Generalized Synthetic Control Method with State-Space Model

Junzhe Shao
Columbia University
js5959@columbia.edu

Mingzhang Yin
The University of Florida
mingzhang.yin@warrington.ufl.edu

Xiaoxuan Cai
The Ohio State University
cai.1083@osu.edu

Linda Valeri
Columbia University
lv2424@columbia.edu

Abstract

Synthetic control method (SCM) is a widely used approach to assess the treatment effect of a point-wise intervention for cross-sectional time-series data. The goal of SCM is to approximate the counterfactual outcomes of the treated unit as a combination of the control units' observed outcomes. Many studies propose a linear factor model as a parametric justification for the SCM that assumes the synthetic control weights are invariant across time. However, such an assumption does not always hold in practice. We propose a generalized SCM with time-varying weights based on state-space model (`GSC-SSM`), allowing for a more flexible and accurate construction of counterfactual series. `GSC-SSM` recovers the classic SCM when the hidden weights are specified as constant. It applies Bayesian shrinkage for a two-way sparsity of the estimated weights across both the donor pool and the time. On the basis of our method, we shed light on the role of auxiliary covariates, on nonlinear and non-Gaussian state-space model, and on the prediction interval based on time-series forecasting. We apply `GSC-SSM` to investigate the impact of German reunification and a mandatory certificate on COVID-19 vaccine compliance.

1 Introduction

The synthetic control method (SCM) is widely used in analyzing the impacts of interventions on the aggregated units [Abadie, 2021]. Synthetic control is often considered as a comparative case study—the impact of an intervention is inferred by comparing the development of outcome variables of interest between a unit subjected to that intervention, referred as the *target*, and a set of units that are comparable to the exposed unit but are not influenced by that intervention, referred as *donors*.

Under the potential outcome framework [Splawa-Neyman et al., 1923, Rubin, 1974], classic SCM imputes the counterfactual outcomes of the target with time-invariant weights of donors' observed outcomes [Abadie et al., 2010]. The weights are often estimated by a regression under convex hull constraints in practice. To justify the regression and the time-invariant linear weights, parametric assumptions are usually made about the real data generating process (DGP) with a linear factor model [Bai, 2009].

A concern of SCM is that the estimated weights are invariant to the permutation of the time index in the pre-treatment period. Such permutation invariance is a result of linear factorization, which could be inconsistent with the sequential nature of the time series data where the order of the time index matters. To cope with the sequential data, a popular approach is the state-space model that adopts a Markov dependence structure. The state-space model has been applied in estimating the

NeurIPS 2022Workshop on Causal Machine Learning for Real-World Impact (CML4Impact 2022).

effect of point interventions [Brodersen et al., 2015, Li and Bühlmann, 2020]. However, applying the state-space model directly to causal inference requires correctly specifying the underlying DGP; a misspecified model may suffer from time-varying unmeasured confounders.

Synthetic control for comparative case study mitigates the confounding problem. Under the spirit of comparative cases analysis, we model the target unit using all the donors as time-varying covariates, and construct the counterfactual series leveraging the similarity among the units. Unconfoundness is delivered by the assumption that the unmeasured confounders are shared by the target and donors, and the relationship between the target and donors can be captured by the time-varying weights.

In this paper, we generalize the classic SCM with time-varying weights of donors. We first propose a a generalized synthetic control with state-space model (`GSC-SSM`) under dynamic linear regression. The estimate of the time-varying donor weights have a closed-form via Kalman filter and smoother for the linear Gaussian case, and we can naturally provide the prediction interval in the context of time-series forecasting. We will further discuss the two-way dynamic sparsity under the Bayesian inference framework. Moreover, the linear factorization can be view as a special case of state-space modeling with transition matrices specified as identity matrix and the variance being zero. `GSC-SSM` can adapt to the true DGP automatically, recovering the classic SCM when the data suggests so. On the synthetic data, we find when the latent state is stationary, `GSC-SSM` will predict counterfactual series similarly to those by classic SCMs; when the hidden state is non-stationary, it has a significantly lower prediction error. We further demonstrate the application of `GSC-SSM` on two observational datasets that manifest non-stationarity.

2 Setup

2.1 Review of SCM and notations

We first introduce the background and notation. For the sake of simplicity, we review the synthetic control under the situation of only one treated unit and no covariates. Let the unit index $j = 1$ be the treated unit (target) and $\mathcal{I}^{(0)} = \{2, ..., N\}$ be the index set for the control units (donors). For the time index set $\{1, ..., T_0, T_0 + 1, ..., T\}$, we define $\mathcal{T}^{pre} = \{1, ..., T_0\}$ as the pre-treatment period, and define $\mathcal{T}^{post} = \{T_0 + 1, ..., T\}$ as the post-treatment period. Here, $T_0 + 1$ is the time point for the one time shock. We denote D_{jt} as the treatment assignment where $D_{jt} = 0$ for all $t \in \mathcal{T}^{pre}$ and units j. For $t \in \mathcal{T}^{post}$, the treatments $D_{jt} = 0$ and $D_{1t} = 1$. Under the SUTVA and consistency assumptions [Rubin, 1980] (see Appendix A for detailed definition and the additional assumptions of exchangeability and positivity we rely on), we have,

$$Y_{jt} = D_{jt}Y_{jt}(1) + (1 - D_{jt})Y_{jt}(0). \tag{1}$$

Our goal is to estimate the causal effect of the policy at time $t > T_0$,

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

The key challenge is to build a counterfactual series for the post-treatment period of the target $\{Y_{1t}(0)\}_{t\in\mathcal{T}^{post}}$. The classic SCM using no covariates imputes the counterfactual by

$$\widehat{Y}_{1t}(0) = \sum_{j\in\mathcal{I}^{(0)}} \hat{\beta}_j Y_{jt}(0), \text{ for } t \in \mathcal{T}^{post}. \tag{2}$$

Here $\boldsymbol{\beta} = \{\beta_j : j \in \mathcal{I}^{(0)}\}'$ where $\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{t\in\mathcal{T}^{pre}} \left(Y_{1t} - \sum_{j\in\mathcal{I}^{(0)}} \beta_j Y_{jt}\right)^2$. It often assumes there is a convex hull constraint on β_j to avoid extrapolation with non-negative $\beta_j \geq 0$ and $\sum_{j\in\mathcal{I}^{(0)}} \hat{\beta}_j = 1$ [Abadie et al., 2010].

The parametric assumption of the data generation follows the interactive fixed effect [Bai, 2009], i.e., $Y_{jt} = \boldsymbol{\lambda}_t'\boldsymbol{\gamma}_j + \tau_t D_{jt} + \epsilon_{jt}$. This serves as a theoretical justification for classic SCM, as the regression over the pre-treatment assumes the existence of β_j such that $\boldsymbol{\gamma}_1 = \sum_{j\in\mathcal{I}^{(0)}} \beta_j \boldsymbol{\gamma}_j$.

2.2 State-space model

A limitation for regression under linear factorization is that it gives the same β_j regardless of any permutation for the time index in the pre-treatment period. In developing `GSC-SSM`, we consider a

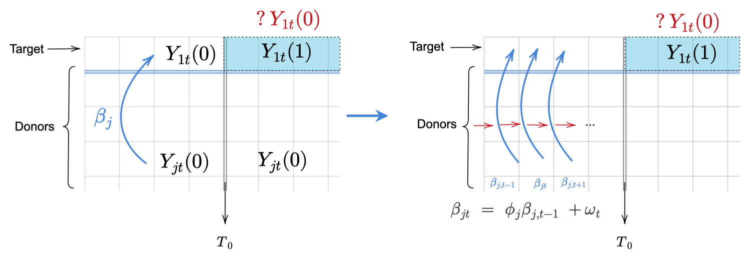

Figure 1: An illustration of `GSC-SSM`, compared to classic SCMs with constant synthetic weights.

situation when β_{jt} is time-varying and the linear factorization assumption does not necessarily hold. For simplicity, we define $\boldsymbol{\beta}_t = \{\beta_{jt} : j \in \mathcal{I}^{(0)}\}'$ which is a $(N-1)\times 1$ vector, $\mathbf{Y}_t = \{Y_{jt} : j \in \mathcal{I}^{(0)}\}$ which is a $1 \times (N-1)$ vector, and the corresponding potential outcomes as $\mathbf{Y}_t(0), \mathbf{Y}_t(1)$. Instead of a static linear combination, we consider a state-space model underlying the potential outcomes,

$$Y_{1t}(0) = f(\mathbf{Y}_{1:t}(0), \boldsymbol{\beta}_{1:t}, Y_{1,1:t-1}(0)), \tag{3}$$

where the hidden state $\boldsymbol{\beta}_t = \phi(\boldsymbol{\beta}_{t-1})$. Under dynamic linear regression, the potential outcomes are modeled as

$$\begin{aligned} Y_{1t}(0) &= \mathbf{Y}_t(0)\boldsymbol{\beta}_t + v_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \mathbf{w}_t, \end{aligned} \tag{4}$$

where

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

We further assume diagonal $\boldsymbol{\Phi} = \mathrm{diag}(\phi_2, ..., \phi_N), \mathbf{Q} = \mathrm{diag}(\omega_2^2, ..., \omega_N^2)$.

The classic SCM (2) can be represented under the framework of (3) as $Y_{1t}(0) = \mathbf{Y}_t(0)\boldsymbol{\beta},\ \boldsymbol{\beta}_t = \boldsymbol{\beta}$, or under the framework of (4) with $\boldsymbol{\Phi} = I$ and $\mathbf{Q} = \mathbf{0}$.

2.3 Inference

There exists a closed-form estimate for the $\boldsymbol{\beta}_t$ in the dynamic linear model. Denote the conditional expectation and covariance of $\boldsymbol{\beta}_t$ for a given period $Y_{1,1:s}$ as $\boldsymbol{\beta}_t^s := \mathbb{E}\left(\boldsymbol{\beta}_t \mid Y_{1,1:s}\right), \mathbf{P}_{t_1,t_2}^s := \mathbb{E}\left[\left(\boldsymbol{\beta}_{t_1} - \boldsymbol{\beta}_{t_1}^s\right)\left(\boldsymbol{\beta}_{t_2} - \boldsymbol{\beta}_{t_2}^s\right)'\right]$. The expectation and variance of the time-varying synthetic control weights can be estimated by the Kalman filtering [Shumway and Stoffer, 2017].

Theorem 1 (Kalman filter) *The synthetic weights and the covariance matrix are estimated as*

$$\boldsymbol{\beta}_t^{t-1} = \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}^{t-1}, \quad \mathbf{P}_t^{t-1} = \boldsymbol{\Phi}\mathbf{P}_{t-1}^{t-1}\boldsymbol{\Phi}' + \mathbf{Q},$$

where we can recursively give

$$\begin{aligned} \boldsymbol{\beta}_t^t = \boldsymbol{\beta}_t^{t-1} + \mathbf{K}_t\left(Y_{1t} - \mathbf{Y}_t(0)\boldsymbol{\beta}_t^{t-1}\right), \mathbf{P}_t^t = \left(I - \mathbf{K}_t\mathbf{Y}_t(0)\right)\mathbf{P}_t^{t-1} \\ \mathbf{K}_t = \mathbf{P}_t^{t-1}\mathbf{Y}_t(0)'\left(\mathbf{Y}_t(0)\mathbf{P}_t^{t-1}\mathbf{Y}_t(0)' + R\right)^{-1}. \end{aligned}$$

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

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

Based on the estimated potential outcome, we can estimate the treatment effect and generate the confidence interval by its variance.

Theorem 2 (estimation and variance of the treatment effect) *The treatment effects $\hat{\tau}_t$ on the target unit and its variance are*

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

2.4 Shrinkage for state-space model

In practice, we will need another source of dynamic sparsity since we are now extending the traditional synthetic control with a time-varying weight, as it may introduce the problem of over-fitting. So, we develope a two-way shrinkage approach. One shrinkage selects which unit should be in the donor pool, and the other one indicates whether ϕ_j and ω_j should be all set at zero in the hidden space equation. The estimation framework is similar to Belmonte et al. [2014], Bitto and Frühwirth-Schnatter [2019] and details can be found in Appendix C.

2.5 Different roles of covariates

The role of covariates is complicated even in the classic synthetic regression. Abadie et al. [2010] suggest that there are p time-invariant auxiliary covariates $\boldsymbol{Z}_j \in \mathbb{R}^p$ that could be helpful to construct the synthetic control as additive factors (usually they are covariates like population, age structure),

$$Y_{jt} = \boldsymbol{\lambda}_t' \boldsymbol{\gamma}_j + \boldsymbol{\theta}_t' \boldsymbol{Z}_j + \tau_t D_{jt} + \epsilon_{jt}.$$

And they assume the existence of β_j^* simultaneously fulfill

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

Botosaru and Ferman [2019] further discusses the role of $\boldsymbol{Z}_j$ and suggests we do not need to simultaneously fulfill both equations once the outcome is matched well.

In our state-space model setting, the role of auxiliary covariates can be even more flexible. Denote the time-varying auxiliary covariates $\boldsymbol{Z}_{jt}$ as a $p \times 1$ vector, and $\mathbf{Z}_t := \{\boldsymbol{Z}_{jt}; \text{for all } j \in \mathcal{I}^{(0)}\}$ as the $p \times (N-1)$ stacked matrix over donors. Now the state-space model can be written as an extension of Equation 4,

$$\begin{aligned} Y_{1t}(0) &= \mathbf{Y}_t(0)\boldsymbol{\beta}_t + \boldsymbol{\Upsilon} \boldsymbol{Z}_{1t} + v_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \boldsymbol{\Gamma}\mathbf{Z}_t + \mathbf{w}_t. \end{aligned} \tag{5}$$

The term $\boldsymbol{\Gamma}\mathbf{Z}_t$ can introduce more flexibility in the hidden-state. Here, $\boldsymbol{\Gamma}$ is a $1 \times p$ vector, meaning the similarity between the units can be affected by the covariate, while $\boldsymbol{\Upsilon}$ should be set to a zero matrix based on the idea of comparative study. The Kalman filter under Equation 5 with fixed $\boldsymbol{\Upsilon}, \boldsymbol{\Gamma}$ has a similar form to Theorem 1, see [Shumway and Stoffer, 2017, Naranjo et al., 2013].

Another way of including the auxiliary covariate is by the idea of Abadie et al. [2010]. We extend the uni-variate dynamic linear regression to a multivariate version. Define all the outcomes need to match as $\mathbf{A}_{1t} = \{Y_{1t}, \boldsymbol{Z}_{1t}\}$, which is a $(p+1) \times 1$ vector, and $\mathbf{A}_{\cdot t} = \{Y_{jt}, \boldsymbol{Z}_{jt}\}$ for all $j \in \mathcal{I}^{(0)}$ as the $(p+1) \times (N-1)$ stacked matrix over donors. The dynamic linear regression is adjusted as

$$\begin{aligned} \mathbf{A}_{1t}(0) &= \mathbf{A}_{\cdot t}(0)\boldsymbol{\beta}_t + \mathbf{v}_t \\ \boldsymbol{\beta}_t &= \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \mathbf{w}_t. \end{aligned}$$

3 Simulation Study

We evaluate the performance of different synthetic control methods for a synthetic panel data with large $T = 1000$ and small $N = 4$. The donors are $\mathcal{I}^{(0)} = \{2, 3, 4\}$, each of these donors $\{Y_{jt}\}_{j \in \mathcal{I}^{(0)}}$ follows a different random walk. We choose random walk to ensure we can not make any valid long-term prediction for the target alone, and hence we can evaluate the performance more accurately.

The DGP of Y_{1t} follows

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

where the treatment effect $\tau_t = 1$ and the observation error $\epsilon_{1t} \sim \mathcal{N}(0, 0.01)$, with intervention $D_{1t} = 1$ when $T > 700$. We consider five different scenarios for $\boldsymbol{\beta}_t = \{\beta_{2t}, \beta_{3t}, \beta_{4t}\}'$. We compare the Mean Squared Prediction Error (MSPE) of the proposed `GSC-SSM` with the classic SCM [Abadie et al., 2010] (`SC-ADH`), and with the SCM without convex hull constrains [Hsiao et al., 2012] (`SC-Hsiao`). The comparison of MSPE is listed in Table 3. `GSC-SSM` has similar MSPE as

Table 1: Results for simulation. GSC-SSM generally gives a lower or similar MSPE.

Scenario	$\boldsymbol{\beta}_t$	GSC-SSM	SC-Hsiao	SC-ADH
constant $\boldsymbol{\beta}_t$	$(0.5, 0.3, 0.2)$	0.00974	0.00974	0.00972
constant + white noise	$(0.5, 0.3, 0.2) +$ i.i.d $\mathbf{w}_t$	0.0708	0.0708	0.0707
change point $T_c = 300$	$(0.5, 0.3, 0.2)$ for $t < T_c$ $(0.2, 0.4, 0.4)$ for $t \geq T_c$	0.0390	0.384	0.563
AR(1)	$\beta_{jt} = \phi_j \beta_{j,t-1} + \omega_j, \phi_j \neq 0, 1$	0.064	0.234	0.330
random walk	$\beta_{jt} = \beta_{j,t-1} + \omega_j$	0.184	0.480	0.445

classic SCMs when the latent state $\boldsymbol{\beta}_t$ is stationary (the first two scenarios). GSC-SSM has significant lower MSPE compared to other two classic SCMs for non-stationary scenarios when the latent state is generated as a random walk or with a change point. Figures including observed simulated Y_{1t}, synthetic counterfactual $\widehat{Y}_{1t}(0)$ (with 95% CI) and estimated $\hat{\tau}_t$ for each scenario can be found in Appendix E.

4 Empirical Analysis

4.1 German reunification data

We apply GSC-SSM to the analysis of classic German reunification data [Abadie et al., 2015]. The purpose of this study is to examine the economic consequences of German reunification in 1990 on West Germany. The donor pools are countries from OECD assumed to be comparable to West Germany. Using GSC-SSM, we have the results in Figure 2.

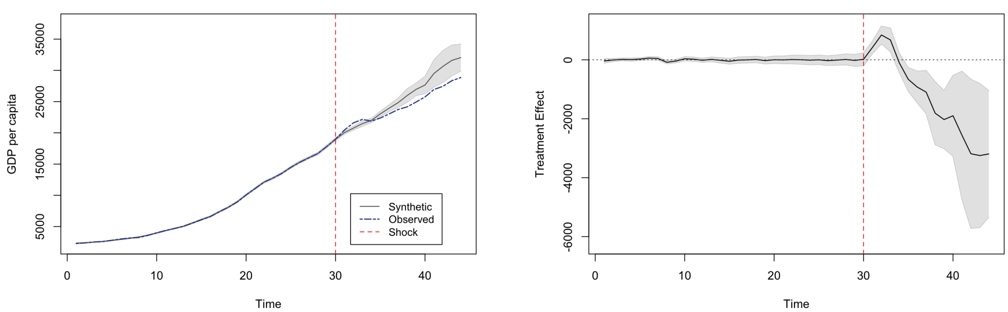

Figure 2: Left: Synthetic counterfactual series $\widehat{Y}_{1t}(0)$ (with 95% CI) and the observed series Y_{1t} for the GDP ; Right: The estimated treatment effect $\hat{\tau}_t = Y_{1t} - \widehat{Y}_{1t}(0)$ of German reunification.

Based on GSC-SSM, we have the estimate of the treatment effect and also the estimate of the hidden state. The estimation of the pre-treatment outcomes are close to the synthetic target and the true target. We notice a pronounced negative effect on West German GDP after the first few years. Such conclusion is consistent with previous literature.

We notice there is a change during the time point $t = 14 \sim 16$ (corresponding to the year of 1974~1976) of the estimated hidden state (see Appendix E). This suggests a change of similarity between the donors and target at those years. The 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 and heterogeneous effect on the GDP of next year for OECD countries. The price of oil had nearly tripled by the time. The embargo ended in March 1974. The existence of oil crisis suggests a time-varying relationship between the donors and the target. GSC-SSM shows a clear variety of effect among donors, causing the weight to shift during those years.

4.2 Vaccination compliance study

We apply GSC-SSM to revisit the study of [Mills and Rüttenauer, 2022]. In this study, we investigate the daily vaccination data of twenty five countries. Nineteen of twenty five countries didn't introduce any mandatory COVID-19 vaccination certificate for entry into public places. These are the donors for the SCM. Further, we have six target countries that introduced such mandatory policy. We revisit this analysis and suggest that a time-varying weight might produce a better pre-treatment fit. In Figure 3, GSC-SSM has a good pre-treatment fit. The estimated treatment effect suggests a significant effect of a mandatory vaccination policy on increasing the vaccination compliance in Israel.

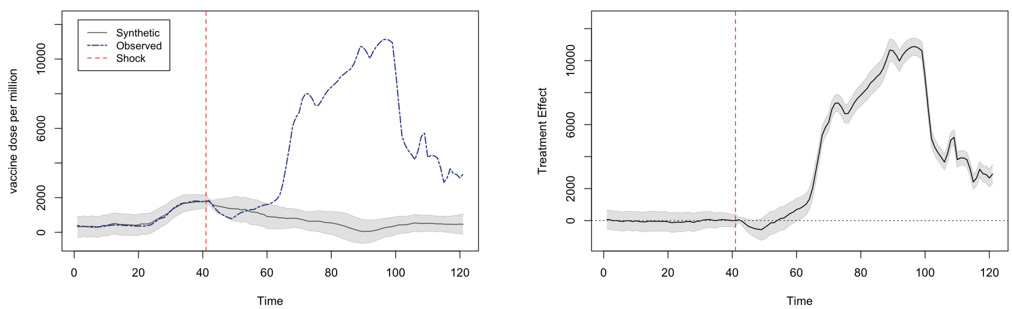

Figure 3: Left: Synthetic counterfactual series $\widehat{Y}_{1t}(0)$ (with 95% CI) and the observed series Y_{1t} for vaccine dose per million of Israel ; Right: The estimated treatment effect $\hat{\tau}_t = Y_{1t} - \widehat{Y}_{1t}(0)$.

5 Discussions

In this paper, we propose GSC-SSM, a novel synthetic control framework based on state-space model. We demonstrate how our new framework can uncover time-series nature when evaluating the effect of point-wise intervention. GSC-SSM is more flexible and generally gives better pre-treatment fit. GSC-SSM can accommodate the common generalizations of classic SCM like more than one target, convex hull assumption, and robustness check methods such as the placebo test. Dynamic linear regression can be further extended to non-linear and non-Gaussian cases. A more detailed discussion of those extensions can be found in Appendix D.

Our work has a close relationship with CausalImpact [Brodersen et al., 2015] and CausalTransfer [Li and Bühlmann, 2020]. They both study the effect of point-wise intervention. However, both CausalImpact and CausalTransfer directly model the time-series. In CausalImpact, the unit specific dynamic similarity is not considered. While CausalTransfer is not built under the context of comparative cases or synthetic control. Both methods do not emphasize the benefits of comparative cases, making them more similar to a state-space model version of interrupted time series analysis [Ferron and Rendina-Gobioff, 2005].

Comparative case studies have their own advantages in terms of reducing the effect of unmeasured confounders and taking advantage of natural experiment. The dynamic of weights is typically required, as the similarity between different units can vary over time. The similarity might evolve as a separate time series or be influenced by an external shock. Our method benefits from both sides: it operates within the framework of a comparative case study while also models the counterfactual with dynamic weights. The simulation and observational studies confirm the necessity of dynamic weights. Furthermore, as the simulation shows, if the true DGP follows a linear factor form, indicating a static weight, GSC-SSM still provide the similar estimation as classic SCM. We also alleviate some of the concerns about the overfitting by discussing a two-way sparsity and whether time-varying weights are necessary.

After the acceptance to this workshop, we became aware of [Klinenberg, 2022], who proposed a similar approach using state-space model and Bayesian shrinkage. Though we address different simulations scenarios of $\boldsymbol{\beta}_t$ which $\boldsymbol{\Phi} \neq I$, the role of auxiliary covariates and different extentions. And our research has different empirical study and findings therein. We hope to conduct thorough comparisons of these approaches in future research.

References

Alberto Abadie. Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects. *Journal of Economic Literature*, 59(2):391–425, June 2021. ISSN 0022-0515. doi: 10.1257/jel.20191450. URL `https://pubs.aeaweb.org/doi/10.1257/jel.20191450`.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 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, June 2010. ISSN 0162-1459. doi: 10.1198/jasa.2009.ap08746. URL `https://doi.org/10.1198/jasa.2009.ap08746`. Publisher: Taylor & Francis _eprint: https://doi.org/10.1198/jasa.2009.ap08746.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510, 2015. ISSN 1540-5907. doi: 10.1111/ajps.12116. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12116`. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12116.

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

Jushan Bai. Panel Data Models with Interactive Fixed Effects. *Econometrica*, 77(4):1229–1279, 2009. ISSN 0012-9682. URL `https://www.jstor.org/stable/40263859`. Publisher: [Wiley, The Econometric Society].

Miguel A. G. Belmonte, Gary Koop, and Dimitris Korobilis. Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, 33(1):80–94, January 2014. ISSN 0277-6693. doi: 10.1002/for.2276. URL `http://eprints.gla.ac.uk/80412/`. Number: 1 Publisher: Wiley.

Angela Bitto and Sylvia Frühwirth-Schnatter. Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, 210(1):75–97, May 2019. ISSN 0304-4076. doi: 10.1016/j.jeconom.2018.11.006. URL `https://www.sciencedirect.com/science/article/pii/S0304407618302070`.

David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017. ISSN 0162-1459. doi: 10.1080/01621459.2017.1285773. URL `https://doi.org/10.1080/01621459.2017.1285773`. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2017.1285773.

Irene Botosaru and Bruno Ferman. On the role of covariates in the synthetic control method. *The Econometrics Journal*, 22(2):117–130, May 2019. ISSN 1368-4221. doi: 10.1093/ectj/utz001. URL `https://doi.org/10.1093/ectj/utz001`.

Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Inferring causal impact using Bayesian structural time-series models. *The Annals of Applied Statistics*, 9(1), March 2015. ISSN 1932-6157. doi: 10.1214/14-AOAS788. URL `https://projecteuclid.org/journals/annals-of-applied-statistics/volume-9/issue-1/Inferring-causal-impact-using-Bayesian-structural-time-series-models/10.1214/14-AOAS788.full`.

Matias D. Cattaneo, Yingjie Feng, and Rocio Titiunik. Prediction Intervals for Synthetic Control Methods. *Journal of the American Statistical Association*, 116(536):1865–1880, October 2021. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2021.1979561. URL `https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1979561`.

Victor Chernozhukov, Kaspar Wüthrich, and Yinchu Zhu. An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls. *Journal of the American Statistical Association*, 116(536):1849–1864, October 2021. ISSN 0162-1459. doi: 10.1080/01621459.2021.1920957. URL `https://doi.org/10.1080/01621459.2021.1920957`. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2021.1920957.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 0035-9246. URL https://www.jstor.org/stable/2984875. Publisher: [Royal Statistical Society, Wiley].

John Ferron and Gianna Rendina-Gobioff. Interrupted Time Series Design. October 2005. doi: 10.1002/0470013192.bsa312. URL https://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa312. Book Title: Encyclopedia of Statistics in Behavioral Science ISBN: 9780470013199 Publisher: American Cancer Society.

Sylvia Frühwirth-Schnatter and Helga Wagner. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, 154(1):85–100, January 2010. ISSN 0304-4076. doi: 10.1016/j.jeconom.2009.07.003. URL https://www.sciencedirect.com/science/article/pii/S0304407609001614.

Matthew D. Hoffman, David M. Blei, Chong Wang, and John Paisley. Stochastic Variational Inference. *Journal of Machine Learning Research*, 14(4):1303–1347, 2013. ISSN 1533-7928. URL http://jmlr.org/papers/v14/hoffman13a.html.

Hsiao, Ching, and Wan. 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. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1230.

Danny Klinenberg. Synthetic Control with Time Varying Coefficients A State Space Approach with Bayesian Shrinkage. *Journal of Business & Economic Statistics*, 0(0):1–12, July 2022. ISSN 0735-0015. doi: 10.1080/07350015.2022.2102025. URL https://doi.org/10.1080/07350015.2022.2102025. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/07350015.2022.2102025.

Shu Li and Peter Bühlmann. Estimating heterogeneous treatment effects in nonstationary time series with state-space models. *arXiv:1812.04063 [stat]*, April 2020. URL http://arxiv.org/abs/1812.04063. arXiv: 1812.04063.

Melinda C Mills and Tobias Rüttenauer. The effect of mandatory COVID-19 certificates on vaccine uptake: synthetic-control modelling of six countries. *The Lancet Public Health*, 7(1):e15–e22, January 2022. ISSN 2468-2667. doi: 10.1016/S2468-2667(21)00273-5. URL https://www.sciencedirect.com/science/article/pii/S2468266721002735.

Arlene Naranjo, A. Alexandre Trindade, and George Casella. Extending the State-Space Model to Accommodate Missing Values in Responses and Covariates. *Journal of the American Statistical Association*, 108(501):202–216, March 2013. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2012.746066. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.2012.746066.

Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974. ISSN 1939-2176. doi: 10.1037/h0037350. Place: US Publisher: American Psychological Association.

Donald B. Rubin. Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75(371):591–593, 1980. ISSN 0162-1459. doi: 10.2307/2287653. URL https://www.jstor.org/stable/2287653. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Robert H. Shumway and David S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer Texts in Statistics. Springer International Publishing, Cham, 2017. ISBN 978-3-319-52451-1 978-3-319-52452-8. doi: 10.1007/978-3-319-52452-8. URL http://link.springer.com/10.1007/978-3-319-52452-8.

Jerzy Splawa-Neyman, D. M. Dabrowska, and T. P. Speed. On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465–472, 1923. ISSN 0883-4237, 2168-8745. doi: 10.1214/ss/1177012031. URL https://projecteuclid.org/journals/statistical-science/volume-5/issue-4/

On-the-Application-of-Probability-Theory-to-Agricultural-Experiments-Essay/
10.1214/ss/1177012031.full. Publisher: Institute of Mathematical Statistics.

Appendix

A Assumption for Equation 1

We need a few more classic assumptions:

Assumption 1 (Stable Unit Treatment Value Assumption(SUTVA)) *D_{jt} is well-defined and there is only one version of the potential outcome, and a certain unit's outcome only depends on its own treatment $Y_{jt}(D_{1:N,t}) = Y_{jt}(D_{jt})$*

Assumption 2 (Consistency) *When the treatment $D_{jt} = D$, $Y_{jt} = Y_{jt}(D)$.*

Assumption 3 (Exchangeability) *The treatment assignment to a unit at a time t does not depend on the unit's potential outcomes, $\{Y_{jt}(0), Y_{jt}(1)\} \perp D_{jt}$*

Assumption 4 (Positivity) $\Pr(D_{jt} = d) > 0$

B Unknown Parameters

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 is a classic solution for this problem[Dempster et al., 1977, Shumway and Stoffer, 2017]. Algorithm:

(i) Initialize by choosing starting values for the parameters in $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\Phi}, \mathbf{Q}, R)$, 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 $\boldsymbol{\beta}_t^n, \mathbf{P}_t^n$ and $\mathbf{P}_{t,t-1}^n, t = 1, \ldots, n$, and calculate some component derived from quasi likelihood $\mathcal{Q}\left(\Theta \mid \Theta^{(j-1)}\right) = \mathbb{E}\left[\ln L(\Theta \mid \boldsymbol{\beta}, Y) \mid Y, \Theta^{(j-1)}\right], \quad .$

(iii) Perform the M-Step: Update the estimates in $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \boldsymbol{\Phi}, \mathbf{Q}, R)$ 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 3 (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

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

C Details for Bayesian shrinkage

Based on Equation 4, and follow the ideas of [Frühwirth-Schnatter and Wagner, 2010, Belmonte et al., 2014, Bitto and Frühwirth-Schnatter, 2019]

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

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

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

We can first rewrite the Equation 4 in an equivalent way:

$$Y_{1t}(0) = \mathbf{Y}_t(0)\boldsymbol{\beta}_C + \mathbf{Y}_t(0)\boldsymbol{\beta}_t^* + v_t$$
$$\boldsymbol{\beta}_t^* = \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1}^* + \mathbf{w}_t, \boldsymbol{\beta}_{t=0}^* = \mathbf{0}$$

We expand the above formula at subject level $\boldsymbol{\beta}_t = \{\beta_{jt} : j \in \mathcal{I}^{(0)}\}$:
for each subject j, first define $\widetilde{\beta}_{jt} = \beta_{jt}^*/\omega_j$ we can have

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

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

- If ω_j is shrunk to 0, but $\beta_{j,C}$ is not shrunk to 0, then the estimator is deterministic, if $\phi_j = 1$, returning 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 to impute the counterfactual.
- If ω_j is not shrunk to 0, but $\beta_{j,C}$ is shrunk to 0, it means a time-varying coefficient 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.

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: $\xi_i^2 \mid \kappa \sim Exp\left(\kappa^2/2\right)$ with $\kappa \sim \mathcal{G}(b_1, b_2)$ and $\widetilde{\beta}_{jt} \mid \widetilde{\beta}_{j,t-1}$ can be given by state-space model.*

And then compute the posterior by MCMC. The details of MCMC can be found in [Belmonte et al., 2014, Bitto and Frühwirth-Schnatter, 2019]. With such method, we can automatically characterize each control subject into the four categories of sparsity.

D Details for extension

D.1 More than one target

We can further generalize our case with more than one target, using a multivariate dynamic regression. Suppose we have a target set $\mathcal{I}^{(1)} = \{1, ..., k, ..., M\}$ with more than one treated units, and donor set $\mathcal{I}^{(0)} = \{1, ..., j, ..., N\}$, we define the multivariate stacked outcome as $\mathbf{C}_{1t} = \{Y_{kt}, k \in \mathcal{I}^{(1)}\}$, $\mathbf{C}_{0t} = \{Y_{jt}, j \in \mathcal{I}^{(0)}\}$, and their corresponding potential outcome $\{\mathbf{C}_{1t}(1), \mathbf{C}_{1t}(0)\}, \{\mathbf{C}_{0t}(1), \mathbf{C}_{0t}(0)\}$. We have the following equation and the estimation is similar to the univariate version.

$$\mathbf{C}_{1t}(0) = \mathbf{C}_{0t}(0)\boldsymbol{\beta}_t + \mathbf{v}_t$$
$$\boldsymbol{\beta}_t = \boldsymbol{\Phi}\boldsymbol{\beta}_{t-1} + \mathbf{w}_t$$

D.2 Constrains and convex hull assumptions

When it comes to the constraints of β_j in the original synthetic control setting, there exists a convex hull constrain to avoid extrapolation. A set of non-negative weight summed up to 1 is also easy and transparent to interpret. In our model, 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.

D.3 Approximate Bayesian inference

When it goes to non-Gaussian and non-linear cases, the estimation becoming challenging since Kalman filter can not be used(though it is still the best linear predictor). Under such situations, approximate Bayesian inference like variational inference [Blei et al., 2017, Hoffman et al., 2013] will be an effective method. And Black Box variational inference [Archer and Park, 2016] can address the issue without specifying a model.

D.4 Prediction Interval

There has been many discussions of prediction interval problems in synthetic controls based on conformal inference [Chernozhukov et al., 2021, Cattaneo et al., 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 considered, and its uncertainty is usually assessed by bootstrapping based method. Abadie et al. [2010] originally suggested a placebo test and leave-one-out method as an evaluation of robustness. We can further implement these robustness check methods and conformal prediction for a better prediction interval.

E Details of Simulation & Empirical Analysis

We presented figures for imputed counterfactuals and estimated treatment effect for each scenario:

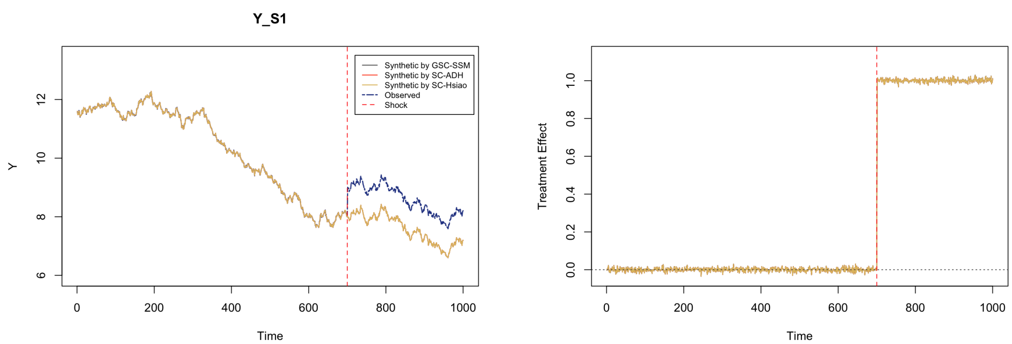

Figure 4: Scenario 1: β_t is constant, the DGP is the same as classic SCMs, `GSC-SCM` gives a similar imputation as the other two methods.

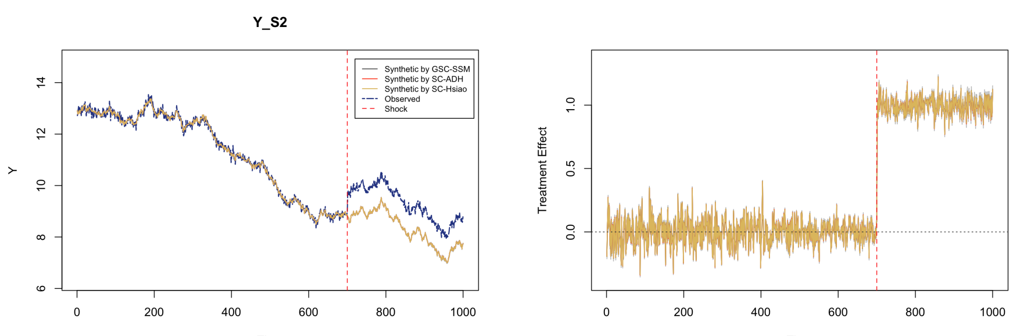

Figure 5: Scenario 2: constant β_t + white noise; In this case β_t is still stationary, we find the three methods give similar imputation.

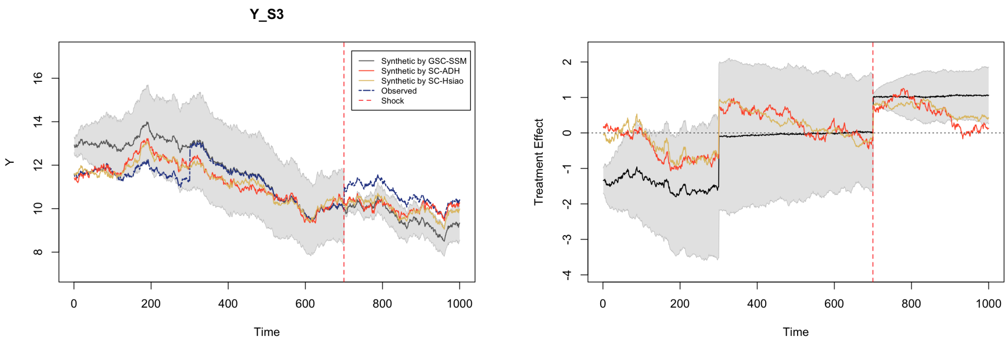

Figure 6: Scenario 3: We add a change point for $\boldsymbol{\beta}_t$ at $T_c = 300$, now the classic SCMs are trying to give a constant $\boldsymbol{\beta}$, resulting in a biased estimation of $\boldsymbol{\beta}_t$ and hence has higher MSPE.

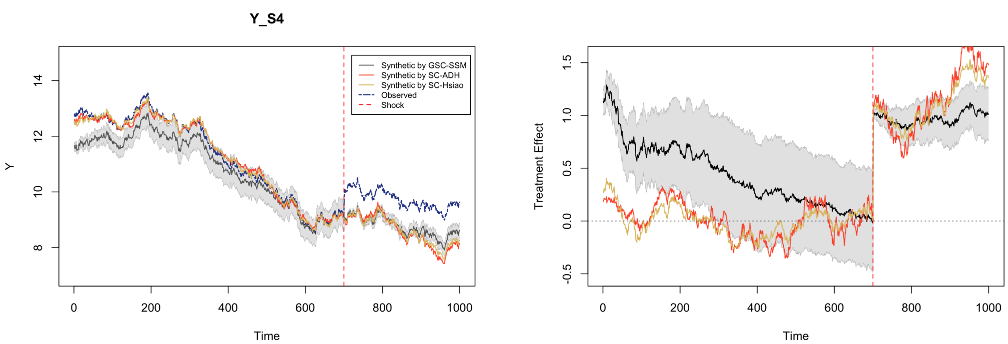

Figure 7: Scenario 4: $\boldsymbol{\beta}_t$ follows an AR(1) process before convergence, `GSC-SSM` gives estimation of treatment effect closer to the ground truth $\tau_t = 1$.

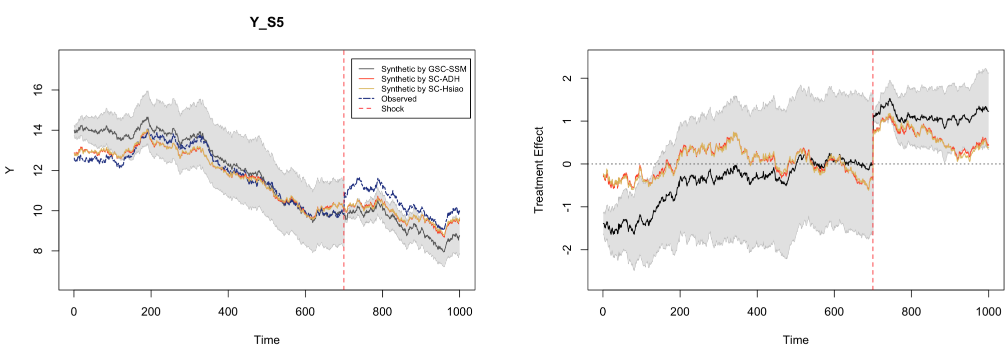

Figure 8: Scenario 5: $\boldsymbol{\beta}_t$ follows a random walk process , `GSC-SSM` gives estimation of treatment effect closer to the ground truth $\tau_t = 1$, but in the long term, we can not make any valid prediction.

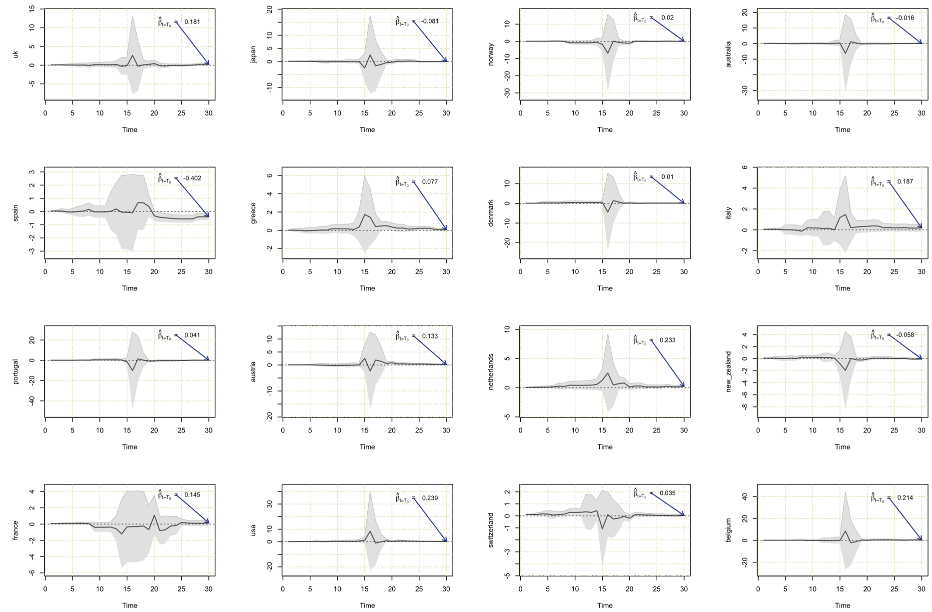

Figure 9: Estimated hidden states of each donor for German reunification data, we noticed the changing during the time point $t = 14 \sim 16$. We also label the value of $\beta_{j,t=T_0}$ for each donor.