with small values. Looking at the data carefully, we observe that the measure of spending on social security/welfare is missing in 1979, 1980, and from 1987 to 1989. Consequently, it is reasonable to treat them as missing. It turns out that this setup does not affect the results.

[14] I update the observations on FDI for missingness from World Development Indicators (2011) from the World Bank. This updated data can be downloaded in http://data.worldbank.org/data-catalog/world-development-indicators. For the remaining missing values in TRADE and FDI, I employed multiple imputation (Rubin 1987) by the R package mice (Van Buuren and Groothuis-Oudshoorn 2011).

As I am not sure which covariate is correlated with the unit effects, I model all of them. Therefore, the simultaneous equation model is given by Equation 23:

$$\begin{aligned}
\ln(\mathrm{SSW}_{jt}) &= \beta_0 + \phi \ln(\mathrm{SSW}_{j(t-1)}) + \beta_1 \mathrm{DEM}_{jt} + \beta_2 \mathrm{GDPPC}_{jt} + \beta_3 \mathrm{UBNPOP}_{jt} \\
&\quad + \beta_4 \mathrm{POP65}_{jt} + \beta_5 \mathrm{TRADE}_{jt} + \beta_6 \mathrm{FDI}_{jt} + \delta_j + \varepsilon_{jt}, \\
\mathrm{DEM}_{jt} &= \zeta_1 + \eta_{1j} + \xi_{1jt}, \\
\mathrm{GDPPC}_{jt} &= \zeta_2 + \eta_{2j} + \xi_{2jt}, \\
\mathrm{UBNPOP}_{jt} &= \zeta_3 + \eta_{3j} + \xi_{3jt}, \\
\mathrm{POP65}_{jt} &= \zeta_4 + \eta_{4j} + \xi_{4jt}, \\
\mathrm{TRADE}_{jt} &= \zeta_5 + \eta_{5j} + \xi_{5jt}, \\
\mathrm{FDI}_{jt} &= \zeta_6 + \eta_{6j} + \xi_{6jt}. \qquad (23)
\end{aligned}$$

This model is estimated with MCMC techniques and implemented in JAGS 3.1.0 called from R (R2jags).[15]

Results of Analysis

The results of the analysis of social spending are displayed in Figures 5–7. Figure 5 presents the intervals of 90 percent level for explanatory variables from the six estimators discussed in the Monte Carlo Simulations section. We can see that, first, in general the BSEM, FE, and MLM provide very similar results for the estimates for these covariates,[16] which suggest that democracy, GDP per capita, and the urban population matter for the outcome variable. In specific, democracy and the urban population have a positive effect while GDP per capita has a negative effect on social spending.

Second, the estimate for democracy under the FEVD is negative, which is extremely different from the results in alternative estimators. The findings in the simulations tell us that this may be misleading as the FEVD is biased for rarely changing variable. Moreover, we also observe that the FEVD provides smaller uncertainty of estimates than the FE (Breusch et al. 2011), which leads to incorrect inferences, e.g., a negative effect of democracy and a positive effect of the aged population and trade.

Finally, the estimates for democracy under the pooled OLS and RE are somewhat different from those under the BSEM, FE, and MLM. As the simulation results suggest, it is possible that the level of democracy is correlated to the unit effects based on the findings in the simulations. This situation also occurs for GDP per capita and the aged population.

---

[15] Vague priors are used for parameters, which are presented in Appendix A. The estimation was performed with three parallel chains of 50,000 iterations each to be conservative. The first half of the iterations were discarded as a burn-in period and five as thinning and thus 15,000 samples were generated. The convergence of MCMC chains is conducted by using the R function superdiag and there is no evidence of non-convergence in these chains.

[16] A Bayesian version of the MLM (BMLM) is estimated for the comparison with the BSEM. The results (not shown here) suggest that the BMLM performs almost the same with the BSEM. Although both BMLM and BSEM perform equally well, like the MLM, the BMLM provides no information on the correlation between the unit effects and covariates.

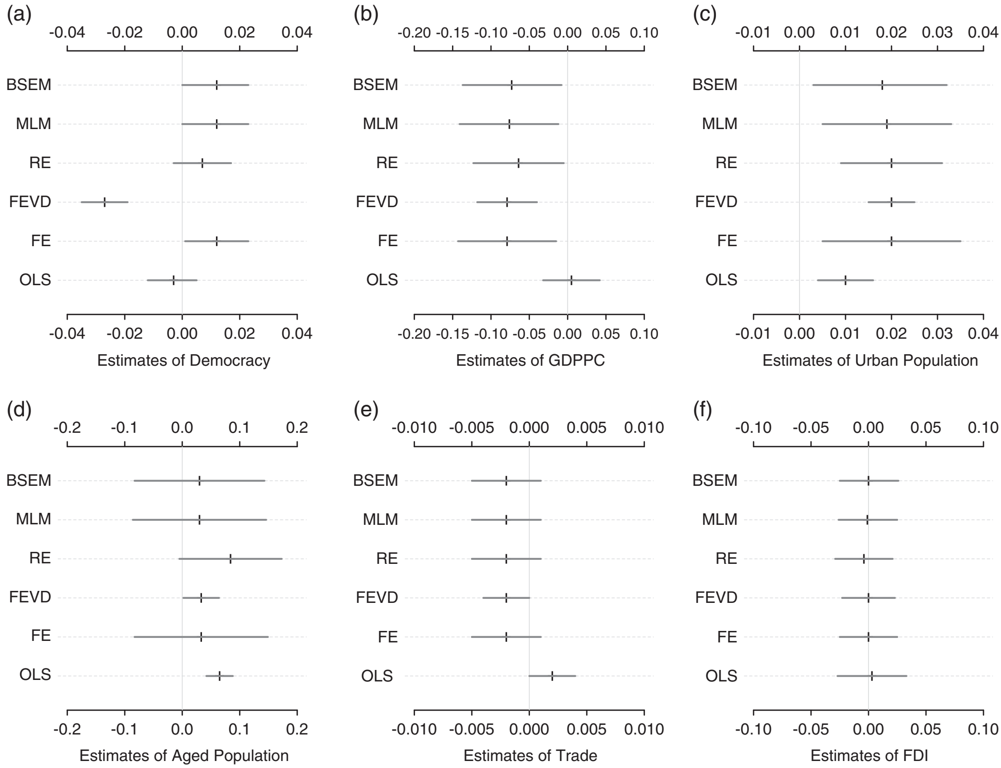

Fig. 5. Determinants of social security spending, 1971–2000
Note: BSEM = Bayesian simultaneous equation model; MLM = multilevel models; RE = random effects; FEVD = fixed-effects vector decomposition; FE = fixed effects; OLS = ordinary least squares; GDPPC = gross domestic production per capita; FDI = foreign direct investment.

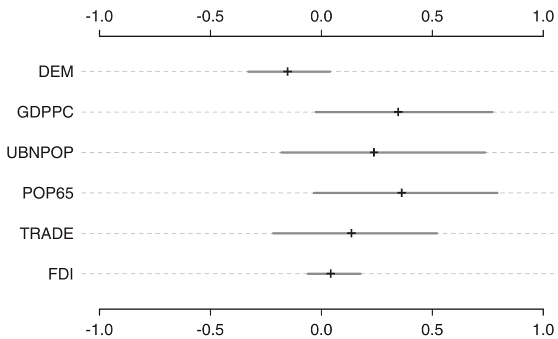

Fig. 6. The 90 percent highest posterior density of estimated correlation between unit effects and covariates
Note: DEM = democracy; GDPPC = gross domestic production per capita; UBNPOP = percentage of population that lives in the urban area; POP65 = percentage of aged population; TRADE = measurement of trade openness; FDI = foreign direct investment.

To check the possibility of correlation between these variables and the unit effects, I rely on the 90 percent credible intervals of the estimated correlations provided by the BSEM presented in Figure 6. As can be seen, the posterior distribution of the estimated correlation between the

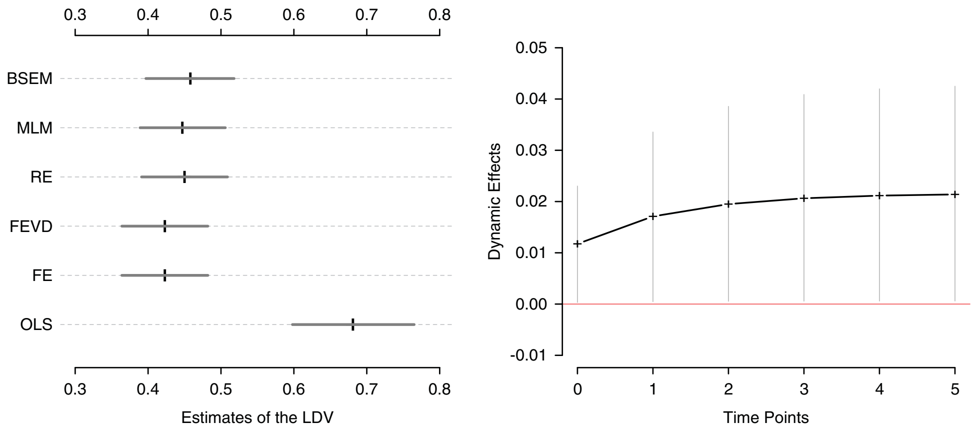

Fig. 7. Estimates of the autoregressive coefficient and dynamic effects of political regimes on social spending. The solid lines represent 90 percent credible intervals

Note: BSEM = Bayesian simultaneous equation model; MLM = multilevel models; RE = random effects; FEVD = fixed-effects vector decomposition; FE = fixed effects; OLS = ordinary least squares; LDV = lagged dependent variable.

unit effect and the level of democracy mainly lies in the negative region, which implies a negative correlation. In contrast, Figure 6 shows that the other five variables are positively correlated with the unit effects, among which the 90 percent credible intervals of correlation for GDPPC and POP65 mainly lie in the positive region although the intervals are wide. In short, the results suggest that the level of democracy, GDP per capita, and the aged population are slightly or moderately correlated to the unit effects. In other words, the results imply that the unobserved, time-invariant features of individual countries simultaneously influence social spending, political development, economic development, and the aged population.

Figure 5 shows that democracy matters for social spending. In a dynamic model, it basically means that democracy has an immediate effect on social provision policy. To see dynamic effects, one should consider the dynamic multiplier, which involves the coefficient of the LDV, ϕ (Keele and Kelly 2006). The left panel in Figure 7 shows the 90 percent intervals of the LDV from the six estimators. As can be seen, the estimates of the LDV from five estimators are almost equivalent except for that from the pooled OLS. It is well known that the estimate of the LDV is biased upward under the pooled OLS, which is observed in Figure 7. Taking the estimates from the BSEM, the right panel in Figure 7 presents the dynamic effects of democracy over six time periods when the Polity scores increase 1 unit. As can be seen, the posterior probability of the effect of democracy on social spending is over 90 percent. This implies that, for a given country, when the level of democracy increases, social spending increases as well. Overall, the evidence shows that democratic regimes have a positive effect on social spending in either the short term or the long term.

CONCLUDING REMARKS

Existing literature argues that democracies pursue more welfare-enhancing policies than non-democracies. Recently, some studies argue that the realization of redistributive policies in democracies may be observed in the long term (Keefer and Khemani 2005; Keefer and Vlaicu 2008). The empirical evidence of how political regimes affect redistributive policies, however, is inconsistent. Following studies of political regimes and redistributive policies in

Latin America, this article examines the dynamic effects of political regimes on social programs by focusing on estimating the effects of rarely changing variables within a Bayesian framework.

The dynamic effects of rarely changing variables such as the level of democracy are often of great interest in the study of political economy. Nevertheless, the classical estimators are problematic when estimating the effects of time-invariant and rarely changing variables along with unit effects, the control of which is important in the analysis of welfare states (Kaufman and Segura-Ubiergo 2001; Ross 2006). This article shows the flexibility and advantages of the Bayesian approach to estimating the coefficients of endogenous explanatory variables along with the parameters of the correlation between the unit effects and covariates. The finite sample properties of the proposed model and alternative estimators that are widely applied to panel data analyses are explored in Monte Carlo simulations. The results show that the proposed model not only performs as well as, or better than alternative estimators in terms of bias and efficiency, but also provides additional information on the degree of the correlation between the unit effects and covariates, which will not be discovered without the proposed model. Applying the BSEM to the study on social spending in Latin America, this article finds evidence that political regimes affect social welfare spending both in the short and long terms.

For future research, first, the proposed model can be applied to studies in which researchers are interested in the long-term effects of political institutions such as the effect of centralized wage bargaining on income inequality (Scheve and Stasavage 2009). Second, it would be useful to model heterogeneous coefficients and account for the correlated predictors at the group level as we have substantive (Western 1998; Beck and Katz 2007) and methodological (Nerlove 1971; Pesaran and Smith 1995; Hsiao Pesaran and Tahmiscioglu 1999) reasons to do so. Finally, the endogenous covariates are not limited to continuous variables, but include discrete variables, which can be modeled by assuming an unobserved latent trait underlying observed discrete variables.

REFERENCES

Anderson, T. W., and Cheng Hsiao. 1981. 'Estimation of Dynamic Models with Error Components'. *Journal of the American Statistical Association* 76(375):598–606.

Anderson, T. W., and Cheng Hsiao. 1982. 'Formulation and Estimation of Dynamic Models Using Panel Data'. *Journal of Econometrics* 18(1):47–82.

Avelino, George, David S. Brown, and Wendy Hunter. 2005. 'The Effects of Capital Mobility, Trade Openness, and Democracy on Social Spending in Latin America, 1980–1999'. *American Journal of Political Science* 49(3):625–41.

Baltagi, Badi H. 2005. *Econometric Analysis of Panel Data*, 3rd ed. New York: John Wiley & Sons, Inc.

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. 'Fitting Linear Mixed – Effects Models Using lme4'. *Journal of Statistical Software* 67(1):1–48.

Beck, Nathaniel. 2001. 'Time-Series-Cross-Section Data: What Have we Learned in the Past Few Years?'. *Annual Review of Political Science* 4(1):271–93.

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

Beck, Nathaniel, and Jonathan N. Katz. 2007. 'Random Coefficient Models for Time-Series-Cross-Section Data: Monte Carlo Experiments'. *Political Analysis* 15(2):182–95.

Beck, Nathaniel, and Jonathan N. Katz. 2011. 'Modeling Dynamics in Time-Series-Cross-Section Political Economy Data'. *Annual Review of Political Science* 14:331–52.

Breusch, Trevor, Michael B. Ward, Hoa Thi Minh Nguyen, and Tom Kompas. 2011. 'On the Fixed-Effects Vector Decomposition'. *Political Analysis* 19(2):123–34.

Brown, David S., and Wendy Hunter. 1999. 'Democracy and Social Spending in Latin America, 1980–92'. *American Political Science Review* 93(4):779–90.

Carlin, Bradley P., and Thomas A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman & Hall/CRC.

Casella, George, and Edward I. George. 1992. 'Explaining the Gibbs Sampler'. *American Statistician* 46(3):167–74.

Chib, Siddhartha, and Edward Greenberg. 1995. 'Understanding the Metropolis-Hastings Algorithm'. *American Statistician* 49(4):327–35.

Cominetti, Rossella. 1996. 'Social Expenditure in Latin America: An Update'. Technical report, CEPAL Technical Department, Santiago, Chile.

Croissant, Yves, and Giovanni Millo. 2008. 'Panel Data Econometrics in R: The PLM Package'. *Journal of Statistical Software* 27(2):1–43.

De Boef, Suzanna, and Luke Keele. 2008. 'Taking Time Seriously'. *American Journal of Political Science* 52(1):184–200.

Enders, Walter. 2004. *Applied Econometric Time Series*. New York: John Wiley & Sons, Inc.

Gelfand, Alan E., and Adrian F. M. Smith. 1990. 'Sampling-Based Approaches to Calculating Marginal Densities'. *Journal of the American Statistical Association* 85(410):398–409.

Gelman, Andrew. 2006. 'Multilevel (Hierarchical) Modeling: What it Can and Cannot Do'. *Technometrics* 48(3):432–35.

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*, 2nd ed. New York: Chapman & Hall/CRC.

Gill, Je. 2008a. *Bayesian Methods: A Social and Behavioral Sciences Approach*, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC press.

Gill, Je. 2008b. 'Is Partial-Dimension Convergence a Problem for Inferences from MCMC Algorithms?'. *Political Analysis* 16(2):153–78.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hsiao, Cheng. 2003. *Analysis of Panel Data*. New York: Cambridge University Press.

Hsiao, Chen, M. Hashem Pesaran, and A. Kamil Tahmiscioglu. 1999. 'Bayes Estimation of Short-Run Coefficients in Dynamic Panel Data Models'. In Chen Hsiao, M. Hashem Pesaran, Kajal Lahiri and Lung Fei Lee (eds), *Analysis of Panels and Limited Dependent Variables*, chapter 11, 268–96. New York, NY: Cambridge University Press.

Huber, Evelyne, John D. Stephens, Thomas Mustillo, and Jenny Pribble. 2008. *Social Policy in Latin America and the Caribbean Dataset, 1960–2006*. Chapel Hill, NC: University of North Carolina.

Huber, Evelyne, John D. Stephens, Thomas Mustillo, and Jenny Pribble. 2008. 'Politics and Social Spending in Latin America'. *The Journal of Politics* 70(2):420–36.

Kaufman, Robert R., and Alex Segura-Ubiergo. 2001. 'Globalization, Domestic Politics, and Social Spending in Latin America: A Time-Series Cross-Section Analysis, 1973–97'. *World Politics* 53(4): 553–87.

Keefer, Philip, and Razvan Vlaicu. 2008. 'Democracy, Credibility, and Clientelism'. *Journal of Law, Economics, and Organization* 24(2):371–406.

Keefer, Philip, and Stuti Khemani. 2005. 'Democracy, Public Expenditures, and the Poor: Understanding Political Incentives for Providing Public Services'. *The World Bank Research Observer* 20(1):1–27.

Keele, Luke, and Nathan J. Kelly. 2006. 'Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables'. *Political Analysis* 14(2):186–205.

Koehler, Elizabeth, Elizabeth Brown, and Sebastien J.-P. Haneuse. 2009. 'On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses'. *The American Statistician* 63(2):155–62.

Lindley, D. V. 1972. *Bayesian Statistics: A Review*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Lunn, David J., Andrew Thomas, Nicky Best, and David Spiegelhalter. 2000. 'WinBUGS – A Bayesian Modeling Framework: Concepts, Structure, and Extensibility'. *Statistics and Computing* 10(4):325–37.

Marshall, Monty G., Keith Jaggers, and Ted Robert Gurr. 2010. *Polity IV Project: Political Regime Characteristics and Transition, 1800–2010*. College Park, MD: University of Maryland.
Meltzer, Allan H., and Scott F. Richard. 1981. 'A Rational Theory of the Size of Government'. *The Journal of Political Economy* 89(5):914–27.
Mundlak, Yair. 1978. 'On the Pooling of Time Series and Cross Section Data'. *Econometrica: Journal of the Econometric Society* 46(1):69–85.
Nerlove, Marc. 1971. 'Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections'. *Econometrica: Journal of the Econometric Society* 39(2):359–82.
Nickell, Stephen. 1981. 'Biases in Dynamic Models with Fixed Effects'. *Econometrica: Journal of the Econometric Society* 49(6):1417–26.
Persson, Torsten, and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.
Pesaran, M. H., and R. Smith. 1995. 'Estimating Long-Run Relationships from Dynamic Heterogeneous Panels'. *Journal of Econometrics* 68(1):79–113.
Plummer, Martyn. 2003. 'Jags: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling'. Proceedings of the 3rd International Workshop on Distributed Statistical Computing, March 20, Vienna, Austria.
Plümper, Thomas, and Vera E. Troeger. 2007. 'Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects'. *Political Analysis* 15(2):124–39.
Plümper, Thomas, Vera E. Troeger, and Philip Manow. 2005. 'Panel Data Analysis in Comparative Politics: Linking Method to Theory'. *European Journal of Political Research* 44(2):327–54.
Robert, Christian. 2001. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. New York: Springer Verlag.
Ross, Michael L. 2006. 'Is Democracy Good for the Poor?'. *American Journal of Political Science* 50(4): 860–74.
Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley & Sons Inc.
Scheve, Kenneth, and David Stasavage. 2009. 'Institutions, Partisanship, and Inequality in the Long Run'. *World Politics* 61(2):215–53.
Sevestre, Patrick, and Alain Trognon. 1985. 'A Note on Autoregressive Error Components Models'. *Journal of Econometrics* 28(2):231–45.
Su, Yu-Sung, and Masanao Yajima. 2012. A Package for Running JAGS from R. R Package ver. 0.03-08. http://cran.r-project.org/web/packages/R2jags.
Tsai, Tsung-han, and Je Gill. 2012. 'Superdiag: A Comprehensive Test Suite for Markov Chain Nonconvergence'. *The Political Methodologist* 19(2):12–18.
Van Buuren, S., and K. Groothuis-Oudshoorn. 2011. 'Mice: Multivariate Imputation by Chained Equations in R'. *Journal of Statistical Software* 45(3):1–66.
Western, Bruce. 1998. 'Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach'. *American Journal of Political Science* 42(4):1233–59.
Wooldridge, Jerey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.
World Development Indicators. 2011. Washington DC: World Bank. http://data.worldbank.org/data-catalog/world-development-indicators.