Time-Aware Synthetic Control

Saeyoung Rho
Columbia University
rho@cs.columbia.edu

Cyrus Illick
Columbia University
cdi2105@columbia.edu

Samhitha Narasipura
Columbia University
sn3145@columbia.edu

Alberto Abadie
MIT
abadie@mit.edu

Daniel Hsu
Columbia University
djhsu@cs.columbia.edu

Vishal Misra
Columbia University
misra@cs.columbia.edu

Abstract

The synthetic control (SC) framework is widely used for observational causal inference with time-series panel data. SC has been successful in diverse applications, but existing methods typically treat the ordering of pre-intervention time indices interchangeable. This invariance means they may not fully take advantage of temporal structure when strong trends are present. We propose Time-Aware Synthetic Control (TASC), which employs a state-space model with a constant trend, while preserving a low-rank structure of the signal. TASC uses the Kalman filter and Rauch–Tung–Striebel smoother: it first fits a generative time-series model with expectation–maximization and then performs counterfactual inference. As an initial demonstration, we apply the TASC to a case study on California's healthcare policy (Proposition 99). Our method showed promising results in placebo tests, indicating its potential applicability in a broader range of healthcare contexts.

1 Introduction

Synthetic Control (SC) is a popular method in observational causal inference. Often described as a natural extension of the Difference-in-Differences (D-in-D, [1]), SC aims to evaluate the effects of an intervention more accurately by creating synthetic counterfactual data. The first application was measuring the economic impact of the 1960's terrorist conflict in Basque Country, Spain (a *target unit*) by combining GDP data from other Spanish regions (*donor units*) prior to the conflict to construct a *synthetic* GDP data for Basque Country in the counterfactual world without the conflict [2]. Unlike D-in-D, which compares the changes in outcomes over time between a treated group and a comparison group, SC builds a synthetic comparison unit as a weighted combination of donors. SC is becoming increasingly popular with an expanding range of applications, including economics [2, 3, 4], political sciences [5, 6], social sciences [7, 8], and healthcare [9, 8, 10].

SC methods assume that time-series panel data arise from a latent variable model, without restricting a relationship among the time-varying latent factors. The linear factor model, widely adopted in SC literature [2, 3, 5], is one example. This model is both flexible and versatile; for example, it can be extended to incorporate autoregressive components. However, this same flexibility leads SC methods built on top of the model to produce identical estimations when the pre-intervention time indices are permuted. While such flexibility avoids imposing strong structural assumptions, it also prevents the model from capturing predictive signals when a learnable trend exists. A key insight is that time-series data often exhibit stable trends, which we explicitly incorporate into the model.

Another key property of time-series panel dataset is that, as the data size increases, the resulting data matrix tends to be approximately low-rank. This phenomenon, well analyzed by [11], becomes more

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Learning from Time-Series for Health.

pronounced when temporal trends are stronger, limiting the movement of latent factors across time points. Building on this insight, numerous SC variants have been proposed to leverage the data's low-rank structure. These methods typically rely on spectral analysis of the data matrix: for example, [12] employs principal component regression, while [13] frames SC as a nuclear-norm minimization problem. However, these approaches are also time-agnostic because shuffling of time indices does not affect the spectrum of a matrix.

Our contribution lies in embedding SC panel data within a state-space model to simultaneously harness both the low-rank and time-series properties of the data. In Section 3, we introduce our model and outline a simple EM approach to learn the model, employing Kalman filtering and Rauch–Tung–Striebel smoothing. In Section 4, we apply our method to a classic health policy evaluation case (Proposition 99) demonstrating enhanced performance and interpretability relative to traditional methods. This approach offers a principled framework for integrating causal inference with time-series analysis, which is particularly valuable in healthcare and clinical research settings.

2 Background

2.1 Synthetic Control Methods

The time-series panel dataset for SC consists of the following components. Let $Y_{i,t} \in \mathbb{R}$ be the observation from i-th unit at time t. We have n (untreated) donor units indexed by $i \in \{2, \dots, n+1\}$ and a (treated) target unit with index 1. The untreated observation matrix is of size $(n+1) \times T$, where the target unit's values after $t > T_0$ is missing due to the treatment happening at time T_0. Figure 1 illustrates the general structure of an SC dataset, where the superscript $-$ denotes the pre-intervention period and the superscript $+$ denotes the post-intervention period.

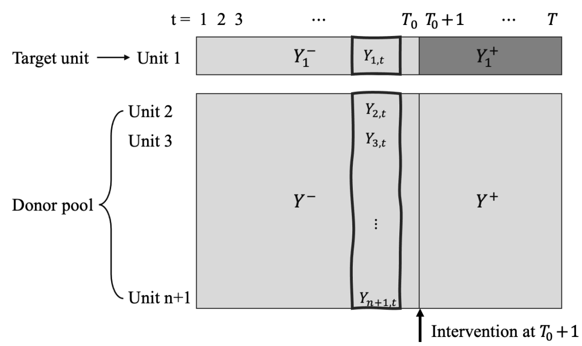

Figure 1: General data structure for synthetic control. The donor pool ($Y = [y_2, \dots, y_{n+1}]^\top$) and the target unit ($y_1$) are divided into pre- and post-intervention periods.

Based on this data structure, we define SC family of methods (Algorithm 1) as follows. SC first learns the relationship between the target unit and donors using the pre-intervention data. For example, this ($\mathcal{M}$) can be a *vertical* regression where the pre-intervention column vectors $Y_{2:n+1,t}$ become input features for the label $Y_{1,t}$, for all $t \in [T_0]$. Then, SC uses this knowledge (f) to project the post-intervention donor data Y^+ and predict $\hat{Y}_1^+$. Finally, the causal effect of the intervention for the target unit is estimated as $Y_1^+ - \hat{Y}_1^+$.

Algorithm 1 Synthetic Control Family of Methods

Data: Target unit's pre-intervention data $Y_1^- \in \mathbb{R}^{T_0}$, Donor data $Y = [Y^-, Y^+] \in \mathbb{R}^{n \times T}$
Result: Counterfactual prediction $\hat{Y}_1^+$, SC weights f
1. Learn $f = \mathcal{M}(Y_1^-, Y^-, Y^+)$ `/* the use of` Y^+ `is optional */`
2. Project $\hat{Y}_1^+ = f(Y^+)$
3. Infer the estimated causal effect of the intervention for the target is $Y_1^+ - \hat{Y}_1^+$

$\mathcal{M}$ refers to the SC learning algorithm. Much of the SC literature has adopted a least squares predictor over the convex scan of Y^-: $f = \arg\min_f ||Y_1^- - f^\top Y^-||^2$ where $\sum_{i=1}^n f_i = 1, 0 \leq f \leq 1 \; \forall i \in [n]$ [2, 3, 5]. The convex scan condition can be replaced by Lasso ($||f||_1$) or Ridge ($||f||_2^2$) regularization [14, 12]. Some approaches use PCA to keep only the top few singular values in data prior to the optimization step [12, 15]. Other variations of synthetic control algorithms focused on issues such as handling multiple treated units [16, 17], dealing with a large the number of donors [4, 18], and correcting biases [19]. See [20] for a detailed survey of these techniques.

2.2 Latent Variable Models for Synthetic Control

The first SC algorithm suggested by [2] assumes a linear factor model

$$Y_{i,t} = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{i,t}, \tag{1}$$

where δ_t is a time trend, $Z_i \in \mathbb{R}^p$ and $\mu_i \in \mathbb{R}^q$ are vector of observed and unobserved predictors, with coefficients θ_t and μ_i, and $\epsilon_{i,t}$ is the noise. This model implies that the signal component of the matrix has a rank no more than $p + q + 1$. When this quantity is considerably smaller than the matrix's full rank, the observation matrix becomes approximately low rank. This is indeed a common case for a factor model, verified in many real-world data [21, 22, 23]. This has inspired a range of SC algorithms to utilize the approximately low-rank structure of data. Several SC algorithms employ simplex constraints or regularizers to minimize the number of active donor units [2, 3, 14, 24]; [18] introduces donor selection step to reduce the number of donors in the first place; [12, 15] uses principal component regression; and [13] frames SC as a problem of nuclear norm minimization.

Another characteristic of the panel data used in SC is its time-series nature. Despite this, many SC algorithm variants remain invariant to permutations of time indices in pre-intervention data. Although the permutation-invariant approach provides robustness by accommodating a wide range of temporal trends, it discards ordering information, whereas explicit modeling strategies can exploit additional structure when meaningful temporal patterns exist. To address this, some researchers have introduced algorithms that assume temporal trend by utilizing state-space models. [25] designed the state vector to include elements such as SC weights, local linear trends, and seasonality. [26, 27] further simplified this structure and only take SC weights as a latent state. The central concept of these approaches is to allow SC weights to change over time, defining the target unit's time series as an observation (scalar) and the SC weights (potentially alongside additional components) as latent states (at least n-dimensional vector, where n is the number of donors). Therefore, these modeling approaches do not necessarily ensure that the observation matrix is approximately low-rank. Furthermore, by not explicitly modeling the stochasticity of the donor pool, these models may not fully leverage the information available in the donor pool.

3 State-Space Model for Time-Aware Synthetic Control

We assume the following state space model:

$$x_t = Ax_{t-1} + q_{t-1} \quad q_{t-1} \sim \mathcal{N}(0, Q), \tag{2}$$

$$y_t = Hx_t + r_t \quad , r_t \sim \mathcal{N}(0, R). \tag{3}$$

where we assume the initial hidden state x_0 to be drawn from $\mathcal{N}(m_0, P_0)$ (i.e., $x_0 \sim \mathcal{N}(m_0, P_0)$.)

The model parameters are $\theta = \{A, H, Q, R, m_0, P_0\}$, where $A \in \mathbb{R}^{d\times d}$, $H \in \mathbb{R}^{(n+1)\times d}$, $Q \in \mathbb{R}^{d\times d}$, $R \in \mathbb{R}^{(n+1)\times(n+1)}$, $m_0 \in \mathbb{R}^d$, $P_0 \in \mathbb{R}^{d\times d}$, and $d \ll min(n, T)$. This is a classical linear Gaussian model, and we set all covariance matrices Q, R, and P_k be positive semi-definite. If desired, we may constrain that the noise covariance matrices Q and R are diagonal with non-zero diagonal entries to reduce the number of parameters. The advantage of this model is that the output $Y = [y_1, \ldots, y_T]$ will be an approximately low-rank matrix (with the approximate rank being d), and the learning algorithm will no longer be agnostic to the permutations of pre-intervention time indices.

Based on this model, we propose an approach for model learning and counterfactual inference: Time-Aware Synthetic Control (TASC). We use an Expectation-Maximization (EM) algorithm to learn both the parameters and the hidden states from the observations y_t. The algorithm begins by randomly initializing the parameters θ, and then alternates between two steps: first, estimating the hidden states while keeping the parameters fixed; and second, updating the parameters while holding the hidden states constant [1]. This EM algorithm is designed using Kalman filter and RTS smoother, and we use only the pre-intervention data to learn the model parameters. For counterfactual inference, we set the variance of the post-intervention target data to infinity, ensuring that the model does not gain any information from that data. This counterfactual estimation phase utilizes post-intervention data, while keeping the model parameters we learned from pre-intervention data. A full description of algorithms are deferred to Appendix B.

[1] See [28] for more information.

3.1 When TASC Model Is Advantageous

A distinctive feature of the TASC approach is its explicit modeling of the trend A. This design choice offers several advantages, though it may introduce limitations in certain cases. First, incorporating A enhances the *interpretability* of the TASC model, as it captures the underlying trend in time-series data. Second, if trend A persists beyond the intervention point, it provides TASC with strong *predictive power*, which is particularly advantageous over longer time horizons. Third, by modeling A, TASC becomes *sensitive to the time orders*, rather than being agnostic to them. Overall, TASC leverages temporal structure when a consistent trend exists but offers limited benefits when the trend is weak or absent (e.g., $A = 0$).

Another benefit of TASC is the approximately low-rank structure, represented as $Y = HX + E$, where HX is an exactly low-rank signal and E denotes noise. This low-rank property commonly appears in real-world datasets, especially as the data size increases. In such cases, Principal Component Analysis (PCA) is a widely adopted technique for extracting low-rank signals. It has been applied in various domains such as image processing, speech and audio analysis, genomics, finance, and sensor data analysis, and has also been utilized in synthetic control methods by [12]. However, the performance of PCA deteriorates in the presence of substantial observational noise, as it assumes that the learned principal components are noise-free in the selected directions. In contrast, TASC may offer improved robustness in extracting signals under noisy conditions by assuming omnidirectional noise (i.e., the noise matrix is full rank).

4 Evaluating Effect of Proposition 99 in California

We illustrate our method on the Proposition 99 case from [3]. Proposition 99 was a policy enacted in California in 1988 that significantly increased the state's cigarette tax. This policy was followed by a noticeable decline in cigarette sales (black line in Figure 2). To assess whether this decline was causally driven by the policy, synthetic control methods can be applied to estimate the counterfactual outcome for California—i.e., what cigarette sales would have looked like had the policy not been implemented. For economic analyses, multiple auxiliary predictors are often used to improve predictive accuracy. For example, [3] incorporate variables such as the average retail price of cigarettes, per capita state personal income, the percentage of the population aged 15–24, and per capita beer consumption, in addition to the target time series (per-capita cigarette sales). However, in our analysis, we intentionally focus on a single predictor, per-capita cigarette sales, to ensure a fair and consistent comparison across different methods. Our goal is not to produce the most accurate estimate of the effect of the policy, but to evaluate the performance of competing methodologies under a controlled setting.

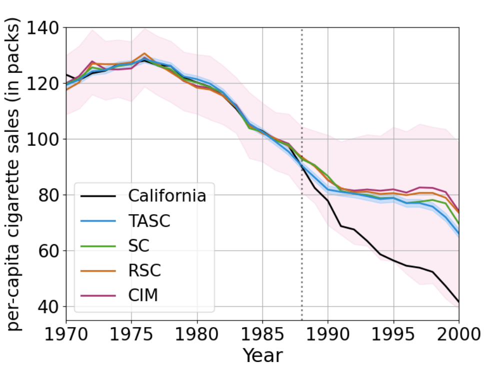

Figure 2: Per-capita cigarette sales in packs in California (black line), and estimated counterfactual values from SC.

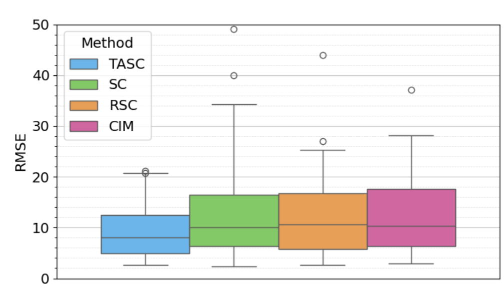

Figure 3: Post-intervention RMSE from placebo test for different SC algorithms.

We compare the estimated counterfactuals by Synthetic Control (SC [3]), Robust Synthetic Control (RSC [12]), Causal Impact Model (CIM [25]), and our TASC [2]. In Figure 2, all predictions lie above the observed trend, with the gap capturing the policy's effect. To test which SC estimates are most reliable, we use a *placebo test*: predicting each donor's time series by constructing a separate SC instance using the remaining donors as the donor pool. A valid SC method should also accurately reproduce untreated outcomes. Figure 3 shows post-intervention root mean squared error (RMSE) from the placebo test. Among these, TASC achieves the lowest RMSE and all three benchmarks show similar median RMSE with SC having the largest variance. This ordering suggests that the TASC estimates in Figure 2 likely yield counterfactual predictions with smaller error and may therefore be closer to the true counterfactual. In Appendix D, we present additional analyses with the Proposition 99 case study.

5 Conclusion and Future Work

In this work, we introduced Time-Aware Synthetic Control using a state-space model. We provide algorithms for model learning and counterfactual inference based on Kalman filtering and RTS smoothing. With this approach, we enforce synthetic control algorithm to be aware of the temporal structure of data, which was not explicitly modeled in many variations of synthetic control algorithms. We demonstrated TASC on the classical Proposition 99 case study, and the placebo test showed promising result that TASC may be a favorable choice in certain datasets over SC or RSC. We suspect that a pronounced temporal trend is the key for TASC to succeed, which is a common case for a lot of health and clinical time-series panel data.

Looking forward, we see several directions for extension. In terms of the model and algorithms, we can incorporate multiple auxiliary time series, improve the parameter learning algorithm, and also consider non-linear state-space models. From a theoretical perspective, it would be valuable to identify conditions under which TASC is preferable to other SC variants, and to determine if these conditions are testable. For the applications, we can apply TASC to clinical trial datasets or electronic health records. With the promising initial result, we envision that TASC could serve as a robust framework for integrating causal inference with time-series modeling.

Acknowledgement

This work was partially supported by ONR (N00014-24-1-2687, N00014-24-1- 2700), and the Columbia-Dream Sports AI Innovation Center PhD Fellowship.

[2] SC uses the relative importance matrix $V = I$ (See [3] for more information, it is equivalent to linear regression with simplex constraint introduced in Section 2.1); RSC uses top $d = 2$ singular values in noise filtering step and Ridge regression coefficient was set to 0.1 (See [12] for more information); $TASC$ uses $d = 2$ for hidden state dimension

References

[1] David Card and Alan B Krueger. Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania. *American Economic Review*, 90(5):1397–1420, 2000.

[2] Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1):113–132, 2003.

[3] 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, 2010.

[4] Alberto Abadie and Jérémy L'Hour. A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834, 2021.

[5] Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510, 2015.

[6] Noémi Kreif, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton. Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics*, 25(12):1514–1528, 2016.

[7] Michael W Robbins, Jessica Saunders, and Beau Kilmer. A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, 112(517):109–126, 2017.

[8] Giacomo Vagni and Richard Breen. Earnings and income penalties for motherhood: estimates for British women using the individual synthetic control method. *European Sociological Review*, 37(5):834–848, 2021.

[9] Kristian Thorlund, Louis Dron, Jay JH Park, and Edward J Mills. Synthetic and external controls in clinical trials–a primer for researchers. *Clinical epidemiology*, pages 457–467, 2020.

[10] Dennis Shen, Anish Agarwal, Vishal Misra, Bjoern Schelter, Devavrat Shah, Helen Shiells, and Claude Wischik. Obtaining personalized predictions from a randomized controlled trial on alzheimer's disease. *Scientific Reports*, 15(1):1671, 2025.

[11] Madeleine Udell and Alex Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.

[12] Muhammad Amjad, Devavrat Shah, and Dennis Shen. Robust synthetic control. *Journal of Machine Learning Research*, 19(22):1–51, 2018.

[13] Susan Athey, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116:1716–1730, 2021.

[14] Nikolay Doudchenko and Guido W Imbens. Balancing, regression, difference-in-differences and synthetic control methods: A synthesis, 2016. NBER Working Paper 22791.

[15] Muhammad Amjad, Vishal Misra, Devavrat Shah, and Dennis Shen. mRSC: Multi-dimensional robust synthetic control. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(2):1–27, 2019.

[16] Arindrajit Dube and Ben Zipperer. Pooling multiple case studies using synthetic controls: An application to minimum wage policies, 2015. IZA Discussion Paper No. 8944.

[17] Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017.

[18] Saeyoung Rho, Andrew Tang, Noah Bergam, Rachel Cummings, and Vishal Misra. Clustersc: Advancing synthetic control with donor selection. In *International Conference on Artificial Intelligence and Statistics*, pages 109–117. PMLR, 2025.

[19] Eli Ben-Michael, Avi Feller, and Jesse Rothstein. The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803, 2021.

[20] Alberto Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, 2021.

[21] James H Stock and Mark W Watson. Forecasting inflation. *Journal of monetary economics*, 44(2):293–335, 1999.

[22] Allan W Gregory and Allen C Head. Common and country-specific fluctuations in productivity, investment, and the current account. *Journal of Monetary Economics*, 44(3):423–451, 1999.

[23] Mario Forni, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554, 2000.

[24] 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, 2021.

[25] 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):247–274, 2015.

[26] Danny Klinenberg. Synthetic control with time varying coefficients a state space approach with bayesian shrinkage. *Journal of Business & Economic Statistics*, 41(4):1065–1076, 2023.

[27] Junzhe Shao, Mingzhang Yin, Xiaoxuan Cai, and Linda Valeri. Generalized synthetic control method with state-space model. In *NeurIPS 2022 Workshop on Causality for Real-world Impact*, 2022.

[28] Simo Särkkä and Lennart Svensson. *Bayesian filtering and smoothing*, volume 17. Cambridge university press, 2023.

A Basic Algorithms

In this section, we provide the full pseudocode for the basic algorithms comprising the EM approach: Kalman Filter (Algorithm 2), RTS Smoother (Algorithm 3), and the M-step (Algorithm 4).

Algorithm 2 Kalman Filter

Input : $y_k \in \mathbb{R}^{n+1}$, previous estimate m_{k-1}, P_{k-1}, current parameter $\theta = \{A, H, Q, R, m_0, P_0\}$
Output : m_k, P_k
$m_{k|k-1} \leftarrow Am_{k-1}$ /* prediction from the previous timestep $k-1$ */
$P_{k|k-1} \leftarrow AP_{k-1}A^\top + Q$ /* prediction from the previous timestep $k-1$ */
$v_k \leftarrow y_k - Hm_{k|k-1}$
$S_k \leftarrow HP_{k|k-1}H^\top + R$
$K_k \leftarrow P_{k|k-1}H^\top S_k^{-1}$ /* Kalman Gain */
$m_k \leftarrow m_{k|k-1} + K_k v_k$ /* Update after observing y_k */
$P_k \leftarrow P_{k|k-1} - K_k S_k K_k^\top$ /* Update after observing y_k */

Algorithm 3 RTS Smoother

Input : Kalman filter estimate m_k, P_k, smooted estimate m^s_{k+1}, P^s_{k+1}, current parameter $\theta = \{A, H, Q, R, m_0, P_0\}$
Output : m^s_k, P^s_k
$m_{k+1|k} \leftarrow Am_k$ /* prediction from Kalman filter estimate m_k */
$P_{k+1|k} \leftarrow AP_kA^\top + Q$ /* prediction from Kalman filter estimate P_k */
$G_k \leftarrow P_kA^\top P^{-1}_{k+1|k}$
$m^s_k \leftarrow m_k + G_k[m^s_{k+1} - m_{k+1|k}]$ /* $m^s_t = m_t$ for the last timestep $t = T_0$ or T */
$P^s_k \leftarrow P_k + G_k[P^s_{k+1} - P_{k+1|k}]G_k^\top$ /* $P^s_t = P_t$ for the last timestep $t = T_0$ or T */
return m^s_k, P^s_k, G_k

Algorithm 4 Parameter Update (M-Step)

Input : current parameter $\theta = \{A, H, Q, R, m_0, P_0\}$, length of the sequence T, RTS parameters m^s_k, P^s_k, G_k for all $k \in \{0, \dots, T\}$, observations y_k for all $k \in \{1, \dots, T\}$
Output : θ'
Define
$\Sigma = \frac{1}{T}\sum_{k=1}^{T} P^s_k + m^s_k {m^s_k}^\top$
$\Phi = \frac{1}{T}\sum_{k=1}^{T} P^s_{k-1} + m^s_{k-1} {m^s_{k-1}}^\top$
$B = \frac{1}{T}\sum_{k=1}^{T} y_k {m^s_k}^\top$
$C = \frac{1}{T}\sum_{k=1}^{T} P^s_k G^\top_{k-1} + m^s_k {m^s_{k-1}}^\top$
$D = \frac{1}{T}\sum_{k=1}^{T} y_k {y_k}^\top$
Update
$A' \leftarrow C\Phi^{-1}$
$H' \leftarrow B\Sigma^{-1}$
$Q' \leftarrow \mathrm{Diag}(\Sigma - 2CA^\top + A\Phi A^\top)$ /* $\mathrm{Diag}(\cdot)$ keeps only the diagonal elements of the input */
$R' \leftarrow \mathrm{Diag}(D - 2BH^\top + H\Sigma H^\top)$
$m'_0 \leftarrow m^s_0$
$P'_0 \leftarrow P^s_0 + (m^s_0 - m_0)(m^s_0 - m_0)^\top$
return $\theta' = \{A', H', Q', R', m'_0, P'_0\}$

B Full description of Time-Award Synthetic Control (TASC) Algorithm

In this section, we disclose the derivation of TASC.

B.1 Pre-intervention Fit

For pre-intervention data, we can take the classical EM approach for a linear gaussian state-space model. The E-step comprises of a forward pass (Kalman filter) and a backward pass (RTS smoothing). This gives us estimates m_k^s and P_k^s to define a lower bound for the posterior probability distribution.

Algorithm 5 $EM_{pre}(Y_{pre}; N)$, EM for Pre-intervention Fit

Data: Y_{pre} where (i, j)-th element is $y_{i,t}$ $\forall (i, t) \in [1 : n + 1] \times [1 : T_0]$ (pre-intervention data from the target and donors)
Result: $\theta = \{A, H, Q, R, m_0, P_0\}$
Initialize $\theta^{(0)}$
for $i \leftarrow 1$ **to** N **do**
  for $k \leftarrow 1$ **to** T_0 **do**
    Update m_k, P_k via Kalman filtering with $\theta^{(i-1)}$; `/* forward pass */`
  end
  for $k \leftarrow T_0 - 1$ **to** 0 **do**
    Update m_k^s, P_k^s, G_k via RTS Smoothing with $\theta^{(i-1)}$; `/* backward pass */`
  end
  Update $\theta^{(i)}$ via the M-step of EM
end
return $\theta^{(N)}$

TASC utilizes Algorithm 5 to learn the parameters θ from the pre-intervention data. With these fixed parameters, TASC uses Kalman filter and RTS smoother to estimate the internal states m_k^s, P_k^s, and then translate these to finally estimate the post-intervention time series. However, this is impossible without a special treatment since the first element of y_k (which belongs to the target unit) is missing. To handle this, we deem that the target unit's data is missing, and separate the donor portion of the data and parameters:

$$y_t = \begin{bmatrix} y_{t,1} \\ y_{t,2} \end{bmatrix}, r_t = \begin{bmatrix} r_{t,1} \\ r_{t,2} \end{bmatrix}, H = \begin{bmatrix} h_1^\top \\ H_2 \end{bmatrix}, \text{ and } R = \begin{bmatrix} r_1 & 0 \\ 0 & R_2 \end{bmatrix},$$

where $y_{t,2}, r_{t,2} \in \mathbb{R}^n$, $H_2 \in \mathbb{R}^{n \times d}$, and $R_2 \in \mathbb{R}^{n \times n}$.

Then, we redefine the observation model (i.e., Equation (3)) to be the following

$$y_{t,2} = H_2 x_t + r_{t,2},$$

where $r_{t,2} \sim \mathcal{N}(0, R_2)$. With this new model, the post-intervention observations will not inform the target-related parameters: h_1 and r_1. This is equivalent to setting $r_1 \to \infty$ in the original model.

Algorithm 6 shows the Kalman filtering process with infinite variance for the target observation (i.e., $r_1 \to \infty$). Note that

$$\begin{aligned} K_k &= P_{k|k-1} H^\top S_k^{-1} \\ &= \begin{bmatrix} 0 & P_{k|k-1}^{-1} H_2^\top (H_2 P_{k|k-1} H_2^\top + R_2)^{-1} \end{bmatrix}, \end{aligned}$$

and the update on m_k is

$$\begin{aligned} m_k &= m_{k|k-1} + K_k v_k \\ &= m_{k|k-1} + P_{k|k-1}^{-1} H_2^\top (H_2 P_{k|k-1} H_2^\top + R_2)^{-1} \, v_{k,2}, \end{aligned}$$

(Note: $v_{k,2} \in \mathbb{R}^n$ is the last n elements of v_k—i.e., $v_{k,2} = y_{k,2} - H_2 m_{k|k-1}$) , and the update on P_k is

$$\begin{aligned} P_k &= P_{k|k-1} + K_k S_k K_k^\top \\ &= P_{k|k-1} + P_{k|k-1} H_2^\top (H_2 P_{k|k-1} H_2^\top + R_2)^{-1} H_2 P_{k|k-1}^\top. \end{aligned}$$

Algorithm 6 Kalman Filter with Infinite Variance

Input : $y_k \in \mathbb{R}^{n+1}$ with the target(first) element missing, previous estimate m_{k-1}, P_{k-1}, current parameter $\theta' = \{A, H, Q, R, m_0, P_0\}$, where $R'_{1,1} = \infty$

Output : m_k, P_k

Define h_1, H_2, R_2 from $H = \begin{bmatrix} h_1^\top \\ H_2 \end{bmatrix}$, $R' = \begin{bmatrix} \infty & 0 \\ 0 & R_2 \end{bmatrix}$

$y_k \leftarrow [h_1^\top m_{k|k-1}, y_{1,k}, \ldots, y_{n,k}]^\top$ /* augment target values */

$m_{k|k-1} \leftarrow A m_{k-1}$ /* prediction from the previous timestep $k-1$ */

$P_{k|k-1} \leftarrow A P_{k-1} A^\top + Q$ /* prediction from the previous timestep $k-1$ */

$v_k \leftarrow y_k - H m_{k|k-1}$ /* the first element is zero */

$S_k \leftarrow H P_{k|k-1} H^\top + R'$

$S_k^{-1} \leftarrow \begin{bmatrix} 0 & 0 \\ 0 & (H_2 P_{k|k-1} H_2^\top + R_2)^{-1} \end{bmatrix}$ /* by Schur Complement */

$K_k \leftarrow P_{k|k-1} H^\top S_k^{-1}$

$m_k \leftarrow m_{k|k-1} + K_k v_k$ /* Update after observing y_k */

$P_k \leftarrow P_{k|k-1} - K_k S_k K_k^\top$ /* Update after observing y_k */

With the infinite variance, post-intervention target time series do not affect the outcome of Kalman filtering, hence we can set it to any value (such as zero for the first element in $y_k \leftarrow [0, y_{2,k}, \ldots, y_{n+1,k}]^\top$). The RTS Smoother algorithm remains the same, as it does not use R or y_k as an input—it only utilizes m_k, P_k estimates from the Kalman filter in addition to the parameters A and Q. As a result, this only changes the Kalman filter part in the post-intervention time steps from Algorithm 2 to Algorithm 6. The full description of TASC is provided in Algorithm 7.

Algorithm 7 TASC($Y; N_1$)

Data: $y_{i,t}\ \forall (i,t) \in [1:n+1] \times [1:T_0]$ and $\forall (i,t) \in [2:n+1] \times [T_0+1:T]$

Result: $\hat{\theta} = \{A, H, Q, R, m_0, P_0\}, \hat{y}_{1,T_0+1}, \ldots, \hat{y}_{1,T}$

Learn $\theta^{N_1} \leftarrow EM_{pre}(Y_{pre}; N_1)$

for $k \leftarrow 1$ **to** T_0 **do**

  Update m_k, P_k via Algorithm 2 with θ^{N_1} /* pre-intervention forward pass */

end

for $k \leftarrow T_0 + 1$ **to** T **do**

  Update m_k, P_k via Algorithm 6 with θ^{N_1} /* post-intervention forward pass */

end

for $k \leftarrow T - 1$ **to** 0 **do**

  Update m_k^s, P_k^s via RTS Smoothing with $\theta^{(i-1)}$ /* backward pass */

end

Define $H = \begin{bmatrix} h_1^\top \\ H_2 \end{bmatrix}$

for $k \leftarrow T_0 + 1$ **to** T **do**

  Predict $\hat{y}_{1,t} \leftarrow h_1^\top m_t^s$ /* Post-intervention target prediction */

end

return $\theta^{N_1}, \hat{y}_{1,T_0+1}, \ldots, \hat{y}_{1,T}$

C Benchmark Synthetic Control Algorithms

In this section, we provide a full algorithm description for the benchmark synthetic control algorithms, SC and RSC.

The classical Synthetic Control (SC) performs a vertical regression with simplex constraint [3]. Algorithm 8 shows the classical SC implemented in our study, where the importance matrix V is set to an identity matrix. Note that our implementation in Proposition 99 study is slightly different from the original analysis (in [3]) because we only use the target time series of interest (per capita tobacco sales in packs) without any additional covariates.

Algorithm 8 Synthetic Control, [3]

Data: Target unit's pre-intervention data $Y_1^- \in \mathbb{R}^{T_0}$, Donor data $Y = [Y^-, Y^+] \in \mathbb{R}^{n \times T}$
Result: Counterfactual prediction $\hat{Y}_1^+$, SC weights f
1. Learn
$f = \arg\min_f ||Y_1^- - f^\top Y^-||^2$ where $0 \leq f \leq 1, \sum_{i=1}^n f_i = 1$ `/* Simplex constraint */`
2. Project $\hat{Y}_1^+ = f(Y^+)$
3. Infer the estimated causal effect of the intervention for the target is $Y_1^+ - \hat{Y}_1^+$

Robust Synthetic Control (RSC, [12]) performs hard singular-value thresholding (HSVT) as a pre-processing step to denoise the observation data. Then, it learns a vertical regression model using the pre-intervention portion of the data, and projects with the post-intervention data for counterfactual inference. Algorithm 9 describes our adoption of the original algorithm with the observation probability $p = 1$ (no missing data).

Algorithm 9 Robust Synthetic Control [12]

Data: Target unit's pre-intervention data $Y_1^- \in \mathbb{R}^{T_0}$, Donor data $Y = [Y^-, Y^+] \in \mathbb{R}^{n \times T}$, Number of singular values to keep d
Result: Counterfactual prediction $\hat{Y}_1^+$, SC weights f
1-1. Denoise
$Y = \sum_{i=1}^{min(n,T)} s_i u_i v_i^\top$ `/* Singular Value Decomposition (SVD) */`
$\tilde{Y} = \sum_{i=1}^{d} s_i u_i v_i^\top$ `/* Hard Singular Value Thresholding (HSVT) */`
1-2. Learn $f = \arg\min_f ||Y_1^- - f^\top \tilde{Y}^-||^2 + \lambda ||f||^2$
2. Project $\hat{Y}_1^+ = f^\top \tilde{Y}^+$
3. Infer the estimated causal effect of the intervention for the target is $Y_1^+ - \hat{Y}_1^+$

D Additional Experiments with Proposition 99 Data

In this section, we provide additional experiments we ran with Proposition 99 data that were not included in the main text due to the space limit. Following the approach from [3], we plot the difference between the observed outcome and the predicted counterfactual for California, alongside those of the control states obtained from the placebo test. The first column of Figure 4 reports the gap between prediction and observation of per-capita cigarette sales (in packs) for California and the 38 control states. Subsequent columns restrict the set of control states based on the relative quality of pre-intervention fit, measured by mean-squared error (MSE). The second column includes only control states whose pre-intervention MSE is no more than 10 times that of California, while the third and fourth columns apply stricter thresholds of 5 and 2 times, respectively.

Notably, the last row (CIM) retains the largest number of control states as stricter thresholds are imposed, indicating that the accuracy of pre-intervention prediction was similar across states. In contrast, the TASC approach (the first row) retains substantially fewer control states under the most stringent threshold compared to SC or RSC. This indicates that the pre-intervention fit for California was more accurate than other states when using TASC. Note that California is one of the most populated states in the US, and hence the collected data (per capita cigarette sales) may have lower variance compared to other states due to averaging effect. In such case, TASC may have learned smaller observation noise variance (R) and yielded more accurate (pre-intervention) fit for California. Indeed, corresponding variance for California was the smallest (2.58, with median 12.95, standard deviation 36.17 and maximum 170.79). Across all specifications, California consistently displays the largest gap, while it is more apparent in the plots in the top right corner (stricter thresholds, TASC method). The estimate effect of policy is similar across methods, diverging only at the end. TASC and RSC shows flatter estimates closer to 2000, whereas SC and CIM estimates keep increasing.

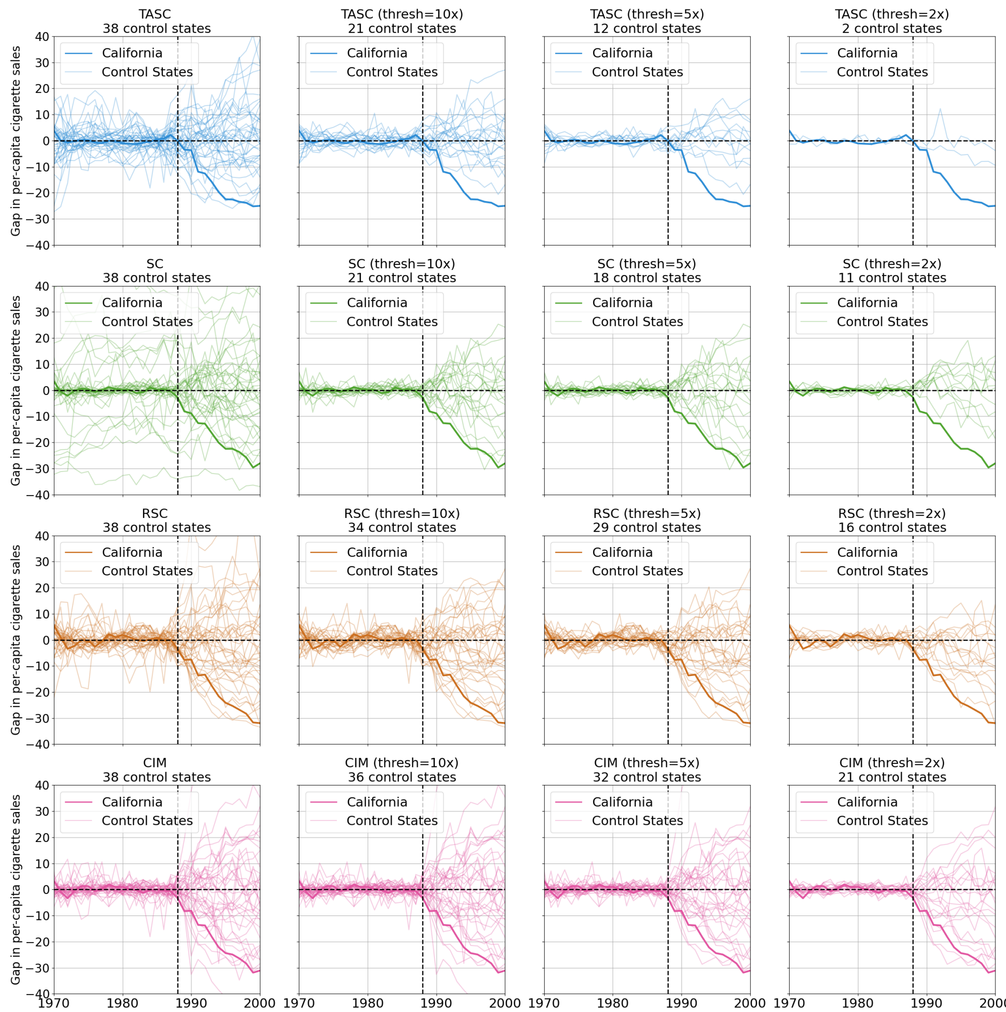

Figure 4: Gap in per-capita cigarette sales (in packs) between the observed outcome and synthetic control predictions (comparable to Figures 4–7 in [3]). Each row represents a different algorithm—TASC, SC, RSC, and CIM (from top to bottom)—while each column applies a different threshold for selecting control units: no threshold, at most 10 times California's pre-intervention error, 5 times, and 2 times (from left to right).