Synthetic Control with Time Varying Coefficients

A State Space Approach with Bayesian Shrinkage

Danny Klinenberg*

Last Updated: 2020-10-18

Abstract

Synthetic control methods are a popular tool for measuring the effects of policy interventions on a single treated unit. In practice, researchers create a counterfactual using a linear combination of untreated units that closely mimic the treated unit. Often times, creating a synthetic control is not possible due to untreated units' dynamic characteristics such as integrated processes or a time varying relationship. In this paper, I propose a new approach to estimate the synthetic control counterfactual by incorporating time varying parameters. This is done using a state space framework and Bayesian shrinkage. The dynamics allow for a closer pre-treatment fit leading to more precise counterfactual estimation. Monte Carlo simulations are performed to investigate the usefulness of the proposed model in a synthetic control setting. I then compare the proposed model to two existing approaches in classic synthetic control case studies. Results suggest the proposed model produces lower mean squared forecast error when dynamic relationships are present. In addition, the model performs similar to existing approaches when no dynamics are present.

*Ph.D. student at University of California, Santa Barbara; dklinenberg@ucsb.edu. I'd like to thank Dick Startz, Jaime Ramirez-Cuellar, Ryan Sherrard, and the UCSB Econometrics reading group for comments. All errors are my own.

1 Introduction

In this paper, I consider the problem of estimating the causal effect of an intervention on an outcome of interest when there is one unit affected by the intervention and the relationship between the unit of interest and all other units may be non-constant. A common approach to this problem is the synthetic control framework. The goal is to construct a counterfactual for the treated unit as a linear combination of untreated units. The synthetic control method has been used in many areas of economics including (but not limited to) the effects of terrorism (Abadie and Gardeazabal 2003), trade policies (Billmeier and Nannicini 2013), natural disasters (Cavallo et al. 2013) and social issues (Powell 2018).

Abadie, Diamond, and Hainmueller (2010) show that if there exists some linear combination of untreated units such that a perfect pretreatment estimate of the treated unit exists, then the asymptotic bias of the estimated treatment effect approaches zero as the pretreatment period increases. Their approach simplifies the time series estimation problem to a cross sectional one. Each time period in the pretreatment is an "observation" used to estimate missing "observations" in the post treatment. This approach limits the scope of the tool to constant relationships between treatment and control units. In a time series setting, this method excludes nonstationary data such as integrated processes. If such heterogeneous relationships exist and are ignored, the test of no intervention effect is extremely oversized (Carvalho 2016). Abadie, Diamond, and Hainmueller (2010) acknowledge this limitation and explicitly warn against the uses of synthetic control when an accurate counterfactual cannot be constructed.

This paper proposes incorporating time varying coefficients into the synthetic control framework to address such situations. An immediate concern with time varying coefficients is the risk of overfitting. Recent advances in macroeconometric forecasting have developed methods to address this concern (Dangl and Halling 2012; Bitto and Frühwirth-Schnatter 2019; Belmonte, Koop, and Korobilis 2014). These methods rely on two ideas: non-centered state space modeling and Bayesian shrinkage. Non-centered state space modeling decomposes time varying parameters into a time varying component and a time invariant component. By decomposing each time varying parameter, Bayesian shrinkage techniques "shrink" irrelevant parameters towards zero increasing out of sample performance. These two techniques allow the proposed model to perform as well as a static-coefficient model when the true data generating process involves only static coefficients and superior otherwise.

First, I propose a time varying parameter model based on recent macroeconometric advances to estimate the counterfactual. Second, I compare a popular state space counterfactual model, Brodersen et al. (2015), to the proposed model in simulation studies. Prior to this paper, testing of this class of model focused on high frequency data including stock prices and inflation rates (Dangl and Halling 2012; Bitto and Frühwirth-Schnatter 2019). This differs from a synthetic control setting where data tends to be yearly or monthly with few pre and post periods and potentially many controls. Finally, I compare the model's performance to Brodersen et al. (2015) and Abadie, Diamond, and Hainmueller (2010) using the German Reunification (Abadie, Diamond, and Hainmueller 2015) and the California Tobacco tax (Abadie, Diamond, and Hainmueller 2010) case studies.

1.1 Related Work

Developments to synthetic control fall into three main categories: i) multiple treatments/outcomes (Xu 2017; Athey et al. 2020; L'Hour 2019), ii) inference (Li 2019; Cattaneo, Feng, and Titiunik 2019; Chernozhukov, Wuthrich, and Zhu 2019), and iii) counterfactual estimation. This paper is focused on counterfactual estimation. For a full review of synthetic controls, the reader is directed to Abadie (2019). For an in depth comparison of multiple synthetic control approaches, the reader is directed to Samartsidis et al. (2019) and Kinn (2018).

An increasingly common issue in the case study literature is more untreated units than pre-treatment periods. For example, Abadie, Diamond, and Hainmueller (2010) considers a situation in which there are 29 untreated units and 17 pretreatment periods. Past researchers have addressed this issue with machine learning techniques (e.g. Doudchenko and Imbens (2016), Athey et al. (2018)). Pang (2010) recommends a model comparison algorithm using Bayes factors while Pang, Liu, and Xu (2020) and Brodersen et al. (2015) incorporate shrinkage priors. These methods place a majority of mass of the prior distribution at zero. These methods force coefficients biased towards zero which allows for the usage of more covariates than observations and better out of sample predictions while avoiding overfitting.

Bayesian methods have been used to estimate the counterfactual in a synthetic control setting. Brodersen et al. (2015) models the counterfactual using a combination of spike and slab priors and linear Gaussian state space modeling. The spike and slab priors are used to perform automatic variable selection. The authors allow for the coefficients to be constant or dynamic. However, they warn of the dangers of overfitting and implausibly large probability intervals with dynamic coefficients (Brodersen et al. 2015).

This paper aims to solve the issues Brodersen et al. (2015) faced when incorporating dynamic coefficients. First, the proposed model incorporates the decomposition of time varying coefficients. Second, the model uses a different set of priors to create the Bayesian LASSO. The decomposition paired with the choice of priors solves the issues of overfitting and implausibly large probability intervals. The proposed model allows for shrinkage on the time varying and time invariant portion of the coefficient. Adding the parameter decomposition and Bayesian Lasso allows for the use of time varying parameters without implausibly large probability intervals.

Pang, Liu, and Xu (2020) develop a Bayesian approach based off of the Xu (2017) linear factors model. Utilizing Bitto and Frühwirth-Schnatter (2019) non-centered parameterization, the authors explicitly model the latent factors as well as allowing for time varying coefficients. Similar to this paper, they employ a state-space Bayesian framework with shrinkage priors. This paper differs in several key ways. The proposed model begins with the established non-centered parameterization framework and adds causal assumptions while Pang, Liu, and Xu (2020) begins with the linear factors models and incorporates time varying components. The different initial points leads to vastly different functional forms. I focus on the case where there is one treated unit and the only predictors are untreated units, while they extend their model to multiple treated units utilizing additional covariates. The final difference between the two papers is scope and purpose. Pang, Liu, and Xu (2020) establish a full Bayesian framework to the synthetic control setting while the purpose of this paper is to introduce time varying parameters to the synthetic control framework and investigate their usefulness. It is not entirely clear the increased flexibility of this model will be beneficial in synthetic control settings given the small sample size. In this manner, the papers can be seen as compliments to one another.

2 Setup

2.1 Potential Outcomes and Parameter of Interest

I define the treatment as an intervention or policy change. Once treated, the unit remains treated indefinitely. The treatment status is known for all units in all time periods.

Let $(Y_{j,t}(0), Y_{j,t}(1))$ represent potential outcomes in the presence and absence of a treatment with $t = 1, ..., T_0 - 1, T_0, T_0 + 1, ...T$ and $j = 0, 1, ..., J$. Denote the period of intervention as T_0. Define the treatment status as $D_j = \{0, 1\}$, where a 1 indicates if the unit is treated in any period. I assume the potential outcomes are random.

Define $Y_{j,t} = (1 - D_{j,t})Y_{j,t}(0) + D_{j,t}Y_{j,t}(1)$ where $D_{j,t} = D_j I(t \geq T_0)$. The researcher observes the following:

$$Y_{j,t} = \begin{cases} Y_{j,t}(1) & D_{j,t} = 1 \\ Y_{j,t}(0) & D_{j,t} = 0 \end{cases} \tag{1}$$

The average treatment effect of the treated unit at each period t is denoted

$\tau_t = \mathbb{E}[Y_{j,t}(1)|D_j = 1] - \mathbb{E}[Y_{j,t}(0)|D_j = 1]$. In order to draw causal inference, I assume conditional independence on past outcomes:

Assumption 1. Conditional Independence on Past Observed Outcomes

$$\{Y_{j,T_0+i}(0), Y_{j,T_0+i}(1)\} \perp\!\!\!\perp D_{j,T_0+i}|Y_{j,1}, \ldots, Y_{j,T_0} \tag{2}$$

for $i \in \{1, \ldots, T - T_0\}$.

The conditional independence assumption uses the full set of pretreatment outcomes to proxy for the unobserved confounders.

Since we are only interested in the treatment effect post intervention, replace t with T_0+i. The estimand can be rewritten as:

$$\begin{aligned}
\tau_{T_0+i} &= \mathbb{E}[Y_{j,T_0+i}(1)|D_j = 1] - \mathbb{E}[Y_{j,T_0+i}(0)|D_j = 1] \\
&= \mathbb{E}[Y_{j,T_0+i}(1)|D_j = 1] - \mathbb{E}[\mathbb{E}[Y_{j,T_0+i}(0)|Y_{j,1}, \ldots, Y_{j,T_0}, D_j = 1]|D_j = 1] \\
&= \mathbb{E}[Y_{j,T_0+i}(1)|D_j = 1] - \mathbb{E}[\mathbb{E}[Y_{j,T_0+i}(0)|Y_{j,1}, \ldots, Y_{j,T_0}, D_j = 0]|D_j = 1] \\
&= \mathbb{E}[Y_{j,T_0+i}(1)|D_j = 1] - \mathbb{E}[g_{T_0+i}(Y_{j,1}, \ldots, Y_{j,T_0})|D_j = 1]
\end{aligned}$$

where $g_{T_0+i}(Y_{j,1}, \ldots, Y_{j,T_0}) = \mathbb{E}[Y_{j,T_0+i}(0)|Y_{j,1}, \ldots, Y_{j,T_0}, D_j = 0]$.

Suppose only one unit, $j = 0$, is treated beginning in period T_0 and remains treated for all $T_0 + i \geq T_0$. Also suppose the other J units are unaffected by the treatment and there is no anticipation effects (SUTVA holds, Rubin (1990))[1]. We observe a sample $\{(y_{j,1}, \ldots, y_{j,T}, d_j)\}_{j=0}^{J}$ from the distribution $(Y_{j,1}, \ldots, Y_{j,T}, D_j)$. Since there is only one treated unit, $\mathbb{E}[Y_{j,T_0+i}(1)|D_j = 1]$ can be estimated with the realization of the data, y_{0,T_0+i}, and $\mathbb{E}[g_{T_0+i}(Y_{j,1}, \ldots, Y_{j,T_0})|D_j = 1]$ can be estimated as $g_{T_0+i}(y_{0,1}, \ldots, y_{0,T_0})$.

The goal is to estimate the sample ATT:

[1] This is a common assumption in the synthetic control literature. Recently, Grossi et al. (2020) have introduced spillover effects in analyzing new light rail transit.

$$\hat{\tau}_{T_0+i} = y_{0,T_0+i} - \hat{g}_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right) \tag{3}$$

An additional metric of interest in this paper is the sample average treatment effect in the post period:

$$\Delta\tau = \frac{1}{T-T_0}\sum_{i=0}^{T-T_0}\tau_{T_0+i}$$

The sample average treatment effect in the post period is estimated as:

$$\hat{\Delta}\tau = \frac{1}{T-T_0}\sum_{i=0}^{T-T_0}\hat{\tau}_{T_0+i}$$

2.2 General Model

There are three sources of information to estimate $y_{0,T_0+i}(0) = g_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)$:

i) Untreated pre-treatment units: $\mathbf{y}$.

ii) Treated pre-treatment units: $\mathbf{y_0} = (y_{0,1},\ldots,y_{0,T_0})$.

iii) Untreated post-treatment units.

Creating the counterfactual relies on utilizing all three sources of information. The parameters are estimated in the pre-period and then used to forecast out in the post-period. I propose using a time-varying coefficient model. Namely:

$$g_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right) = \sum_{j=1}^{J+1}\beta_{j,T_0+i}y_{j,T_0+i} \tag{4}$$

where $y_{J+1,T_0+i} = 1$ for all t (intercept). With a time varying structure, a perfect fit can be made for each $y_{0,t}$. More so, there exists an infinite combination of perfect matches.

The model would have no out of sample predictive ability. The perfect fit would have no differentiation between signal and noise. To account for this, I add structure to the time varying coefficients through state space modeling. The coefficients are modeled as random walks. Incorporating (4) into a state space framework and adding the random walk component yields:

$$g_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)=\sum_{j=1}^{J+1}\beta_{j,T_0+i}y_{j,T_0+i}+\epsilon_{T_0+i} \qquad \epsilon_{T-0+i}\sim\mathcal{N}(0,\sigma^2) \tag{5}$$

$$\beta_{j,T_0+i}=\beta_{j,T_0+i-1}+\eta_{j,T_0+i} \qquad \eta_{j,T_0+i}\sim\mathcal{N}(0,\theta_j) \quad \forall j \tag{6}$$

$$\beta_{j,0}\sim\mathcal{N}(\beta_j,\theta_jP_{jj}) \qquad \forall j \tag{7}$$

It is common in state space literature to model time varying parameters as random walks (Dangl and Halling 2012; Belmonte, Koop, and Korobilis 2014; Bitto and Frühwirth-Schnatter 2019). The choice of random walk allows for a useful decomposition of the coefficients into time-varying and constant components.

The parameters of the model are $\mathbf{v}=\{\sigma^2,\beta_1,\ldots\beta_{J+1},\theta_1,\ldots\theta_{J+1}\}$ with ϵ_t and $\eta_{j,t}$ assumed independent of all other unknowns. P_{jj} is a hyperparameter set to ensure the initial distribution is disperse. In practice, P_{jj} is set to a very large value creating a diffuse initial state (Durbin and Koopman 2012).

3 Estimation of Parameters and Counterfactual

I estimate the parameters using a Bayesian approach. $g_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)$ will be estimated as a conditional distribution on the observable:

$$g_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)\sim f_{T_0+i}(y_{0,T_0+i}|\mathbf{y},\mathbf{y_0}) \tag{8}$$

This distribution can be rewritten as a model average on parameters:

$$f_{T_0+i}(y_{0,T_0+i}|\mathbf{y},\mathbf{y_0}) = \int_{v\in V} f_{T_0+i}(y_{0,T_0+i}|\mathbf{y},\mathbf{y_0},v)Pr(v|\mathbf{y},\mathbf{y_0})dv \tag{9}$$

Given the state space setup, $f_{T_0+i}(y_{0,T_0+i}|\mathbf{y},\mathbf{y_0},v) \sim \mathcal{N}\left(\sum_{j=1}^{J+1}\beta_{j,T_0+i}y_{j,T_0+i},\sigma^2\right)$. To estimate the posterior parameters, I incorporate a variant of the Bayesian Lasso (Park and Casella 2008). I model the shrinkage using the global-local shrinkage framework (Polson and Scott 2011b). This framework is designed to apply stronger amounts of shrinkage for smaller parameter estimates allowing larger parameter estimates to "escape" leading to better in-sample and out-of-sample fits.

3.1 Reparameterization

Equation (6) can be rewritten to decompose $\beta_{j,t}$ into a time varying and constant components:[2]

$$\begin{aligned}
&\beta_{j,t} = \beta_j + \tilde{\beta}_{j,t}\sqrt{\theta_j}\\
&\tilde{\beta}_{j,t} = \tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t} \qquad \tilde{\eta}_{j,t} \sim N(0,1)\\
&\tilde{\beta}_{j,0} \sim N(0,P_{jj})
\end{aligned}$$

with priors provided below. β_j can now be interpreted as the time invariant component of $\beta_{j,t}$ and $\sqrt{\theta_j}\tilde{\beta}_{j,t}$ the time varying component. $\sqrt{\theta_j}$ is defined as the root of θ_j and allowed to take both positive and negative values. Defining $\sqrt{\theta_j}$ in this manner allows 0 to be an interior point in the prior distribution. This is a desirable feature when performing Bayesian shrinkage (Bitto and Frühwirth-Schnatter 2019). The absolute value of $\sqrt{\theta_j}$ is the standard deviation of the time varying coefficient. Substituting the reformulation back into the original equation yields the proposed state space model.

[2]See Frühwirth-Schnatter and Wagner (2010) for more details.

Assumption 2. The Time Varying Parameter Bayesian Lasso (BL-TVP) takes the following form:

$$y_{0,t}(0) = \sum_{j=1}^{J+1} \left(\beta_j + \tilde{\beta}_{j,t}\sqrt{\theta_j}\right) y_{j,t} + \epsilon_t \qquad \epsilon_t \sim N(0, \sigma^2) \tag{10}$$

$$\tilde{\beta}_{j,t} = \tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t} \qquad \tilde{\eta}_{j,t} \sim N(0, 1) \tag{11}$$

$$\tilde{\beta}_{j,0} \sim N(0, P_{jj}) \tag{12}$$

$$\pi(\mathbf{v}) \tag{13}$$

where $\pi(\mathbf{v})$ represents the prior distribution of the parameters with the specific distributions defined below. Equations (10) - (13) constitute the model. This setup is commonly known as the *non-centered parameterization of state space models*. This formulation allows estimation of the time varying and time invariant component of the coefficients individually. The effect of each control unit can be summarized into one of the four categories: (i) time varying non-zero, (ii) time invariant, (iii) time varying centered at zero, and (iv) time invariant zero coefficients (irrelevant).

Notice if $\sqrt{\theta_j} = 0$ for all j, the model is a Bayesian version of the LASSO estimator discussed in Kinn (2018). Setting $\sqrt{\theta_j} = 0$ and restricting β such that $\beta_j \in [0, 1]$ and $\sum_j \beta_j = 1$ yields a parametric version of Abadie and Gardeazabal (2003) synthetic control model. Similarly, if the data generating process is spurious (e.g. $\beta_j = \sqrt{\theta_j} = 0$ for all j), then (10) - (13) collapses to a local level model.

There are 2(J+1)+1 parameters to be estimated: the J+1 time invariant coefficients (i.e. β_j's), the J+1 time varying coefficients (i.e. $\sqrt{\theta_j}$'s) and the variance (σ^2).

3.1.1 Bayesian Shrinkage Priors

I set up the prior distribution for coefficients $\beta = [\beta_1, \beta_2, ..., \beta_{J+1}]$ with variances $\alpha^2 = [\alpha_1^2, \alpha_2^2, ..., \alpha_{J+1}^2]$ as a *global-local* shrinkage prior:

$$\beta|\alpha^2, \lambda^2 \sim \mathcal{N}_{J+1}(0_{J+1}, \lambda^2 diag[\alpha_1^2, ...\alpha_{J+1}^2])$$

$$\alpha_j^2 \sim \pi(\alpha_j^2)$$

$$\lambda^2 \sim \pi(\lambda^2)$$

This prior formulation has gained popularity in the Bayesian framework due to it's attractive shrinkage properties (Makalic and Schmidt (2016), Polson and Scott (2011b)). λ^2 controls the overall complexity of the model while α_j^2 produces individual shrinkage. This formulation allows for strong shrinkage on small coefficients while leaving larger coefficients relatively unshrunk.

α_j^2 is assigned an exponential distribution with rate 1. The hierarchical formulation of β and α^2 are identical to a priori independent Laplace priors. Such a prior forms the Bayesian LASSO proposed by Park and Casella (2008). Park and Casella (2008) showed this choice of priors leads to posterior performance similar to the frequentist machine learning approach LASSO (Tibshirani 1996).

λ^2 is represented as a half-Cauchy distribution with mean 0 and scale parameter 1. The half-Cauchy is used for the global shrinkage prior because of the flexibility and better behavior near 0 compared to alternatives (Polson and Scott 2011a). In addition, the half-Cauchy has significant amounts of mass at the point 0 leading to better shrinkage properties.

Like the Laplace distribution, the half-Cauchy has a hierarchical representation where $\lambda^2|\zeta_\beta$ follows an inverse gamma with shape parameter 1/2 and rate $1/\zeta_\beta$. The hierarchical parameter, ζ_β, follows an inverse gamma with shape parameter 1/2 and rate parameter 1. The prior distribution for $\beta = [\beta_1, \beta_2, ..., \beta_{J+1}]$ with variances $\alpha^2 = [\alpha_1^2, \alpha_2^2, ..., \alpha_{J+1}^2]$ are:

$$\beta|\alpha^2, \lambda^2 \sim \mathcal{N}_{J+1}(0_{J+1}, \lambda^2 diag[\alpha_1^2, ...\alpha_{J+1}^2]) \tag{14}$$

$$\alpha_j^2 \sim exp\,(1) \tag{15}$$

$$\lambda^2|\zeta_\beta \sim InverseGamma\left(\frac{1}{2}, \frac{1}{\zeta_\beta}\right) \tag{16}$$

$$\zeta_\beta \sim InverseGamma\left(\frac{1}{2}, 1\right) \tag{17}$$

Traditionally, variances have been defined by the inverse gamma distribution. However, the inverse gamma does not allow for effective shrinkage given its support. Frühwirth-Schnatter and Wagner (2010) provide an in depth argument for the use of the normal distribution as an alternative. Briefly, the inverse gamma prior performs poorly in terms of shrinkage due to 0 being an extreme value in the distribution. This limits the amount of mass which can be placed at 0 in turn limiting the amount of shrinkage. This becomes a problem when there is believed to be many parameters equal to zero. The normal distribution allows for mass at zero avoiding this problem. Similarly to β, assign the prior of $\sqrt{\theta} = \left[\sqrt{\theta_1}, \sqrt{\theta_2}, ..., \sqrt{\theta_{J+1}}\right]$ with variances $\xi^2 = \left[\xi_1^2, \xi_2^2, ...\xi_{J+1}^2\right]$ as:

$$\sqrt{\theta}|\xi^2, \kappa^2 \sim \mathcal{N}_{J+1}(0_{J+1}, \kappa^2 diag[\xi_1^2, ...\xi_{J+1}^2]) \tag{18}$$

$$\xi_j^2 \sim exp\,(1) \tag{19}$$

$$\kappa^2|\zeta_{\sqrt{\theta}} \sim InverseGamma\left(\frac{1}{2}, \frac{1}{\zeta_{\sqrt{\theta}}}\right) \tag{20}$$

$$\zeta_{\sqrt{\theta}} \sim InverseGamma\left(\frac{1}{2}, 1\right) \tag{21}$$

σ^2 is defined as $\frac{1}{\sigma^2} \sim Gamma(a_1, a_2)$ with *shape* hyperparameter a_1 and *scale* hyperparameter a_2. If $\sqrt{\theta_j} = 0$ for all j, the model collapses to a time invariant estimation with the Bayesian LASSO performing shrinkage. A derivation of the posterior distributions can be found in the appendix.

4 The Posterior Estimation (MCMC)

In order to draw the counterfactual, the posterior distribution must be calculated: $P(\tilde{\beta}, \beta, \alpha^2, \lambda^2, \sqrt{\theta}, \xi^2, \kappa^2, \zeta_\beta, \zeta_{\sqrt{\theta}}, \sigma^2 | \mathbf{y_0}, \mathbf{y})$. With values drawn from the posterior distribution, $\hat{g}_{T_0+i}\left(y_{0,1}(0), \dots, y_{0,T_0}(0)\right)$ can then be estimated for $T_0 + i \geq T_0$. A closed form does not exist for the posterior. Therefore, I implement the Gibbs sampler. After a sufficiently large initial sample, or burn in, the draws from the conditional posterior will be simulations of the joint posterior.

The posterior estimation can be broken into three main steps:

(i) Estimation of $\tilde{\beta}|\beta, \alpha^2, \lambda^2, \sqrt{\theta}, \xi^2, \kappa^2, \zeta_\beta, \zeta_{\sqrt{\theta}}, \sigma^2, \mathbf{y_0}, \mathbf{y}$.

(ii) Estimation of the parameters: $P(\beta, \alpha^2, \lambda^2, \sqrt{\theta}, \xi^2, \kappa^2, \zeta_\beta, \zeta_{\sqrt{\theta}}, \sigma^2 | \mathbf{Y_0}, \mathbf{Y})$.

(iii) Estimation of $\hat{g}_{T_0+i}\left(y_{j,1}, \dots, y_{j,T_0}\right)$ for $t \geq T_0$.

4.1 Estimation of $\tilde{\beta}|\beta, \alpha^2, \lambda, \sqrt{\theta}, \xi^2, \kappa^2, \zeta_\beta, \zeta_{\sqrt{\theta}}, \sigma^2, \mathbf{y_0}, \mathbf{y}$

Draw $\tilde{\beta}$ using Durbin and Koopman (2002) for the state space model. First, rewrite equations (5), (6), and (7) as:

$$y'_{0,t}(0) = \sum_{j=1}^{J+1} \tilde{\beta}_{j,t} z_{j,t} + \epsilon_t \qquad \epsilon_t | \sigma^2 \sim N(0, \sigma^2) \tag{22}$$

$$\tilde{\beta}_{j,t} = \tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t} \qquad \tilde{\eta}_{j,t} \sim N(0,1) \tag{23}$$

$$\tilde{\beta}_{j,0} \sim N(0, P_{jj}) \tag{24}$$

where $z_{j,t} = \sqrt{\theta_j} y_{j,t}$, $y'_{0,t} = y_{0,t} - \sum_{j=1}^{J+1} \beta_j y_{j,t}$.

Many algorithms have been proposed to simulate latent variables in a state space framework. I use the method proposed by Durbin and Koopman (2002). First run the Kalman filter and smoother given the data and parameters to produce $\tilde{\beta}^*_t$. Then simulate new $\tilde{\beta}^+_{j,t}$ and $y'^+_{0,t}$ for all j using equations (22), (23), and (24). Then run the Kalman filter and smoother on $y'^+_{0,t}(0)$ and $\tilde{\beta}^+_{j,t}$ for all j producing $\tilde{\beta}^{*,+}_t$. The new simulated draw of $\tilde{\beta}_t$ (denoted $\tilde{\beta}'_t$) is $\tilde{\beta}'_t = \tilde{\beta}^*_t - \tilde{\beta}^+_t + \tilde{\beta}^{*,+}_t$.

4.2 Estimation of the Parameters: $P(\beta, \alpha^2, \lambda, \sqrt{\theta}, \xi^2, \kappa^2, \zeta_\beta, \zeta_{\sqrt{\theta}}, \sigma^2 | y_0)$

Attempting to sample $P(\beta, \alpha^2, \lambda^2, \sqrt{\theta}, \xi^2, \kappa^2, \zeta_\beta, \zeta_{\sqrt{\theta}}, \sigma^2 | \mathbf{y_0}, \mathbf{y})$ would lead to the same problem as before: no analytic posterior exists. Rather than sampling all parameters at once, I will sample the parameters as blocks. The sampling distributions are derived in the appendix.

4.2.1 Sample β and $\sqrt{\theta}$

Block draw β and $\sqrt{\theta}$ from the normal conditional posterior:

$$\mathcal{N}_{2(J+1)}\left((\tilde{\mathbf{y}}^T\tilde{\mathbf{y}} + \sigma^2 V^{-1})^{-1}\tilde{\mathbf{y}}^T\mathbf{y_0}, \sigma^2(\tilde{\mathbf{y}}^T\tilde{\mathbf{y}} + \sigma^2 V^{-1})^{-1}\right) \tag{25}$$

Where:

$$\tilde{\mathbf{y}} = \begin{pmatrix} y_{1,1} & y_{2,1} & \dots & y_{J+1,1} & \tilde{\beta}_{1,1}y_{1,1} & \tilde{\beta}_{2,1}y_{2,1} & \dots & \tilde{\beta}_{J+1,1}y_{J+1,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{1,T_0-1} & y_{2,T_0-1} & \dots & y_{J+1,T_0-1} & \tilde{\beta}_{1,T_0-1}y_{1,T_0-1} & \tilde{\beta}_{2,T_0-1}y_{2,T_0-1} & \dots & \tilde{\beta}_{J+1,T_0-1}y_{J+1,T_0-1} \end{pmatrix} \tag{26}$$

$$V = diag\left[\lambda^2\alpha_1^2, \lambda^2\alpha_2^2, ..., \lambda^2\alpha_{J+1}^2, \kappa^2\xi_1^2, \kappa^2\xi_2^2, \ldots, \kappa^2\xi_{J+1}^2\right] \tag{27}$$

Sampling from sparse matrices can lead preset matrix inversion techniques to fail. To avoid such failures, I implement the algorithm proposed by Bhattacharya, Chakraborty, and Mallick (2016).

4.2.2 Sample α^2

Draw α^2 using the fact $\frac{1}{\alpha_j^2}$ each have independent inverse-Gaussian (IG) conditional priors:

$$IG\left(\sqrt{\frac{2\lambda^2}{\beta_j^2}}, 2\right) \text{ for j=1,...J+1} \tag{28}$$

4.2.3 Sample λ^2

Draw λ^2 from the conditional inverse gamma prior:

$$InverseGamma\left(shape = \frac{J+1}{2}, rate = \frac{1}{\zeta_\beta} + \frac{1}{2}\sum_{j=1}^{J+1}\frac{\beta_j^2}{\alpha_j^2}\right) \tag{29}$$

4.2.4 Sample ζ_β

Draw ζ_β from the conditional inverse gamma prior:

$$InverseGamma\left(1, 1 + \frac{1}{\lambda^2}\right) \tag{30}$$

4.2.5 Sample ξ^2

Draw ξ^2 using the fact $\frac{1}{\xi_j^2}$ each have independent inverse-Gaussian (IG) conditional priors:

$$IG\left(\sqrt{\frac{2\kappa^2}{\theta_j}}, 2\right) \text{ for j=1,...J+1} \tag{31}$$

4.2.6 Sample κ^2

Draw κ^2 from the conditional gamma prior:

$$InverseGamma\left(shape = \frac{J+1}{2}, rate = \frac{1}{\zeta_{\sqrt{\theta}}} + \frac{1}{2}\sum_{j=1}^{J+1}\frac{\sqrt{\theta_j}^2}{\xi_j^2}\right) \tag{32}$$

4.2.7 Sample $\zeta_{\sqrt{\theta}}$

Draw $\zeta_{\sqrt{\theta}}$ from the conditional inverse gamma prior:

$$InverseGamma\left(1, 1 + \frac{1}{\kappa^2}\right) \tag{33}$$

4.2.8 Sample σ^2

Draw σ^2 from the posterior distribution:

$$InverseGamma\left(a_1 + \frac{T_0 - 1}{2}, a_2 + \frac{\sum_{t=1}^{T_0-1}\left(y_{0,t} - \sum_{j=1}^{J+1}\left(\beta_j + \tilde{\beta}_{j,t}\sqrt{\theta_j}\right) y_{j,t}\right)^2}{2}\right) \tag{34}$$

Frühwirth-Schnatter and Wagner (2010) note an identification problem arises when using the non-centered parameterization. There is no way to distinguish between $\sqrt{\theta_j}\tilde{\beta}_{j,t}$ and $(-\sqrt{\theta_j})(-\tilde{\beta}_{j,t})$. This problem is referred to as *label switching problem*. This issue is a common occurrence in Bayesian estimation when a distribution is multi-modal, as is the case with the square root of a variance. To solve this identification problem, Frühwirth-Schnatter and Wagner (2010) suggest a random sign change at the end of each iteration of the Gibbs Sampler. With 50% chance, the signs on $\tilde{\beta}$ and $\sqrt{\theta}$ are switched. Both Belmonte, Koop, and Korobilis (2014) and Bitto and Frühwirth-Schnatter (2019) employ this method.

A final note of interest is the formulation of λ^2 (and κ^2). The conditional distribution of λ^2 relies on $\sum_{j=1}^{J+1} \alpha_j^2$ where each posterior α_j^2 relies on β_j. This direct reliance on β_j in the conditional distributions can lead to scaling issues. Data bigger in magnitude can dominate the distribution of λ^2. The issue of scaling is common in both parametric and nonparametrics shrinkage estimation. To account for this, **all covariates except the intercept are scaled to mean zero variance one** prior to analysis.

4.3 Sample of $\hat{g}_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)$ for $t \geq T_0$.

After a sufficiently large burn in period, use the proceeding draws to calculate $\hat{g}_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)$ for $T_0 + i \geq T_0$. Namely, perform the following steps:

(1) Simulate $\tilde{\beta}_{j,t} = \tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t}$ for all j. Use $\tilde{\beta}'_{j,T_0-1}$ simulated in section 4.1 as an initial value. Each iteration of the Gibbs sampler will create a new $\tilde{\beta}'_{j,T_0-1}$.

(2) Using the simulated $\tilde{\beta}_{j,t}$, predict $\hat{g}_{T_0+i}\left(y_{0,1}(0),\ldots,y_{0,T_0}(0)\right)$ as:

$$\hat{g}_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right) = \sum_{j=1}^{J+1}\left(\beta_j + \tilde{\beta}_{j,t}\sqrt{\theta_j}\right) y_{j,t} + \epsilon_t$$

drawing $\epsilon_t \sim N(0,\sigma^2)$. Each iteration of the Gibbs sampler will produce new parameter and state values.

4.4 Estimate $\hat{\tau}_{T_0+i}$ and $\hat{\Delta}_\tau$

After sampling $\hat{g}_{T_0+i}\left(y_{0,1},\ldots,y_{0,T_0}\right)$, the sample average treatment effect on the treated is:

$$\hat{\tau}_{T_0+i} = y_{0,T_0+i} - \hat{g}_{T_0+i}\left(y_{0,1}, \ldots, y_{0,T_0}\right) \tag{35}$$

for $i = \{0, \ldots, T - T_0\}$. The sample average treatment effect in the post period is then calculated as:

$$\hat{\Delta}_\tau = \frac{1}{T - T_0} \sum_{i=0}^{T-T_0} \hat{\tau}_{T_0+i} \tag{36}$$

Sampling from the Gibbs Sampler creates an empirical distribution for the treatment effects. Statistical testing can then be performed with the distribution. A major benefit of this approach is valid probability intervals.

5 Monte Carlo Simulation

The proposed model provides two changes to the synthetic control framework: time varying parameters and Bayesian Lasso shrinkage. To better understand which addition is contributing and how, I perform Monte Carlo simulations comparing *BL-TVP* to Brodersen et al. (2015). I compare each model with and without time varying parameters. The four variations are:

1) the original Brodersen et al. (2015) model (*CI*),

2) Brodersen et al. (2015) with time varying coefficients (*CI-TVP*),

3) The proposed model without time varying coefficients (*BL*),

4) The proposed model (*BL-TVP*).

BL is a simplification of *BL-TVP* in which $\sqrt{\theta_j} = 0$ for all j. *CI* and *CI-TVP* are run using the R package `CausalImpact` (Brodersen et al. 2015). *BL* is run using the `BayesReg` R package (Makalic and Schmidt 2016).

The simulation is based off of Kinn (2018). Assume the following data generating process:

$$\begin{aligned} y_{j,t}(0) &= \xi_{j,t} + \psi_{j,t} + \epsilon'_{j,t} & \text{j=1,..,J} \\ y_{0,t}(0) &= \sum_{j=1}^{J} w_{j,t}(\xi_{j,t} + \psi_{j,t}) + \epsilon'_{1.t} \end{aligned}$$

for t=1,..,T where ξ_{jt} is the trend component, $\psi_{j,t}$ is a seasonality component, and $\epsilon'_{jt} \sim N(0, \sigma^2)$. Specifically, $\xi_{jt} = c_j t + z_j$ where $c_j,\ z_j \in \mathbb{R}$. This will allow for each observation to have a unit-specific time varying confounding factor and a time-invariant confounding factor. Seasonality will be represented as $\psi_{j,t} = \gamma_j sin\left(\frac{\pi t}{\rho_j}\right)$. The explicit data generating process is:

$$\begin{aligned} y_{j,t}(0) &= c_j t + z_j + \gamma_j sin\left(\frac{\pi t}{\rho_j}\right) + \epsilon'_{j,t} & \text{j=1,..,J} \\ y_{0,t}(0) &= \sum_{j=1}^{J} w_{j,t}\left(c_j t + z_j + \gamma_j sin\left(\frac{\pi t}{\rho_j}\right)\right) + \epsilon'_{1.t} \end{aligned}$$

The treatment begins at period T_0. The treatment effect is set to 0 for all periods.

This paper proposes testing two scenarios: (i) deterministic continuous varying coefficients with no treatment effect, and (ii) constant coefficients with no treatment effect. These scenarios will provide insight on the point prediction accuracy (via mean squared error) and the credibility interval size. Within each scenario, the pre-treatment time length and donor pool will be varied. The full time frame will be 34 periods (e.g. $T = 34$).

5.1 Deterministic Continuous Varying Coefficients

To simulate continuous varying coefficients, $c_{1,t}$ and $c_{2,t}$ are defined .75 and .25 respectively. All other $c_{j,t}$ are randomly drawn from U[0,1]. In order to avoid $y_{1,t}$ and $y_{2,t}$ crossing, set $z_1 = 25$ and $z_2 = 5$. Finally, define $w_{1,t} = .2 + .6\frac{t}{T}$ and $w_{2,t} = 1 - w_{1,t}$ in the time varying case.

To summarize, the parameters of this simulation are:

1) $c_{1,t} = .75$, $c_{2,t} = .25$, and $c_{j,t} \sim U[0,1]$ for all $j \notin \{1,2\}$.

2) $z_1 = 25$, $z_2 = 5$ and z_j is sampled from $\{1,2,3,4,...,50\}$.

3) $\epsilon'_{j,t} \sim N(0,1)$.

4) $T = 34$.

5) $\gamma_j = 4$.

6) $\rho_j = 20$.

7) $w_{1,t} = .2 + .6\frac{t}{T}$, $w_{2,t} = 1 - w_{1,t}$, and $w_{j,t} = 0$ for all else (Time Varying).

The data generating process for the time varying coefficient case can be rewritten in recursive form:

$$
\begin{aligned}
y_{0,t}(0) &= \sum_{j=1}^{J} w_{j,t}\left(c_j t + z_j + \gamma_j sin\left(\frac{\pi t}{\rho_j}\right)\right) + \epsilon'_{1.t} \\
w_{1,t} &= w_{1,t-1} + \frac{.6}{T} \\
w_{2,t} &= w_{2,t-1} - \frac{.6}{T} \\
w_{j,t} &= w_{j,t-1} \qquad\qquad\qquad\qquad\qquad\qquad j \notin \{1,2\}
\end{aligned}
$$

with initial conditions:

$$
\begin{aligned}
w_{1,0} &= .2 \\
w_{2,0} &= .8 \\
w_{j,0} &= 0 \quad j \notin \{1,2\}
\end{aligned}
$$

5.2 Constant Coefficients

The setup for constant coefficients is identical to deterministic continuous varying coefficients except point (6) is replaced by (6'):

(6') $w_{1,t} = .2$, $w_{2,t} = 1 - w_{1,t}$, and $w_{j,t} = 0$ for all else (Time Invariant).

5.3 Model Testing and Comparison

I will compare the mean squared error (MSE) in pre and post treatment. Mean squared error encompasses the paper's main goal of estimation. However, an empirical researcher also needs to understand the inference aspect. Since these models are Bayesian, inference derives from the posterior predictive distribution. This is easily calculated from each iteration of the Gibbs sampler. I summarize the results using the 95% credible interval spread (PI Spread) and post treatment coverage of the 95% credibility interval (95% PI). Each measurement is defined as:

$$\text{MSE} \equiv \frac{1}{T - T_0} \sum_{t=T_0}^{T} \left(y_{0,t} - \hat{y}_{0,t}\right)^2$$

$$\text{PI Spread} \equiv \frac{1}{T} \sum_{t=1}^{T} \left(\hat{y}_{0,t}^{.975} - \hat{y}_{0,t}^{.025}\right)$$

$$95\% \text{ PI} \equiv \frac{1}{T - T_0} \sum_{t=T_0}^{T} I\left(y_{0,t} \in \left[\hat{y}_{0,t}^{.025}, \hat{y}_{0,t}^{.975}\right]\right)$$

where $\hat{y}_{0,t}(0)$ is the median of the posterior predictive density created by each model specification and $\hat{y}_{0,t}^{.025}(0)$ and $\hat{y}_{0,t}^{.975}(0)$ are the 2.5^{th} and 97.5^{th} quantiles of the posterior estimations. The results for mean squared error are presented in Table 1, the credible interval spread is summarized in Table 2, and posterior predictive density is summarized in Table 3.

5.4 Results

5.4.1 Mean Squared Error

CI-TVP creates a perfect pre-treatment fit in all eight simulation studies. This becomes evident when focusing on the mean squared error in the post treatment. *BL* and *BL-TVP* maintain smaller post-treatment mean squared errors in both time varying parameters and time invariant parameters when $T_0 = 17$. When parameters are time invariant, *BL* and *BL-TVP* have a post-treatment mean squared error magnitudes smaller than both version of *CI*. With time varying parameters, *BL* performs worse than *CI-TVP* but significantly better than *CI*. *BL-TVP* produces a post-treatment mean squared error 6 times smaller than *CI*. When *BL-TVP* ranked first or second smallest for all four simulations in which $T_0 = 17$.

When $T_0 = 5$ and $J = 17$, *CI*, *BL*, and *BL-TVP* all perform similarly in terms of mean squared error. *CI-TVP* produces an post treatment mean squared error 3-4 times larger than the other models. In the case of dynamic coefficients, no model is able to recreate a good counterfactual. There simply is not enough data to identify the complex data generating process. In the event of $T_0 = 5$, $J = 5$ and constant weights, time invariant models greatly outperform time varying models. This would be a situation in which the assumption of constant parameters is easier to argue. However, no model performs well with dynamic weights in this setting.

BL-TVP had a lower post treatment mean squared error compared *CI* 6 of the 8 simulations. Similarly, *BL-TVP* had a lower post treatment mean squared error compared *CI-TVP* 6 of the 8 simulations.

5.4.2 95% Credible Interval Spread

BL-TVP credibility interval spread is close in magnitudes to *BL* with $T_0 = 17$. *Bl* and *BL-TVP* maintain tighter credibility intervals than *CI* and *CI-TVP* when $T_0 = 17$. *Bl-TVP* produces slightly larger credibility intervals to *BL* when the data generating process consists of constant parameters and slightly smaller credibility intervals when the data generating

Table 1: Simulation Study of Point Estimates

			Pre-Treament MSE				Post-Treatment MSE			
T_0	J	Coefficient Type	CI	CI-TVP	BL	BL-TVP	CI	CI-TVP	BL	BL-TVP
17	17	Constant	0.68	**0**	0.584	0.276	6.267	8.921	**2.179**	3.066
17	17	Dynamic	7.296	**0**	1.917	0.451	446.492	97.509	153.098	**73.554**
17	5	Constant	0.779	**0**	0.776	0.673	5.392	6.136	**2.353**	3.147
17	5	Dynamic	9.223	**0**	6.545	4.954	455.998	241.129	281.808	**170.251**
5	17	Constant	0.11	**0**	0.095	0.03	5.108	8.827	**5.108**	5.938
5	17	Dynamic	0.354	**0**	0.132	0.018	1502.307	**1389.461**	1513.255	1441.438
5	5	Constant	0.161	**0**	0.331	0.046	7.294	21.612	**4.298**	13.581
5	5	Dynamic	0.446	**0**	0.604	0.037	1536.796	**1335.843**	1626.007	1366.189

* Median results of 1,000 monte carlo simulations with T=34.
† Each simulation of BL-TVP is run 3000 times with a 1500 burn-in.
‡ All other models are run according to presets.
§ The preset Causal Impact model was used as described in Brodersen et al. 2015.
¶ Cells with lowest MSE per simulation and period are bolded.
** CI: Causal Impact
†† CI-TVP: Causal Impact with Time Varying Parameters
‡‡ BL: Bayesian Lasso
§§ BL-TVP: Bayesian Lasso with Time Varying Parameters

process consists of time varying parameters. However, the differences are minuscule. *BL-TVP* maintains smaller credibility intervals than *CI-TVP* in all cases except $T_0 = 5$, $J = 5$.

Table 2: Simulation Study of Credibility Interval Spread Over Whole Sample

T_0	J	Coefficient Type	CI	CI-TVP	BL	BL-TVP
17	17	Constant	6.549	15.866	**5.496**	6.757
17	17	Dynamic	31.638	34.84	23.703	**20.211**
17	5	Constant	16.251	48.427	**19.153**	55.258
17	5	Dynamic	40.203	135.679	38.444	**80.141**
5	17	Constant	**14.41**	37.86	15.726	46.696
5	17	Dynamic	37.273	39.093	**34.7**	59.603
5	5	Constant	**NA**	NA	NA	NA
5	5	Dynamic	NA	NA	**NA**	NA

* Median results of 1,000 monte carlo simulations with T=34.
† Each simulation of BL-TVP is run 3000 times with a 1500 burn-in.
‡ All other models are run according to presets.
§ The preset Causal Impact model was used as described in Brodersen et al. 2015.
¶ Cells with lowest credible interval spread per simulation are bolded
** CI: Causal Impact
†† CI-TVP: Causal Impact with Time Varying Parameters
‡‡ BL: Bayesian Lasso
§§ BL-TVP: Bayesian Lasso with Time Varying Parameters

To add context to the results, consider the simulation in which $T_0 = 17$, $J = 17$, and the coefficients are constants. A researcher may be interested in the minimal average effect that can be detected in the post period. At the 95% probability level, *BL* identifies an average treatment effect over the post period of 23% or more and *BL-TVP* can identify an effect of 28.8% or more. In contrast, *CI* can identify an effect of 35% or more while *CI-TVP* can identify a 66% of the average treatment effect[3]. This demonstrates *BL-TVP* ability to perform similarly to time invariant parameter models when the data generating process only includes time invariant parameters.

5.4.3 95% Coverage

All models achieve optimal coverage in the post treatment period with constant coefficients. However, only *CI-TVP* consistently covers 100% of the post treatment period.

CI-TVP maintains full coverage in the post treatment period. This is primarily due to large credibility intervals. *BL-TVP* suffers from low coverage in all dynamic settings. However, *BL-TVP* achieves significantly higher coverage than *CI* and *BL*. This demonstrates *BL-TVP* can serve as an intermediate alternative between *CI* and *CI-TVP*. The model achieves better coverage in a time varying parameter scenario than *CI* but does not suffer from improbably large credibility intervals in time invariant parameter settings like *CI-TVP*.

6 Empirical Results

BL-TVP is now compared to the leading synthetic control model, Abadie, Diamond, and Hainmueller (2010) (*SC*), and Brodersen et al. (2015) original model (*CI*). The models are compared in two classic synthetic control scenarios: West Germany reunification (Abadie, Diamond, and Hainmueller 2015) and California Tobacco Tax (Abadie, Diamond, and Hainmueller 2010).

[3] Values calculated from additional simulations not included in paper. Simulations are available upon request.

Table 3: Simulation Study of Coverage of Models

			Pre-Treament Coverage				Post-Treatment Coverage			
T_0	J	Coefficient Type	CI	CI-TVP	BL	BL-TVP	CI	CI-TVP	BL	BL-TVP
17	17	Constant	1	1	0.824	1.000	1.000	1.00	0.941	1.000
17	17	Dynamic	1	1	1.000	1.000	0.471	1.00	0.706	0.824
17	5	Constant	1	1	0.706	0.765	1.000	1.00	0.882	0.941
17	5	Dynamic	1	1	0.824	0.824	0.529	1.00	0.294	0.412
5	17	Constant	1	1	1.000	1.000	1.000	1.00	1.000	1.000
5	17	Dynamic	1	1	1.000	1.000	0.276	1.00	0.241	0.931
5	5	Constant	1	1	1.000	1.000	1.000	1.00	1.000	1.000
5	5	Dynamic	1	1	1.000	1.000	0.207	0.31	0.172	0.483

[*] Median results of 1,000 monte carlo simulations with T=34.
[†] Each simulation of BL-TVP is run 3000 times with a 1500 burn-in.
[‡] All other models are run according to presets.
[§] The preset Causal Impact model was used as described in Brodersen et al. 2015.
[¶] Coverage is defined using a 95% credibility interval.
[**] CI: Causal Impact
[††] CI-TVP: Causal Impact with Time Varying Parameters
[‡‡] BL: Bayesian Lasso
[§§] BL-TVP: Bayesian Lasso with Time Varying Parameters

6.1 West Germany Reunification

Abadie, Diamond, and Hainmueller (2015) studied the economic effects of the reunification of East and West Germany in 1990. The effect is measured using Purchasing Power Parity adjusted GDP per capita set to 2002 U.S. Dollars. The analysis focuses on the time period 1960-2003 with 16 OECD countries. This equates to $T = 1990 - 1960 = 30$, $T_0 = 2003 - 1990 = 13$, and $J = 16$.

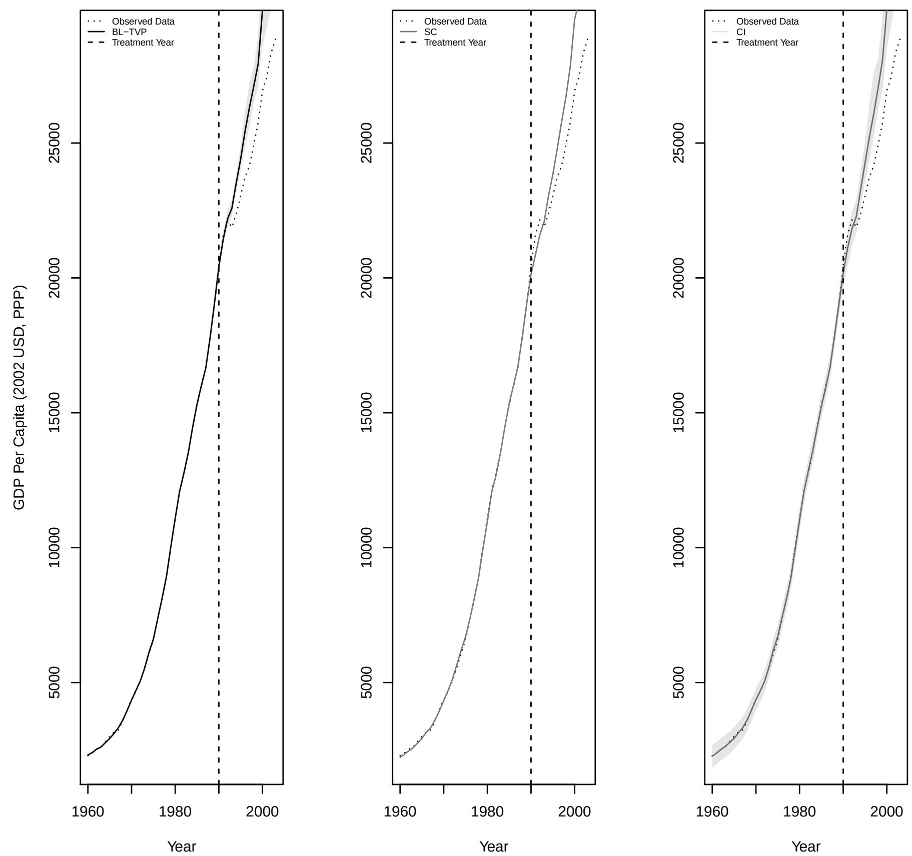

Figure 1: West Germany vs Synthetic Control West Germany obtained from BL-TVP (left), Synthetic Control (center) and Causal Impact (right) with 95% credibility intervals (when applicable).

All three methods match West Germany's pre-treatment period closely and produce similar effects. In the case of CI and BL-TVP, the significance is at the 95% credibility interval.

To further demonstrate, I compare the treatment effects BL-TVP to SC and CI. I rescale the graphs such that 0 is equal to the observed data:

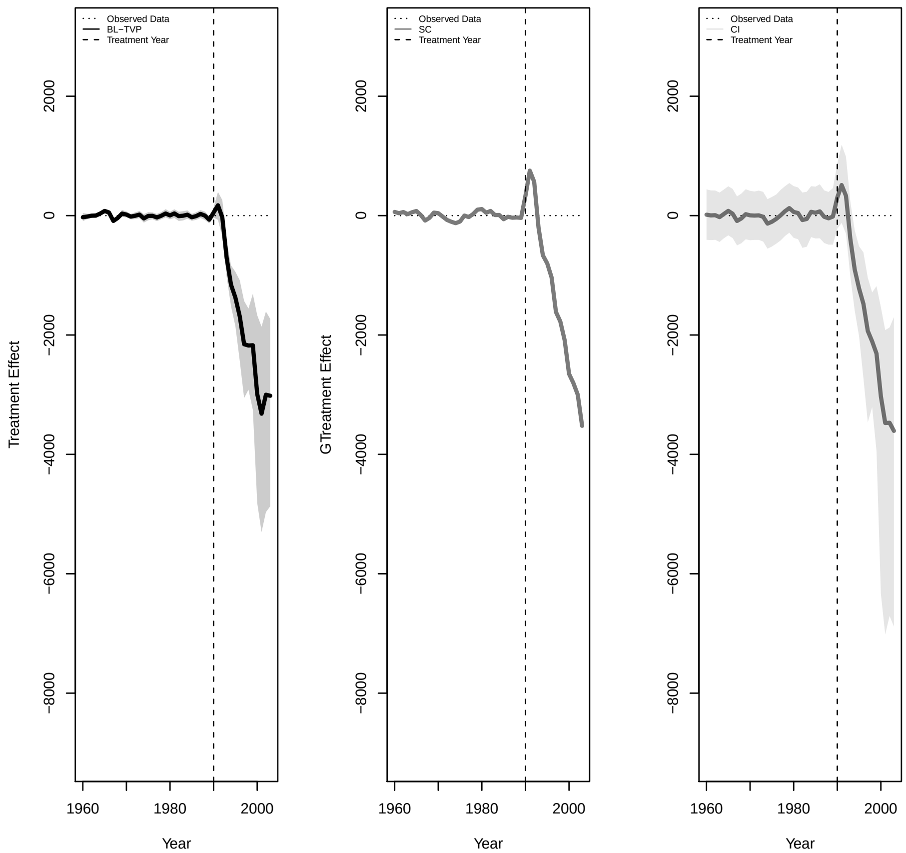

Figure 2: Synthetic Control West Germany vs West Germany Treatment Effects obtained from BL-TVP (left), Synthetic Control (center) and Causal Impact (right) with 95% credibility intervals (when applicable).

Notice BL-TVP closely fits the pretreatment period similarly to SC and CI. In addition, the post-treatment effect closely mirrors the two models, with overlap in the confidence intervals for all post-treatment periods between CI and BL-TVP. When compared to SC, BL-TVP suggests slightly stronger negative effects in the outcome immediately following reunification. However, the long term effects are similar between the two models.

The pre-treatment fit, measured in Mean Squared Error, is significantly tighter than CI and SC. When performing synthetic control analysis, the increased tightness in the pre-period serves as assurance of accurate inference.

Notice all three models produce similar average treatment effects. This demonstrates that the non-centered parameterization can perform as well as pre-existing models in the event that the parameters are not time varying.

Table 4: Average Reduction in West Germany's Annual Per Capita GDP (1990-2003)

Model	Average Decrease	2.5th Percentile Decrease	97.5th Percentile Decrease
BL-TVP	1713.7	975.2	2605.2
CI	1627.1	616.4	3255.7
SC	1322.0	NA	NA

* CI: Causal Impact
† SC: Synthetic Control
‡ BL-TVP: Bayesian Lasso with Time Varying Parameters

6.2 California Tobacco Tax

Abadie, Diamond, and Hainmueller (2010) studied the impacts of a major tobacco tax implemented only in California. The outcome variable was per capita cigarette sales (in packs). The analysis focuses on the time period 1970-2000 with California the single treated unit.The policy was passed in January 1989 leading to 19 years of pre-treatment. 38 other states were used as control units. This equates to $T_0 = 1989 - 1970 = 19$, $T = 2000 - 1970 = 30$, and $J = 32$.

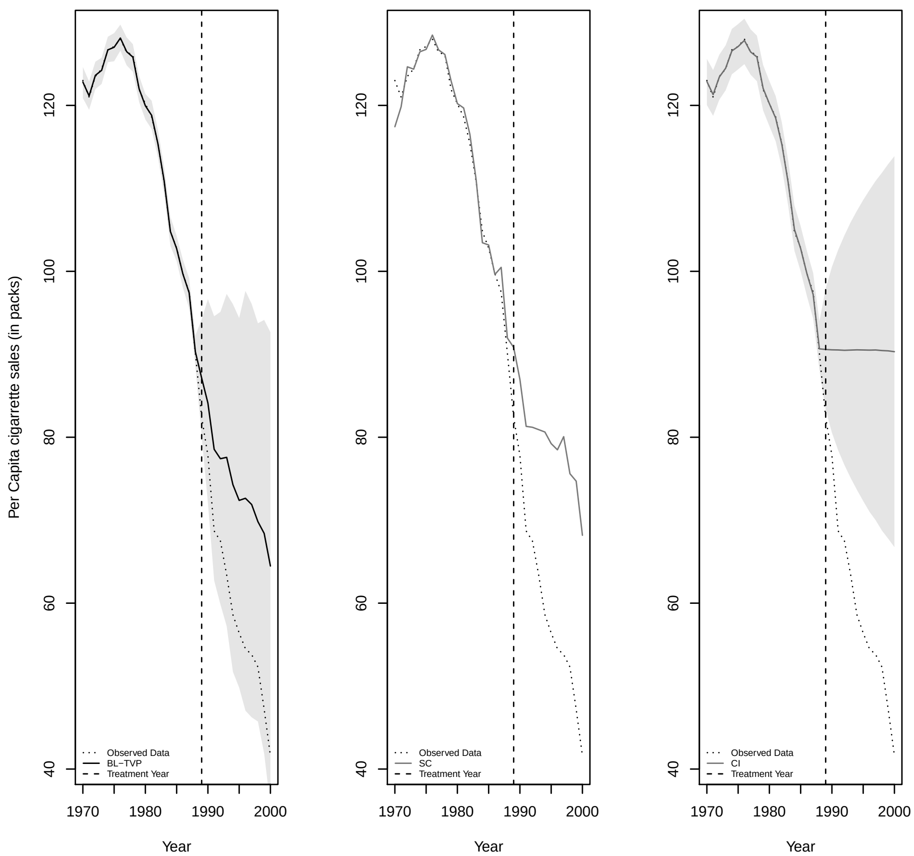

Figure 3: California vs Synthetic Control California obtained from BL-TVP (left), Synthetic Control (center) and Causal Impact (right) with 95% credibility intervals (when applicable).

BL-TVP and SC follow comparable paths while CI diverges immediately following the treatment. This divergence could be indicative of overfitting in the pretreatment. The credibility intervals for BL-TVP are significantly tighter than CI.

The benefits of BL-TVP can be seen even clearer when focusing solely on the treatment effect. The graphs are normalized such that 0 is the observed California. BL-TVP produces tighter intervals in the pre-treatment than CI. However, BL-TVP closely follows SC in the post treatment, something CI fails to do. This demonstrates BL-TVP is able to perform similarly to SC in situations in which there are no time varying coefficients.

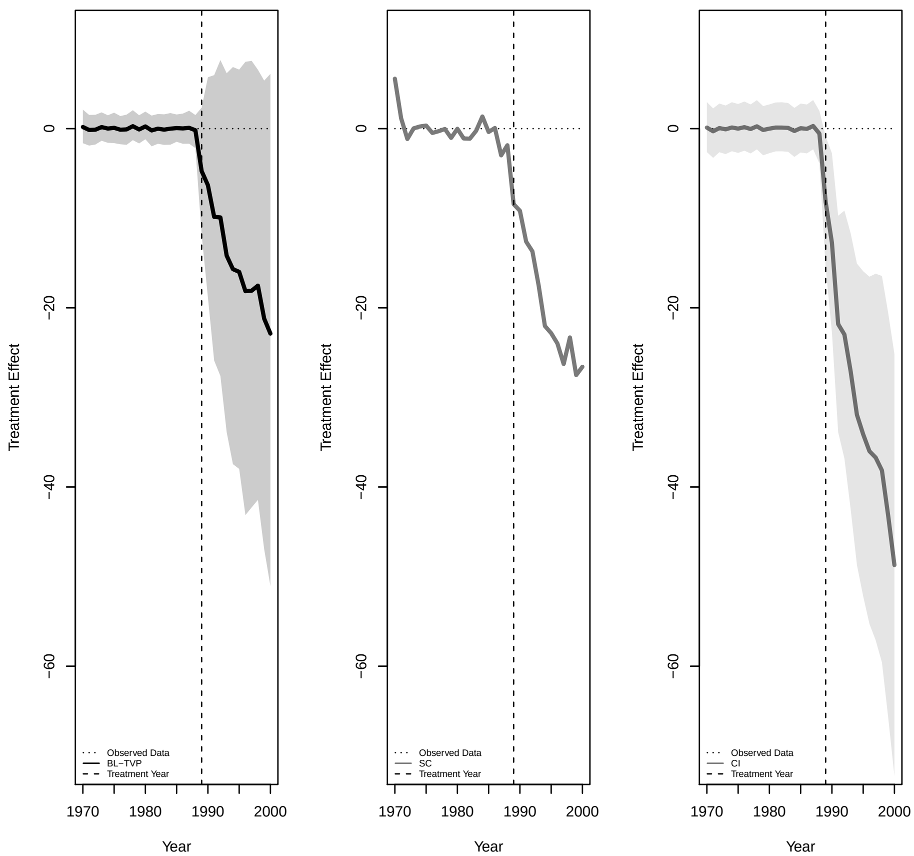

Figure 4: Synthetic Control California vs California obtained from BL-TVP (left), Synthetic Control (center) and Causal Impact (right) with 95% credibility intervals (when applicable).

Similar to West Germany, BL-TVP produces tighter pre-treatment estimates. In both cases, the difference between BL-TVP and SC were a matter of magnitudes.

However, inference differs between the proposed model and the existing models. Both *CI* and *SC* suggest statistically significant benefits from the intervention while *BL-TVP* does not.

7 Conclusion

I propose a time varying parameter approach to synthetic control. Building off of recent advances macroeconometric forecasting, BL-TVP allows empirical researchers to incorpo-

Table 5: Average Reduction in California's Per Capita Cigarette Sales (Post Intervention)

Model	Average Decrease	2.5th Percentile Decrease	97.5th Percentile Decrease
BL-TVP	13.3	-5.8	32.3
CI	27.9	12.2	43.6
SC	18.1	NA	NA

* CI: Causal Impact

† SC: Synthetic Control

‡ BL-TVP: Bayesian Lasso with Time Varying Parameters

rate additional flexibility with reduced risk of overfitting in a synthetic control framework. Simulation results suggest the additional flexibility allows for smaller mean squared forecast errors while increasing coverage compared to existing approaches. The proposed model also performs similarly to establish synthetic control approaches in two classic examples. In the event of time varying relationships between treated and control units, BL-TVP is an easy to implement model that captures the relationship.

8 Work Cited

Abadie, Alberto. 2019. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," 44.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505. https://doi.org/10.1198/jasa.2009.ap08746.

———. 2015. "Comparative Politics and the Synthetic Control Method: COMPARATIVE POLITICS AND THE SYNTHETIC CONTROL METHOD." *American Journal of Political Science* 59 (2): 495–510. https://doi.org/10.1111/ajps.12116.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32. https://doi.org/10.1257/000282803321455188.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2018. "Matrix Completion Methods for Causal Panel Data Models." *arXiv:1710.10251 [Econ, Math, Stat]*, September. http://arxiv.org/abs/1710.10251.

———. 2020. "Matrix Completion Methods for Causal Panel Data Models." *arXiv:1710.10251 [Econ, Math, Stat]*, June. http://arxiv.org/abs/1710.10251.

Belmonte, Miguel A. G., Gary Koop, and Dimitris Korobilis. 2014. "Hierarchical Shrinkage in Time-Varying Parameter Models: Hierarchical Shrinkage in Time-Varying Parameter Models." *Journal of Forecasting* 33 (1): 80–94. https://doi.org/10.1002/for.2276.

Bhattacharya, Anirban, Antik Chakraborty, and Bani K. Mallick. 2016. "Fast Sampling with Gaussian Scale-Mixture Priors in High-Dimensional Regression." *arXiv:1506.04778 [Stat]*, June. http://arxiv.org/abs/1506.04778.

Billmeier, Andreas, and Tommaso Nannicini. 2013. "Assessing Economic Liberalization Episodes: A Synthetic Control Approach." *Review of Economics and Statistics* 95 (3): 983–1001. https://doi.org/10.1162/REST_a_00324.

Bitto, Angela, and Sylvia Frühwirth-Schnatter. 2019. "Achieving Shrinkage in a Time-Varying Parameter Model Framework." *Journal of Econometrics* 210 (1): 75–97. https://doi.org/10.1016/j.jeconom.2018.11.006.

Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2015. "Inferring Causal Impact Using Bayesian Structural Time-Series Models." *The Annals of Applied Statistics* 9 (1): 247–74. https://doi.org/10.1214/14-AOAS788.

Carvalho, Carlos. 2016. "The Perils of Counterfactual Analysis with Integrated Processes." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2894065.

Cattaneo, Matias D., Yingjie Feng, and Rocio Titiunik. 2019. "Prediction Intervals for Synthetic Control Methods." *arXiv:1912.07120 [Econ, Stat]*, December. http://arxiv.org/abs/1912.07120.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *THE REVIEW OF ECONOMICS AND STATISTICS*, 13.

Chernozhukov, Victor, Kaspar Wuthrich, and Yinchu Zhu. 2019. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." *arXiv:1712.09089 [Econ, Stat]*, November. http://arxiv.org/abs/1712.09089.

Dangl, Thomas, and Michael Halling. 2012. "Predictive Regressions with Time-Varying Coefficients." *Journal of Financial Economics* 106 (1): 157–81. https://doi.org/10.1016/j.jfineco.2012.04.003.

Doudchenko, Nikolay, and Guido Imbens. 2016. "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis." w22791. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w22791.

Durbin, J., and S. J. Koopman. 2012. *Time Series Analysis by State Space Methods*. 2nd ed. Oxford Statistical Science Series 38. Oxford: Oxford University Press.

———. 2002. "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis." *Biometrika* 89 (3): 603–16. https://doi.org/10.1093/biomet/89.3.603.

Frühwirth-Schnatter, Sylvia, and Helga Wagner. 2010. "Stochastic Model Specification Search for Gaussian and Partial Non-Gaussian State Space Models." *Journal of Econometrics* 154 (1): 85–100. https://doi.org/10.1016/j.jeconom.2009.07.003.

Grossi, Giulio, Patrizia Lattarulo, Marco Mariani, Alessandra Mattei, and Özge Öner. 2020. "Synthetic Control Group Methods in the Presence of Interference: The Direct and Spillover Effects of Light Rail on Neighborhood Retail Activity." *arXiv:2004.05027 [Econ, Stat]*, June. http://arxiv.org/abs/2004.05027.

Kinn, Daniel. 2018. "Synthetic Control Methods and Big Data." *arXiv:1803.00096 [Econ]*, February. http://arxiv.org/abs/1803.00096.

L'Hour, Alberto Abadie Jeremy. 2019. "A Penalized Synthetic Control Estimator for Disaggregated Data," 53.

Li, Kathleen T. 2019. "Statistical Inference for Average Treatment Effects Estimated by Synthetic Control Methods." *Journal of the American Statistical Association*, December, 1–16. https://doi.org/10.1080/01621459.2019.1686986.

Makalic, Enes, and Daniel F. Schmidt. 2016. "High-Dimensional Bayesian Regularised Regression with the BayesReg Package." *arXiv:1611.06649 [Stat]*, December. http://arxiv.org/abs/1611.06649.

Pang, Xun. 2010. "Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-Sectional Data: A Bayesian Multilevel Model with AR(p) Errors." *Political Analysis* 18 (4): 470–98. https://doi.org/10.1093/pan/mpq019.

Pang, Xun, Licheng Liu, and Yiqing Xu. 2020. "Bayesian Predictive Synthesis for Causal Inference with TSCS Data: A Multilevel State-Space Factor Model with Hierarchical Shrinkage Priors," 54.

Park, Trevor, and George Casella. 2008. "The Bayesian Lasso." *Journal of the American Statistical Association* 103 (482): 681–86. https://doi.org/10.1198/016214508000000337.

Polson, Nicholas G., and James G. Scott. 2011a. "On the Half-Cauchy Prior for a Global Scale Parameter." *arXiv:1104.4937 [Stat]*, September. http://arxiv.org/abs/1104.4937.

———. 2011b. "Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction*." In *Bayesian Statistics 9*, edited by José M. Bernardo, M. J. Bayarri, James O. Berger, A. P. Dawid, David Heckerman, Adrian F. M. Smith, and Mike West, 501–38. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199694587.003.0017.

Powell, David. 2018. "Imperfect Synthetic Controls:" 55.

Rubin, Donald B. 1990. "Formal Mode of Statistical Inference for Causal Effects." *Journal of Statistical Planning and Inference* 25 (3): 279–92. https://doi.org/10.1016/0378-3758(90)90077-8.

Samartsidis, Pantelis, Shaun R. Seaman, Anne M. Presanis, Matthew Hickman, and Daniela De Angelis. 2019. "Assessing the Causal Effect of Binary Interventions from Observational Panel Data with Few Treated Units." *Statistical Science* 34 (3): 486–503. https://doi.org/10.1214/19-STS713.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1): 267–88. http://www.jstor.org/stable/2346178.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25 (1): 57–76. https://doi.org/10.1017/pan.2016.2.

A Appendix

A.1 Deriving Distributions for the Gibbs Sampler

The derivations are based off of Park and Casella (2008). Notable changes have been made for this specific application. Namely, the model is larger, β and $\sqrt{\theta}$ are not conditioned on σ^2, and the hierarchical structure is redefined to be a *global-local* shrinkage estimator. Park and Casella (2008) use a hierarchical formulation where the local shrinkage is dependent on the global shrinkage. Park and Casella (2008) also use an inverse gamma distribution to represent the global shrinkage while this paper opts to use a half Cauchy distribution.

Recall:

$$y_{0,t}(0) = \sum_{j=1}^{J+1} \left(\beta_j + \tilde{\beta}_{j,t}\sqrt{\theta_j}\right) y_{j,t}(0) + \epsilon_t$$

The conditional prior of $\mathbf{y_0}$ is defined as $\mathcal{N}\left(\mathbf{y}\beta_j + (\mathbf{y} * \tilde{\beta}_j)\sqrt{\theta_j}), \sigma^2 I\right)$ where $*$ denotes element wise multiplication. Conditional on α_i^2 and ξ_i^2, the model follows a standard linear regression with normal priors. Textbook tools can be used to derive the distributions for the Gibbs sampler. The joint density is defined as:

$$\begin{aligned}
f(\mathbf{y_0}|\beta, \sqrt{\theta}, &\sigma^2)\pi(\sigma^2)\pi(\lambda^2)\pi(\kappa^2) \prod_{j=1}^{J+1} \pi(\beta_j|\alpha_j^2, \lambda^2)\pi(\alpha_j^2)\pi(\sqrt{\theta_j}|\xi_j^2, \kappa^2)\pi(\xi_j^2) = \\
&\frac{1}{(2\pi\sigma^2)^{\frac{T_0-1}{2}}}\ exp\left\{\frac{-1}{2\sigma^2}\left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right)^T \left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right)\right\} \\
&\frac{a_2^{a_1}}{\Gamma(a_1)}(\sigma^2)^{-a_1-1} exp\left\{-\frac{a_2}{\sigma^2}\right\} \frac{\frac{1}{\zeta_\beta}^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)}(\lambda^2)^{-\frac{1}{2}-1} exp\left\{-\frac{\frac{1}{\zeta_\beta}}{\lambda^2}\right\} \frac{\frac{1}{\zeta_{\sqrt{\theta}}}^{1/2}}{\Gamma\left(\frac{1}{2}\right)}(\kappa^2)^{\frac{-1}{2}-1} exp\left\{-\frac{\frac{1}{\zeta_{\sqrt{\theta}}}}{\kappa^2}\right\} \\
&\frac{1^{1/2}}{\Gamma(1/2)}\zeta_\beta^{-\frac{1}{2}-1} exp\left\{\frac{-1}{\zeta_\beta}\right\} \frac{1^{1/2}}{\Gamma(1/2)}\zeta_{\sqrt{\theta}}^{-\frac{1}{2}-1} exp\left\{\frac{-1}{\zeta_{\sqrt{\theta}}}\right\} \\
&\prod_{j=1}^{J+1} \frac{1}{(2\pi\alpha_j^2\lambda^2)^{\frac{1}{2}}} exp\left\{\frac{-1}{(2\alpha_j^2\lambda^2)}\beta_j^2\right\} exp\left\{-\alpha_j^2\right\} \frac{1}{(2\pi\xi_j^2\kappa^2)^{\frac{1}{2}}} exp\left\{\frac{-1}{(2\xi_j^2\kappa^2)}\sqrt{\theta}_j^2\right\} exp\left\{\xi_j^2\right\}
\end{aligned}$$

A.1.1 Conditional Distribution of β and $\sqrt{\theta}$

To solve for the conditional distribution of β and $\sqrt{\theta}$, drop the terms that don't involve β and $\sqrt{\theta}$. This only leaves 3 exponential terms:

$$exp\left\{\frac{-1}{2\sigma^2}\left(\mathbf{y_0}-\mathbf{y}\beta-(\mathbf{y}*\tilde{\beta})\sqrt{\theta}\right)^T\left(\mathbf{y_0}-\mathbf{y}\beta-(\mathbf{y}*\tilde{\beta})\sqrt{\theta}\right)\right\}$$
$$\prod_{j=1}^{J+1} exp\left\{\frac{-1}{(2\alpha_j^2\lambda^2)}\beta_j^2\right\} exp\left\{\frac{-1}{(2\xi_j^2\kappa^2)}\sqrt{\theta_j}^2\right\}$$

Combining exponents yields:

$$exp\left\{\frac{-1}{2\sigma^2}\left(\mathbf{y_0}-\mathbf{y}\beta-(\mathbf{y}*\tilde{\beta})\sqrt{\theta}\right)^T\left(\mathbf{y_0}-\mathbf{y}\beta-(\mathbf{y}*\tilde{\beta})\sqrt{\theta}\right)+\sum_{j=1}^{J+1}\frac{-\sigma^2}{(2\alpha_j^2\lambda^2)}\beta_j^2+\sum_{j=1}^{J+1}\frac{-\sigma^2}{(2\xi_j^2\kappa^2)}\sqrt{\theta_j}^2\right\}$$

Define:

$$\tilde{\mathbf{y}}=\left[\mathbf{y},\mathbf{y}*\tilde{\beta}\right]_{T_0-1,2(J+1)}$$

,

$$\Theta=\left[\beta,\sqrt{\theta}\right]_{2(J+1),2(J+1)}$$

and

$$D=diag\left[\lambda^2\alpha_1^2,...\lambda^2\alpha_{J+1}^2,\kappa^2\xi_1^1,\ldots,\kappa^2\xi_{J+1}^2\right]_{2(J+1),2(J+1)}$$

.

Focusing solely on the exponential term and rearranging yields:

$$\frac{-1}{2\sigma^2}\left[(\mathbf{y_0}-\tilde{\mathbf{y}}\Theta)^T(\mathbf{y_0}-\tilde{\mathbf{y}}\Theta)+\Theta^T\sigma^2V^{-1}\Theta\right]$$

Multiplying out and rearranging yields:

$$\frac{-1}{2\sigma^2}\left[\mathbf{y_0}^T\mathbf{y_0} - 2\mathbf{y_0}\tilde{\mathbf{y}}\Theta + \Theta^T(\tilde{\mathbf{y}}^T\mathbf{y_0} + \sigma^2 V^{-1})\Theta\right]$$

Focus solely on the terms within the brackets including Θ for a moment. Setting $A = \tilde{\mathbf{y}}^T\tilde{\mathbf{y}} + \sigma^2 V^{-1}$ and completing the square yields:

$$\begin{aligned}&\left(\Theta - A^{-1}\tilde{\mathbf{y}}^T y\right)^T A\left(\Theta - A^{-1}\tilde{\mathbf{y}}^T y\right)\\ &+ y^T\left(I - \tilde{\mathbf{y}}A^{-1}\tilde{\mathbf{y}}^T\right)y\end{aligned}$$

Therefore, the part of the conditional distribution that relies on Θ can be written as:

$$\frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{\frac{-1}{2\sigma^2}\left(\Theta - A^{-1}\tilde{\mathbf{y}}^T\mathbf{y_0}\right)^T A\left(\Theta - A^{-1}\tilde{\mathbf{y}}^T\mathbf{y_0}\right)\right\}$$

which can be summarized as Θ conditionally distributed as:

$$\mathcal{N}\left(A^{-1}\tilde{\mathbf{y}}^T\mathbf{y_0}, \sigma^2 A^{-1}\right)$$

A.1.2 Conditional Distribution of σ^2

Now, I will derive the conditional distribution for σ^2. Returning to the joint probability, drop all terms that do not include σ^2:

$$\begin{aligned}&\frac{1}{(\sigma^2)^{\frac{T_0-1}{2}}}\ exp\left\{\frac{-1}{2\sigma^2}\left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right)^T\left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right)\right\}\\ &(\sigma^2)^{-a_1-1} exp\left\{-\frac{a_2}{\sigma^2}\right\}\end{aligned}$$

Rearranging yields:

$$(\sigma^2)^{-\frac{T_0-1}{2}-a_1-1}\ exp\left\{\frac{-1}{2\sigma^2}\left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right)^T\left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right) - \frac{a_2}{\sigma^2}\right\}$$

which is an inverse gamma distribution without the scaling term. Therefore, σ^2 is conditionally inverse gamma with *shape* parameter $\frac{T_0-1}{2} + a_1$ and *scale* parameter

$\frac{1}{2}\left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right)^T \left(\mathbf{y_0} - \mathbf{y}\beta - (\mathbf{y} * \tilde{\beta})\sqrt{\theta}\right) + a_2.$

A.1.3 Conditional Distribution of α_j^2 and ξ_j^2

Focusing only on terms involving α_j^2, the conditional distribution is:

$$\frac{1}{(\alpha_j^2)^{1/2}} exp\left\{\frac{-1}{(2\alpha_j^2\lambda^2)}\beta_j^2 - \alpha_j^2\right\}$$

Park and Casella (2008) note that by setting $\frac{1}{\alpha_j^2} = \zeta^2$, the density can be rewritten proportionally as an inverse Gaussian:

$$\begin{aligned}
(\zeta^2)^{-3/2} exp\left\{-\left(\frac{\beta_j^2\zeta_j^2}{2\lambda^2} + \alpha_j^2\right)\right\} &\propto (\zeta^2)^{-3/2} exp\left\{\frac{-\beta_j^2}{2\zeta^2\lambda^2}\left[\zeta^2 - \sqrt{\frac{2\lambda^2}{\beta_j^2}}\right]^2\right\} \\
&= (\zeta^2)^{-3/2} exp\left\{\frac{-2}{2\zeta^2\frac{2\lambda^2}{\beta_j^2}}\left[\zeta^2 - \sqrt{\frac{2\lambda^2}{\beta_j^2}}\right]^2\right\}
\end{aligned}$$

This is one of many parameterizations of the Inverse Gaussian distribution. The Inverse Gaussian distribution can be written as:

$$f(x) = \sqrt{\frac{\lambda'}{2\pi}}\, x^{-3/2} exp\left\{-\frac{\lambda'(x-\mu')^2}{2(\mu')^2 x}\right\}$$

with mean parameter μ' and scale parameter λ.

Therefore, $\frac{1}{\alpha_j^2}$ is conditionally distributed Inverse Gaussian with mean parameters $\frac{2\lambda^2}{\beta_j^2}$ and scale parameter $\lambda' = 2$. ξ_j^2 is derived following the same steps.

A.1.4 Conditional Distribution of λ^2 and κ^2

Focusing solely on λ^2 in the joint distribution yields:

$$\left(\lambda^2\right)^{-\frac{J+2}{2}-1} exp\left\{\left(-\frac{\sum_{j=1}^{J+1}\alpha_j^2}{2} - \frac{1}{\zeta_\beta}\right)\frac{1}{\lambda^2}\right\}$$

which is proportional to an inverse gamma distribution with *shape* parameter $\frac{J+1}{2}$ and *rate* parameter $\frac{1}{\zeta_\beta} + \frac{1}{2}\sum_{j=1}^{J+1} \frac{\beta_j^2}{\alpha_j^2}$.

Similarly, κ^2 will follow an inverse gamma distribution with *shape* parameter $\frac{J+1}{2}$ and *rate* parameter $\frac{1}{\zeta_{\sqrt{\theta}}} + \frac{1}{2}\sum_{j=1}^{J+1} \frac{\sqrt{\theta_j}^2}{\xi_j^2}$.

A.1.5 Sample $\zeta_{\sqrt{\theta}}$ and ζ_β

Finally, ζ_β will follow an inverse gamma with shape 1 and rate $1 + \frac{1}{\lambda^2}$. Similarly, $\zeta_{\sqrt{\theta}}$ will follow an inverse gamma with shape 1 and rate $1 + \frac{1}{\kappa^2}$.