Shrinkage Among Time Varying Weights in Counterfactual Analysis

Danny Klinenberg

Last Updated: 2020-04-17

Abstract

Shrinkage will be introduced to time varying parameters in a synthetic control framework using state space models utilizing the Bayesian Lasso. This class of state space models has been shown to produce better out of sample predictions in the presence of time varying coefficients as compared to constant coefficient models. In addition, these time varying coefficient models perform similarly to constant coefficient models without the presence of time varying coefficients. The introduction of time varying parameter shrinkage state space models to a synthetic control framework should allow for valid inference in situations where the current tools are inadmissable. A simulation study comparing time varying parameter shrinkage (TVPS) state space models to traditional synthetic control will be conducted. The contribution of this paper will introduce TVPS state space models to the synthetic control framework.

1 Introduction

Synthetic control is a panel data approach to analyze the effect of a treatment on very few treated units. In many cases, there is only one treated unit. Uses of synthetic control have

included the effect of joining the European Union on economic growth (Campos, Coricelli, and Moretti 2019) and the effect of terrorism on GDP (Abadie and Gardeazabal 2003). The intuitive idea behind synthetic control methods is to construct a counterfactual for a treated unit using a weighted average of the untreated units. The weights are assumed to be constant in pre and post treatment. If the constructed, or synthetic, observation fits the treated observation "well before" the intervention, then the synthetic observation is said to have worked and qualifies as a counterfactual. "Fits the treated observation" means the synthetic control is almost perfectly matching in the pretreatment. "Well before" has been defined differently by different authors, but a general consensus is at least ten pre-periods. A causal interpretation can then be made on the effect of the treatment in post treatment periods by comparing the synthetic unit to the actual treated unit.

Causal inference in a synthetic control framework is jeopardized when the weights used for the counterfactual are non-constant. This would lead to the synthetic control not "fitting the treated observation". Common diagnostic checks on synthetic control output should alert the researcher of such violations. With such a violation, the synthetic control would be inadmissible. One solution is to explicitly account for the changing relationship using dynamic coefficient state space models. The addition of changing weights can lead to the synthetic control "fitting well" while the constant weights assumptions is violated. However, using all dynamic coefficients will lead to overfitting and implausibly large probability intervals (Brodersen et al. 2015). Recent developments in time varying parameter state space models have introduced shrinkage estimators to reduce dynamic coefficients to static ones in the event of overfitting. Time varying parameter shrinkage (TVPS) state space models have been utilized in macroeconometrics for inflation predictions and stock returns, but not yet synthetic control ((Belmonte, Koop, and Korobilis 2014) and (Frühwirth-Schnatter and Wagner 2010)). This research introduces TVPS state space models to the synthetic control framework. The introduction will be done through a simulation study.

2 Working Literature Review

In many research endeavors, a true counterfactual does not exist. In addition, there are relatively few treated observation. Therefore, traditional tools such as difference in differences, regression discontinuity, or simple randomization cannot be applied. Synthetic control was proposed as an alternative in which observations unaffected by a policy are pooled together to create a counterfactual for the treated observations. This was first popularly employed in Abadie and Gardeazabal (2003) and formalized in a followup paper (Abadie, Diamond, and Hainmueller 2010) (referred to as *ADH Synthetic Control*). If there exists constant weights bounded between 0 and 1 that sum to unity such that a weighted average of control observations match the treated observation's covariates well in pre-treatment periods, then that average is used as a counterfactual in the post treatment periods.

Recent developments in synthetic control have focused on three goals: the bounded weights assumption, adding more treated observations, and matching on covariates. Doudchenko and Imbens (2016) first noted that ADH synthetic control mirrors machine learning processes. They then suggested using an alternative machine learning process, elastic net, to produce counterfactuals. Similarly, Athey et al. (2018) approached the synthetic control problem from a machine learning perspective using matrix completion methods. The idea viewed developing a counterfactual as a missing data problem in a matrix. Both methods do not require the weights sum to unity and be bounded. Xu (2017) and Powell (2018) generalized ADH synthetic controls to include many treated observations with treatment heterogeneity and start dates. Finally, Kaul et al. (2018) compared creating weights through matching on covariates (ADH Synthetic Control) to only matching on the outcome and showed that matching on covariates provided little benefit to estimation.

Fully parametric methods have also been proposed to analyze counterfactuals. Scott and Varian (2013) present Bayesian Structural Time Series for prediction. This method uses a Bayesian variable selection process, spike and slab, to identify relevant variables for pre-

diction. Like synthetic controls, Scott and Varian (2013) performs variable selection by shrinking the coefficient on irrelevant covariates to 0. This was then extended by Brodersen et al. (2015) as an alternative for synthetic control. Both Scott and Varian (2013) and Brodersen et al. (2015) warn about the use of time varying coefficients in their methods. Their method focuses on shrinkage among the coefficients only. This means the shrinkage determines if a variable should be included or not. This method does not differentiate between static and dynamic inclusion. This is set exogenously by the researcher. Assuming all relevant variables are dynamic leads to both overfitting and implausibly large probability intervals.

In many situations, time varying coefficients better capture the relationship between variables. Dangl and Halling (2012) compared the out of sample predictability on the S&P 500 monthly returns of time varying coefficient state space models and constant coefficient state space models. The authors found that the time varying coefficient models had both economic and statistical improvements. Belmonte, Koop, and Korobilis (2014) then added a shrinkage estimator, the Bayesian Lasso, to the dynamic components of the state space model. This was done to curb overfitting while gaining the benefits of time varying coefficients. The approach found relative success with the main advantage being that the TVPS state space model performed equally well to the constant parameter state space model in constant situations. Bitto and Frühwirth-Schnatter (2019) recently proposed using a *double gamma* prior to achieve shrinkage in TVPS state space models. This proposal has shown significant improvements in one step ahead forecasts compared to constant parameter models. Nonparametric solutions to shrinkage of time varying coefficients have also been proposed. One example was Kapetanios and Zikes (2018) proposed time varying lasso.

2.1 Linear Gaussian State Space Modeling

Identifying time varying coefficients can be thought of as a latent variable estimation problem. State space modeling is a time series concept that allows for modeling latent variables explicitly. This means modeling unobserved components like time trends, seasonality, and time varying coefficients. A state space model is composed of an observation equation and transition equation. A general form of these equations follows:

$$y_t = Z_t \alpha_t + \epsilon_t \quad \text{observation equation}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \text{transition equation}$$

$$\alpha_0 \sim N(a_0, P_0)$$

where $\epsilon_t \sim N(0, \sigma_t^2)$ and $\eta_t \sim N(0, Q_t)$ are independent of all unknown factors. y_t is the observed data and α_t is a combination of observed data (e.g. control variables) and unobserved components (e.g. trend and cycle). In the case of a scalar output, y_t, with m variables and r time varying components, Z_t would be a 1 x m dimensional matrix, α_t a m x 1 matrix, and ϵ_t a scalar. α_{t+1} would also be a m x 1 matrix, T_t an m x m matrix, R_t a m x r matrix and Q_t an r x r matrix. Finally, a_0 is m x 1 and P_0 is m x m. Linear Gaussian state space models are structural models. The assumptions necessary for linear Gaussian state space models are:

1) $\epsilon_t \sim N(0, \sigma_t^2)$ and $\eta_t \sim N(0, Q_t)$. The errors are also assumed to be serially uncorrelated. This is because they are meant to be random disturbances within the model.

2) The errors must be normal.

3) the transition equations can be of lag order 1. Any additional lag orders can be rewritten as order 1 using the state space framework.

Recent developments in time varying parametric estimation have extended shrinkage to both

the static and dynamic portion of coefficients. This means coefficients are biased towards being static or irrelevant, similarly to frequentest shrinkage.

This proposal applies Belmonte, Koop, and Korobilis (2014) method to synthetic control. The method focused on using the Bayesian Lasso first proposed by Park and Casella (2008) to shrink the static and dynamic portion of coefficients. This is a direct extension of ADH synthetic control: Kinn (2018) argued synthetic control is a restricted version of Lasso similarly to how Doudchenko and Imbens (2016) compared elastic net to synthetic control. Bayesian Lasso is identical to Lasso under independent Laplace priors (Park and Casella 2008). This methodology applies Bayesian Lasso to time varying parameters.

3 The Model

3.1 Initial Setup

The model proposed by Belmonte, Koop, and Korobilis (2014) will be implemented. Suppose the state space model is defined as:

$$y_{1,t} = \sum_{j=2}^{J+1} \beta_{j,t} y_{j,t} + \epsilon_t \qquad\qquad \epsilon_t | \sigma^2 \sim N(0, \sigma^2) \qquad\qquad (1)$$

$$\beta_{j,t} = \beta_{j,t-1} + \eta_{j,t} \qquad\qquad \eta_{j,t} \sim N(0, \theta_j) \quad \forall j \qquad\qquad (2)$$

$$\beta_{j,0} \sim N(\beta_j, \theta_j P_{jj}) \qquad\qquad\qquad \forall j \qquad\qquad (3)$$

where P_{jj} is a hyperparameter. This specification of θ_j lends itself to a useful interpretation: θ_j governs the dynamics of $\beta_{j,t}$ [Bitto and Frühwirth-Schnatter (2019).

The errors are assumed to be independent of one another and independent of all leads and lags. The errors between coefficients are assumed to be independent (e.g. $cov(\eta_{j,t}, \eta_{i,t}) = 0$

for $i \neq j$). This assumption is to keep the model relatively parsimonious (Belmonte, Koop, and Korobilis (2014), Bitto and Frühwirth-Schnatter (2019)). If $\theta_j = 0$, then β_j is a static coefficient. Past counterfactual analysis papers have introduced shrinkage to β_j in a synthetic control framework (Brodersen et al. 2015). This paper will maintain shrinkage of β_j while adding shrinkage to θ_j in a synthetic control framework.

The transition equation can be rewritten to decompose β_j into a time varying and constant components (Frühwirth-Schnatter and Wagner 2010).

$$\beta_{j,t} = \beta_j + \tilde{\beta}_{j,t}\sqrt{\theta_j} \tag{4}$$

$$\tilde{\beta}_{j,t} = \tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t} \qquad\qquad \tilde{\eta}_{j,t} \sim N(0,1) \tag{5}$$

$$\tilde{\beta}_{j,0} \sim N(0, P_{jj}) \tag{6}$$

To verify these representations of $\beta_{j,t}$ are equal, note:

$$\beta_{j,t} - \beta_{j,t-1} = (\beta_j + \sqrt{\theta_j}\tilde{\beta}_{j,t}) - (\beta_j + \sqrt{\theta_j}\tilde{\beta}_{j,t-1}) \quad \text{Plugging in (4)}$$

$$= \sqrt{\theta_j}(\tilde{\beta}_{j,t} - \tilde{\beta}_{j,t-1}) \qquad\qquad \text{Regroup}$$

$$= \sqrt{\theta_j}(\tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t} - \tilde{\beta}_{j,t-1}) \qquad\qquad \text{Plug in (5)}$$

$$= \sqrt{\theta_j}\tilde{\eta}_{j,t} \qquad\qquad \text{Simplify}$$

Notice that $\tilde{\eta}_{j,t} \sim N(0,1)$. Therefore $\sqrt{\theta_j}\tilde{\eta}_{j,t} \sim N(0, \theta_j)$ which is $\eta_{j,t}$.

β_j can now be interpreted as the point estimate of $\beta_{j,t}$ and $\sqrt{\theta_j}$ the time varying portion. The advantage of this formulation is that shrinkage estimation can be performed on both aspects of the coefficient. Substituting the reformulation back into the original equation yields the state space model:

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

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

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

Frühwirth-Schnatter and Wagner (2010) refer to this setup as the *non-centered parameterization of state space models*. This is extremely useful because the problem of *variance selection* has now been recast as one of *variable selection*. Variable selection problems are far better understood and applied. However, there are some additional precautions that must be made when working with non-centered parameterization of state space models. One such issue is an identification problem arises in that $\sqrt{\theta_j}\tilde{\beta}_{j,t}$ can be replaced by $(-\sqrt{\theta_j})(-\tilde{\beta}_{j,t})$ without affecting the likelihood function. The appropriate solution to this issue is discussed in Frühwirth-Schnatter and Wagner (2010) and implemented in the Gibbs sampler.

The setup allows for four possibilities:

i) constant coefficient (β_j not shrunk to 0 but $\sqrt{\theta_j}$ shrunk to 0)

ii) irrelevant coefficient (β_j shrunk to 0 and $\sqrt{\theta_j}$ shrunk to 0)

iii) small time-varying coefficient (β_j shrunk to 0 but $\sqrt{\theta_j}$ not shrunk to 0)

iv) time-varying coefficient (β_j not shrunk to 0 and $\sqrt{\theta_j}$ not shrunk to 0)

3.2 The Priors

The goal of this research is to utilize Bayesian Lasso because of it's direct relationship to ADH synthetic control. Park and Casella (2008) showed the Bayesian Lasso is identical to Lasso under independent LaPlace priors which is identical to a mean 0 normal distribution with variance defined as exponential. Following in suit, assign the priors for β and τ_j^2 as:

$$| ^2, ^2 \sim N(0, \sigma^2 diag[\tau_2^2, ... \tau_J^2])$$

$$\tau_j^2 | \lambda^2 \sim exp\left(\frac{\lambda^2}{2}\right)$$

λ^2 can then be calculated via MLE or with it's own prior. Staying true to (Park and Casella 2008), set $\lambda^2 \sim Gamma(b_1, b_2)$, with b_1 and b_2 hyperparameters to be set.

Traditionally, variances have been defined by the inverse gamma distribution. However, the inverse gamma does not allow for effective shrinkage given it's support. Frühwirth-Schnatter and Wagner (2010) provide an in depth argument for the use of the normal distribution as an alternative. Briefly, inverse gammas perform poorly in terms of shrinkage and the normal distribution does not. Similarly to β_j, assign the prior of $\sqrt{\theta}$ as:

$$\sqrt{\theta} | \xi^2 \sim N(0, \sigma^2 diag[\xi_2^2, ... \xi_J^2])$$

$$\xi_j^2 | \kappa^2 \sim exp\left(\frac{\kappa^2}{2}\right)$$

where $\kappa^2 \sim Gamma(c_1, c_2)$ with c_1 and c_2 as hyperparameters.

Finally σ^2 is defined as $\frac{1}{\sigma^2} \sim Gamma(a_1, a_2)$ with a_1 and a_2 as hyperparameters. Again, the purpose for defining the priors as such is to recreate the Bayesian Lasso. The Bayesian Lasso is identical to the frequentest Lasso with the priors described above and ADH synthetic control is a restricted version of Lasso (Kinn 2018).

4 Identifying Assumptions

Thus far, this paper has focused on the estimation technique. In order to gain causal inference, two assumptions must be implemented:

i) The control time series are unaffected by the treatment. If this were to be violated, the causal estimates could be biased.

ii) The dynamic relationship between the treated variable and the controls established in the pretreatment periods does not change.

These are not the only assumptions being made within the model. Every prior choice, hyperparameter, and state space formulation are also assumptions being made.

5 Methods

6 Monte Carlo Simulation Data

The simulation is restricted to the outcomes of the observed units, without considering covariates. A growing body of literature has supported synthetic control analysis without covariates. Athey and Imbens (2017) and Doudchenko and Imbens (2016) argue the outcomes tend to be far more important than covariates in terms of predictive power. They further argue that minimizing the difference between treated outcomes and control outcomes prior to treatment tend to be sufficient to construct a synthetic control. Kaul et al. (2018) showed covariates become redundant when all lagged outcomes are included in ADH approach. Botosaru and Ferman (2019) showed the counterfactual estimated by using only pre-treatment outcomes were very close to the original ADH. Brodersen et al. (2015) opt to omit covariates. Finally, both Kinn (2018) and Samartsidis et al. (2019) do not use covariates in their model comparisons.

For the purpose of this paper, the argument that covariates follow the same time varying weight structure as the outcome would be hard to rationalize theoretically or empirically. Because of this, the simulation opts to avoid covariates entirely.

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

$$y_{j,t}(0) = \xi_{j,t} + \psi_{j,t} + \epsilon_{j,t} \qquad\qquad \text{j=1,..,J}$$

$$y_{1,t}(0) = \sum_{j=2}^{J} w_{j,t}(\xi_{j,t} + \psi_{j,t}) + \epsilon_{1.t}$$

for t=1,..,T where ξ_{jt} is the trend component, ψ_{jt} is the 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 $\psi_{j,t} = \gamma_j sin\left(\frac{\pi t}{\rho_j}\right)$. Parallel trends are created when $c_j = c \ \forall \ j$ and $\gamma_j = 0 \ \forall \ j$. The explicit data generating process is:

$$y_{j,t}(0) = c_j t + z_j + \gamma_j sin\left(\frac{\pi t}{\rho_j}\right) + \epsilon_{j,t} \qquad\qquad \text{j=2,..,J}$$

$$y_{1,t}(0) = \sum_{j=2}^{J} w_{j,t}\left(c_j t + z_j + \gamma_j sin\left(\frac{\pi t}{\rho_j}\right)\right) + \epsilon_{1.t}$$

Following Kinn (2018), a sparse set of controls will have nonzero weights. This means properly identifying the correct controls will be important for an accurate counterfactual. The treatment begins at period T_0. The treatment effect is initially set to 0.

This paper proposes one scenario to test continuous time varying weights.

6.1 Deterministic Continuous Varying Weights

To simulate continuous varying weights, $c_{2,t}$ and $c_{3,t}$ are defined .75 and .25 respectively. All other $c_{j,t}$ are randomly drawn from U[0,1]. In order to avoid $y_{2,t}$ and $y_{3,t}$ from crossing, set $z_2 = 25$ and $z_3 = 5$. In addition, set $\psi_{j,t} = 0$ for all j,t. Finally, define $w_{2,t} = .2 + .6\frac{t}{T}$ and $w_{3,t} = 1 - w_{2,t}$. In order to compare to ADH, the sum of the weights was sets to unity. This ensure the convex hull assumption is met.

To summarize, the parameters of this simulation are:

1) $c_{2,t} = .75$, $c_{3,t} = .25$, and $c_{j,t} \sim U[0,1]$ for all $j \notin \{2,3\}$

2) $z_2 = 25$, $z_3 = 5$ and z_j is sampled from $\{1, 2, 3, 4, ..., 50\}$.

3) $\epsilon_{j,t} \sim N(0, 1)$.

4) T = 50, $T_0 = 30$.

5) J = 51.

6) $w_{2,t} = .2 + .6\frac{t}{T}$, $w_{3,t} = 1 - w_{2,t}$, and $w_{j,t} = 0$ for all else

7) $\gamma_j = 0 \ \forall j$.

The data generating process can be rewritten in recursive form:

$$y_{1,t}(0) = \sum_{j=2}^{J} w_{j,t} \left(c_j t + z_j + \gamma_j sin \left(\frac{\pi t}{\rho_j} \right) \right) + \epsilon_{1.t}$$

$$w_{2,t} = w_{2,t-1} + \frac{.6}{T}$$

$$w_{3,t} = w_{2,t-1} - \frac{.6}{T}$$

$$w_{j,t} = w_{j,t-1} \qquad\qquad\qquad j \notin \{1, 2, 3\}$$

with initial conditions:

$$w_{2,0} = .2$$

$$w_{3,0} = .8$$

$$w_{j,0} = 0 \quad j \notin \{1, 2, 3\}$$

6.2 Model Testing and Comparison

This simulation will test the accuracy of the estimates of the treatment effect and the accuracy of the inference (significant or not). The treatment effect sizes will be tested at 0%, 0.1%, 1%, 10%, and 100% similarly to Brodersen et al. (2015). These treatment effects will

be calculated by defining $Y_{1,t}(1) = \rho Y_{1,t}(0)$ for $\rho \in \{1, 1.001, 1.01, 1.1, 2\}$. For inference, a causal effect is assumed only if 95% of the posteriod probability interval excludes 0.

In order to compare the TVPS state space model to ADH synthetic control, the median observation of the posterior distribution at each post treatment period will be used. ADH synthetic control does not have a confidence interval, so all comparisons must be done as point estimates. This test will compare the recovered treatment effect size versus the actual. Zero percent, 0.1%, 1%, 10%, and 100% will be used for treatment effect sizes with $Y_{1,t}(1)$ defined as before.

7   Conclusion

This proposal adds shrinkage among time varying weights to counterfactual analysis. The addition of shrinkage among time varying weights will extend the scope of synthetic control to data previously restricted from analysis as well as add to the very limited existing literature of state space models in counterfactual analysis. Future research will include extending the model to multiple outcome variables (e.g. GDP and unemployment).

Work Cited and References

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.

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.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31 (2): 3–32. https://doi.org/10.1257/jep.31.2.3.

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.

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.

Botosaru, Irene, and Bruno Ferman. 2019. "On the Role of Covariates in the Synthetic Control Method." *The Econometrics Journal*, January, utz001. https://doi.org/10.1093/ectj/utz001.

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.

Campos, Nauro F., Fabrizio Coricelli, and Luigi Moretti. 2019. "Institutional Integration and Economic Growth in Europe." *Journal of Monetary Economics* 103 (May): 88–104. https://doi.org/10.1016/j.jmoneco.2018.08.001.

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: Na-

tional Bureau of Economic Research. https://doi.org/10.3386/w22791.

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.

Kapetanios, George, and Filip Zikes. 2018. "Time-Varying Lasso." *Economics Letters* 169 (August): 1–6. https://doi.org/10.1016/j.econlet.2018.04.029.

Kaul, Ashok, Stefan Kloßner, Gregor Pfeifer, and Manuel Schieler. 2018. "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together with Covariates," 24.

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

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.

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

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.

Scott, Steven, and Hal Varian. 2013. "Bayesian Variable Selection for Nowcasting Economic Time Series." w19567. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w19567.

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.