

Shrinkage Among Time Varying Weights in Counterfactual Analysis

Danny Klinenberg

Last Updated: 2020-04-15

Abstract

I plan to introduce shrinkage to time varying parameters in a synthetic control framework using state space models. I plan to do this using the Bayesian Lasso. Past state space models focus on shrinkage of the point estimate. This means shrinkage on the constant part of the coefficients. I introduce models that include shrinkage on the time-varying part of the coefficients to the synthetic control framework. I referred to these models as Time Varying Parameter Shrinkage (TVPS) state space models. I plan to conduct a simulation study comparing TVPS state space models to traditional synthetic control. My objective is to demonstrate this class of shrinkage estimators perform just as well as synthetic control in a traditional synthetic control scenario as well as in situations where synthetic controls assumptions are not met. My contribution is the introduction of 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 2019a) and effect of terrorism on GDP (Abadie and Gardeazabal 2003). The intuitive idea behind synthetic control methods is to construct a counterfactual for the treated observations using a weighted average of the untreated. 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 periods. A causal interpretation can then be made on the effect of the treatment in following periods by comparing the synthetic observation to the actual.

Causal inference in a synthetic control framework is jeopardized when the weights used for the counterfactual are non-constant. This means that the synthetic control does not “fit the treated observation”. Common diagnostic checks on synthetic control output should alert the researcher to violations. The problem then becomes many answers to important questions are unattainable. One solution is to explicitly account for the changing relationship using dynamic coefficients in 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. 2015a). 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)). My contribution is introducing this class of models to the synthetic control framework. The introduction will be done through a simulation study.

2 Working Literature Review

2.1 Current State of Counterfactual Analysis: ADH Synthetic Control

In many research endeavors, a true counterfactual does not exist. In addition, there are relatively few treated observations. This means a researcher cannot use traditional tools such as difference in differences, regression discontinuity, or simple randomization. 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. ADH is a powerful tool that has spurred a whole line of literature.

Recent developments in synthetic control have focused on three goals: the bounded assumption, adding more treated observations, and matching on covariates. (Doudchenko and Imbens 2016) first noted that ADH synthetic control mirror 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, n.d.) generalized ADH synthetic controls to include many treated observations with treatment heterogeneity and start dates. Finally, (Kaul et al., n.d.) compared creating weights through matching on covariates (ADH Synthetic Control) to only matching on the outcome. This paper showed that matching on covariates provided little benefit to estimation.

Suppose there is an outcome of interest, Y_{it} . Suppose there are $N+1$ observations observed over T years with one observation being exposed to treatment. Without loss of generality, assume observation $i = 1$ is treated starting at time T_0 . Furthermore, define the treatment D such that $D_{1t} = 1 * I(t \geq T_0)$ with treatment effect $treat_{1,t}$. Using the potential outcomes framework, assume there are two states of the world: untreated $Y_{it}(0)$ and treated $Y_{it}(1) = treat_{it}D_{it} + Y_{it}(0)$. The objective is to measure:

$$treat_{1t} = Y_{1t}(1) - Y_{1t}(0)$$

Furthermore, assume that $Y_{it}(0)$ comes from the following data generating process:

$$Y_{it}(0) = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it}$$

where δ_t is an unknown common time factor constant across units, θ_t is a $(1 \times r)$ of unknown parameters, Z_t is a $(r \times 1)$ vector of observed covariates that are not affected by treatment, λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is a $(F \times 1)$ vector of unobserved factor loadings, and ϵ_{it} is the error term.

Unfortunately, $Y_{1t}(1)$ and $Y_{1t}(0)$ are never observed at the same time. Therefore, a counterfactual must be created. This is done through a convex set of the control observations over the pre-treatment period. The control observations are referred to as the *donor pool*. Formally, the donor pool is defined as $Y_{jt}(0) \forall j \neq 1$. If a study involved the policy implications on a country's GDP, the control observations would be other countries' GDP.

Suppose there exists a weighting matrix $W = (w_2, \dots, w_{n+1})'$ such that $w_i \in [0, 1]$ and $\sum_{i=2}^{n+1} w_i = 1$. ADH bounds the weights between 0 and 1 and sums them to unity to avoid extrapolation bias¹. Choosing an optimal weighting matrix, w^* , would create an unbiased estimator. Namely:

¹This assumption has been thoroughly investigated. There are countless papers questioning this assumption. A good starting point is (Doudchenko and Imbens 2016).

ADH Assumption: The estimator of $Y_{i1}(0)$ is unbiased if :

$$\sum_{i=2}^{n+1} w_i^* Z_i = Z_1$$

$$\sum_{i=2}^{n+1} w_i^* \mu_i = \mu_1$$

Using w^* would yield:

$$Y_{1,1}(0) = \sum_{i=2}^{n+1} w_i^* Y_{i,1}(0)$$

$$Y_{1,2}(0) = \sum_{i=2}^{n+1} w_i^* Y_{i,2}(0)$$

.

.

.

$$Y_{1,T_0}(0) = \sum_{i=2}^{n+1} w_i^* Y_{i,T_0}(0)$$

The equality above implies that the weighted average of the other observations creates a perfect counterfactual for the treated unit. On an intuitive level, the notion of constant weights means the relationship between the treated unit and controls is constant throughout the analysis period. If a synthetic control analysis is being employed to study country GDP (as is quite common), each country cannot have a significant change throughout the analysis. This means no major governmental shifts, policy reforms, or economic booms/busts. Given that the synthetic control literature recommends a long pre-treatment window of ten periods and post of five, countries cannot undergo major changes in 15 periods. Since most country level data is reported at the yearly level, this translates to a 15 year constant relationship between the treatment and control.

Mathematically, the intuition is describing a situation where w^* holds with noise. This means $Y_{1,t} \neq \sum_{i=2}^{n+1} w_i^* Y_{i,t}(0)$ for some t. This would be a violation of the ADH assumptions. One way to account for this error is to allow time specific weights: $Y_{1,t} = \sum_{i=2}^{n+1} w_{i,t}^* Y_{i,t}(0)$. An immediate concern with this strategy is identification. Instead of having T_0 equations and n coefficients (e.g. $[w_2^*, \dots, w_{n+1}^*]$), now there is 1 equation with n coefficients (e.g. $[w_{2,t}^*, \dots, w_{n+1,t}^*]$). Another immediate concern of using all time varying coefficients is overfitting ((Bitto and Frühwirth-Schnatter 2019), (Belmonte, Koop, and Korobilis 2014)). Overfitting is when the model too closely fits the data rather than uncovering the true data generating process. This leads to erroneous out of sample predictions. In a Bayesian framework, it also leads to impractically large prediction intervals (Scott and Varian, n.d.). To account for the danger of overfitting, shrinkage can be applied. Shrinkage in a regression is when coefficients are biased towards zero to decrease mean squared error at the expense of increased bias. Shrinkage estimators have gained extreme popularity in statistic and machine learning. I refer to models that apply shrinkage to time varying coefficients as Time Varying Parameter Shrinkage (TVPS) models. I plan to apply TVPS models to a synthetic control framework.

2.2 Linear Gaussian State Space Modeling

The problem described above is a latent variable estimation problem. State space modeling is a time series concept that allows for modeling latent variables. This means modeling unobserved components like time trends, seasonality, and time varying coefficients. Thinking back to synthetic control, this would be the

elements of μ_i . A state space model is composed of an observation equation and transition equation. A general form of these equations are:

$$\begin{aligned} y_t &= Z_t \alpha_t + \epsilon_t && \text{observation equation} \\ \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t && \text{transition equation} \\ \alpha_0 &\sim N(a_0, P_0) \end{aligned}$$

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 $1 \times m$ dimensional matrix, α_t a $m \times 1$ matrix, and ϵ_t a scalar. α_{t+1} would also be a $m \times 1$ matrix, T_t an $m \times m$ matrix, R_t a $m \times r$ matrix and Q_t and $r \times r$ matrix. Finally, a_0 is $m \times 1$ and P_0 is $m \times m$. 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 mutually and 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.

(Scott and Varian, n.d.) proposed using state space models in combination with Spike and Slab variable selection for high dimensional time series analysis. This was then extended by (Brodersen et al. 2015a) as an alternative for synthetic control. Both (Scott and Varian, n.d.) and (Brodersen et al. 2015a) 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.

Recent developments in time varying parametric estimation has 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. I apply (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

I plan to use the model proposed by (Belmonte, Koop, and Korobilis 2014). Suppose the state space model is defined as:

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

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

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

where P_{jj} is a hyperparameter. This specification of θ_j lends itself to a useful interpretation: θ_j govern the dynamics of $\beta_{j,t}$.

The errors are assumed to be independent of one another and independent of all leads and lags. Notice, I am also assuming the errors between coefficients are 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 (e.g. (Brodersen et al. 2015a)). This paper will add shrinkage to θ_j in that framework.

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

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

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

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

To verify these are equal, note:

$$\begin{aligned} \beta_{j,t} - \beta_{j,t-1} &= (\beta_j + \sqrt{\theta_j} \tilde{\beta}_{j,t}) - (\beta_j + \sqrt{\theta_j} \tilde{\beta}_{j,t-1}) && \text{Plugging in (4)} \\ &= \sqrt{\theta_j} (\tilde{\beta}_{j,t} - \tilde{\beta}_{j,t-1}) && \text{Regroup} \\ &= \sqrt{\theta_j} (\tilde{\beta}_{j,t-1} + \tilde{\eta}_{j,t} - \tilde{\beta}_{j,t-1}) && \text{Plug in (5)} \\ &= \sqrt{\theta_j} \tilde{\eta}_{j,t} && \text{Simplify} \end{aligned}$$

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 $\eta_{j,t}$. This shows that equation (2) and (5) are equivalent.

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. Plugging the reformulation back into the original equation yields the state space model:

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

(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 paper is to utilize Bayesian Lasso because of it's direct relationship to ADH synthetic control. Recall that (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. This direct relationship to ADH synthetic control provides a nice segway into the TVPS state space model literature. Following in suit, I choose:

$$\begin{aligned}\beta|\tau^2, \sigma^2 &\sim N(0, \sigma^2 \text{diag}[\tau_2^2, \dots, \tau_j^2]) \\ \tau_j^2|\lambda^2 &\sim \exp\left(\frac{\lambda^2}{2}\right)\end{aligned}$$

λ^2 can then be calculated via MLE or with it's own prior. Staying true to (Park and Casella 2008), I choose $\lambda^2 \sim \text{Gamma}(b_1, b_2)$, 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 , I define $\sqrt{\theta_j}$ as:

$$\begin{aligned}\sqrt{\theta}|\xi^2 &\sim N(0, \sigma^2 \text{diag}[\xi_2^2, \dots, \xi_j^2]) \\ \xi_j^2|\kappa^2 &\sim \exp\left(\frac{\kappa^2}{2}\right)\end{aligned}$$

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

All that's left is to define σ^2 . I set $\frac{1}{\sigma^2} \sim \text{Gamma}(a_1, a_2)$ 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 frequentist Lasso with the priors described above and ADH synthetic control is a restricted version of Lasso (Kinn 2018).

4 Identifying Assumptions

Up until this point, this paper has focused on the estimation technique. In order to gain causal inference, two assumptions must be implemented. These are not the only assumptions being made within the model. Every prior choice, hyperparameter, and state space formulation are also assumptions being made. However, the following two assumptions are necessary for the results from TVPS state space model to be causal:

- i) The control time series are unaffected by the treatment. If his were to be violated, the causal estimates would be biased.
- ii) The dynamic relationship between the treated variable and the controls established in the pretreatment periods does not change.

5 Monte Carlo Simulation Data

The simulation is restricted to the outcomes of the observed units, without considering underlying 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.,

n.d.) also shows that covariates become redundant when all lagged outcomes are included in ADH approach. (Botosaru and Ferman 2019) show that the counterfactual estimated by using only pre-treatment outcomes is very close to the original ADH. (Brodersen et al. 2015a) 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:

$$\begin{aligned} y_{j,t}(0) &= \xi_{j,t} + \psi_{j,t} + \epsilon_{j,t} & j=1,\dots,J \\ y_{1,t}(0) &= \sum_{j=2}^J w_{j,t}(\xi_{j,t} + \psi_{j,t}) + \epsilon_{1,t} \end{aligned}$$

for $t=1,\dots,T$ 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 $\gamma_j = 0 \forall j, t$. The explicit data generating process is:

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

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.

5.1 Model Testing and Comparison

There are two components to successful inference in synthetic control: accurate estimates of the treatment effect and accurate inference (significant or not). I plan to test both of these by simulating treatment effect sizes at 0%, 0.1%, 1%, 10%, and 100% similarly to (Brodersen et al. 2015a). These treatment effects will be calculated by defining $Y_{1,t}(1) = \rho Y_{0,t}(0)$ for $\rho \in \{1, 1.001, 1.01, 1.1, 2\}$. For inference, I will conclude a causal effect only if 95% of the posterior probability interval excludes 0. I am purposefully not using the cumulative effect because I am interested in how the accuracy of the method changes over the post-treatment period.

In order to compare TVPS state space model to ADH synthetic control output, 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. Again, I will use 0%, 0.1%, 1%, 10%, and 100% for treatment effect sizes. I will define $Y_{1,t}(1)$ as before.

5.2 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, I set $z_2 = 25$ and $z_3 = 5$. I set $\psi_{j,t} = 0$ for all j, t . Finally, I define $w_{2,t} = .2 + .6 \frac{t}{T}$ and $w_{3,t} = 1 - w_{2,t}$. In order to compare to ADH, I set the sum of the weights to unity to ensure the convex hull assumption is met in their method.

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 $\{1, 2, 3, 4, \dots, 50\}$.
- 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) $\gamma_j = 0 \forall j$.

Notice that given this setup, the data generating process can be rewritten in recursive form:

$$\begin{aligned}
y_{1,t}(0) &= \sum_{j=2}^J w_{j,t} \left(c_j t + z_j + \gamma_j \sin \left(\frac{\pi t}{\rho_j} \right) \right) + \epsilon_{1,t} \\
w_{2,t} &= w_{2,t-1} + \frac{.6}{T} \\
w_{3,t} &= w_{2,t-1} - \frac{.6}{T} \\
w_{j,t} &= w_{j,t-1} \quad j \notin \{1, 2, 3\}
\end{aligned}$$

with initial conditions:

$$\begin{aligned}
w_{2,0} &= .2 \\
w_{3,0} &= .8 \\
w_{j,0} &= 0 \quad j \notin \{1, 2, 3\}
\end{aligned}$$

6 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. This also adds 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/10.1257/00028280321455188>.
- 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>.
- 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>.
- 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. 2015a. “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>.
- . 2015b. “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. 2019a. “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>.
- . 2019b. “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: 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.
- 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>.
- Kaul, Ashok, Stefan Klobner, Gregor Pfeifer, and Manuel Schieler. n.d. “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>.
- Pesaran, M. Hashem, Andreas Pick, and Mikhail Pranovich. 2013. “Optimal Forecasts in the Presence of Structural Breaks.” *Journal of Econometrics* 177 (2): 134–52. <https://doi.org/10.1016/j.jeconom.2013.04.002>.
- Powell, David. n.d. “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 L, and Hal Varian. n.d. “Predicting the Present with Bayesian Structural Time Series,” 21.

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>.