

Change-Point Detection and Regularization in Time Series Cross-Sectional Data Analysis

Jong Hee Park¹ and Soichiro Yamauchi²

¹Department of Political Science and International Relations, IR Data Center, Seoul National University, Seoul, South Korea. E-mail: jongheepark@snu.ac.kr

²Department of Government, Harvard University, Cambridge, MA, USA. E-mail: syamauchi@g.harvard.edu

Abstract

Researchers of time series cross-sectional data regularly face the change-point problem, which requires them to discern between significant parametric shifts that can be deemed structural changes and minor parametric shifts that must be considered noise. In this paper, we develop a general Bayesian method for change-point detection in high-dimensional data and present its application in the context of the fixed-effect model. Our proposed method, hidden Markov Bayesian bridge model, jointly estimates high-dimensional regime-specific parameters and hidden regime transitions in a unified way. We apply our method to Alvarez, Garrett, and Lange's (1991, *American Political Science Review* 85, 539–556) study of the relationship between government partisanship and economic growth and Allee and Scalera's (2012, *International Organization* 66, 243–276) study of membership effects in international organizations. In both applications, we found that the proposed method successfully identify substantively meaningful temporal heterogeneity in parameters of regression models.

Keywords: Bayesian inference, change-point detection, regularization, shrinkage, high-dimensional data

1 Introduction

Many of the datasets encountered by political scientists in applied research take the form of repeated observations from a fixed number of subjects, the most well-known of which is time series cross-sectional (TSCS) data. When examining these time series data, researchers frequently run into the change-point problem because an underlying theory predicts that the effect of some variables will change or because researchers are concerned that unknown breaks will result in model misspecification and omitted variable bias. In any scenario, identifying change points in regression coefficients has a significant impact on substantive findings.

Detecting change points in regression parameters requires researchers to distinguish between major parametric shifts that can be interpreted as structural changes and minor parametric shifts that must be interpreted as noise. There are two statistical challenges that arise when attempting to identify major change points. First, two “unknowns” in the change-point problem (change points and regime-dependent parameters) must be *jointly* estimated. Separate estimates, such as data segmentation and separate model fitting, introduce the danger of overfitting or incorrect data splitting. Second, due to the possibility of rank deficiency in subsample data, parameter regularization must be considered in conjunction with the joint estimation of change points and regime-dependent parameters.

In this paper, we propose a new Bayesian method for joint estimation of change points and regime-specific regression parameters in high-dimensional data. Our proposed method combines Bayesian methods for parameter regularization, change-point detection, and variable selection. We first introduce the hidden Markov Bayesian bridge model (HMBB), which combines a Bayesian bridge model for parameter regularization with a hidden Markov model for multiple change-point detection. In this paper, we present HMBB in the context of TSCS data because TSCS data have been the core of dynamic model development in political science literature (Beck 2001; Beck *et al.* 1993;

Political Analysis (2023)
vol. 31: 257–277
DOI: [10.1017/pan.2022.23](https://doi.org/10.1017/pan.2022.23)

Published
7 October 2022

Corresponding author
Jong Hee Park

Edited by
Jeff Gill

© The Author(s) 2022. Published by Cambridge University Press on behalf of the Society for Political Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Beck and Katz 1995; Beck and Katz 2011; Box-Steffensmeier *et al.* 2014; Brandt and Freeman 2006; Hazlett and Wainstein 2022; Imai and Kim 2021; Pang, Liu, and Xu 2022; Western and Kleykamp 2004; Wucherpfennig *et al.* 2021).

Having said that, our work closely follows the development of change-point models in political science, economics, and statistics. In political science, Beck (1983) introduced the idea of change points as a special case of time-varying parameter models with a sharp change. Afterward, Western and Kleykamp (2004) elaborated on the benefits of using a Bayesian modeling method to the change-point problem. According to Western and Kleykamp, the Bayesian approach “combines the advantages of diagnostic and parametric approaches but addresses their limitations. . . . Like diagnostic methods, the Bayesian analysis treats the timing of change as uncertain and the location of a change point as a parameter to be estimated. . . . Like parametric models, the Bayesian model yields statistical inferences about regression coefficients. However, these inferences reflect prior uncertainty about the location of the change point that is unaccounted for in conventional models” (355). After Western and Kleykamp (2004), Spirling (2007b) developed the concept of Bayesian change-point modeling in the setting of a limited dependent variable based on Carlin, Gelfand, and Smith (1992). On the other hand, Park (2010, 2011b, 2012) extended Chib’s (1998) multiple-change-point model to binary, ordinal, count, and panel data cases. Blackwell (2018) relaxed the restriction of the fixed change-point numbers in over-dispersed count data models by employing Fox *et al.*’s (2011) hierarchical Dirichlet process approach. Furthermore, Kent, Wilson, and Cranmer (2022) presented a change-point detection method using a permutation-based parameter distribution.¹

We will discuss the change-point problem in regression models with a large number of predictors in the sections that follow. Following that, we provide a fixed-effects HMBB for TSCS data and describe our proposed estimation and model diagnostic procedures. We demonstrate the proposed method’s performance on simulated data. Then, we apply our method to Alvarez, Garrett, and Lange’s (1991) study of the relationship between government partisanship and economic growth, as well as Allee and Scalera’s (2012) study of membership effects in international organizations (IOs). Our proposed method is freely available as an open-source R package *BridgeChange* (<https://github.com/jongheepark/BridgeChange>).

2 Problem

The change-point problem in regression models with a large number of predictors is illustrated in Figure 1. The difficulty of detecting substantial from minor changes in time-varying characteristics is illustrated in panel (a). The ground truth shows two major shifts (vertical gray bars) in two groups of parameters (A and C). All parameters, however, have changed slightly. The number and location of major changes, as well as regime-specific parameter values, are the two main quantities of importance in the change-point analysis. We need a principled method to identify major changes (vertical gray bars) from minor changes (local fluctuations). It is also important to note that the regime shifts in panel (a) are abrupt, but not deterministic. Separate regression after data splitting (or using a period dummy regression model) does not adequately represent the data generating process with stochastic regime transitions.

Panel (b) in Figure 1 reveals a case of rank deficiency in the change-point analysis even when the pooled data have full rank. Despite the fact that the entire sample has $N > K$, where N denotes the number of observations and K denotes the number of predictors, one of the subsamples identified

¹ Change-point models have been widely used in applied work in political science such as the study of court (Hendershot *et al.* 2013; Pang *et al.* 2012; Park 2011a), terrorist attacks (Brandt and Sandler 2010; Santifort, Sandler, and Brandt 2013), congress (Smith *et al.* 1999; Wawro and Katznelson 2014), civil war (Cederman, Gleditsch, and Wucherpfennig 2017), war casualties (Spirling 2007a), and international conflict (Nieman 2015).

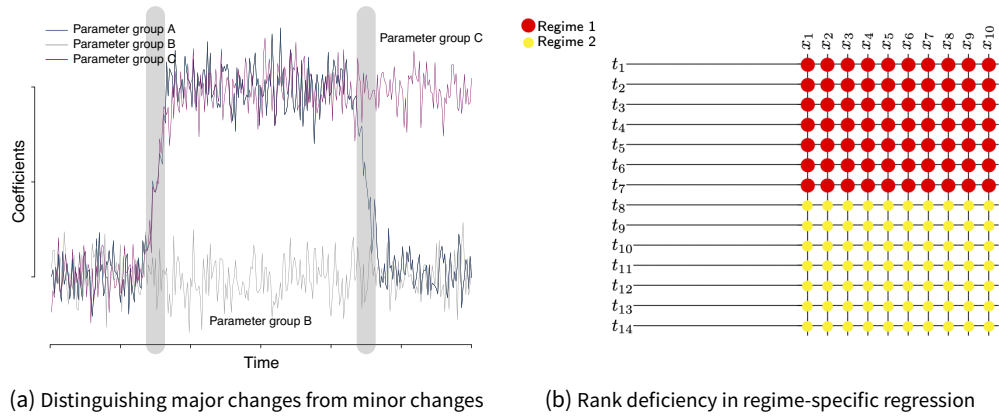


Figure 1. Illustration of the change-point problem in regression models with a large number of predictors.

by a hidden regime may have $N_m \leq K$, where N_m denotes the number of observations in regime m and K denotes the number of predictors.

In order to address the problems in Figure 1, we need a statistical method that combines change-point models with high-dimensional regression models. Recently, there has been a surge of high-dimensional change-point detection methods in frequentist approaches (e.g., Chan, Yau, and Zhang 2014; Frick, Munk, and Sieling 2014; Lee *et al.* 2018; Lee, Seo, and Shin 2016). Most of these methods focus on simple cases of high-dimensional change-point problems in which only a small subset of parameters are time-varying or the case under consideration is limited to a single-break case.

Our strategy is to take a full advantage of recent developments in regularization methods in the statistics literature (Carvalho, Polson, and Scott 2010; Chernozhukov *et al.* 2017; Fan and Li 2001; Hoerl and Kennard 1970; Park and Casella 2008; Polson 2012; Tibshirani 1996; Tibshirani *et al.* 2004; Zou and Hastie 2005). In particular, we emphasize that a Bayesian shrinkage approach to high-dimensional regression provides an effective framework for regularizing high-dimensional model parameters while also providing a credible measure of estimation uncertainty (Kyung *et al.* 2010; Polson and Scott 2010).

3 Method

We introduce our proposed method for examining change-point effects in regression models with a large number of predictors in this section. An example procedure for implementing the proposed method is as follows:

1. model specification based on a theory and available data,
2. model fitting using multiple HMBSs with a varying number of break points,
3. model diagnostic using the Watanabe–Akaike Information Criterion (WAIC),
4. posterior summary of hidden state transitions and time-varying parameters.

3.1 Bridge Estimator

We begin with the Bridge estimator for parameter regularization (Frank and Friedman 1993; Fu 1998). The Bridge estimator is motivated by the penalized likelihood formulation shown below:

$$\hat{\beta}_{\text{bridge}} = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \sum_{t=1}^n (y_t - \mathbf{x}_t^T \beta)^2 + \nu \sum_{j=1}^p |\beta_j|^\alpha \right\}, \tag{1}$$

where $0 < \alpha \leq 2$ (Frank and Friedman 1993; Fu 1998). The above formula has an interesting feature in that the popular lasso estimator and ridge regression can be obtained as special cases of

this estimator when $\alpha = 1$ and $\alpha = 2$, respectively. The variable selection feature of the bridge regression is asymptotic when $0 < \alpha \leq 1$ (Frank and Friedman 1993; Huang, Horowitz, and Ma 2008). Because of this generality, Equation (1) has garnered an increasing attention in the statistical literature (Armagan 2009; Fan and Li 2001; Huang *et al.* 2009; Huang *et al.* 2008; Liu *et al.* 2007; Polson, Scott, and Windle 2014).

We employ Polson *et al.*'s (2014) Bayesian approach of the bridge model in this research. A significant novelty in Polson *et al.*'s (2014) Bayesian method is the use of Lévy processes to create joint priors for β_j and local shrinkage parameters (λ_j). A combined prior distribution of the regression parameter b and the local shrinkage parameter $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_j)$ is represented as follows using scale mixes of normal representation:

$$p(\beta, \Lambda | \tau, \alpha) \propto \prod_{j=1}^p \exp\left(-\frac{\beta_j^2}{2\tau^2 \lambda_j}\right) p(\lambda_j), \tag{2}$$

where $p(\lambda_j)$ is the density of $2S_{\alpha/2}$ and S_{α} is the Lévy alpha-stable distribution.

Equation (2) increases the efficiency of the Markov chain Monte Carlo (MCMC) method by allowing posterior samples of local shrinkage parameters ($\Lambda = \text{diag}(\lambda_1, \dots, \lambda_j)$) to be sampled independently of global shrinkage parameter sampling (τ). The dependence of the sampling of the global shrinkage parameter on the sample of the local shrinkage parameter has been cited as a significant restriction of Bayesian shrinkage models (Hans 2010; Polson *et al.* 2014).

In the case of a linear regression model with β as regression slope parameters and σ^2 as a residual variance parameter, the posterior distribution of the Bayesian bridge model can be written as

$$\begin{aligned} p(\beta, \sigma^2, \Lambda, \alpha, v | \mathbf{y}, \mathbf{X}) &\propto p(\mathbf{y} | \beta, \sigma^2) p(\beta, \Lambda | \tau, \alpha) p(\sigma^2) p(\alpha) p(v) \\ &\propto \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)^\top (\mathbf{y} - \mathbf{X}\beta)\right] \prod_{j=1}^p \exp\left(-\frac{\beta_j^2}{2\tau^2 \lambda_j}\right) p(\lambda_j) \\ &\quad \times \left(\frac{1}{\sigma^2}\right)^{\frac{d_0}{2}+1} \exp\left(-\frac{b_0}{2\sigma^2}\right) v^{d_0-1} \exp(-d_0 v). \end{aligned} \tag{3}$$

3.2 HMBB

Now, we allow the Bayesian bridge model's parameters to vary in response to a hidden state transition. To be more precise, let \mathbf{S} denote a vector of hidden state variables where s_t is an integer-valued hidden state variable at t

$$\mathbf{S} = \{(s_1, \dots, s_n) : s_t \in \{1, \dots, M\}, t = 1, \dots, n\}, \tag{4}$$

and \mathbf{P} as a forward moving $M \times M$ transition matrix where \mathbf{p}_i is the i th row of \mathbf{P} and M is the total number of hidden states. For example, if we assume a single break, $M = 2$ and \mathbf{P} is a 2×2 transition matrix. For efficient sampling of hidden states, we adopt Chib's (1998) non-ergodic transition of hidden states where a hidden state variable starts from state 1 and moves forward to the terminal state (M).

From the above description, we can develop a fixed-effects HMBB for TSCS data using the group-demeaned data ($\bar{\mathbf{y}}, \bar{\mathbf{x}}$):

$$\bar{y}_{it} = \begin{cases} \bar{\mathbf{x}}_{it}^\top \beta_1 + \varepsilon_{it}, & \varepsilon_{it} \sim \mathcal{N}(0, \sigma_1^2) \text{ for } t_0 \leq t < \tau_1 \\ \vdots & \vdots \\ \bar{\mathbf{x}}_{it}^\top \beta_M + \varepsilon_{it}, & \varepsilon_{it} \sim \mathcal{N}(0, \sigma_M^2) \text{ for } \tau_{M-1} \leq t < T \end{cases} \tag{5}$$

where τ_m is the break point between regime $m - 1$ and regime m . In order to allow many predictors, we use the Bayesian bridge prior for β as shown in Equation (2). The posterior distribution of the resulting model is

$$\begin{aligned} \rho(\beta, \sigma^2, \Lambda, \alpha, \tau, \mathbf{P} | \bar{\mathbf{y}}, \bar{\mathbf{x}}) &= \int \rho(\bar{\mathbf{y}}_1 | \bar{\mathbf{x}}_1, \beta, \sigma_1^2, \Lambda_1, \alpha_1, \tau_1) \\ &\times \prod_{t=2}^T \sum_{m=1}^M \rho(\bar{\mathbf{y}}_t | \bar{\mathbf{Y}}_{t-1}, \bar{\mathbf{X}}_{t-1}, \beta_m, \sigma_m^2, \Lambda_m, \alpha_m, \tau_m, \mathbf{P}) \\ &\times \rho(s_t = m | s_{t-1}, \beta, \sigma^2, \Lambda, \alpha, \tau, \mathbf{P}) \rho(\mathbf{P}) \rho(\beta, \Lambda) \rho(\sigma^2) \rho(\alpha) \rho(\tau) d\mathbf{S}, \end{aligned} \tag{6}$$

where $\bar{\mathbf{y}}_t = (\bar{y}_{1t}, \dots, \bar{y}_{nt})$ and $\bar{\mathbf{x}}_t = (\bar{x}_{1t}, \dots, \bar{x}_{nt})$. $\bar{\mathbf{Y}}_{t-1}$ and $\bar{\mathbf{X}}_{t-1}$ indicate all the group-demeaned data up to $t - 1$. The subscript m in model parameters $(\beta, \sigma^2, \Lambda, \alpha, \nu)$ indicates hidden states it belongs to.²

If we ignore the Markov property of hidden regimes, we can simplify Equation (6) as

$$\rho(\Theta | \mathbf{D}) = \prod_{t=1}^T \sum_{m=1}^M \rho(s_t = m) \times \rho(\Theta_m | \mathbf{D}_t)$$

by setting $\mathbf{D}_t = (\bar{\mathbf{y}}_t, \bar{\mathbf{x}}_t)$ and $\Theta = (\beta, \sigma^2, \Lambda, \alpha, \nu)$. The posterior distribution takes a form of a mixture distribution. From this, it becomes clear that HMBB estimates can be considered as weighted averages of Bayesian bridge regression model estimates fitted to subsets of data partitioned by known change points. Because we do not know the location and number of change points in reality, the hidden state variable is added as a latent variable and sampled from data in HMBB. The sampling algorithm is discussed in Appendix A.

3.3 Model Diagnostics using WAIC

After fitting multiple HMBBs with a varying number of breaks (or different model specifications), researchers must assess the model's fit to observed data. Model checking is a critical step in Bayesian analysis in general. This is especially true in the case of change-point analysis, where the break points are unknown.

We recommend the WAIC for model diagnostics of HMBB because of its low computational cost.³ WAIC is a fully Bayesian estimate of model uncertainty. WAIC approximates the expected

2 A random-effects HMBB can be written in a similar way by letting parameters of the random-effects model vary across subjects:

$$y_{it} = \begin{cases} \mathbf{x}_{it}^T \beta_1 + \mathbf{w}_{it}^T \mathbf{b}_i + \varepsilon_{it}, & \mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}_1), \varepsilon_{it} \sim \mathcal{N}(0, \sigma_1^2) & \text{for } t_0 \leq t < \tau_1 \\ \vdots & \vdots & \vdots \\ \mathbf{x}_{it}^T \beta_M + \mathbf{w}_{it}^T \mathbf{b}_i + \varepsilon_{it}, & \mathbf{b}_i \sim \mathcal{N}(0, \mathbf{D}_M), \varepsilon_{it} \sim \mathcal{N}(0, \sigma_M^2) & \text{for } \tau_{M-1} \leq t < T \end{cases}$$

Then, the Bayesian bridge prior is used as the prior distribution of β . The posterior distribution of the resulting model will take the following form:

$$\begin{aligned} \rho(\beta, \mathbf{D}, \sigma^2, \Lambda, \alpha, \tau, \mathbf{P} | \mathbf{y}, \mathbf{X}, \mathbf{W}) &= \int \rho(\mathbf{y}_1 | \mathbf{X}_1, \mathbf{W}_1, \beta_1, \mathbf{b}_1, \mathbf{D}_1, \sigma_1^2, \Lambda_1, \alpha_1, \tau_1) \\ &\prod_{t=2}^T \sum_{m=1}^M \rho(\mathbf{y}_t | \mathbf{Y}_{t-1}, \mathbf{X}_{t-1}, \mathbf{W}_{t-1}, \beta_m, \mathbf{b}_i, \mathbf{D}_m, \sigma_m^2, \Lambda_m, \alpha_m, \tau_m, \mathbf{P}) \\ &\rho(s_t = m | s_{t-1}, \beta, \mathbf{b}_i, \mathbf{D}, \sigma^2, \Lambda, \alpha, \tau, \mathbf{P}) \\ &\rho(\mathbf{P}) \rho(\beta, \Lambda) \rho(\mathbf{D}) \rho(\sigma^2) \rho(\alpha) \rho(\tau) d\mathbf{b}_i d\mathbf{S}. \end{aligned} \tag{7}$$

Although we do not discuss a random-effects HMBB in this paper, it is available in `BridgeChange`.

3 We also examined the approximate log marginal likelihood (Chib 1995). However, in our test, the approximate log marginal likelihood of HMBB does not show a satisfactory result. The software to compute the approximate log marginal likelihood of HMBB is available in `BridgeChange`. We discuss the computational algorithm to estimate the approximate log marginal likelihood of HMBB in the Supplementary Material.

log pointwise predictive density by subtracting a bias for the effective number of parameters from the sum of log pointwise predictive density. Using Gelman, Hwang, and Vehtari’s (2014) formula, a WAIC of HMBB with M latent states (\mathcal{M}_M) is

$$\text{WAIC}_{\mathcal{M}_M} = -2 \left(\underbrace{\sum_{t=1}^n \log \left[\frac{1}{G} \sum_{g=1}^G p(y_t | \beta^{(g)}, \sigma^{2,(g)}, \mathbf{A}^{(g)}, \alpha^{(g)}, \tau^{(g)}, \mathbf{P}^{(g)}, \mathcal{M}_M) \right]}_{\text{the expected log pointwise predictive density}} - \underbrace{\sum_{t=1}^n V_{g=1}^G \left[\log p(y_t | \beta^{(g)}, \sigma^{2,(g)}, \mathbf{A}^{(g)}, \alpha^{(g)}, \tau^{(g)}, \mathbf{P}^{(g)}, \mathcal{M}_M) \right]}_{\text{bias for the effective number of parameters}} \right),$$

where G is the MCMC simulation size, $V[\cdot]$ indicates a variance, and $\theta^{(g)}$ are the g th simulated outputs for θ .

4 Simulation Study

4.1 Simulated Data

We construct 24 sets of TSCS data with varied group sizes ($n = (10, 20)$), time lengths ($t = (30, 60)$), predictor sizes ($k = (20, 30)$), and break numbers ($m = (0, 1, 2)$) to evaluate the validity of our proposed method. We set $\mathbf{x}_t \sim \mathcal{N}_k(0, \mathbf{I}_k)$ and

$$\begin{cases} \beta_1 \sim \mathcal{N}(2, 1), & \text{for } t_0 \leq t < \tau_1, \\ \beta_2 \sim \mathcal{N}(-2, 1), & \text{for } t_1 \leq t < \tau_2 \text{ if } m \geq 1, \\ \beta_2 \sim \mathcal{N}(2, 1), & \text{for } t_2 \leq t < \tau_3 \text{ if } m \geq 2. \end{cases}$$

From this, we generate \mathbf{y}_t as

$$\mathbf{y}_t = \underbrace{\mathbf{x}_t \beta_{s_t}}_{\text{systematic component}} + \underbrace{\alpha + \text{CHOL}(\sigma_{s_t} \omega_t \mathbf{I}_t)}_{\text{stochastic component}} \epsilon_t,$$

$\alpha \sim \mathcal{N}(0, 5)$, (individual effects)
 $\sigma_M = (\sqrt{2}, \sqrt{3}, \sqrt{2})$, (state-level heterogeneity)
 $\omega_t \sim \mathcal{N}(0, 1)$, (contemporaneous shocks)
 $\epsilon_t \sim \mathcal{N}(0, 1)$, (observation error)

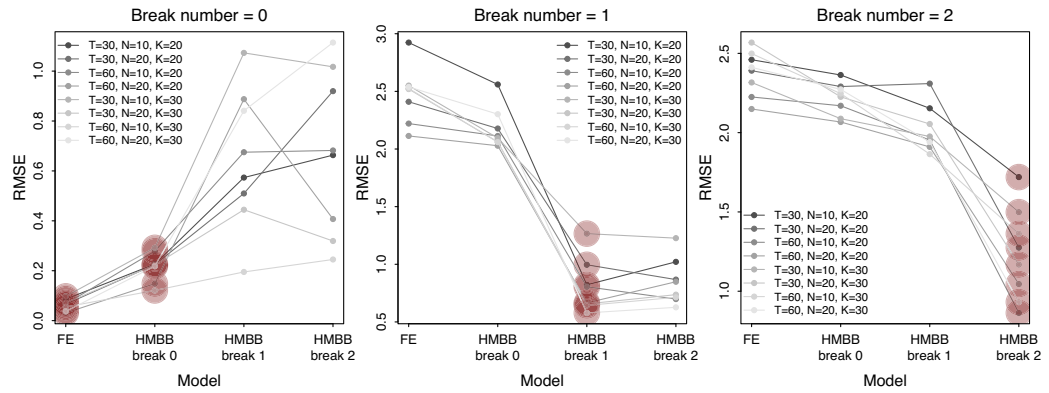
where CHOL indicates the Cholesky factorization. That is, the observed data are generated by the systematic component, time-invariant group-level factors, time-varying contemporaneous shocks, and time-varying observation error.

The fixed-effects HMBB of BridgeChange uses pre-transformed data for the input. The pre-transformation is done by `p1m` package in R (Croissant and Millo 2008). For the one-way fixed-effects HMBB, the data are transformed into group-centered (either by time or group) data. For the two-way fixed-effects HMBB, the data are transformed into doubly group-centered data.

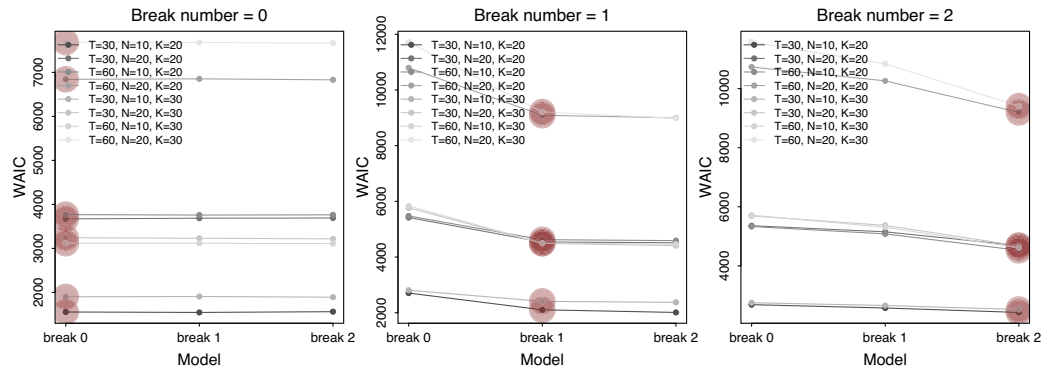
4.2 Simulation Results

Figure 2 summarizes the results of the simulation. Panel (a) shows the root-mean-square errors (RMSEs) of the frequentist fixed-effects model using `p1m` package (FE) and three fixed-effects HMBBs with break numbers of 0, 1, and 2 (HMBB breaks 0, 1, and 2). The formula for RMSE is

$$\text{RMSE} = \sqrt{K^{-1} T^{-1} \sum_{k=1}^K \sum_{t=1}^T (\hat{\beta}_{k,t} - \beta_{k,t}^{\text{true}})^2}.$$



(a) RMSE: True states are marked by brown circles.



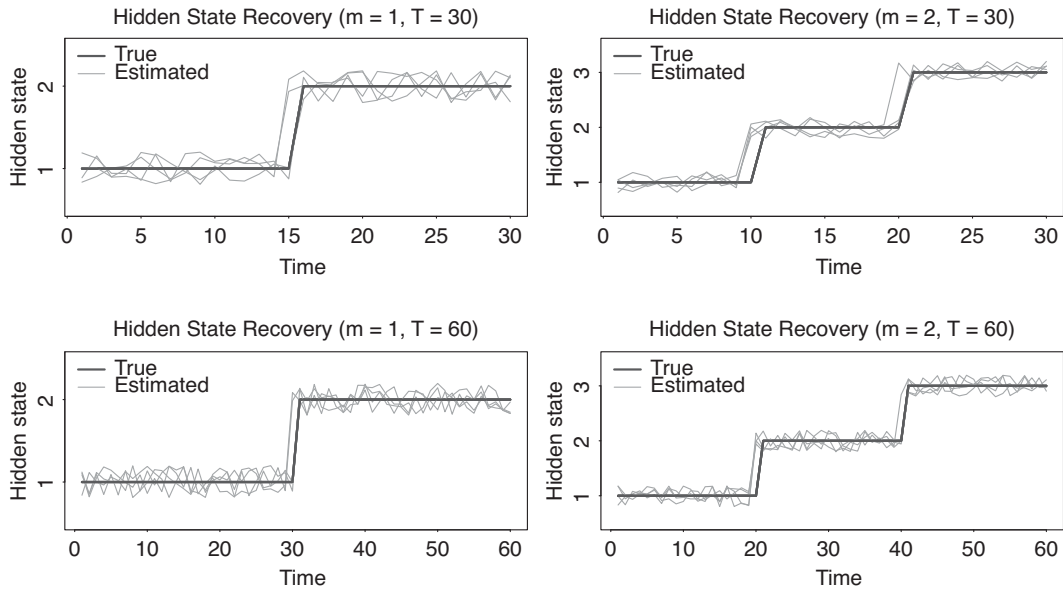
(b) WAIC: True states are marked by brown circles.

Figure 2. Simulation outcomes from 24 sets of TSCS data. The brown circles indicate the true break numbers. Panel (a) indicates the root-mean-square error of time-varying parameters. Panel (b) is WAIC. A lower WAIC score indicates a good predictive accuracy.

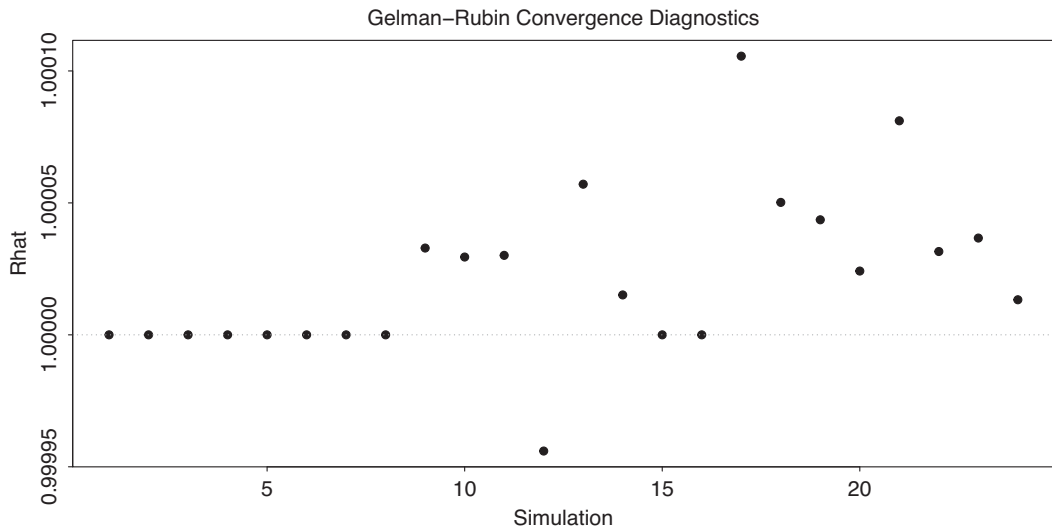
The brown circles indicate the true break numbers in panel (a). When the true break value is 0 (left), the fixed-effects model and the fixed-effects HMBB with no break have the lowest RMSEs. When the true break number is 1 (center), the RMSEs of no-break models (FE or HMBB break 0) are much greater than those of HMBB with multiple breaks, indicating a poor model fit. Panel (b) in Figure 2 also shows that WAIC scores successfully identify true models (brown circles).

One interesting pattern in Figure 2 is the *over-detection* of hidden states by both RMSE and WAIC. When the ground truth is time series data with a single break (the middle column), RMSEs and WAIC scores sometimes favor two-break models over one break models, which are closer to the ground truth. RMSEs and WAIC scores do not, however, favor models with fewer breaks than the ground truth. That is, there is no sign of under-detection of hidden states by HMBB.

Generally speaking, under-detection is a significantly more problematic issue than over-detection. Figure 4 illustrates this point. We compare the simulation results with one break (a) and two breaks (b) for $n = 20, t = 60,$ and $k = 30$. The ground truth of panel (a) is time series data with a single break, and the ground truth of panel (b) is time series data with two breaks. Thus, the right column of panel (a) shows the case of the hidden state over-detection, and the left column of panel (b) shows the case of the hidden state under-detection. The left column of panel (a) and the right column of panel (b) show the cases where the number of hidden states is correctly assumed. It is clear that substantive results obtained using either of the two models in panel (a) are self-evidently similar, whereas substantive results obtained using the two models in panel (b) are vastly different. The under-detected model (the left column of panel (b)) fails to capture an upward parameter shift in $t > 40$.



(a) Hidden state recovery: Bright thin lines are estimates.

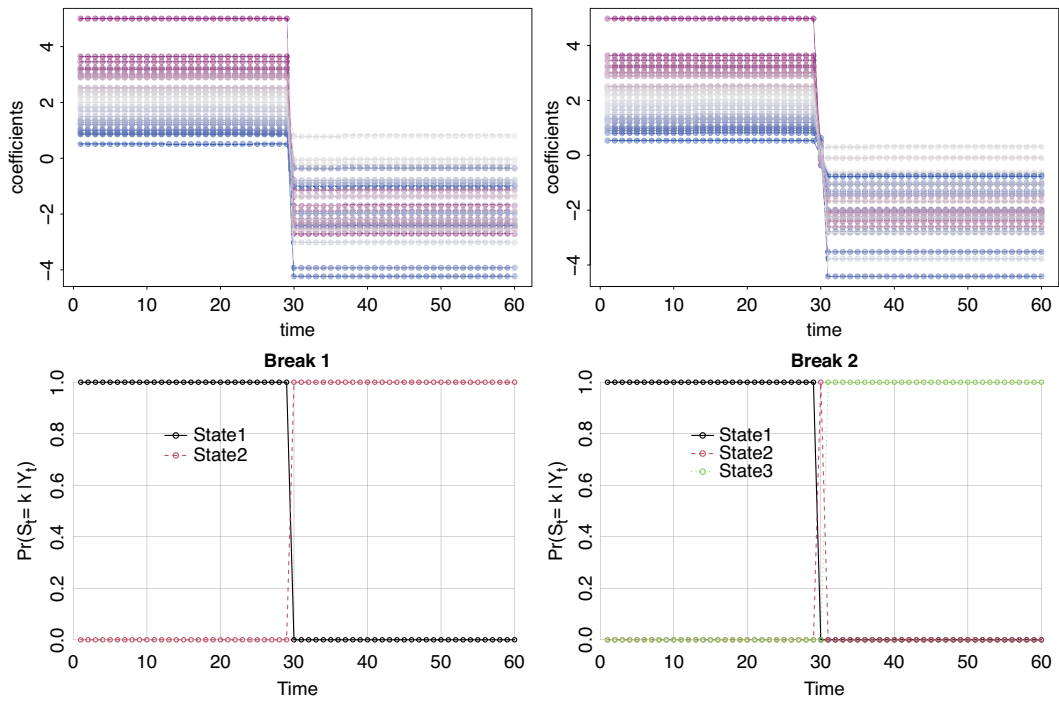


(b) Convergence diagnostics using stabilized Gelman-Rubin statistics

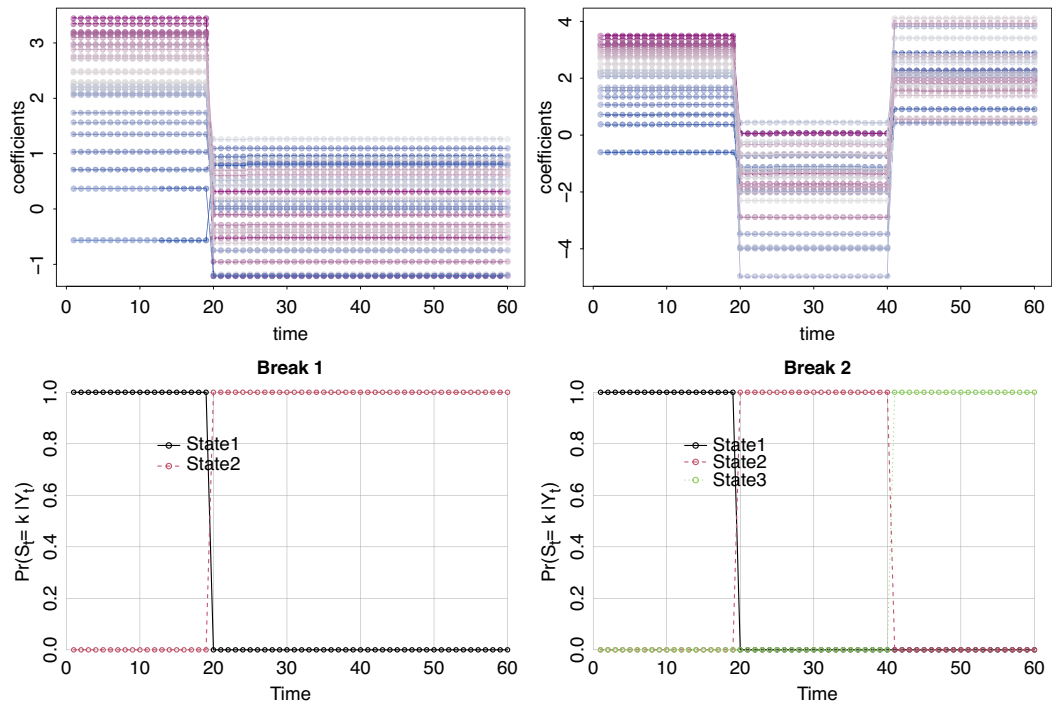
Figure 3. Simulation outcomes from 24 sets of TSCS data. Panel (a) shows recovered hidden states (gray) over true states (black). We jittered hidden state estimates for easy comparison. Panel (b) is a stabilized Gelman-Rubin statistics (Vats and Knudson 2021). The values close to 1 indicate good convergence.

Figure 3 shows additional simulation results. Panel (a) compares recovered hidden states (bright thin lines) with the ground truth (thick lines). Panel (a) clearly demonstrates that HMBB successfully uncovers hidden state structures under various setups. Panel (b) shows convergence diagnostics of the simulation results using a stabilized Gelman-Rubin statistics (Vats and Knudson 2021). The values that are close to 1 indicate good convergence of Markov chains. All Gelman-Rubin statistics are close to 1.

To summarize, Bayesian model diagnostics with WAIC give a solid framework for avoiding the problem of under-detection. However, WAIC frequently exposes researchers to the problem of over-detection. Comparing hidden state transitions between multiple models, as we did in Figure 4, is the best way to check for the over-detection problem. We will cover a more practical guidance in greater detail when we examine applications.



(a) The ground truth is one break. The right plot is from the over-detected model.



(b) The ground truth is two breaks. The left plot is from the under-detected model.

Figure 4. Over-detection (a) and under-detection (b) of hidden states: The top plot in each panel shows the posterior estimates of time-varying parameters, which is computed by $p(\beta_{k,t} | \mathbf{y}) = \sum_{m=1}^M p(\beta_k, s_t = m | \mathbf{y})$. The bottom plot shows hidden state probabilities ($p(s_t | \mathbf{y})$). The data are simulated from $n = 20$, $t = 60$, and $k = 30$.

5 Applications

In this section, we apply our proposed method to two studies in political science. The first example is Alvarez *et al.*'s (1991) study of partisan sources of economic growth, and the second example is Allee and Scalera's (2012) study on the IO membership effects.

Table 1. WAIC scores of HMBBs on Alvarez *et al.* (1991): The estimation is based on 10,000 MCMC runs after discarding the first 10,000 MCMC runs.

Model	Break 0	Break 1	Break 2
No interaction	678	665	663
Partial interaction	668	646	642
Full interaction	670	664	654

5.1 Alvarez *et al.*'s (1991) Study of Partisan Politics and Economic Growth

Alvarez *et al.* (1991) investigate how government partisanship and the level of labor union centralization affect economic growth in advanced countries using a longitudinal dataset of 16 OECD countries. They discovered that centralized labor organizations have a conditional effect on economic growth: Centralized labor organizations are “conducive to better economic performance when the Left was politically powerful.” In contrast, weaker union movements “had desirable consequences for growth and inflation when governments were dominated by rightist parties” (551). Since then, the growth-promoting effect of left-party government and inclusive labor has received ongoing attention in comparative political economy literature (Beck *et al.* 1993; Boix 1997; Franzese Jr. 2002; Garrett 1998; Rueda 2008; Scruggs 2001; Soskice and Iversen 2000; Western 1998).

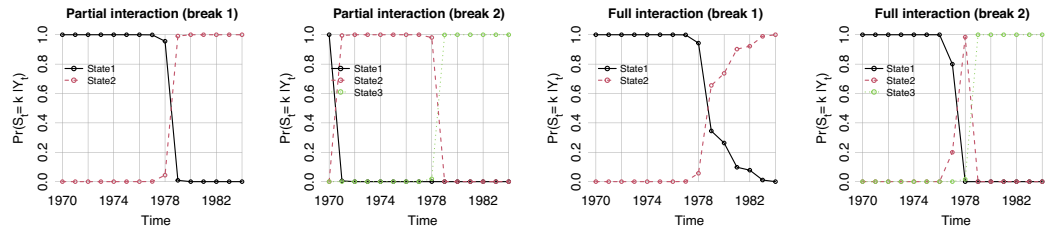
The annual growth rate observed at country i and year t is the dependent variable of Alvarez *et al.* (1991). The independent variables are lagged growth rate (`lagg1`), weighted OECD demand (`opengdp`), weighted OECD export (`openex`), weighted OECD import (`openimp`), cabinet composition of left-leaning parties (`leftc`), and the degree of labor organization encompassment (`central`).⁴ We add a year-fixed effect to Alvarez *et al.*'s (1991) partial interaction model,

$$y_{it} = \alpha_t + \beta_1 \text{opengdp}_{it} + \beta_2 \text{openimp}_{it} + \beta_3 \text{openex}_{it} + \beta_4 \text{left}_{it} + \beta_5 \text{central}_{it} + \beta_6 \text{leftc}_{it} \times \text{central}_{it} + \beta_7 \text{lagg1} + \varepsilon_{it}.$$

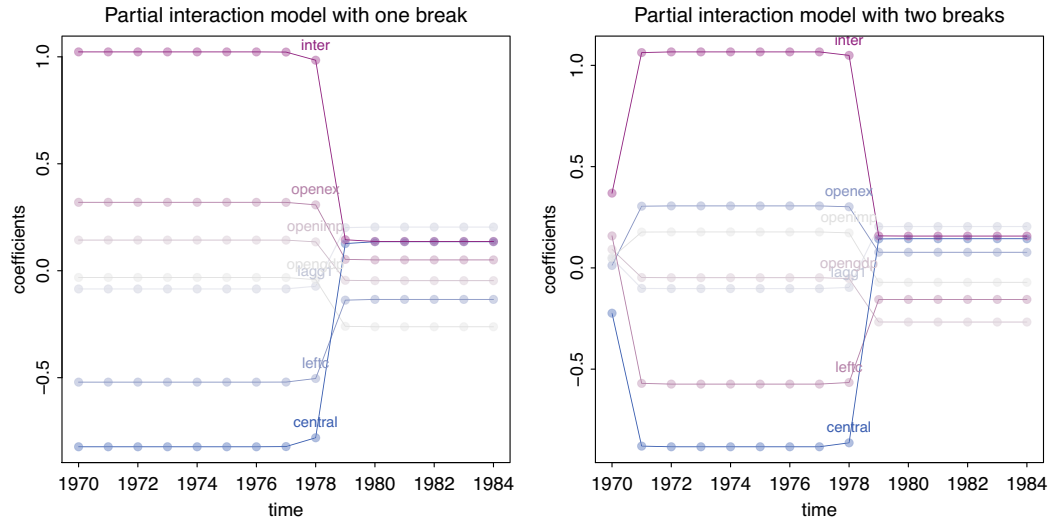
Our main goal in the replication is to check whether the key explanatory variables (`central`, `leftc`, and `leftc × central`) have time-varying effects during the sample period. We consider three different model specifications (no interaction model, a partial interaction model, and a full interaction model) using the original input variables of Alvarez *et al.* (1991). Then, given the short duration of the TSCS data ($T = 15$), we set the top limit of breaks to two for each model specification, providing nine models to compare. The panel method employed is the one-way year fixed effects. The country fixed effects are not used due to the time-invariant predictor (`central`).

The WAIC scores of the tested models are listed in Table 1. The partial interaction model with two breaks has the lowest WAIC score (642), followed by the same model with one break (646). However, as illustrated in the center-left in panel (a) of Figure 5, the initial regime of the two-break partial interaction models lasts only one year and appears to be redundant given the absence of a comparable pattern in the other models. Panel (b) of Figure 5 compares posterior estimates of time-varying parameters of the single-break partial interaction HMBB (left) with the two-break partial interaction HMBBs (right). Except the starting point difference, the two models produce almost identical posterior estimates of time-varying parameters. As a result, we will interpret the data using the single-break partial interaction model.

⁴ For the replication of Alvarez *et al.* (1991), we use the `ag1` dataset in the R package `pcse` (Bailey and Katz 2011). Alvarez *et al.*'s (1991) original dataset begins in 1969, whereas `ag1` dataset begins in 1970. The variable names shown in this section are the same as in `pcse`.



(a) Hidden state probabilities ($p(s_t | \mathbf{y})$) of four HMBB models based on Alvarez, Garrett, and Lange (1991): A common break point across all four models is between 1978 and 1979.



(b) Posterior estimates of time-varying parameters in the partial interaction model ($p(\beta_{k,t} | \mathbf{y})$) based on Alvarez, Garrett, and Lange (1991): The left plot is generated using a single break HMBB and the right plot is drawn using a two break HMBB.

Figure 5. Hidden state transitions and time-varying movements of parameters in Alvarez *et al.*'s (1991) partial interaction model.

Panel (b) of Figure 5 shows the posterior estimates of time-varying parameters for a single-break HMBB (left) and a two-break HMBB (right). As expected, the left plot, generated using a single-break HMBB, exhibits a pattern identical to that of the right plot, generated using a two-break HMBB except the initial regime in the right plot.

Last, we compare HMBB estimates with conventional fixed-effects estimates in Figure 6. Several interesting patterns are worth noting. First, conventional fixed-effects estimates take the form of weighted averages of regime-specific HMBB values for certain covariates (e.g., *central*, *inter*, *lagg1*, and *leftc*), but not for others. Second, the interaction term (*inter*) and its two constituent terms (*central* and *leftc*) exhibit the most pronounced parametric change, which adds an interesting twist to Alvarez *et al.*'s (1991) original conclusion. *Only until 1978 did the left party government have a growth-promoting influence in the presence of centralized labor organizations.* After that, the effect waned significantly, signaling the start of a new era, the era of neo-liberal reform (Frieden 2020; Helleiner 1994).

5.2 Allee and Scalera's (2012) Study of Membership Effects in International Organizations

In our second example, we revisit Allee and Scalera's (2012) study on the divergent effects of membership in IOs from 1950 to 2006. Rose (2003) was the first to challenge the conventional wisdom about the GATT/trade-promoting WTO's effects. Rose concluded from an analysis of bilateral trade data spanning 175 countries and 50 years that "the GATT/WTO seems to have a huge effect on trade if one does not hold other things constant; the multilateral trade regime

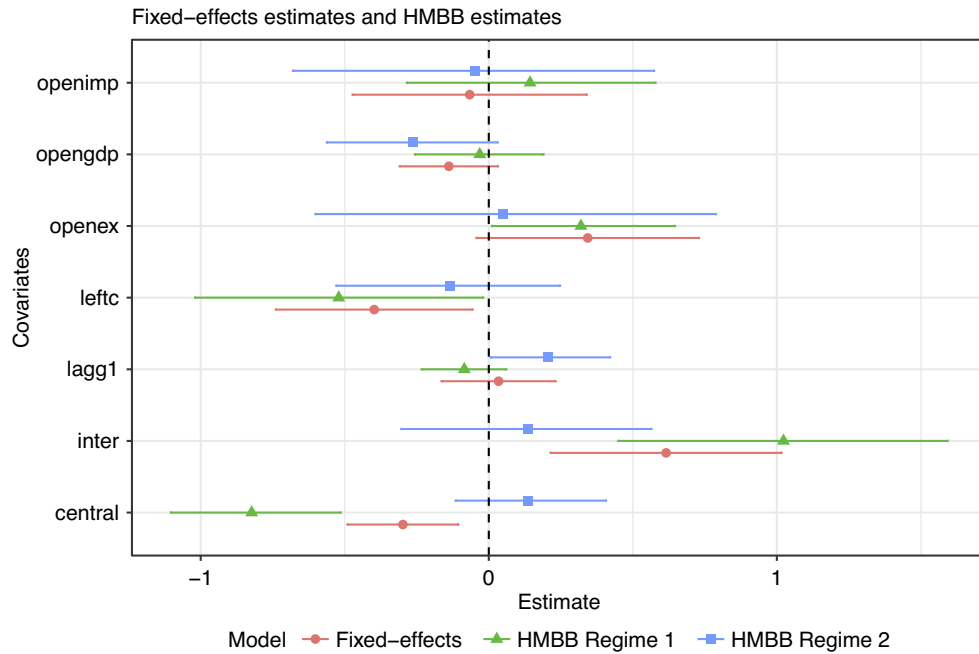


Figure 6. Comparison of parameter estimates: Fixed-effects estimates are obtained by the least squares method and the panel robust standard error of MacKinnon and White (1985). HMBB estimates are obtained from a single-break model. The detected break point is between 1978 and 1979.

matters, *ceteris non paribus*” (emphasis original, 111). This discovery sparked a flood of following studies amending or questioning the null effect (e.g., Goldstein, Rivers, and Tomz 2007; Gowa and Kim 2005; Park 2012; Subramanian and Wei 2007; Tomz, Goldstein, and Rivers 2007). Allee and Scalera (2012) is one of recent amendments that settles conflicting evidence regarding the effects of GATT/WTO membership on trade.

The key explanatory variable in Allee and Scalera (2012) is the type of accession, which is classified into three categories: early accession, automatic accession, and rigorous accession. According to Hypothesis 1 in Allee and Scalera (2012), rigorous accession must have a greater trade-promoting effect than other types of accession because “the more rigorous a state’s accession to an IO, and thus the more policy changes required to join, the greater the benefits it will receive from membership” (243). Temporal heterogeneity is one of main concerns in Allee and Scalera (2012). They argue “although rigorous and early joiners should benefit from membership, we [Allee and Scalera] expect those benefits to be most pronounced in the years after accession and to fade over time” (260). They deal with the temporal effect heterogeneity by employing a “counter” variable that counts how many years have passed.

Our goal of replication is to check temporal heterogeneity in the key explanatory variables of Hypothesis 1 in Allee and Scalera (2012) using HMBB.⁵

⁵ More specifically, we chose Column (5) and Column (6) models in Table 4 of the original paper. In these two models, the outcome variable is total national trade (imports and exports). The model formula of Column (5) model is

$$y \sim \lnpop1 + gled_{gdppc} + totalcont + polity + domestic1_9 + rigorous + colonial + earlymem.$$

The model formula of Column (6) model is

$$y \sim \lnpop1 + gled_{gdppc} + totalcont + polity + domestic1_9 + rigorous + rigorouscounter + colonial + colonialcounter + earlymem + earlymemcounter.$$

Table 2. WAIC scores of HMBBs on Allee and Scalera's (2012) Column (5) model and Column (6) model: The estimation is based on 10,000 MCMC runs after discarding the first 10,000 MCMC runs.

Formula	Break 0	Break 1	Break 2
Column (5)	23,671	23,301	22,272
Column (6)	23,459	22,881	22,078

In words, we examine whether the effect of accession type on a country's total national trade may vary over time as a result of economic shocks, decaying effects of accession types, or network effects of IO membership. Due to the significant fraction of missing observations in the original data, we use Ranjit's (2016) imputed data. The specification for the panel is identical to that in the original publication (a two-way fixed-effects at the country and year level).

The results of Bayesian model diagnostics using WAIC for the two models in our replication are summarized in Table 2. The two-break Column (6) model has the lowest WAIC score (22,078) among the six models in comparison. Given the possibility of the over-detection, we further examine hidden state transitions and time-varying movements of parameters in Figure 7.

Panel (a) of Figure 7 clearly shows that either 1964 or 2007 is estimated as change points across different HMBB specifications. Panel (b) demonstrates that a single-break HMBB of Column (6) shows a starkly different picture from a two-break HMBB of Column (6). Compare the two covariates at the top in panel (b). While the effect of (log) population (`lnpop1`) is not affected by hidden regime transitions, the effect of the level of economic development (`gled_gdppc`) shows dramatic shifts over time.

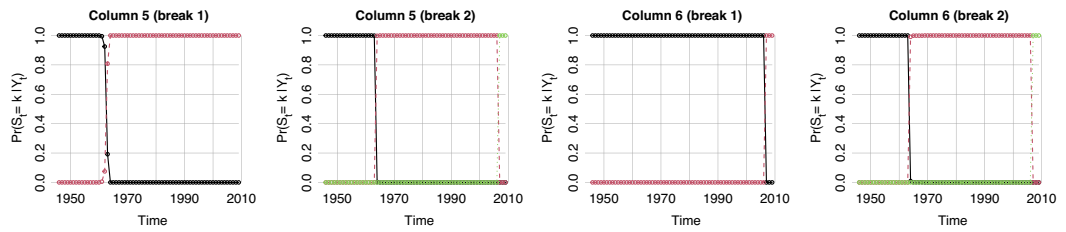
Figure 8 compares parameter estimates of the conventional fixed effects with HMBB estimates. Several interesting patterns are worth noting. First, the two key explanatory variables (`rigorous` and `rigorouscounter`) show strong time-dependent effects. In Regime 1 (1946–1964), the marginal effect of `rigorous` accession on country's total national trade is statistically indistinguishable from 0. This is very much in accordance with reality. Between 1946 and 1954, the proportion of sample countries with the status of rigorous accession is zero. After 1955, the fraction gradually grew, and by 1964, only 12 countries (1%) had achieved rigorous accession. This historical pattern in the effect of rigorous accession is unobservable using traditional fixed-effect estimation methods.

Second, a substantial decline in `rigorouscounter`, the interaction of `rigorous` with its “counter,” following the second break (2007) can be attributed to the economic crisis of 2008–9. The greater negative trend in `rigorouscounter` following 2007 implies that countries that have a lengthy history of rigorous admission have borne the brunt of the economic crisis.

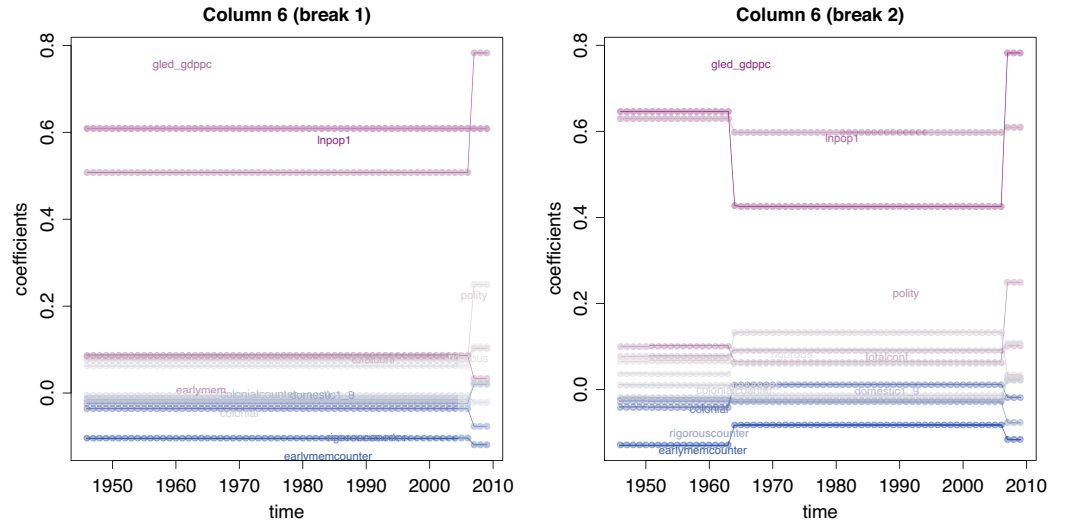
Last, as illustrated in Figure 6, there is no consistent pattern in the association between conventional fixed-effects and HMBB estimates. In some circumstances (e.g., `gled_gdppc`), the

Explanations for the included predictors are as follows:

- `lnpop1`: log of population,
- `gled_gdppc`: Gross Domestic Product (GDP), per capita,
- `totalcont`: total shared borders, or contiguity,
- `polity`: Democracy measure from the Polity dataset,
- `domestic1_9`: a weighted measure of domestic political conflict,
- `rigorous`: equals 1 if a current member joined the GATT/WTO via rigorous accession procedures, and 0 otherwise.
- `rigorouscounter`: number of years since a current rigorous-accession member joined the GATT/WTO via rigorous accession,
- `colonial`: equals 1 if a current member joined the GATT/WTO via Article 26:5(c), or similar post-colonial accession norms, and 0 otherwise,
- `colonialcounter`: number of years since a current colonial-accession member joined the GATT/WTO via colonial accession,
- `earlymem`: equals 1 if a current member joined the GATT during the early years of the trade regime, and 0 otherwise,
- `earlymemcounter`: number of years since an early member joined the GATT/WTO.



(a) Hidden state probabilities ($p(s_t | \mathbf{y})$) of four HMBB models based on Allee and Scalera (2012): The two estimated break points of the last model (favored by WAIC) are 1964 and 2007.



(b) Posterior estimates of time-varying parameters in the partial interaction model ($p(\beta_{k,t} | \mathbf{y})$) based on Allee and Scalera (2012)

Figure 7. Hidden state transitions and time-varying movements of parameters in Allee and Scalera (2012)

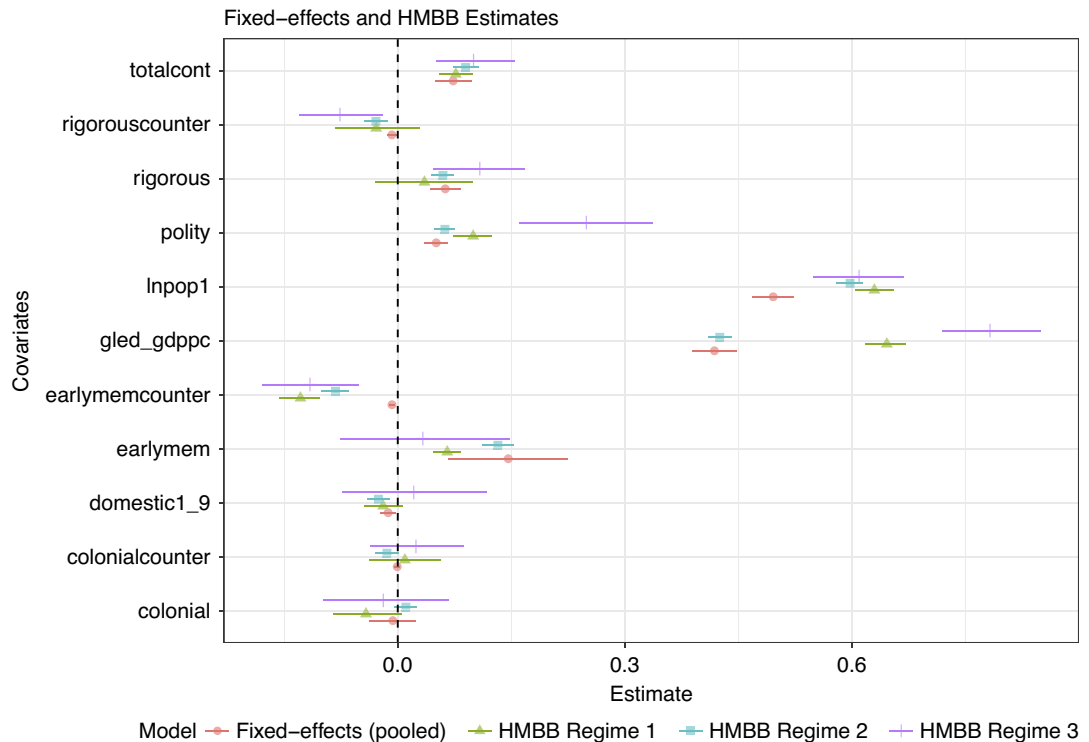


Figure 8. Comparison of parameter estimates based on Allee and Scalera (2012): Fixed-effects estimates are obtained by the least-squares method. HMBB estimates are obtained from a two-break model. The detected break points are 1964 and 2007.

conventional fixed-effects estimates are close to the regime 2 estimates, but not in others (e.g., `lnpop1` and `earlymemcounter`).

6 Discussion

A typical way to use our method for TSCS data will be as follows. First, researchers build a set of regression models for the change-point analysis. Next, researchers fit HMBBs with a varying number of breaks ranging from 0 to an upper limit researchers deem ideal given the model, data, and theory. Researchers make an informed decision regarding the best-fitting model by analyzing the WAIC scores, hidden state transitions, and parameter changes of several HMBBs.

Once researchers have selected an appropriate HMBB using the above procedure, they will have $K \times M$ regime-specific regression coefficients to analyze. When either K or M is large, it might be difficult to interpret all the time-varying information. Although Bayesian shrinkage approaches outperform variable selection methods such as spike and slab prior models in terms of computational efficiency, they lack the sparsity property. That is, in Bayesian shrinkage approaches, parameter estimates are never exactly zero. Thus, it would be beneficial for researchers to discriminate between “strong” (i.e., statistically distinct from 0) and “weak” (i.e., statistically indistinguishable from 0) signals across all $K \times M$ parameters.

In this scenario, we can employ the decoupled shrinkage and selection (DSS) method that minimizes an ℓ_0 -type loss function on pre-regularized posterior distributions (Hahn and Carvalho 2015). The DSS loss function of the HMBB for regime m can be constructed as follows using HMBB estimations as shrinkage inputs:

$$\mathcal{L}(\gamma_m) = \arg \min_{\gamma_m} \underbrace{\|\mathbf{X}_m \beta_m^* - \mathbf{X}_m \gamma_m\|_2^2}_{\text{squared prediction loss}} + \underbrace{\lambda \|\gamma_m\|_0}_{\text{parsimony penalty}}, \quad (8)$$

where $\mathbf{X}_m \beta_m^*$ is the fitted value of the fixed-effects HMBB at regime m . λ is a nonnegative regularization parameter, and γ is the new slope parameters that minimize the loss function.

To find the optimum of Equation (8), we take the popular approach of the ℓ_1 surrogation of the ℓ_0 problem. Furthermore, to better target the ℓ_0 solution, we use the weight vector ($\widehat{w}_{j,m} = \frac{1}{|\widehat{\gamma}_{j,m}|^\delta}$) where $\delta > 0$ and $\widehat{\gamma}_{j,m}$ is the root N -consistent estimate of $\gamma_{j,m}$ as suggested by Zou (2006):

$$\beta_m^{\text{DSS}} = \arg \min_{\gamma_m} \|\mathbf{X}_m \beta_m^* - \mathbf{X}_m \gamma_m\|_2^2 + \lambda \sum_{j=1}^p \widehat{w}_{j,m} |\gamma_{j,m}|. \quad (9)$$

Table 3 summarizes the DSS results using a single-break HMBB fitted on the full interaction model of Alvarez *et al.* (1991). Notably, regime-specific parameters that are near to zero are forced to zero, which helps researchers concentrate on nonzero DSS estimates for succinct interpretation of HMBB results.

7 Conclusion

We presented a Bayesian strategy for detecting and estimating change points in regression models with a high number of predictors in this article. The suggested model unifies a variety of statistical methods (a Bayesian shrinkage method, a change-point model, and sparse regression) to enable researchers to perform effective and consistent inference on time-varying parameters in a variety of TSCS datasets. We concentrate on the fixed-effects method in this article because it is one of the most often-used ways to analyze TSCS data in political science. However, our proposed strategy has a far broader application than the fixed-effect model. Our software package includes a tool for constructing a regression model from univariate time series data as well as multilevel models with changing intercepts or slopes. We intend to expand the presented strategy to models with discrete response data.

Table 3. Variable selection of HMBB estimates using the DSS method: DSS indicates the sparsified estimates of HMBB outputs. The employed model is a single-break HMBB fitted on the full interaction model of Alvarez *et al.* 1991.

	Regime 1		Regime 2	
	HMBB	DSS	HMBB	DSS
lagg1	0.04	0.00	0.02	0.01
opengdp	0.09	0.12	-0.08	-0.07
openex	0.05	0.00	0.22	0.23
openimp	0.05	0.00	0.02	0.00
leftc	0.19	0.15	0.05	0.04
central	-0.06	-0.01	-0.13	-0.13
lagg1-opengdp	0.06	0.00	0.06	0.07
lagg1-openex	0.02	0.00	0.20	0.19
lagg1-openimp	0.06	0.11	-0.28	-0.27
lagg1-leftc	-0.00	0.00	0.03	0.03
lagg1-central	-0.12	-0.11	-0.16	-0.15
opengdp-openex	-0.04	0.00	-0.11	-0.11
opengdp-openimp	-0.04	0.00	0.03	0.01
opengdp-leftc	0.09	0.11	0.01	0.00
opengdp-central	-0.00	0.00	0.13	0.15
openex-openimp	-0.02	0.00	-0.08	-0.06
openex-leftc	0.06	0.00	-0.07	-0.05
openex-central	-0.00	0.00	0.04	0.03
openimp-leftc	0.09	0.00	0.08	0.07
openimp-central	-0.04	0.00	0.02	0.01
leftc-central	0.29	0.31	0.15	0.15

While we are confident in our proposed method’s performance in regular TSCS data settings, we would like to point out a few limitations. First, HMBB is best suited to changes in parameter values that are quite abrupt or significant. Dynamic linear models are better equipped to handle slow, cyclical, or evolutionary changes (West and Harrison 1997). Second, HMBB determines whether all parameters have a common break. Assume that only a small subset of parameters changes over time, whereas the remainder remain constant, or that parameters exhibit heterogeneous change points. HMBB would be incapable of identifying these parameter-specific change points precisely. We are currently developing change-point regression models that allow for parameter-specific break detection and regularization using recent breakthroughs in Bayesian statistics, such as Hahn *et al.* (2018). Finally, HMBB uses the Bayesian bridge model as a baseline model for shrinkage. It is worth trying to combine different shrinkage methods, such as the horseshoe prior (Carvalho *et al.* 2010), with change-point models.

Appendix A. Sampling Algorithm

We first implement centering of data (group-centering in the case of panel data) to make parameters have a common range.

1. Sampling $p(\beta|\alpha, \Lambda, \sigma^2, \tau, \mathbf{P}, \mathbf{S}, \mathbf{y})$: If $n_m > p$, the posterior of β follows the multivariate normal distribution, which is given by

$$\beta_m | \sigma^2, \lambda_m, \alpha_m, \tau, \mathbf{P}, \mathbf{S}, \mathbf{y}_m \sim \mathcal{N}_p \left(\frac{\mathbf{V}\mathbf{X}'_m \mathbf{y}_m}{\sigma_m^2}, \mathbf{V} = \left(\mathbf{X}'_m \mathbf{X}_m + \frac{\sigma_m^2}{\tau^2} \lambda_m \mathbf{I} \right)^{-1} \right). \tag{10}$$

If $n_m \leq p$, we use Bhattacharya, Chakraborty, and Mallick's (2016) algorithm:

- (a) For each regime, sample $\mathbf{u}_m \sim \mathcal{N}(0, \mathbf{D}_m)$ and $\delta_m \sim \mathcal{N}(0, \mathbf{I}_{n_m})$ where $d_{j,j} = \sqrt{\frac{\lambda_{m,j} \sigma_m^2}{\tau_m}}$ is the j th diagonal entry of \mathbf{D}_m matrix and n_m is the number of observations at regime m .
 - (b) Set $\nu_m = \mathbf{X}_m \mathbf{u}_m + \delta_m$.
 - (c) Solve $(\mathbf{X}_m \mathbf{D}_m \mathbf{X}_m' + \mathbf{I}_{n_m}) \omega_m = (\frac{\mathbf{y}_m}{\sigma_m^2} - \nu_m)$.
 - (d) Set $\beta_m = \mathbf{u}_m + \mathbf{D}_m \mathbf{X}_m' \omega_m$.
2. Sampling $p(\beta_0 | \mathbf{A}, \beta, \alpha, \tau, \sigma^2, \mathbf{P}, \mathbf{S}, \mathbf{y})$: We separately estimate the intercepts for each regime to remove any discrepancy in regression slopes in each simulation.

$$\beta_{0m} \leftarrow \bar{y}_m - \bar{\mathbf{X}}_m^T \beta_m,$$

where

$$\bar{y}_m = \frac{\sum_{t=1}^n \mathbf{1}\{s_t = m\} y_t}{\sum_{t=1}^n \mathbf{1}\{s_t = m\}}, \quad \text{and} \quad \bar{\mathbf{X}}_{m,j} = \frac{\sum_{t=1}^n \mathbf{1}\{s_t = m\} \mathbf{X}_{m,tj}}{\sum_{t=1}^n \mathbf{1}\{s_t = m\}}. \quad (11)$$

3. Sampling $p(\alpha | \mathbf{A}, \beta, \sigma^2, \tau, \mathbf{P}, \mathbf{S}, \mathbf{y})$: We use a Griddy Gibbs sampler (Tanner 1996) for the sampling of α because α is univariate and its support is bounded by (0, 2].
4. Sampling $p(\tau | \mathbf{A}, \beta, \alpha, \sigma^2, \mathbf{P}, \mathbf{S}, \mathbf{y})$: Sample v first and then transform v to τ .

$$v_m \sim \text{Gamma}(c, d),$$

$$\tau_m = v^{-\frac{1}{\alpha_m}},$$

where $c = c_0 + p/\alpha_m$ and $d = d_0 + \sum_{j=1}^p |\beta_{j,m}|^{\alpha_m}$.

5. Sampling $\mathbf{S} | \mathbf{A}, \beta, \alpha, \tau, \sigma^2, \mathbf{P}, \mathbf{y}$: Sample \mathbf{S} recursively using Chib's (1998) algorithm.
6. Sampling from $\mathbf{P} | \mathbf{A}, \beta, \alpha, \tau, \sigma^2, \mathbf{S}, \mathbf{y}$:

$$p_{kk} \sim \text{Beta}(a_0 + j_{k,k} - 1, b_0 + j_{k,k+1}),$$

where p_{kk} is the probability of staying when the state is k , and $j_{k,k}$ is the number of jumps from state k to k , and $j_{k,k+1}$ is the number of jumps from state k to $k + 1$.

Appendix B. How to Use BridgeChange

```
## install and load package
devtools::install_github("jongheepark/BridgeChange")
library(BridgeChange)

## get agl data and make plm input parameters
library(pcse)
data(agl)
index = c("country", "year") ## group id first and time id second
model = "within"             ## "between" or "pooling"
effect = "time"              ## "individual" or "twoway"

## formula: regular lm formula
formula <- growth ~ lagg1 + opengdp + openex + openimp + leftc + central
+ inter
```

```

## model fitting: no break, one break, and two break model
agl.cp0 <- BridgeFixedPanel(formula=formula, data = agl,
                             model = model, index = index, effect = effect,
                             mcmc= 1000, burn = 1000, verbose=1000,
                             n.break = 0)
agl.cp1 <- BridgeFixedPanel(formula=formula, data = agl,
                             model = model, index = index, effect = effect,
                             mcmc= 1000, burn = 1000, verbose=1000,
                             n.break = 1)
agl.cp2 <- BridgeFixedPanel(formula=formula, data = agl,
                             model = model, index = index, effect = effect,
                             mcmc= 1000, burn = 1000, verbose=1000,
                             n.break = 2)

## model selection by WAIC
waic <- WaicCompare(list(agl.cp0, agl.cp1, agl.cp2))
plotWaic(waic)

## plot state transitions
par(mfrow=c(1,2))
plotState(agl.cp1)
plotState(agl.cp2)

## plot parameter movements
par(mfrow=c(1,3))
dotplotRegime(agl.cp0, hybrid = FALSE)
dotplotRegime(agl.cp1, hybrid = FALSE)
dotplotRegime(agl.cp2, hybrid = FALSE)

## HMBB-DSS method
agl.dss <- BridgeFixedPanelHybrid(formula=formula, data = agl,
                                   model = model, index = index, effect = effect,
                                   mcmc= 1000, burn = 1000, verbose=1000,
                                   n.break = 1)

## plot DSS parameter movements
dotplotRegime(agl.dss, hybrid = TRUE)

```

Acknowledgments

We appreciate valuable comments from Le Bao, David Carlson, Taeryon Choi, Max Goplerud, Kosuke Imai, Heeseok Oh, and three anonymous reviewers. We would like to thank seminar participants of the 2018 International Society for Bayesian Analysis meeting, the 2019 Asian Political Methodology meeting, the 2019 Midwest Political Science Association meeting, and a symposium on Bayesian political methodology at Washington University in St. Louis, 2019. Jong Hee Park was supported by the SNU 10-10 project of Seoul National University.

Data Availability Statement

Replication code for this article is available in Park and Yamauchi (2022) at <https://doi.org/10.7910/DVN/MCQTYC>.

Supplementary Material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2022.23>.

References

- Allee, T. L., and J. E. Scalera. 2012. "The Divergent Effects of Joining International Organizations: Trade Gains and the Rigors of WTO Accession." *International Organization* 66 (2): 243–276.
- Alvarez, R. M., G. Garrett, and P. Lange. 1991. "Government Partisanship, Labor Organization, and Macroeconomic Performance." *American Political Science Review* 85 (2): 539–556.
- Armagan, A. 2009. "Variational Bridge Regression." In *Proceedings of the 12th International Conference on Artificial Intelligence and Statistics*, Vol. 5, 17–24.
- Bailey, D., and J. N. Katz. 2011. "Implementing Panel-Corrected Standard Errors in R: The pcse Package." *Journal of Statistical Software, Code Snippets* 42 (1): 1–11.
- Beck, N. 1983. "Time-Varying Parameter Regression Models." *American Journal of Political Science* 27 (3): 557–600.
- Beck, N. 2001. "Time-Series-Cross-Section Data: What Have We Learned in the Past Few Years?" *Annual Review of Political Science* 4: 271–293.
- Beck, N., and J. Katz. 1995. "What to Do (and Not to Do) with Time-Series Cross-Sectional Data." *American Political Science Review* 89: 634–647.
- Beck, N., and J. N. Katz. 2011. "Modeling Dynamics in Time-Series-Cross-Section Political Economy Data." *Annual Review of Political Science* 14: 331–352.
- Beck, N., J. N. Katz, R. M. Alvarez, G. Garrett, and P. Lange. 1993. "Government Partisanship, Labor Organization, and Macroeconomic Performance: A Corrigendum." *American Political Science Review* 87 (4): 945–948.
- Bhattacharya, A., A. Chakraborty, and B. K. Mallick. 2016. "Fast Sampling with Gaussian Scale Mixture Priors in High-Dimensional Regression." *Biometrika* 103 (4): 985–991.
- Blackwell, M. 2018. "Game Changers: Detecting Shifts in Overdispersed Count Data." *Political Analysis* 26 (2): 230–239.
- Boix, C. 1997. "Political Parties and the Supply Side of the Economy: The Provision of Physical and Human Capital in Advanced Economies, 1960–90." *American Journal of Political Science* 41 (3): 814–845.
- Box-Steffensmeier, J. M., J. R. Freeman, M. P. Hitt, and J. C. Pevehouse. 2014. *Time Series Analysis for the Social Sciences, Analytical Methods for Social Research*. Cambridge: Cambridge University Press.
- Brandt, P. T., and J. R. Freeman. 2006. "Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis." *Political Analysis* 14 (1): 1–36.
- Brandt, P. T., and T. Sandler. 2010. "What Do Transnational Terrorists Target? Has It Changed? Are We Safer?" *Journal of Conflict Resolution* 54 (2): 214–236.
- Carlin, B. P., A. E. Gelfand, and A. F. M. Smith. 1992. "Hierarchical Bayesian Analysis of Change-point Problems." *Applied Statistics* 41 (2): 389–405.
- Carvalho, C. M., N. G. Polson, and J. G. Scott. 2010. "The Horseshoe Estimator for Sparse Signals." *Biometrika* 97: 465–480.
- Cederman, L.-E., K. S. Gleditsch, and J. Wucherpfennig. 2017. "Predicting the Decline of Ethnic Civil War: Was Gurr and for the Right Reasons?" *Journal of Peace Research* 54 (2): 262–274.
- Chan, N., C. Y. Yau, and R.-M. Zhang. 2014. "Group LASSO for Structural Break Time Series." *Journal of the American Statistical Association* 109 (506): 590–599.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey. 2017. "Double/Debiased/Neyman Machine Learning of Treatment Effects." *American Economic Review* 107 (5): 261–265.
- Chib, S. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90 (432): 1313–1321.
- Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics* 86 (2): 221–241.
- Croissant, Y., and G. Millo. 2008. "Panel Data Econometrics in R: The plm Package." *Journal of Statistical Software* 27 (2): 1–43.
- Fan, J., and R. Li. 2001. "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties." *Journal of the American Statistical Association* 96 (456): 1348–1360.
- Fox, E. B., E. B. Sudderth, M. I. Jordan, and A. S. Willsky. 2011. "A Sticky HDP-HMM with Application to Speaker Diarization." *Annals of Applied Statistics* 5 (2A): 1020–1056.
- Frank, I. E., and J. H. Friedman. 1993. "A Statistical View of Some Chemometrics Regression Tools." *Technometrics* 35: 109–148.
- Franzese, R. J., Jr. 2002. *Macroeconomic Policies of Developed Democracies*. Cambridge University Press.
- Frick, K., A. Munk, and H. Sieling. 2014. "Multiscale Change Point Inference." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76 (3): 495–580.

- Frieden, J. A. 2020. *Global Capitalism: Its Fall and Rise in the Twentieth Century*. New York: W. W. Norton.
- Fu, W. J. 1998. "Penalized Regressions: The Bridge versus the Lasso." *Journal of Computational and Graphical Statistics* 7 (3): 397–416.
- Garrett, G. 1998. *Partisan Politics in the Global Economy*. Cambridge: Cambridge University Press.
- Gelman, A., J. Hwang, and A. Vehtari. 2014. "Understanding Predictive Information Criteria for Bayesian Models." *Statistics and Computing* 24 (6): 997–1016.
- Goldstein, J., D. Rivers, and M. Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* 61: 37–67.
- Gowa, J. S., and S. Y. Kim. 2005. "An Exclusive Country Club: The Effects of the GATT on Trade, 1950–94." *World Politics* 57 (4): 453–478.
- Hahn, P. R., and C. M. Carvalho. 2015. "Decoupling Shrinkage and Selection in Bayesian Linear Models: A Posterior Summary Perspective." *Journal of the American Statistical Association* 110 (509): 435–448.
- Hahn, P. R., C. M. Carvalho, D. Puelz, and J. He. 2018. "Regularization and Confounding in Linear Regression for Treatment Effect Estimation." *Bayesian Analysis* 13 (1): 163–182.
- Hans, C. 2010. "Model Uncertainty and Variable Selection in Bayesian Lasso Regression." *Statistics and Computing* 20 (2): 221–229.
- Hazlett, C., and L. Wainstein. 2022. "Understanding, Choosing, and Unifying Multilevel and Fixed Effect Approaches." *Political Analysis* 30 (1): 46–65.
- Helleiner, E. 1994. *States and the Reemergence of Global Finance*. Ithaca: Cornell University Press.
- Hendershot, M. E., M. S. Hurwitz, D. N. Lanier, and R. L. Pacelle. 2013. "Dissensual Decision Making: Revisiting the Demise of Consensual Norms within the U.S. Supreme Court." *Political Research Quarterly* 66 (2): 467–481.
- Hoerl, A. E., and R. W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12 (1): 55–67.
- Huang, J., J. L. Horowitz, and S. Ma. 2008. "Asymptotic Properties of Bridge Estimators in Sparse High-Dimensional Regression Models." *The Annals of Statistics*: 587–613.
- Huang, J., S. Ma, H. Xie, and C.-H. Zhang. 2009. "A Group Bridge Approach for Variable Selection." *Biometrika* 96 (2): 339–355.
- Imai, K., and I. S. Kim. 2021. "On the Use of Two-Way Fixed Effects Regression Models for Causal Inference with Panel Data." *Political Analysis* 29 (3): 405–415.
- Kent, D., J. D. Wilson, and S. J. Cranmer. 2022. "A Permutation-Based Change-point Technique for Monitoring Effect Sizes." *Political Analysis* 30 (2): 167–178.
- Kyung, M., J. Gill, M. Ghosh, and G. Casella. 2010. "Penalized Regression, Standard Errors, and Bayesian Lassos." *Bayesian Analysis* 5 (2): 369–412.
- Lee, S., Y. Liao, M. H. Seo, and Y. Shin. 2018. "Oracle Estimation of a Change Point in High Dimensional Quantile Regression." *Journal of the American Statistical Association* 113 (523): 1184–1194.
- Lee, S., M. H. Seo, and Y. Shin. 2016. "The Lasso for High Dimensional Regression with a Possible Change Point." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78 (1): 193–210.
- Liu, Y., H. H. Zhang, C. Park, and J. Ahn. 2007. "Support Vector Machines with Adaptive L_q Penalty." *Computational Statistics & Data Analysis* 51 (12): 6380–6394.
- MacKinnon, J., and H. White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29 (3): 305–325.
- Nieman, M. D. 2015. "Moments in Time: Temporal Patterns in the Effect of Democracy and Trade on Conflict." *Conflict Management and Peace Science* 33 (3): 1–21.
- Pang, X., B. Friedman, A. D. Martin, and K. M. Quinn. 2012. "Endogenous Jurisprudential Regimes." *Political Analysis* 20 (4): 417–436.
- Pang, X., L. Liu, and Y. Xu. 2022. "A Bayesian Alternative to Synthetic Control for Comparative Case Studies." *Political Analysis* 30 (2): 269–288.
- Park, J. H. 2010. "Structural Change in U.S. Presidents' Use of Force." *American Journal of Political Science* 54 (3): 766–782.
- Park, J. H. 2011a. "Analyzing Preference Changes using Hidden Markov Item Response Theory Models." In *Handbook of Markov Chain Monte Carlo: Methods and Applications*, edited by G. Jones, S. Brooks, A. Gelman, and X.-L. Meng. London: CRC Press.
- Park, J. H. 2011b. "Change-point Analysis of Binary and Ordinal Probit Models: An Application to Bank Rate Policy under the Interwar Gold Standard." *Political Analysis* 19 (2): 188–204.
- Park, J. H. 2012. "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models." *American Journal of Political Science* 56 (4): 1040–1054.
- Park, J. H., and S. Yamauchi. 2022. "Replication Data for: Change-Point Detection and Regularization in Time Series Cross Sectional Data Analysis." Harvard Dataverse V1. <https://doi.org/10.7910/DVN/MCQTYC>.
- Park, T., and G. Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103 (482): 681–686.
- Polson, N. G. 2012. "Local Shrinkage Rules, Lévy Processes, and Regularized Regression." *Journal of the Royal Statistical Society (Series B)* 74: 287–311.

- Polson, N. G., and J. G. Scott. 2010. "Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction." *Bayesian Statistics* 9: 501–538.
- Polson, N. G., J. G. Scott, and J. Windle. 2014. "The Bayesian Bridge." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76 (4): 713–733.
- Ranjit, L. 2016. "Replication Data for: How Multiple Imputation Makes a Difference." <https://doi.org/10.7910/DVN/CRLKIF>.
- Rose, A. 2003. "Do We Really Know that the WTO Increases Trade?" *American Economic Review* 94 (1): 98–114.
- Rueda, D. 2008. "Left Government, Policy, and Corporatism: Explaining the Influence of Partisanship on Inequality." *World Politics* 60 (3): 349–389.
- Santfort, C., T. Sandler, and P. T. Brandt. 2013. "Terrorist Attack and Target Diversity: Changepoints and Their Drivers." *Journal of Peace Research* 50 (1): 75–90.
- Scruggs, L. 2001. "The Politics of Growth Revisited." *Journal of Politics* 63 (1): 120–140.
- Smith, C. E., R. D. Brown, J. M. Bruce, and L. M. Overby. 1999. "Party Balancing and Voting for Congress in the 1996 National Election." *American Journal of Political Science* 43 (3): 737–764.
- Soskice, D., and T. Iversen. 2000. "The Nonneutrality of Monetary Policy with Large Price or Wage Setters." *Quarterly Journal of Economics* 115: 265–284.
- Spirling, A. 2007a. "'Turning Points' in Iraq: Reversible Jump Markov Chain Monte Carlo in Political Science." *The American Statistician* 61 (4): 315–320.
- Spirling, A. 2007b. "Bayesian Approaches for Limited Dependent Variable Change Point Problems." *Political Analysis* 15 (4): 387–405.
- Subramanian, A., and S.-J. Wei. 2007. "The WTO Promotes Trade, Strongly but Unevenly." *Journal of International Economics* 72: 151–175.
- Tanner, M. A. 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. New York: Springer.
- Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 58: 267–288.
- Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight. 2004. "Sparsity and Smoothness via the Fused Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67 (1): 91–108.
- Tomz, M., J. Goldstein, and D. Rivers. 2007. "Do We Really Know that the WTO Increases Trade? Comment." *American Economic Review* 97 (5): 2005–2018.
- Vats, D., and C. Knudson. 2021. "Revisiting the Gelman–Rubin Diagnostic." *Statistical Science* 36 (4): 518–529.
- Wawro, G. J., and I. Katznelson. 2014. "Designing Historical Social Scientific Inquiry: How Parameter Heterogeneity Can Bridge the Methodological Divide between Quantitative and Qualitative Approaches." *American Journal of Political Science* 58 (2): 526–546.
- West, M., and J. Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. New York: Springer.
- Western, B. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach." *American Journal of Political Science* 42 (4): 1233–1259.
- Western, B., and M. Kleykamp. 2004. "A Bayesian Change Point Model for Historical Time Series Analysis." *Political Analysis* 12 (4): 354–374.
- Wucherpennig, J., A. Kachi, N.-C. Bormann, and P. Hunziker. 2021. "A Fast Estimator for Binary Choice Models with Spatial, Temporal, and Spatio-Temporal Interdependence." *Political Analysis* 29 (4): 570–576.
- Zou, H. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101 (476): 1418–1429.
- Zou, H., and T. Hastie. 2005. "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 67 (2): 301–320.