Generalized Synthetic Control Method with State-Space Model

A Preprint

Junzhe Shao
Department of Biostatistics
Columbia University, Mailman School of Public Health
722 West 168th St. NY, NY 10032
js5959@columbia.edu

June 2, 2022

Abstract

To assess the treatment effect of a point-wise intervention, synthetic control is commonly used. The goal of the synthetic control method is to approximate the counterfactual outcomes of the treated unit with a linear combination of the control units' observed outcomes. Many studies have proposed an interactive fixed effect model as a parametric justification for synthetic control methods. In practice, the validity of such assumptions is rarely considered. The goal of this paper is to establish a new parametric generalization of the synthetic control approach when the factorization assumption can be relaxed. Using a state-space model for a time-varying weight, we demonstrate the model's flexibility. In our setting, the static linear combination is now considered a hidden state. When the hidden space is specified as constant, the model is reduced to the classic synthetic control approach automatically. We applied our method to the study of German reunification, and the effect of the oil crisis confirmed the advantages of our method. Based on our new method, we also bring insight into the prediction interval based on time-series forecasting, dynamic sparsity for variable selection, and the different roles of auxiliary covariates.

Keywords panel data · synthetic control · causal inference · state-space model · sparsity in time series

1 Introduction

Synthetic control is a popular method for analyzing the impacts of aggregate interventions, such as those that influence some big units. See a review from(Abadie 2021). Synthetic control is often considered as a comparative case study: according to its major idea, the impact of an intervention may be inferred by comparing the development of outcome variables of interest between a unit subjected to that intervention, which are usually referred as *target.* and a set of units that are comparable to the exposed unit but are not influenced by that intervention, which are usually referred as *donors.*

Under the framework of potential outcome(Rubin 1974), synthetic control tries to impute the counterfactual outcomes of the target with a time-invariant weighted linear combination of the observed outcome from the donors. The process can be treated as a regression with some constrains in practice(Abadie, Diamond, and Hainmueller 2010). To justify the above regression and the time-invariant linear weight, we usually made parametric assumptions about the real data generation process (DGP) with a linear factor model (or so-called interactive fixed effect model proposed in (Bai 2009)). Such model contains a unit-specific factor interacted with time-varying coefficients and synthetic control is actually regression over the unit-specific factor. Abadie et al. further make convex hull assumptions of that linear combination with all non-negative weights which summed up to 1. With such assumptions we can have a better interpretability and avoid extrapolation. Note that at this stage we do not include auxiliary covariates and the detailed discussion will be left to 6.4.

Following such framework, a lot of inference method (Hsiao, Ching, and Wan 2012)(Ben-Michael, Feller, and Rothstein 2020)(Doudchenko and Imbens 2017)under different DGP assumptions and constrains were made. In (Doudchenko and Imbens 2017), they proposed another way of imputing the counterfactual outcomes with "horizontally" with lagged outcomes without convex hull assumptions. Following the spirit of (Bai 2009) which directly modeling the latent factors instead of regressing, (Xu 2017) proposed a general framework for synthetic control with estimation of latent factors. Under his framework difference-in-difference can be view as a special case of interactive fixed effect model and the target can be relaxed to more than one. (Athey et al. 2017) further proposed a matrix completion method with better computation efficacy and can incorporate different matrix structure and staggered adoption. It was based on matrix factorization and use nuclear norm as regularization.

All these methods based on the assumptions of linear factorization. In (Shi et al. 2021), they evaluate the assumptions of synthetic control under a fine-grained model framework. Under their setting, factorization is a natural result of invariance assumptions for the fine-grained model and that partially explained why synthetic control methods usually work for a static weight and study some big units. In terms of the assumptions for factorization, the basic idea is decomposing the latent cause of target into two subsets: one contains causes of the outcome that are invariant across states but can vary across time, and the other contains causes that vary across states but hold constant across time. As a result, these methods based on the invariance assumption will somehow provide a "static" regression and estimation result is identical to any permutation of the time index. This may subject to the time-series nature. In practice, we can observe some both time-varying and unit specific cause especially when it comes to data collected more frequently, for example, the effect of temperature on people's mood, the effect of some software feature design on the subscription rate. Those effect can be non-stationary and also unique to each subject, and therefore can hardly be factorized. So, the assumption of linear factorization can be further relaxed.

When we go backs to the purpose of synthetic control, the questions fall in the area of causal inference for time series data. See a review from (Moraffah et al. 2021). More particularly, synthetic control is indeed a heterogeneous treatment effect estimation for non-stationary time series. There exists many previous works in estimating point-wise intervention. And state-space model is a powerful tool to model time series data. In (Brodersen et al. 2015) made state-space model assumptions of one single unit. and (Li and Bühlmann 2020) developed a more general method. However, direct state-space modeling suffers from time-varying unmeasured confounders. And usually require a correctly specified model, making the inference extremely difficult. While following the idea of synthetic control, comparative case study can provide us a solution. Under the spirit of comparative cases, we can actually model our target using all the donors as time-varying covariates. We ensure the unconfoundness once we assume some unmeasured confounding are shared by target and donor, and the effect can be model as a weighted combination. If we can build a state-space model based dynamic regression based on comparative cases, we may discover some more time-series nature and do not necessarily need the invariance assumption. Moreover, the linear factorization can be view as a special case of state-space modeling with transition matrix specified as identical and the variance is zero. This ensure when the data set is suitable for classic synthetic control, our state-space modeling approach should also work.

In this article, we generalized the classic synthetic control methods based on factorization with a time-varying weighted linear combination of donors. We will first propose a state-space modeling synthetic control under dynamic linear regression with linear and Gaussian assumption. There exist a closed-form estimation via Kalman filter/smoother for Gaussian linear case and we can naturally give out the prediction interval in the context of time-series forecasting. Usually, there are many donors, and we often made some sparsity assumptions that only a few observed control subjects are selected into the donor pool to avoid overfitting. A heuristic method is using variable based on cross-validation of pre-treatment period. We will further discuss the dynamic sparsity in the Bayesian inference frame work on time series data.(Belmonte, Koop, and Korobilis 2014)(Bitto and Frühwirth-Schnatter 2019).

Based on the features of the state-space model, we may use effective inference techniques such as black box variational inference to the situation of non-linearity and non-Gaussianity. The classic synthetic control method allows for model auxiliary covariates. In addition, we will explain the various roles of covariates in our state-space model. And we can show the our framework can be adapted to more different settings like more than one target.

The following sections of this article are organized as follows: Section 2 reviews the setup of classic synthetic control and introduce our set up of generalized synthetic control with state-space modeling, with discussion of its assumptions and estimation. Section 3 summaries our model characteristics and compares our new setting with related work in causal inference of panel data. Section 4 evaluates the performance of the new method with a simulation study. Section 5 is the empirical analysis for German Reunification where we discover some new findings. Section 6 is a discussion, where we will discuss our method limitations, possible constrains to the coefficient, prediction interval, black box variational inference for non-linear/Gaussian state-space model and the role of covariates. Section 7 is the conclusion.

2 Setup

2.1 Review and basic set up for synthetic control

We first introduce the background and some basic notation. For the sake of simplicity, We implement the synthetic control under the situation of only one treated unit and no covariates. Unit index $j = 1$ is the treated unit(target), we have unit index set for control units(donors) $\mathcal{I}^{(0)} = \{2, ..., N\}$. For time index set, $\{1, ..., T_0, ..., T\}$, We define $\mathcal{T}^{pre} = \{1, ..., T_0\}$ for pre treatment, and similarly define $\mathcal{T}^{post} = \{T_0 + 1, ..., T\}$, Here $T_0 + 1$ is the time point for that one time shock.

We are trying to estimate the causal effect of the policy at time $t > T_0$, it means we are estimating the treatment effect, under potential outcome framework, for $t \in \mathcal{T}^{post}$:

$$\tau_t = Y_{1t}^{(1)} - Y_{1t}^{(0)}$$

The key question is to build a counterfactual series for the post-treatment period of the target$\{Y_{1t}^{(0)}, t \in \mathcal{T}^{post}\}$ A traditional synthetic control without covariate for impute the conterfactual is doing the following:

$$\hat{Y}_{1t}^{(0)} = \sum_{j \in \mathcal{I}^{(0)}} \hat{\beta}_j Y_{jt}, \text{ for } t \in \mathcal{T}^{post}$$

where

$$\hat{\beta}_j = \arg\min_{\beta_j} \sum_{j \in \mathcal{I}^{(0)}} Loss(Y_{1t} - \beta_j Y_{jt}), \text{ for } t \in \mathcal{T}^{pre}$$

Here, we can chose different loss functions, for example we can adding L_1 or L_2 norm to the loss function. in the original version, there exist convex hull constrains on β_j to avoid extrapolation with non negative β_j and $\sum_{j \in \mathcal{I}^{(0)}} \hat{\beta}_j = 1$. (Abadie, Diamond, and Hainmueller 2010)

The parametric assumption of the data generation follows the interactive fixed effect form: (Bai 2009); Here We define τ_t is the treatment effect, D_{jt} is the treatment assignment.

where

$$\begin{aligned}
&D_{jt} = 0, \text{ for all } t \in \mathcal{T}^{pre}, \text{ for all the units} \\
&D_{1t} = 1, \text{ for } t \in \mathcal{T}^{post} \\
&D_{jt} = 0, \text{ for all } t \in \mathcal{T}^{post}, j \in \mathcal{I}^{(0)}
\end{aligned}$$

then linear factorization follows:

$$Y_{jt} = \boldsymbol{\lambda}_t^T \boldsymbol{\gamma}_j + \tau_t D_{jt} + \epsilon_{jt}$$

And this can be a theoretical justification for classic synthetic control, as the regression over pre-treatment assumes the existence of β_j

$$\boldsymbol{\gamma}_1 = \sum_{j \in \mathcal{I}^{(0)}} \beta_j \boldsymbol{\gamma}_j$$

2.2 State-space model of the weighted linear combination

Based on the interactive fixed effect form, the β_j will hold constant for all the time points. A nature question for regression under linear factorization is it gives out the same β_j regardless of any permutation for the time series from pre-treatment.In this novel setting, we consider a situation where the assumptions of interactive fixed effect or linear factorization does not necessarily hold. We will made several assumptions:

Assumption 1 (conditional exogeneity assumption) *We define $\mathcal{L}_{jt}$ as the latent matrix, ϵ_{jt} is the prediction error.*

$$Y_{jt} = \mathcal{L}_{jt} + \tau_t D_{jt} + \epsilon_{jt}$$

while

$$\epsilon_{jt} \perp D_{jt} \mid \mathcal{L}_{jt} \text{ and } \mathbb{E}(\epsilon_{jt} \mid D_{jt}) = 0$$

This assumption mainly define a deterministic latent part and make conditional exogeneity assumption of the treatment D, Similar assumptions have been made in the (Athey et al. 2017).

Then we give out our weighted combination, instead of a static linear combination, our new set up adopt a dynamic hidden state weightβ_{jt}. We define the followings for simplicity: $\boldsymbol{\beta}_t = \{\beta_{jt} : j \in \mathcal{I}^{(0)}\}^T$, $\mathbf{Y}_t = \{Y_{jt} : j \in \mathcal{I}^{(0)}\}$, $\mathbf{L}_t = \{\mathcal{L}_{jt} : j \in \mathcal{I}^{(0)}\}$. And we can make the following assumption:

Assumption 2 (state-space modeling) *We assume there is exist a state-space modeling of the the latent matrix:*

$$\mathcal{L}_{1t} = f(\mathbf{L}_t, \boldsymbol{\beta}_t), \text{ with hidden state } \boldsymbol{\beta}_t = \phi(\boldsymbol{\beta}_{t-1})$$

State space model is very flexible, we first consider a special case of the above case with *dynamic linear regression* under Gaussian assumptions.

Assumption 3 (dynamic linear regression) *Under gaussian linearity, we model $\mathcal{L}_{1t}$ as dynamic linear regression of $\mathcal{L}_{jt}$ in donors.*

$$\begin{aligned} \mathcal{L}_{1t} &= \mathbf{L}_t \boldsymbol{\beta}_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \mathbf{w}_t \end{aligned}$$

Based on the regression above, In $t \in \mathcal{T}^{pre}$, we can model all the outcomes as

$$\begin{aligned} Y_{1t} &= \mathbf{Y}_t \boldsymbol{\beta}_t + v_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \mathbf{w}_t \end{aligned}$$

And the variance term should follows:

$$\left\{ \begin{array}{c} v_t \\ \mathbf{w}_t \end{array} \right\} \sim \text{ i.i.d. } \mathcal{N}\left[0, \left(\begin{array}{cc} R & 0 \\ 0 & \mathbf{Q} \end{array} \right)\right]$$

Note that in more genery case, the transition matrix $\boldsymbol{\Phi}$ can be time-varying $\boldsymbol{\Phi}_t$. Here we can define it as time-invariant while still keep enough flexibility. We can further make assumptions on the structure of $\boldsymbol{\Phi}$ and $\mathbf{Q}$

Assumption 4 (No interference) *In our case we assume no interference between all the subjects in the hidden state:*

$$\begin{aligned} \boldsymbol{\Phi} &= \operatorname{diag}(\phi_2, ..., \phi_N) \\ \mathbf{Q} &= \operatorname{diag}(\omega_2^2, ..., \omega_N^2) \end{aligned}$$

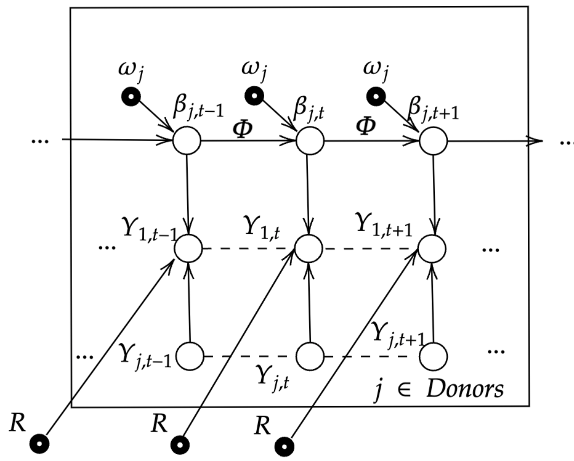


Figure 1: A DAG for our model

If we assign $\mathbf{\Phi} = \mathbf{1}$, $\mathbf{Q} = \mathbf{0}$. It will return to the normal static regression.

2.3 Estimation for hidden state

Given Asumptions 3,4, there exist a closed form estimation for the $\boldsymbol{\beta}_t$ in dynamic linear model. based on (Shumway and Stoffer 2017). State estimation based on a given period $Y_{1,1:s}$ is an essential component of parameter estimation for dynamic linear regression. When $s < t$, the problem is called forecasting or prediction. When $s = t$, the problem is called filtering, and when $s > t$, the problem is called smoothing. In addition to these estimates, we would also want to measure their precision. The solution to these problems is accomplished via the Kalman filter and smoother.

We hereby propose the following notations for the sake of simplicity, conditional expectation and covariance of $\boldsymbol{\beta}_t$ for a given period $Y_{1,1:s}$ can be written as:

$$\begin{aligned}\boldsymbol{\beta}_t^s &= \mathbb{E}\left(\boldsymbol{\beta}_t \mid Y_{1,1:s}\right)\\ \mathbf{P}_{t_1,t_2}^s &= \mathbb{E}\left[\left(\boldsymbol{\beta}_{t_1} - \boldsymbol{\beta}_{t_1}^s\right)\left(\boldsymbol{\beta}_{t_2} - \boldsymbol{\beta}_{t_2}^s\right)'\right]\end{aligned}$$

We first give out the filtering:

Theorem 1 (Kalman filter)

$$\begin{aligned}\beta_t^{t-1} &= \mathbf{\Phi}\beta_{t-1}^{t-1},\\ \mathbf{P}_t^{t-1} &= \mathbf{\Phi}\mathbf{P}_{t-1}^{t-1}\mathbf{\Phi}' + \mathbf{Q}\end{aligned}$$

where we can recursively give:

$$\begin{aligned}\beta_t^t &= \beta_t^{t-1} + \mathbf{K}_t\left(Y_{1t} - \mathbf{Y}_t\beta_t^{t-1}\right)\\ \mathbf{P}_t^t &= \left(I - \mathbf{K}_t\mathbf{Y}_t\right)\mathbf{P}_t^{t-1}\\ \mathbf{K}_t &= \mathbf{P}_t^{t-1}\mathbf{Y}_t'\left(\mathbf{Y}_t\mathbf{P}_t^{t-1}\mathbf{Y}_t' + R\right)^{-1}\end{aligned}$$

Prediction for $t > t_0$ is accomplished recursively with initial conditions $\boldsymbol{\beta}_{t_0}^{t_0}$ and $\mathbf{P}_{t_0}^{t_0}$. for forecasting of counterfactual,for $t \in \mathcal{T}^{post}$:

$$\hat{Y}_{1t}^{(0)} := \mathbb{E}\left(Y_{1t}^{(0)} \mid Y_{1,1:t-1}\right) = \mathbf{Y}_t\boldsymbol{\beta}_t^{t-1}$$

Then we can define the one-step-ahead forecasts error:

$$e_t = Y_{1t}^{(0)} - \mathbb{E}\left(Y_{1t}^{(0)} \mid Y_{1,1:t-1}\right) = Y_{1t}^{(0)} - \mathbf{Y}_t \boldsymbol{\beta}_t^{t-1}$$

This is also called the *innovations* in the literature of state-space model, and has the following properties:

Property 1 $\mathbb{E}(e_t) = 0$*, for any* $s < t$*,* e_t *and* Y_s *are uncorrelated,* e_t *and* e_s *are uncorrelated.* e_t *is a linear function of* $Y_{1,1:t}$*.* $\mathrm{Var}(e_t) = \mathbf{Y}_t \mathbf{P}_t^{t-1} \mathbf{Y}_t' + R$*.*

Based on the properties of innovations above, we can give out the estimated treatment effect $\hat{\tau}_t$

Theorem 2 (estimation and variance of treatment effect)

$$\hat{\tau}_t = Y_{1t}^{(1)} - \hat{Y}_{1t}^{(0)} = Y_{1t} - \mathbf{Y}_t \boldsymbol{\beta}_t^{t-1}$$
$$\mathrm{Var}(\hat{\tau}_t) = \mathrm{Var}(e_t) = \mathbf{Y}_t \mathbf{P}_t^{t-1} \mathbf{Y}_t' + R$$

Next we consider the problem of obtaining estimators from the entire observed data set $Y_{1,1:n}$. Smoothing, as implied by the preceding comments, suggests that each estimated value is a function of the present, future, and past, while the filtered estimator is dependent on the present and past. The forecast depends only on the past, as usual.

Theorem 3 (Kalman smoother)

$$\boldsymbol{\beta}_{t-1}^n = \boldsymbol{\beta}_{t-1}^{t-1} + \mathbf{J}_{t-1}\left(\boldsymbol{\beta}_t^n - \boldsymbol{\beta}_t^{t-1}\right),$$
$$\mathbf{P}_{t-1}^n = \mathbf{P}_{t-1}^{t-1} + \mathbf{J}_{t-1}\left(\mathbf{P}_t^n - \mathbf{P}_t^{t-1}\right)\mathbf{J}_{t-1}',$$

where

$$\mathbf{J}_{t-1} = \mathbf{P}_{t-1}^{t-1} \boldsymbol{\Phi}' \left[\mathbf{P}_t^{t-1}\right]^{-1}.$$

2.4 Unknown parameters for state-space model

As is shown above, we will need initial value $\boldsymbol{\beta}_{t=0} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ for the starting of recursive estimation, We define the unknown parameters set as $\Theta = (\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\Phi}, \mathbf{Q}, R)$. In most of the case, Θ is unknown unless we have some prior knowledge about the dynamic regression.An EM algorithm can be a possible solution for this problem, see(Dempster, Laird, and Rubin 1977)and(Shumway and Stoffer 2017). The details of EM alogrithm for our case is include in Appendix 9.1.

2.5 Sparsity assumptions and variable selection:

Sparsity assumption is often made in the context of sythetic control, as the control units to build the donors are usually redundant,from the view of Matrix Completion, the latent matrix $\mathcal{L}_{jt}$ are assumed to be low-rank. We will first make a assumption of only a subset of donors $\mathcal{I}_s^{(0)} \subseteq \mathcal{I}^{(0)}$ to be the true $\mathbf{Y}_t$ goes into the state-space modeling.

2.5.1 Cross-validation based selection:

A heuristic way is to do a cross validation:

We randomly select M time points $t_m \in \mathcal{T}^{pre}$, $m \in \{1, ..., M\}$, Then we can divide $\mathcal{T}^{pre} = \{\mathcal{T}^{train}, \mathcal{T}^{valid}\}$, where $\mathcal{T}_{(m)}^{train} = \{1, ..., t_m\}$, $\mathcal{T}_{(m)}^{valid} = \{t_m + 1, ..., T_0\}$.

Then calculate RMSE of $\hat{Y}_{1t,(m)}^{(0)}$ for $\mathcal{T}_{(m)}^{valid}$ and optimize the best subset by greedy search for $\mathcal{I}_s^{(0)} \subseteq \mathcal{I}^{(0)}$ in all M.

2.5.2 Shrinkage for state-space model

Cross-validation approach might be cumbersome and in general it is not a good idea with small amount of data, which is very common in the setting of synthetic control. We also need another source of dynamic sparsity since we are now extending our model As discussed before, the parametric assumption of Assumption 3 making the synthetic control method into a problem of time-varying parameters(TVP)models. The variable selection problem of TVP models has been discussed in the literature of macroeconomics and ecnometrics. Modeling state space models is a complex undertaking since one must pick which components to include in the model and whether these components are fixed or time-varying (Frühwirth-Schnatter and Wagner 2010). So in our setting, we will have to include two source of sparsity: one in which unit should be selected into donors, and the other one indicating whether ϕ_j and ω_j should all set ot zero in the hidden space equation from Assumption 3.Shrinkage for TVP models is usually established using a Bayesian framework, their work can automatically do variable selection and reducing time-varying parameters to static ones if the model is overfitting. (Belmonte, Koop, and Korobilis 2014)(Bitto and Frühwirth-Schnatter 2019).

Based on the Assumption 3, follow the ideas of (Frühwirth-Schnatter and Wagner 2010)and(Belmonte, Koop, and Korobilis 2014)(Bitto and Frühwirth-Schnatter 2019).

$$\begin{aligned} Y_{1t} &= \mathbf{Y}_t \boldsymbol{\beta}_t + v_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \mathbf{w}_t \end{aligned}$$

Proposition 1 *We decompose the $\boldsymbol{\beta}_t$ into two parts: a constant $\boldsymbol{\beta}_C$ and a time-varying $\boldsymbol{\beta}_t^*$ with inital values of zero,*

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_C + \boldsymbol{\beta}_t^*$$

We can first rewrite the equation of Assumption 3 in a equivalent way:

$$\begin{aligned} Y_{1t} &= \mathbf{Y}_t \boldsymbol{\beta}_C + \mathbf{Y}_t \boldsymbol{\beta}_t^* + v_t \\ \boldsymbol{\beta}_t^* &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1}^* + \mathbf{w}_t \\ \boldsymbol{\beta}_{t=0}^* &= \mathbf{0} \end{aligned}$$

Then with Assumption 4, we can expand the above formula at subject level $\boldsymbol{\beta}_t = \{\beta_{jt} : j \in \mathcal{I}^{(0)}\}$:
for each subject j, first define

$$\widetilde{\beta}_{jt} = \frac{\beta_{jt}^*}{\omega_j}$$

we can have

$$\begin{aligned} Y_{1t} &= \sum_{j=2}^{N} Y_{jt} \beta_{j,C} + \sum_{j=2}^{N} Y_{jt} \omega_j \widetilde{\beta}_{jt} + v_t \\ \widetilde{\beta}_{jt} &= \phi_j \widetilde{\beta}_{j,t-1} + c_t \\ \widetilde{\beta}_{j,0} &= 0 \text{ , with } c_t \sim \text{i.i.d } \mathcal{N}(0,1) \end{aligned}$$

Based ont the shrinkage properties of ω_j and $\beta_{j,C}$, here follows different sceanrios:

- If ω_j is shrunk to 0 , but $\beta_{j,C}$ is not shrunk to 0 , then the estimator is deterministic, if $\phi_j = 1$, it returns to the classic synthetic control without convex hull constrains.
- If ω_j is shrunk to 0 , and $\beta_{j,C}$ is shrunk to 0 , then unit j is irrelevant for impute the conterfactual.
- If ω_j is not shrunk to 0 , but $\beta_{j,C}$ is shrunk to 0 , it means a time-varying coeffcient starting at zero.
- If ω_j is not shrunk to 0 , and $\beta_{j,C}$ is not shrunk to 0 , it is an unrestricted time-varying coefficient for unit j.

And then give out the hierarchical mixtures of normal prior:

Proposition 2 *for $\beta_{j,C}$:*

$$\beta_{j,C} \mid \sigma_j^2 \sim \mathcal{N}(0, \sigma_j^2)$$

and σ_j^2 follows:

$$\sigma_j^2 \mid \lambda \sim Exp(\lambda^2/2)$$

and

$$\lambda^2 \sim \mathcal{G}(a_1, a_2)$$

$$\omega_j \mid \xi_i^2 \sim \mathcal{N}\left(0, \xi_i^2\right),$$

also with exponential mixing density:

$$\xi_i^2 \mid \kappa \sim Exp\left(\frac{\kappa^2}{2}\right)$$

with

$$\kappa \sim \mathcal{G}(b_1, b_2)$$

and $\widetilde{\beta}_{jt} \mid \widetilde{\beta}_{j,t-1}$ *can be give out by state-space model.*

And then compute the posterior by MCMC. The details of MCMC can be found in (Bitto and Frühwirth-Schnatter 2019). With such method we can automatically characterize each control subject into the four categories of sparisty. In Appendix 9.2, we attached the simulation result under the same set of their paper, which can address our issue of sparsity.

3 Model charasteristics and comparison with related work

After introducing our set up of generalized synthetic control method, we can make a brief comparison with the existing method, and summarize some characteristics of our model. In general, there are two different strains of inference method under the context of synthetic control or similar linear factorization: one follows the regression method from (Abadie and Gardeazabal 2003) (Abadie, Diamond, and Hainmueller 2010), and many research works has been made on adaptions of loss function or different constrains, ususally, such methods do not assume the subject number N to be infinite. The other strain follows directly estimating latent factors based on (Bai 2009), the latent factors λ_t and γ_j can be estimated one the sample size is large, (Xu 2017) gives illustration of interactive fixed models under the context of synthetic control. Eventually, the linkage between matrix completion and synthetic control was established. Under the context of matrix competion, we ususally assuming factorization and the low-rank properties.

The key idea of our method is inspired by the regression strain, and state-space model gives out a way to capture more dynamic information and have a well-established literature of estiamtion and variable selection. the following figure could be an illustration of such generalization:

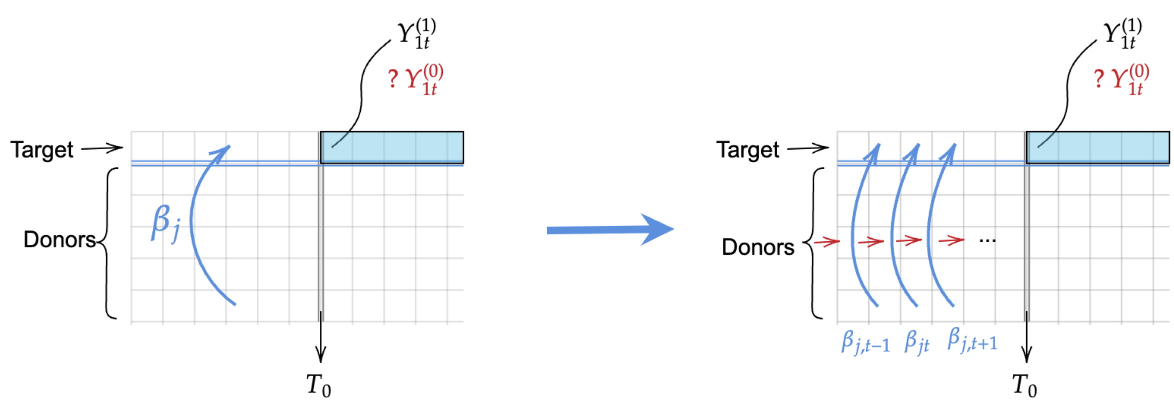


Figure 2: illustraion for our model

The red arrow in the above graph differ our method from all other method based on the assumption of linear factorization, making our method more similar to the time-series forecasting. And we are preliminary focusing on the "fat matrix" with large T but relatively small N. Therefore our work has close relationship with CausalImpact(Brodersen et al. 2015) and CausalTransfer (Li and Bühlmann 2020).These two method both study the effect of point-wise intervention. However, both CausalImpact and CausalTransfer are trying directly modeling the time-series, and in CausalImpact, the author mentioned directly using a classic synthetic control method to aggregate the control group into one trajectory, and served as covariates for the further modeling. The unit specific dynamic similarity is not consider. While in CausalTransfer, it is not build on the context of comparative cases. Both of these method did not emphasized the benefits of comparative cases, making them more similar to a state-space model version of interrupted time series (Ferron and Rendina-Gobioff 2005).

Comparative cases study has its own benefits of reducing the effect of unmeasured confounder and take the advantage of some experiment. Taking the classic tobacco sales, it is almost impossible we are trying to directly modeling the California tobacco sale with some observed variables as the number of latent causes is enormous. But if we are assuming some latent causes are shared by both California and other control states, and the effect of such latent cause can be approximated based on the similarity. Then we do not need to consider such latent causes. Therefore, our method takes the advantage of both side: our method goes under the framework of comparative case study, and also modeling our counterfactual in a dynamic weight.

The dynamic weight is usually essential. Intuitively, the similarity of different units can varies between time. It could either evolve as a seperate time series, or be affected by some external shock. Our simulation and empirical analysis of German reunification data further confirms it is necessary. Based on our discussion upon sparsity assumptions and variable selection, we do not need to worry about the overfitting issue too much, and if the true data generation process follows a linear factor form, our model should give out the right estimation.

Note that, in (Pang, Liu, and Xu 2022), they also consider a dynamic weight under the context of synthetic control. Their model can be expressed in our notation as

$$
\begin{gathered}
Y_{jt} = \mathbf{X}_{jt}\boldsymbol{\beta}_{jt} + \gamma_j'\mathbf{f}_t + \epsilon_{jt} \\
\boldsymbol{\beta}_{jt} = \boldsymbol{\beta} + \boldsymbol{\alpha}_i + \boldsymbol{\xi}_t \\
\boldsymbol{\xi}_t = \Phi_\xi \boldsymbol{\xi}_{t-1} + \boldsymbol{e}_t, \text{and } \mathbf{f}_t = \Phi_f \mathbf{f}_{t-1} + \boldsymbol{\nu}_t
\end{gathered}
$$

$\mathbf{X}_{jt}$ are auxiliary covariates. While the latent factor are still written in a factor form but they are trying to directly model the latent factor $\mathbf{f}_t$ with a state-space model, and $\boldsymbol{\beta}_{jt}$ is decomposed as a two way fixed effect adding another state space model. The estimation procedure follows the strain of (Bai 2009) and (Xu 2017). The underlying assumption is complicated. Instead, our method perform a basic and straightforward dynamic regression based on our assumption 3 in the spirit of synthetic control, thus we do not need to model the components in detail.

4 Simluations

Based on the spirit of state-space model, we will set up a series of simulations with with a "fat matrix," large $T = 1000$ and small $N = 4$. To study the effect of a time-varying effect. I will introduce 5 different scenario of $\boldsymbol{\beta}_t$, and evaluate the performance of our new method.

The donors are $\mathcal{I}^{(0)} = \{2, 3, 4\}$, for each of these donors $Y_{jt}, j \in \mathcal{I}^{(0)}$ follow a different random walk. We choose random walk for them to avoid any other information learned directly from the time-series nature of Y_{jt} and hence we can evaluate the effect of comparative cases study more accurately.

$$
\begin{gathered}
Y_{2t} = Y_{2,t-1} + \mathcal{N}(0, 0.1), \text{with } Y_{2,0} = 10 \\
Y_{3t} = Y_{3,t-1} + \mathcal{N}(0, 0.1), \text{with } Y_{3,0} = 12 \\
Y_{4t} = Y_{4,t-1} + \mathcal{N}(0, 0.1), \text{with } Y_{4,0} = 15
\end{gathered}
$$

The data generation of process Y_{1t} follows:

$$
Y_{1t} = \beta_{2t}Y_{2t} + \beta_{3t}Y_{3t} + \beta_{4t}Y_{4t} + \tau_t D_{1t} + \epsilon_{1t}
$$

,

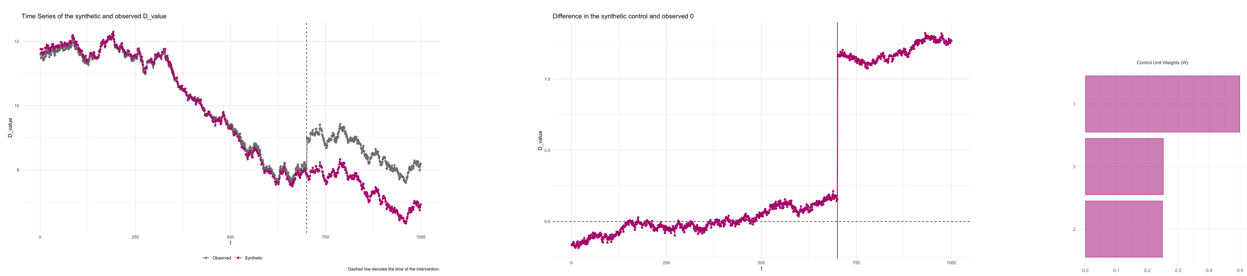
where our treatment effect $\tau_t = 1$ and observation error$\epsilon_{1t} \sim \mathcal{N}(1, 0.01)$,

I will include the different sceanrios for $\boldsymbol{\beta}_t = \{\beta_{2t}, \beta_{3t}, \beta_{4t}\}$, and for each simulation, We will compare the generalized synthetic control with state-space model(Referred as GSC-SSM later) and the classic method from (Abadie, Diamond, and Hainmueller 2010)(Referred as SC-ADH later). We will give out the plot of $Y_{1:4,t}$ and the ground truth for$\boldsymbol{\beta}_t = \{\beta_{2t}, \beta_{3t}, \beta_{4t}\}$ followd by its estimation. Based on the estimated hidden state, we give out the prediction of the counterfactual series after treatment through Kalman filter/smoother. The salmon red indicate the ground truth counterfactual based on our DGP. We should expect, in the case with constant $\boldsymbol{\beta}_t$. The performance of these two methods should be comparable. While when it comes to time-varying coefficient, the SC-ADH will struggle. When the $\boldsymbol{\beta}_t$ follows some random pattern, then the prediction interval for SSM should also be large.

R package `tidysynth` is used to implement SC-ADH, and GSC-SSM is based on R package `KFAS`(Helske 2017)

4.1 Scenario 1: $\boldsymbol{\beta}_t$ hold constant.

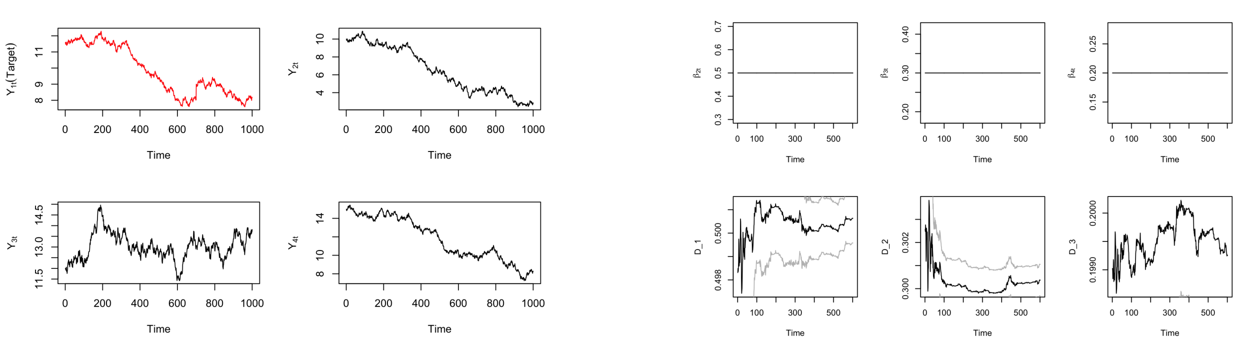
With $\beta_{2t} = 0.3$, $\beta_{3t} = 0.2$, $\beta_{4t} = 0.5$. That means we go back to the classic setting of interactive fixed effect. For SC-ADH:



(a) Imputed conterfactual (b) Estimated treatment effect (c) Weight

Figure 3: SC-ADH for Scenario 1

The fit of SC-ADH is good which is expected.



(a) outcomes (b) ground truth vs estimated weight

Figure 4: GSC-SSM for Scenario 1

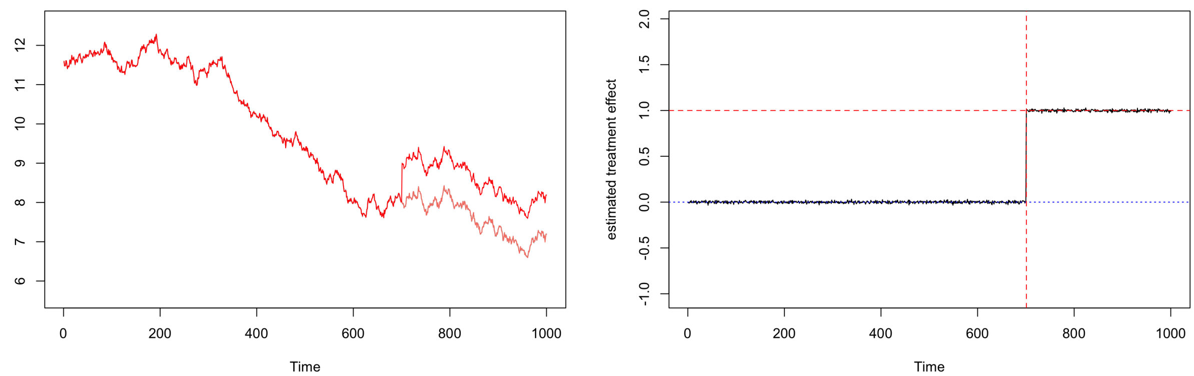


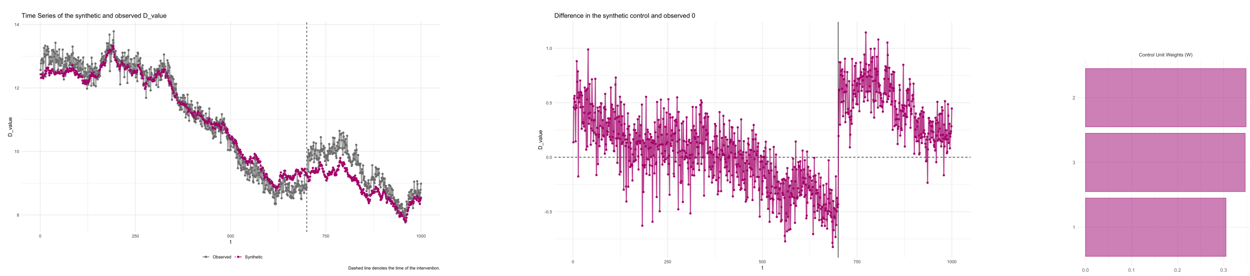
Figure 5: GSC-SSM for Scenario 1

We can tell fit is good and comparable to the SC-ADH, the variance of estimation is smaller and estimation of each hidden state is more accurate.

4.2 Scenario 2: β_t with a white noise.

constant β_{jt} plus white noise $\beta_{2t} = 0.3 + N(0, 0.02)$, $\beta_{3t} = 0.3 + N(0, 0.2)$, $\beta_{4t} = 0.4 + N(0, 0.02)$,

For SC-ADH

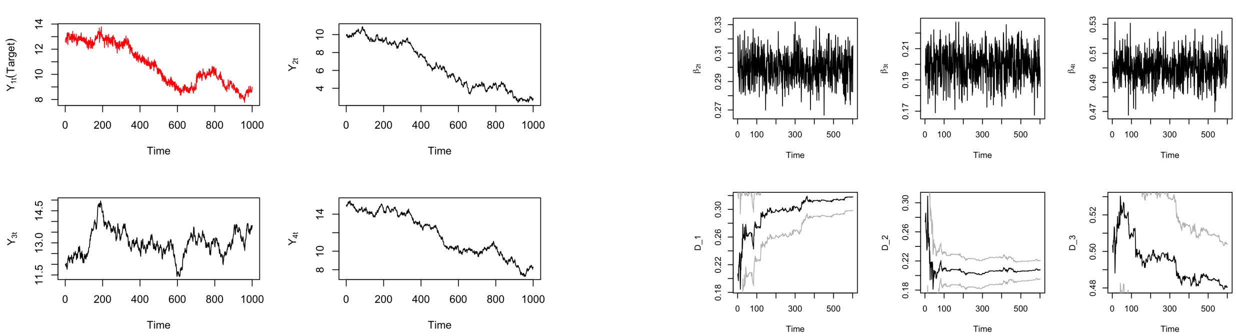


(a) Imputed conterfactual (b) Estimated treatment effect (c) Weight

Figure 6: SC-ADH resultsfor Scenario 2

The estimation of SC-ADH is noisy and the fit is not so good.

For GSC-SSM:



(a) outcomes (b) ground truth vs estimated weight

Figure 7: GSC-SSM estimated hidden state for Scenario 2

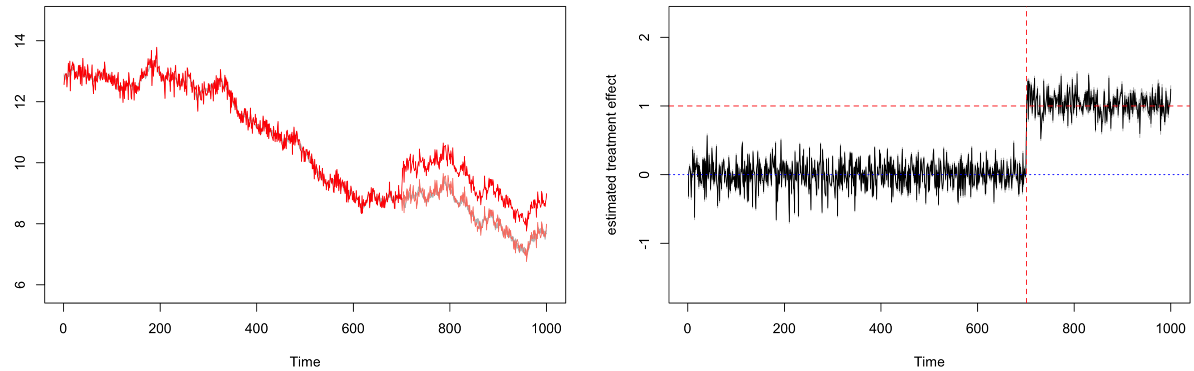


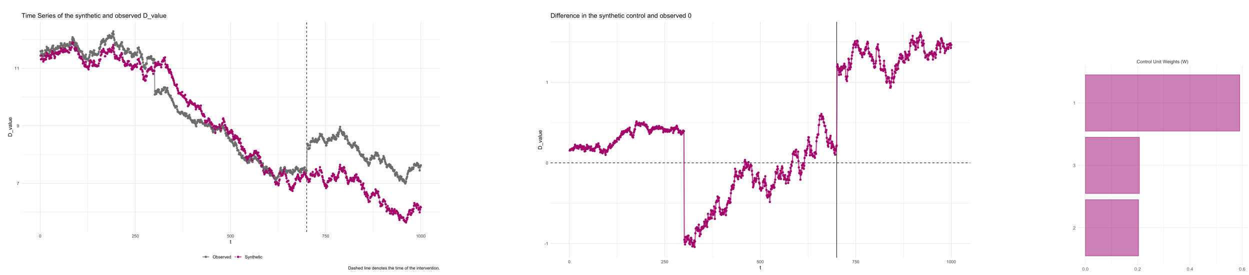
Figure 8: GSC-SSM results for Scenario 2

The estimation from GSC-SSM is good. We give out the correct hidden state and treatment effect.

4.3 Scenario 3: β_t as periodic stable.

for $t < 301$, $\beta_t = \{0.3, 0.2, 0.5\}$, for $t \geq 301$,$\beta_t = \{0.2, 0.2, 0.4\}$. In this case, we are simulating a pointwise intervention directly on β_t at pre-treatment time. And we will see it is possible in the emprical analysis.

For SC-ADH:

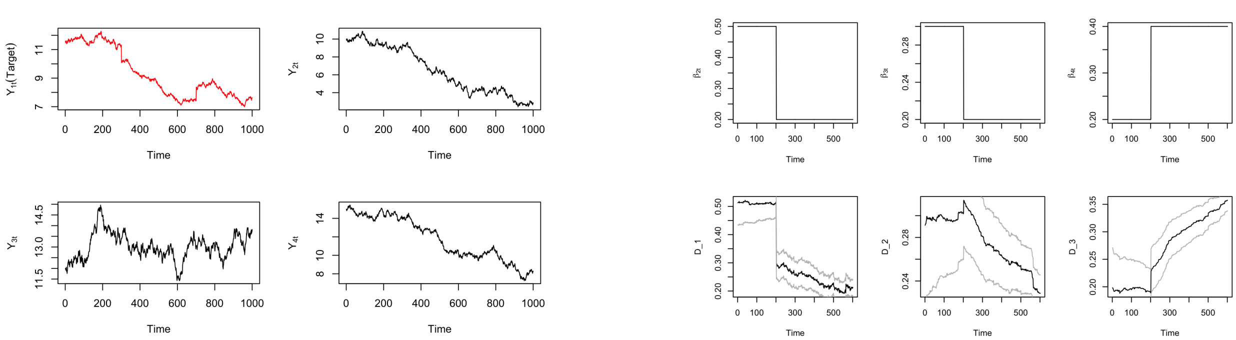


(a) Imputed conterfactual (b) Estimated treatment effect (c) Weight

Figure 9: SC-ADH resultsfor Scenario 3

The SC-ADH can not address this issue well. Periodic stable coefficient will introduce bias.

For GSC-SSM:



(a) outcomes (b) ground truth vs estimated weight

Figure 10: GSC-SSM estimated hidden state for Scenario 3

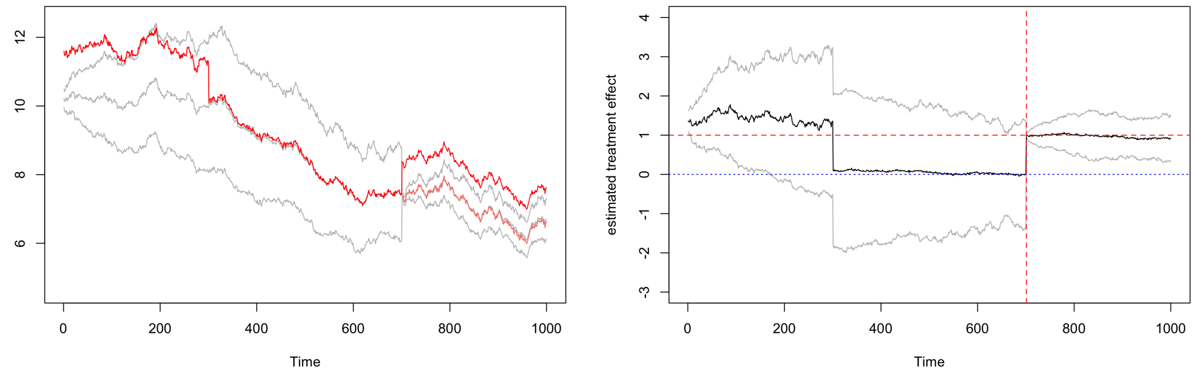


Figure 11: GSC-SSM results for Scenario 3

The GSC-SSM successfully recovers the shock of hidden states and it gives correct prediction with wider prediction interval. We did not try any change point detection at these scenario, adding change point detection could also be very useful.

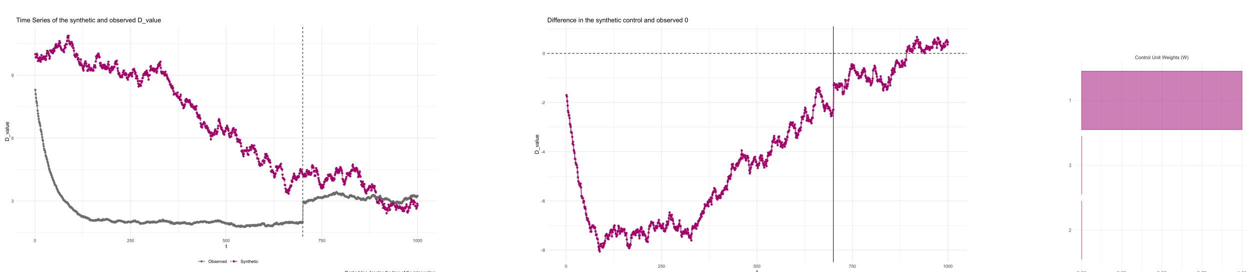
4.4 Scenario 4: $\boldsymbol{\beta}_t$ as AR(1) process.

$\boldsymbol{\beta}_t$ follows an AR(1) process

$$\begin{aligned}
\beta_{2t} &= 0.99\beta_{2,t-1} + N(0, 1 \times 10^{-8}), \text{with } \beta_{2,0} = 0.2 \\
\beta_{3t} &= 0.97\beta_{3,t-1} + N(0, 1 \times 10^{-8}), \text{with } \beta_{3,0} = 0.4 \\
\beta_{4t} &= 1.001\beta_{4,t-1} + N(0, 1 \times 10^{-8}), \text{with } \beta_{4,0} = 0.1)
\end{aligned}$$

$\boldsymbol{\beta}_t$ is change drastically over the period, indicating the similarity between Y_{1t} and Y_{2t}, Y_{3t} is decreasing, while the similarity between Y_{1t} and Y_{4t} is increasing.

For SC-ADH:

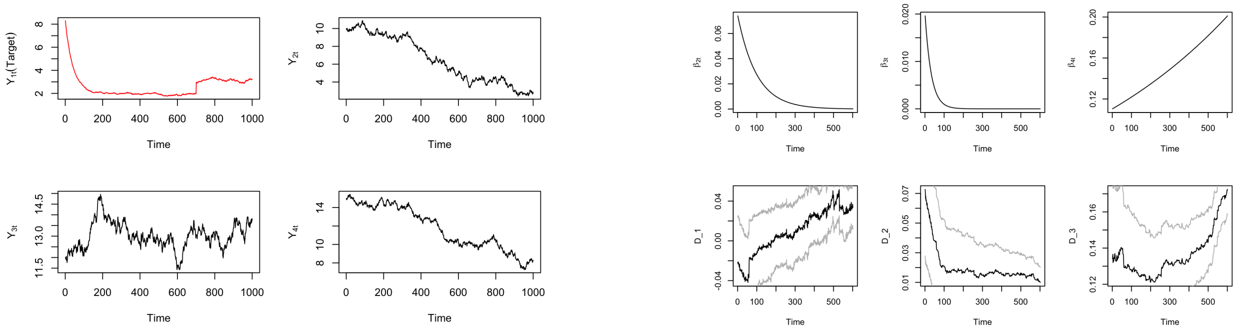


(a) Imputed conterfactual (b) Estimated treatment effect (c) Weight

Figure 12: SC-ADH resultsfor Scenario 4

The SC-ADH gives out wrong prediction which is expected.

For GSC-SSM:



(a) outcomes (b) ground truth vs estimated weight

Figure 13: GSC-SSM estimated hidden state for Scenario 4

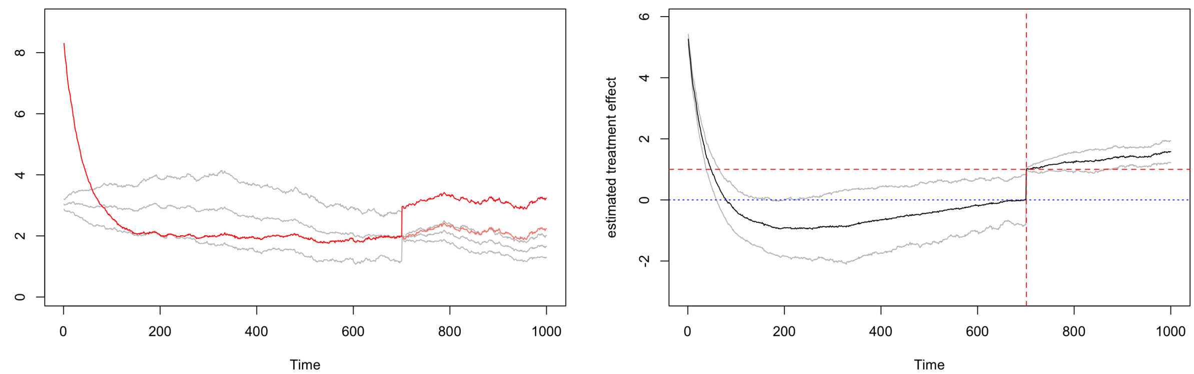


Figure 14: GSC-SSM results for Scenario 4

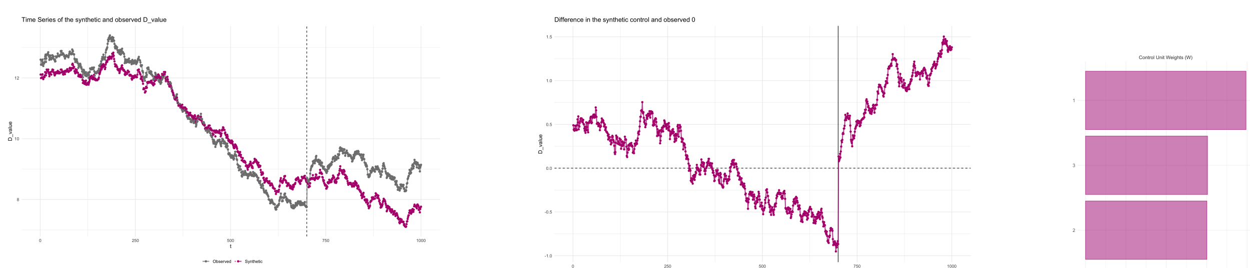
The GSC-SSM gives out correct prediction of hidden state, and it gives out a relative better prediction, the ground truth conterfactual falls in the prediction interval.

4.5 Scenario 5: β_t as a random walk process.

$\boldsymbol{\beta}_t$ a random walk.

$$\beta_{2t} = \beta_{2,t-1} + N(0, 1 \times 0.002), \text{with } \beta_{2,0} = 0.3$$
$$\beta_{3t} = \beta_{3,t-1} + N(0, 1 \times 0.002), \text{with } \beta_{3,0} = 0.3$$
$$\beta_{4t} = \beta_{4,t-1} + N(0, 1 \times 0.002), \text{with } \beta_{4,0} = 0.4$$

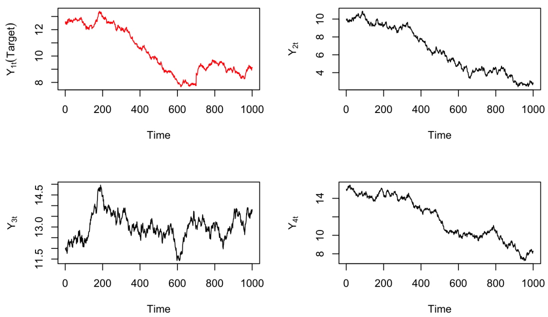
For SC-ADH:



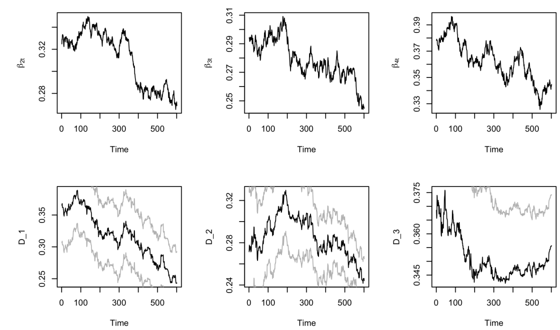
(a) Imputed conterfactual (b) Estimated treatment effect (c) Weight

Figure 15: SC-ADH resultsfor Scenario 5

Tt is almost possible as each data point has a different weight. The prediction is wrong.



(a) outcomes



(b) ground truth vs estimated weight

Figure 16: GSC-SSM estimated hidden state for Scenario 5

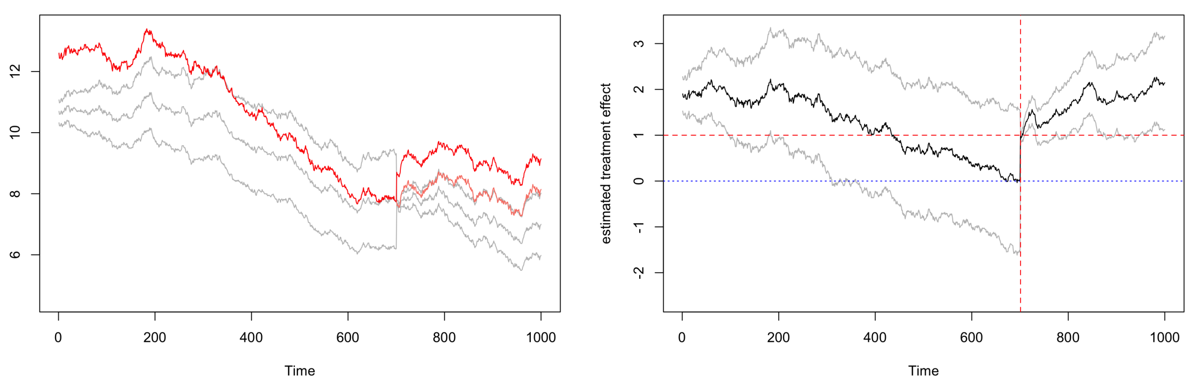
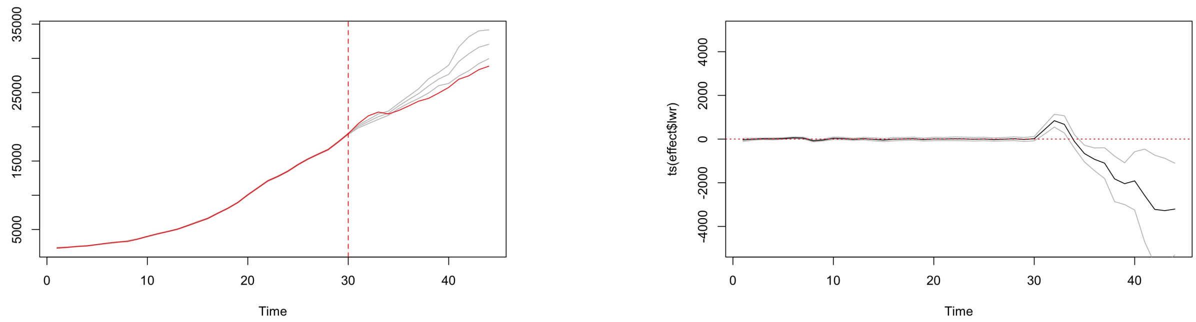


Figure 17: GSC-SSM results for Scenario 5

For GSC-SSM: the long term prediction of a random walk is not possible. But we can still give out a correct short term prediction. The result is good enough in terms of solving practical problems.

5 Emprical Analysis

We try to implement our GSC-SSM with the analysis of German reunification. The purpose of this study is to examine the economic consequences of German reunification in 1990 on West Germany. The real per capita GDP and GDP growth rate of West Germany is the most important variable of relevance. The donor pools are country from OECD assumed to be comparable to West Germany. Using GSC-SSM, we have the following result:



(a) Imputation of Counterfactuals (b) Estimated treated Effect

Figure 18: Synthetic control for German Reunification

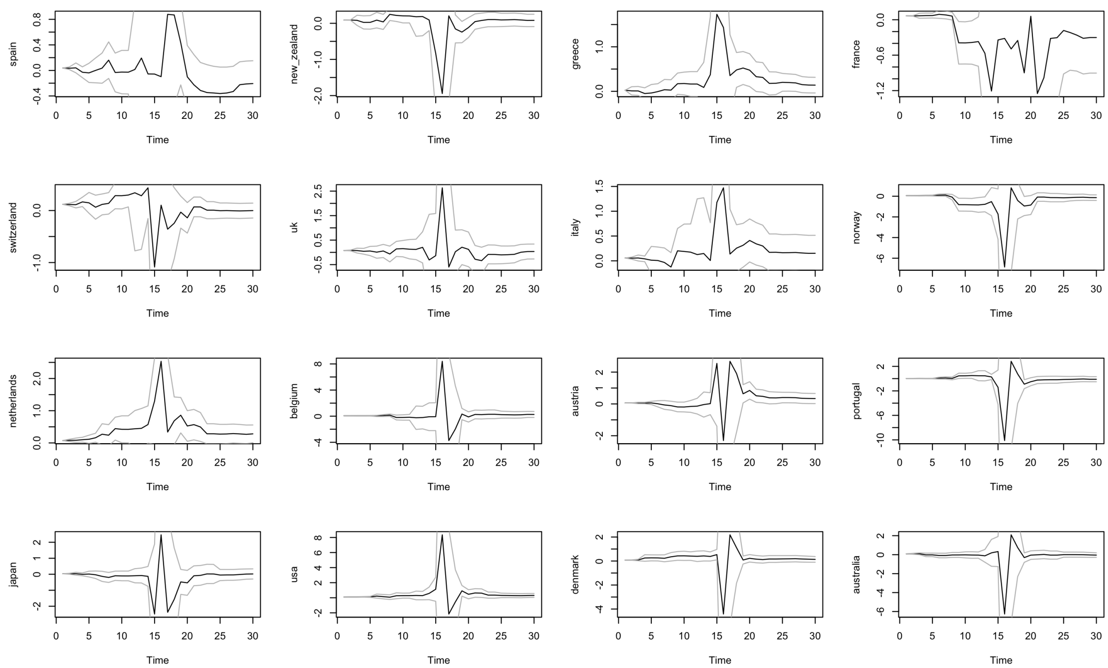
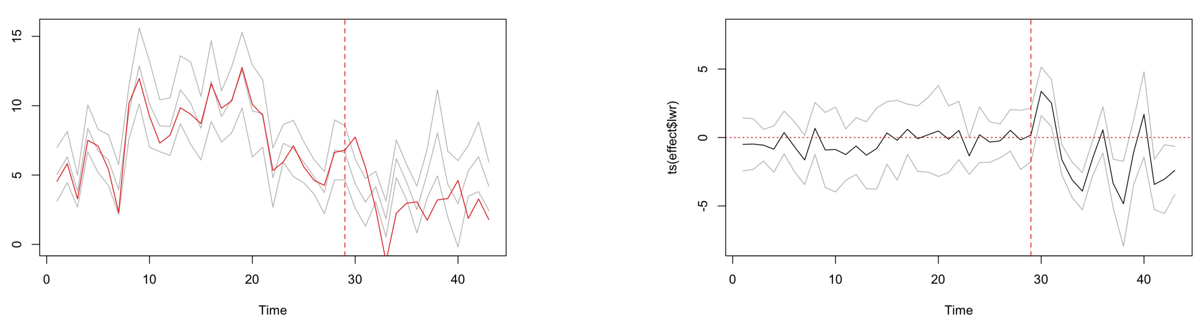


Figure 19: Estimated hidden states for each country



(a) red: observation (b) Estimated treated Effect

Figure 20: Synthetic control for German Reunification, GDP rate

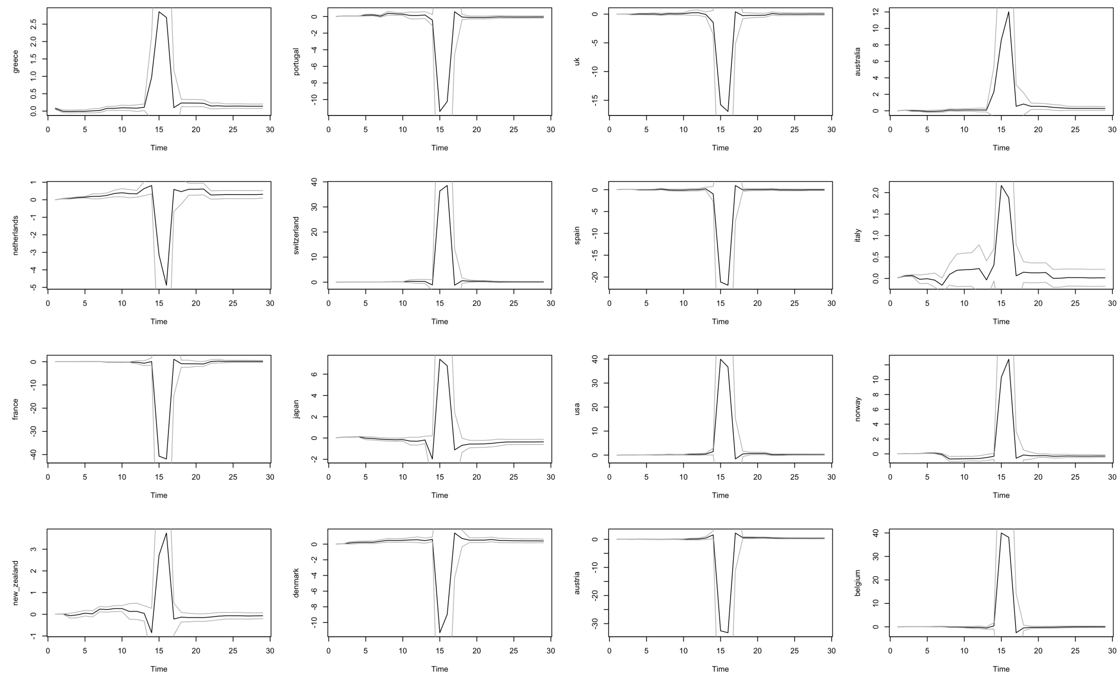


Figure 21: Estimated hidden states for GDP

Based on our GSC-SSM, we can give out the estimation of treatment effect and also the estimation of hidden state. The estimation of pre-treatment is good. And it gives out similary result consistent with previous literature. Surprisingly, we notice a huge change during the time point $t = 14 \sim 16(t = 13 \sim 15$ for GDP growth) indicating the year of 1974~1976. The estimated suggested a changing of similarity between the donors and target at those year. The huge change can be a further proof of the advantage of our method.

The 1973 oil crisis, often known as the first oil crisis, began in October 1973, and it has great impact but heterogeneous effect on the GDP of next year for OECD countries. The Organization of Arab Petroleum Exporting Countries, led by Saudi Arabia, declared an oil embargo. The embargo was aimed against countries who had backed Israel during the Yom Kippur War. The embargo was initially imposed on Canada, Japan, the Netherlands, the United Kingdom, and the United States, but it was eventually extended to Portugal, Rhodesia, and South Africa. The price of oil had nearly tripled by the time the embargo ended in March 1974. The existence of oil crisis could verified our observation. And our method shows a clear variety of effect among donors, causing the weight to shift during those years.

Note that, if we consider the bayesian variable selection, see 9.2, we will find that the weight become constant: this is also reasonable, as the selection gives out some units more similar to the target, and they react in proportion to the unmeassured exogenous event like oil crisis!

6 Discussion

State-space model is not a panacea. The estimation of state-saoce model is challenging, and ususally require a long term period. Also there are many things we have not addressed but are important in the synthetic control literatures and some future directions of improvement:

6.1 Constrains and convex hull assumptions

The contrains of β_j in the orginal synthetic control setting is aiming to achieve an convex hull assumptions. A set of non-negative summed up to 1 weightis is easy and transparent to interpret. The convex hull assumption avoid the issue of extrapolation. In our model based on Assumption 3, the constrain is not easy to set up, as adding non-negative constrain making the state-space model non-linear and we can not have a closed-form estimation. But we can still address the issue under Bayesian inference with techniques like MCMC or using variational inference approach.

6.2 Black Box Variational Inference

As disccused, when Assumption 3 do not hold, the estimation becoming challenging since kalman filter can not be used. Under non-linear or non-Guassian situation, Variational Inference see(Blei, Kucukelbir, and McAuliffe 2017) (Hoffman et al. 2013) might be a effective methods. and Black Box variational inference (Archer and Park 2016) can address the issue of estimation, but it might require a very large data set, which may not be feasible for classic questions of synthetic control.

6.3 Prediction Interval

There has been many discussion of prediction interval problems in synthetic controls(Chernozhukov, Wüthrich, and Zhu 2021)(Cattaneo, Feng, and Titiunik 2021). Our method gives out a prediction interval based on time-series forecasting, but it is only a small proportion of randomness. The randomness of wrongly specified model is not consider, and its uncertainty is usually assessed by bootstrapping based method. Abadie et al, originally suggested a Placebo test and leave-one-out method as a evaluation of robustness. We can further implement these robustness check methods. Also based on our variable selection methods, we may partially relieve some concern about wrongly specified model.

6.4 Covariates

The role of covariates is complicated even in the classic synthetic regression. Abadie et al, suggested that some auxiliary covariates could be helpful to construct the synthetic control, The formula including is given as

$$Y_{jt} = \boldsymbol{\lambda}_t^T \boldsymbol{\gamma}_j + \theta_t Z_j + \tau_t D_{jt} + \epsilon_{jt}$$

and we assume the exisistence of β_j^* simultaneously fulfill

$$\boldsymbol{\gamma}_1 = \sum_{j \in \mathcal{I}^{(0)}} \beta_j^* \boldsymbol{\gamma}_j \text{ and } Z_1 = \sum_{j \in \mathcal{I}^{(0)}} \beta_j^* Z_j$$

In (Botosaru and Ferman 2019) further discussed the role of Z_j and suggested sometime we do not need to simultaneously fulfill both equation once the outcome is matched well.

In our state-space model setting, the role of auxiliary covariates can be even more flexible, a general framework with a auxiliary covariates can be written as below as an extension of Assumption 3:

$$\begin{aligned} Y_{1t} &= \mathbf{Y}_t \boldsymbol{\beta}_t + \Upsilon \mathbf{u}_t + v_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \boldsymbol{\Gamma} \mathbf{u}_t + \mathbf{w}_t \end{aligned}$$

Follow the spirit of comparative case, we can assume $\Upsilon = 0$, unless we have prior knowledge about $\mathbf{u}_t$ and it is unique to our treatment. While the $\boldsymbol{\Gamma}\mathbf{u}_t$ can introduce more flexible in the state-space model. And the estimation under this equation is similar.

Other way of including the auxilary covariate can follow the idea of Abadie et al, the univariate dynamic linear regression can be easily extended to a multivariate version, define all the outcomes need to match:$\mathbf{A}_{1t} = \{Y_{1t}, \mathbf{u}_{1t}\}$. and $\mathbf{A}_{\cdot t}$ is the stacked version in all j from donors. the Assumption 3 can be adapted as:

$$\begin{aligned} \mathbf{A}_{1t} &= \mathbf{A}_{\cdot t} \boldsymbol{\beta}_t + \mathbf{v}_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \mathbf{w}_t \end{aligned}$$

6.5 More than one target

As we propsed the multivariate dynamic regression above, we can further generalize our case with more than one target. Supposed we have a target set $\mathcal{I}^{(1)} = \{1, ..., k, ...M\}$, and donor set $\mathcal{I}^{(0)} = \{1, ..., j, ...N\}$, with a little abuse of notation, we give out the stacked outcome as $\mathbf{C}_{1t} = \{Y_{kt}, k \in \mathcal{I}^{(1)}\}$, $\mathbf{C}_{0t} = \{Y_{jt}, j \in \mathcal{I}^{(0)}\}$l. And we have:

$$\begin{aligned} \mathbf{C}_{0t} &= \mathbf{C}_{1t} \boldsymbol{\beta}_t + \mathbf{v}_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \mathbf{w}_t \end{aligned}$$

Based on the idea of synthetic interventions(Agarwal, Shah, and Shen 2021)and transferrablity assumption we can also give out the following equation:

$$\begin{aligned} \mathbf{C}_{1t} &= \mathbf{C}_{0t} \boldsymbol{\beta}_t + \mathbf{v}_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi} \boldsymbol{\beta}_{t-1} + \mathbf{w}_t \end{aligned}$$

7 Conclusion

In this paper, we propose a novel synthetic control framework based on a state-space model, and we demonstrate how our new framework can uncover more time-series nature when evaluating the effect of point-wise intervention. The new framework is more flexible and can be reduced to traditional synthetic control. We evaluate its performance and posit that this method should work best with long periods of time and few units. We discover the shock of the oil crisis in the empirical analysis, which validates the benefit of our method. We also demonstrate how our framework can be expanded to accommodate more complex settings.

8 References

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59 (2): 391–425. https://doi.org/10.1257/jel.20191450.

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.

Agarwal, Anish, Devavrat Shah, and Dennis Shen. 2021. "Synthetic Interventions." http://arxiv.org/abs/2006.07691.

Archer, Evan, and Il Memming Park. 2016. "BLACK BOX VARIATIONAL INFERENCE FOR STATE SPACE MODELS," 11.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2017. "Matrix Completion Methods for Causal Panel Data Models," 50.

Bai, Jushan. 2009. "Panel Data Models with Interactive Fixed Effects." *Econometrica* 77 (4): 1229–79. http://www.jstor.org/stable/40263859.

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

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2020. "The Augmented Synthetic Control Method." http://arxiv.org/abs/1811.04170.

Bitto, Angela, and Sylvia Frühwirth-Schnatter. 2019. "Achieving Shrinkage in a Time-Varying Parameter Model Framework." *Journal of Econometrics*, Annals issue in Honor of John Geweke "Complexity and Big Data in Economics and Finance: Recent Developments from a Bayesian Perspective," 210 (1): 75–97. https://doi.org/10.1016/j.jeconom.2018.11.006.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. "Variational Inference: A Review for Statisticians." *Journal of the American Statistical Association* 112 (518): 859–77. https://doi.org/10.1080/01621459.2017.1285773.

Botosaru, Irene, and Bruno Ferman. 2019. "On the Role of Covariates in the Synthetic Control Method." *The Econometrics Journal* 22 (2): 117–30. 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." *Annals of Applied Statistics* 9: 247–74.

Cattaneo, Matias D., Yingjie Feng, and Rocio Titiunik. 2021. "Prediction Intervals for Synthetic Control Methods." http://arxiv.org/abs/1912.07120.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2021. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." *Journal of the American Statistical Association* 116 (536): 1849–64. https://doi.org/10.1080/01621459.2021.1920957.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38. http://www.jstor.org/stable/2984875.

Doudchenko, Nikolay, and Guido W. Imbens. 2017. "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." arXiv. https://doi.org/10.48550/arXiv.1610.07748.

Ferron, John, and Gianna Rendina-Gobioff. 2005. "Interrupted Time Series Design." *Encyclopedia of Statistics in Behavioral Science*, October. https://doi.org/10.1002/0470013192.bsa312.

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.

Helske, Jouni. 2017. "KFAS: Exponential Family State Space Models in R." *Journal of Statistical Software* 78 (June): 1–39. https://doi.org/10.18637/jss.v078.i10.

Hoffman, Matthew D., David M. Blei, Chong Wang, and John Paisley. 2013. "Stochastic Variational Inference." *Journal of Machine Learning Research* 14 (4): 1303–47. http://jmlr.org/papers/v14/hoffman13a.html.

Hsiao, Ching, and Wan. 2012. "A PANEL DATA APPROACH FOR PROGRAM EVALUATION: MEASURING THE BENEFITS OF POLITICAL AND ECONOMIC INTEGRATION OF HONG KONG WITH MAINLAND CHINA - Hsiao - 2012 - Journal of Applied Econometrics - Wiley Online Library." 2012. https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1230.

Li, Shu, and Peter Bühlmann. 2020. "Estimating Heterogeneous Treatment Effects in Nonstationary Time Series with State-Space Models." http://arxiv.org/abs/1812.04063.

Moraffah, Raha, Paras Sheth, Mansooreh Karami, Anchit Bhattacharya, Qianru Wang, Anique Tahir, Adrienne Raglin, and Huan Liu. 2021. "Causal Inference for Time Series Analysis: Problems, Methods and Evaluation." *Knowledge and Information Systems* 63 (12): 3041–85. https://doi.org/10.1007/s10115-021-01621-0.

Pang, Xun, Licheng Liu, and Yiqing Xu. 2022. "A Bayesian Alternative to Synthetic Control for Comparative Case Studies." *Political Analysis* 30 (2): 269–88. https://doi.org/10.1017/pan.2021.22.

Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701. https://doi.org/10.1037/h0037350.

Shi, Claudia, Dhanya Sridhar, Vishal Misra, and David M. Blei. 2021. "On the Assumptions of Synthetic Control Methods." http://arxiv.org/abs/2112.05671.

Shumway, Robert H., and David S. Stoffer. 2017. *Time Series Analysis and Its Applications: With R Examples.* Springer Texts in Statistics. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-52452-8.

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.

9 Appendix

9.1 Details of EM algorithm

(i) Initialize by choosing starting values for the parameters in $\{\mu_0, \Sigma_0, \Phi, Q, R\}$, say $\Theta^{(0)}$, and compute the incomplete-data likelihood, $-\ln L_Y\left(\Theta^{(0)}\right)$.

On iteration $j, (j = 1, 2, \ldots)$:

(ii) Perform the E-Step: Using the parameters $\Theta^{(j-1)}$, obtain the smoothed values $\beta_t^n, \mathbf{P}_t^n$ and $\mathbf{P}_{t,t-1}^n, t = 1, \ldots, n$, and calculate some compenont derived from quasi likelihood $\mathcal{Q}\left(\Theta \mid \Theta^{(j-1)}\right) = \mathbb{E}\left[\ln L(\Theta \mid \beta, Y) \mid Y, \Theta^{(j-1)}\right]$, .

(iii) Perform the M-Step: Update the estimates in $\{\mu_0, \Sigma_0, \Phi, Q, R\}$ from the quasi likelihood $\Theta^{(j)} = \arg\max_{\boldsymbol{\Theta}} \mathcal{Q}\left(\Theta \mid \Theta^{(j-1)}\right)$, obtaining $\Theta^{(j)}$.

(iv) Compute the incomplete-data likelihood, $-\ln L_Y\left(\Theta^{(j)}\right)$.

(v) Repeat Steps (ii)-(iv) to convergence.

Theorem 4 (Asymptotic Distribution of the Estimators for Θ_0) *Under general conditions, let $\widehat{\Theta}_n$ be the estimator of Θ_0 obtained by maximizing the innovations likelihood, $L_Y(\Theta)$,. Then, as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\Theta}_n - \Theta_0\right) \xrightarrow{d} \mathcal{N}\left[0, \mathcal{I}\left(\Theta_0\right)^{-1}\right]$$

where $\mathcal{I}(\Theta)$ is the asymptotic information matrix given by

$$\mathcal{I}(\Theta) = \lim_{n\to\infty} n^{-1}\mathbb{E}\left[-\partial^2 \ln L_Y(\Theta)/\partial\Theta\partial\Theta'\right]$$

9.2 Illustration of dynamic sparsity

cited from Bitto and Frühwirth-Schnatter (2019), the problem they are trying to resolve is exactly the same, so I directly quote their simulation for a illustration of dynamic sparsity:

$$\begin{aligned}\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{w}_t, &\quad \mathbf{w}_t \sim \mathcal{N}_d(\mathbf{0}, \mathbf{Q}) \\ y_t = \mathbf{x}_t\boldsymbol{\beta}_t + v_t, &\quad v_t \sim \mathcal{N}\left(0, \sigma_t^2\right)\end{aligned}$$

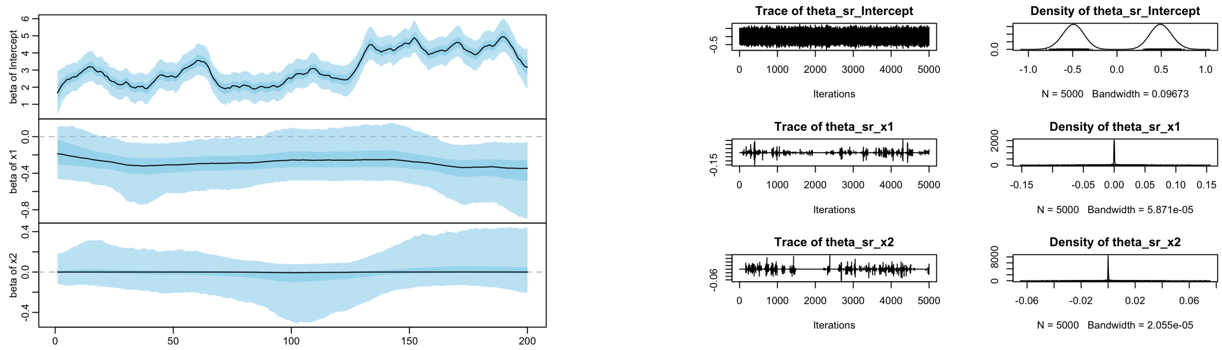
Under diagnol $\mathbf{Q}$:

$$\beta_{jt} = \beta_{j,t-1} + w_{jt}, \quad w_{jt} \sim \mathcal{N}\left(0, \omega_j\right)$$

"To illustrate our methodology for simulated data, we generated 100 univariate time series of length $T = 200$ from a TVP model where $d = 3, \{x_{1t}\} \equiv 1, \{x_{jt}\} \sim \mathcal{N}(0, 1)$ for $j = 2, 3, \sigma^2 = 1, (\beta_1, \beta_2, \beta_3) = (1.5, -0.3, 0)$ and $(\omega_1, \omega_2, \omega_3) = (0.02, 0, 0)$. For each time series, β_{1t} is a strongly time-varying coefficient, β_{2t} is a constant, but significant coefficient, and β_{3t} is an insignificant coefficient. As shrinkage priors on β_j and $\sqrt{\theta_j}$the hierarchical Bayesian Lasso prior (that is $a^\tau = a^\xi = 1$) under the hyperparameter setting $a_1 = a_2 = b_1 = b_2 = 0.001$."

For each of the 100 simulated time series, MCMC estimation is drawing $M = 30,000$ samples after a burn-in of length $30,000$

The result gives out the posterior:



(a) Esitmated hidden state

(b) Estimated variance distribution

Figure 22: Simulation result

We apply our method with the German Reunification data, under Bayesian Shrinkage, We use `shrinkTVP` package:

```
res <- shrinkTVP(as.formula(f_g), data = german_wider[1:30,], a_xi = 1, a_tau = 1)
```

The calculated result: All of the ω_j are 0, indicating constant.

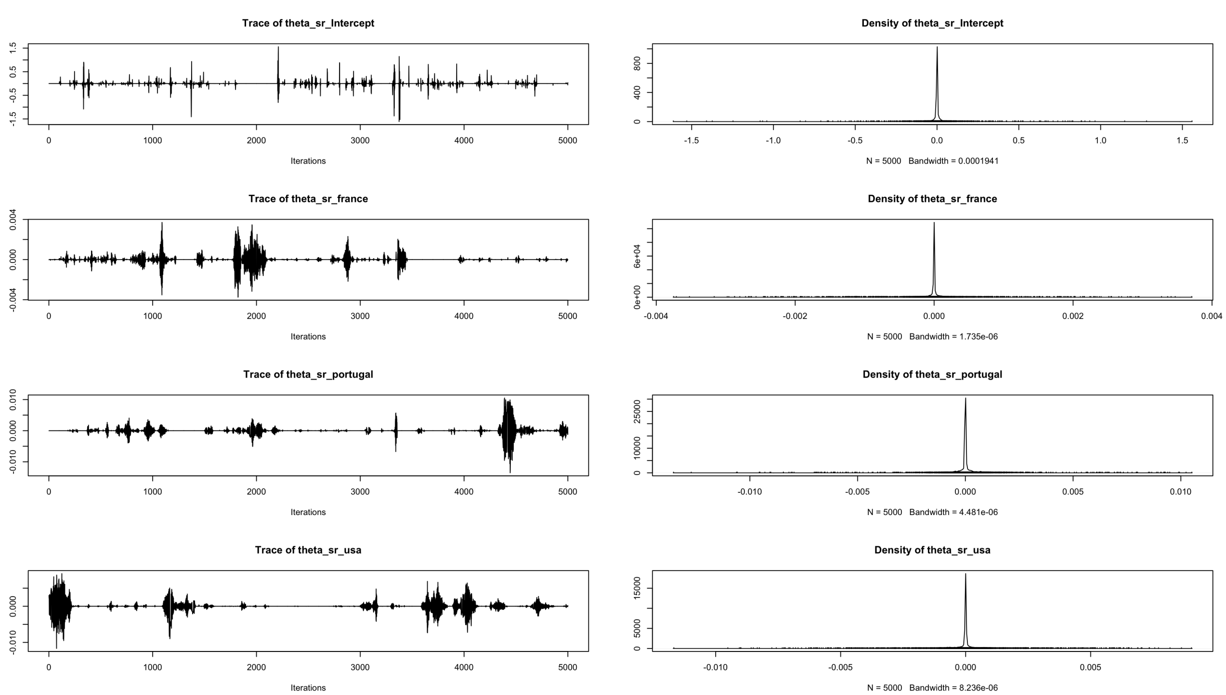


Figure 23: simluated omega

The non-zero weights:

```
> beta[abs(beta) > 0.1]
       beta_mean_usa     beta_mean_denmark beta_mean_netherlands
           0.2273707             0.1126319             0.3663825
    beta_mean_austria
           0.2552494
```

We give out similar result under the framework of Bayesian Shrinkage.