Proximal Causal Inference for Synthetic Control with Surrogates

Jizhou Liu
University of Chicago

Eric J. Tchetgen Tchetgen
University of Pennsylvania

Carlos Varjão
Amazon.com, Inc

Abstract

The synthetic control method (SCM) has become a popular tool for estimating causal effects in policy evaluation, where a single treated unit is observed. However, SCM faces challenges in accurately predicting post-intervention potential outcomes had, contrary to fact, the treatment been withheld, when the pre-intervention period is short or the post-intervention period is long. To address these issues, we propose a novel method that leverages post-intervention information, specifically time-varying correlates of the causal effect called "surrogates", within the synthetic control framework. We establish conditions for identifying model parameters using the proximal inference framework and apply the generalized method of moments (GMM) approach for estimation and inference about the average treatment effect on the treated (ATT). Interestingly, we uncover specific conditions under which exclusively using post-intervention data suffices for estimation within our framework. Through a synthetic experiment and a real-world application, we demonstrate that our method can outperform other synthetic control methods in estimating both short-term and long-term effects, yielding more accurate inferences.

1 INTRODUCTION

The synthetic control method (SCM), first introduced by Abadie and Gardeazabal (2003), has become a widely used baseline method in empirical research across economics and social sciences. It is designed to estimate the impact of an intervention experienced by a single unit by predicting the counterfactual scenario of what would have occurred if the intervention had not taken place, utilizing outcomes from unaffected units. As formalized in Abadie et al. (2010), the approach involves matching the treated unit to a weighted average of control units, referred to as "synthetic control", such that the resulting synthetic control matches the pre-intervention outcome trajectory over time for the treated units. Despite its significance in policy evaluation literature, the SCM may struggle to accurately reproduce the trajectory of the outcome for the treated unit when faced with a limited number of pre-intervention periods or structural breaks during a long pre- and post-intervention period (see Abadie (2021)). In response to these challenges, this paper proposes a novel method that utilizes post-intervention information to enhance the precision of the estimated causal effects.

Our approach leverages time-varying correlates of the causal effect that can be integrated into the synthetic control framework. Contrary to standard SCM, which requires donor units to remain unaffected by the intervention, our proposed model repurposes correlates as ideal surrogates of causal effects to the extent that they are predictive of treatment effects. This innovative method aims to enhance the performance of SCM in various situations, including estimating long-term effects and handling short pre-intervention periods, by incorporating additional post-intervention information, specifically surrogates.

We first establish the conditions required for identifying model parameters within the proximal causal inference framework. We then employ the generalized method of moments (GMM) approach to estimate and conduct statistical inference on the average treatment effect on the treated (ATT). In the pre-intervention period, we rely on the donor series to learn the "synthetic control", conceptually similar to the classic synthetic control method (see Shi et al. (2021)). In the post-intervention period, our method augments the estimation of synthetic weights and the causal effect series with surrogates. Our identification and estimation strategy is based on a joint set of moment equa-

tions that utilize data from both periods. Under this framework, we highlight the possibility of using only post-treatment data for estimation. This is potentially feasible as outcomes of treatment comprise of two components: one influenced by the latent factors affecting donors, and the other by those affecting surrogates. Because our method utilizes observations for both surrogates and donors, it allows for simultaneous learning of these components.

1.1 Related Work

The Synthetic Control Frameworks. The synthetic control literature has primarily focused on frameworks that assume either a perfect match of the synthetic control to the treated unit in the pretreatment period (see for example Abadie et al. (2010)) or a perfect match of the underlying unobserved factor loadings (see for example Ferman (2021); Amjad et al. (2018); Powell (2018); Ferman and Pinto (2021)). However, it is rarely the case that the perfect match assumption holds exactly in the observed data. Furthermore, models that assume perfect match of underlying factor loadings may be subject to bias due to the measurement error that arises in the synthetic control regression of the treated unit on donor outcomes. To address this issue, some researchers have proposed using lagged control unit outcomes as instrumental variables Ferman and Pinto (2021) or denoising the data through singular value thresholding Amjad et al. (2018), under an assumption of independent idiosyncratic error terms across units and time. In contrast, our paper employs the proximal causal inference framework proposed by Shi et al. (2021), which offers several advantages, such as effectively addressing the measurement error issue and providing convenient access to off-the-shelf Generalized Method of Moments (GMM) approaches for estimation and inference of average treatment effects on the treated (ATT). Recent advancements in the field of proximal causal inference include Miao et al. (2018); Shi et al. (2020); Cui et al. (2023); Miao et al. (2020).

Augmented Synthetic Control Method. Our method contributes to the expanding literature on methodological advancements that enhance traditional synthetic control methods by incorporating supplementary estimation techniques or additional information. Ben-Michael et al. (2021) propose an augmented synthetic control method that employs ridge regression to rectify disparities in pre-treatment outcomes between treated units and un-augmented synthetic control estimators. Kellogg et al. (2021) advocate for the integration of synthetic control estimators and nearest-neighbor matching estimators in a weighted average to mitigate interpolation and extrapolation bi-

ases. A number of prior studies have utilized penalization schemes to regularize synthetic control weights (for example, Abadie and L'Hour (2021), Doudchenko and Imbens (2016), Chernozhukov et al. (2021)).

2 SETUP AND PRELIMINARIES

Let Y_t denote the observed outcome for the target unit, W_t denote a $(N \times 1)$ vector of the observed outcome for donors, and X_t denote a $(K \times 1)$ vector of observed surrogates for time periods $t = 1, \ldots, T$. Denote by $Y_t(0)$ the outcome of the target unit at time period t that would be observed in the absence of the intervention, and by $Y_t(1)$ the outcome of the target unit at time period t that would be observed in the presence of the intervention. The (observed) outcome and potential outcomes are related to treatment assignment by the relationship

$$Y_t = \mathbf{1}\left(t > T_0\right) Y_t(1) + \mathbf{1}\left(t \le T_0\right) Y_t(0) , \quad (1)$$

where $1 < T_0 < T$ denotes the time period when the target unit receives treatment. In this paper, we focus on estimation and inference on the so-called average treatment effect on the treated (ATT), which is given by:

$$\tau = E\left[\frac{1}{T - T_0} \sum_{t=T_0}^{T} Y_t(1) - Y_t(0) \right] . \quad (2)$$

Suppose that $Y_t(0)$ and W_t follow an interactive fixed effect model (IFEM).

$$\begin{aligned} Y_t(0) &= \lambda_t' \beta + \epsilon_{Y,t} \\ W_t' &= \lambda_t' \Gamma + \epsilon_{W,t}' , \end{aligned} \quad (3)$$

where λ_t is a $(F \times 1)$ vector of common factors, for $t = 1, \ldots, T$, β is an $(F \times 1)$ vector of factor loadings, Γ is an $(F \times N)$ matrix of factor loadings, $\epsilon_{Y,t}$ and $\epsilon_{W,t}$ are the corresponding error terms. Similarly, we consider a synthetic surrogate model in post intervention period based on a similar factor model:

$$\begin{aligned} Y_t(1) - Y_t(0) &= \rho_t' \theta + \delta_t \\ X_t' &= \rho_t' \Phi + \epsilon_{X,t}' , \end{aligned} \quad (4)$$

where ρ_t is a $(K \times 1)$ vector of common factors, for $t = 1, \ldots, T$, and θ is an $(K \times 1)$ vector of factor loadings, X_t is an $(H \times 1)$ vector of surrogates, Φ is a $(K \times H)$ matrix of factor loadings, δ_t and $\epsilon_{X,t}$ are the corresponding error terms.

The IFEM model in Equation (3) is commonly assumed in the synthetic control literature (see, for example, Xu, 2017; Ferman and Pinto, 2021; Shi et al., 2021). In this paper, we consider an additional surrogate model, as in Equation (4), which is analogous to

Equation (3). The motivation is to leverage extra post-intervention information sometimes available in the dataset to capture post-intervention period variation in causal effects, using surrogates that are highly correlated with post-intervention outcomes. This property is manifested in Equation (4), where both surrogates X_t and causal effects $Y_t(1) - Y_t(0)$ are driven by the same set of latent factors, as assumed in the classical synthetic control approach in Equation (3). To identify and estimate parameters in the model, we need a few more assumptions.

Assumption 2.1 (Conditional independence)
The error terms are conditional mean independent and independent of each other. $E(\epsilon_{Y,t} \mid \lambda_t) = E(\epsilon_{W,t} \mid \lambda_t) = E(\delta_t \mid \rho_t) = E(\epsilon_{X,t} \mid \rho_t) = 0$ *for* $t = 1, \ldots, T$.

Further, we make the following assumption that is key to identification of a synthetic control and the time series of post-intervention causal effects:

Assumption 2.2 (Existence of weights) *There exist a $(N \times 1)$ vector of weights α such that $\Gamma\alpha = \beta$, and a $(H \times 1)$ vector of weights γ such that $\Phi\gamma = \theta$.*

Remark 2.1 *A necessary condition for the existence of such a vector of weights α (or γ) is that the number of donors exceeds the number of latent factors λ_t (or the number of surrogates exceeds the number of latent factors ρ_t), i.e. $N \geq F$ (or $H \geq K$).* ∎

Next, we make the following assumption to identify synthetic weights α and coefficients of the surrogate γ, which can later be used for identifying causal effects:

Assumption 2.3 (Existence of proxies) *We observe proxies $\{Z_{0,t}, Z_{1,t}\}$ such that*

$$Z_{0,t} \perp\!\!\!\perp \{Y_i(0), W_t\} \mid \lambda_t \text{ for any } t \leq T_0 \ ,$$
$$\{Z_{0,t}, Z_{1,t}\} \perp\!\!\!\perp \{Y_i(1), W_t, X_t\} \mid \lambda_t, \rho_t \text{ for any } t > T_0 \ .$$

Assumption 2.3 implies that λ_t encompasses the entire set of confounders between $Z_{0,t}$ and $\{Y_i(0), W_t\}$, while (λ_t, ρ_t) encompasses the entire set of confounders between $\{Z_{0,t}, Z_{1,t}\}$ and $\{Y_i(1), W_t, X_t\}$. As argued in Shi et al. (2021), $Z_{0,t}$, which denotes proxy variables for outcomes of donors, is frequently available in synthetic control settings. For instance, a suitable choice of proxy variables might be the outcome of units that are potential donor candidates but were excluded from the donor pool. Similarly, proxy variables for surrogates, denoted by $Z_{1,t}$, could be potentially unused surrogate candidates.

In order to ensure unique identification of the synthetic control and surrogate weights, we introduce the following completeness condition.

Assumption 2.4 (Completeness) *For any square integrable function g, if*

$$E[g(W_t) \mid Z_{0,t} = z_{0,t}, t \leq T_0] = 0$$

for all $z_{0,t}$, then $g(W_t) = 0$ for any $t \leq T_0$. For any square integrable function h, if

$$E[h(W_t, X_t) \mid Z_{0,t} = z_{0,t}, Z_{1,t} = z_{1,t}, t > T_0] = 0$$

for all $z_{0,t}, z_{1,t}$, then $h(W_t, X_t) = 0$ for any $t > T_0$.

Assumption 2.4 is formally known as completeness condition which characterizes the informativeness of $Z_{0,t}$ about W_t (similarly for $\{Z_{0,t}, Z_{1,t}\}$ and $\{W_t, X_t\}$ in the post-intervention period), in the sense that any infinitesimal variation in W_t is captured by variation in $Z_{0,t}$ such that no information has been lost through projection of W_t on $Z_{0,t}$ in the pre-treatment period.

Building on Assumptions 2.3-2.4, we present a practical approach for identifying and estimating causal effects by finding all necessary variables (donors, surrogates, and proxies). The initial step involves identifying as many appropriate donors and surrogates as possible to ensure $N > F$ and $H > K$. To find suitable donors, we adhere to the standard approach in synthetic control literature. For instance, in the German reunification example, Abadie (2021) considered a set of industrialized countries as potential donors for West Germany. Regarding surrogates, strong candidates should be highly predictive of magnitude of the causal effect for the target unit, which may include covariates that potentially mediate the causal effect of interest for the target unit in the post-intervention period and outcomes or covariates of other units affected by the treatment. Subsequently, we select N_0 donors to construct synthetic controls and H_0 surrogates to predict causal effects. The remaining $N - N_0$ donors and $H - H_0$ surrogates can serve as $Z_{0,t}$ and $Z_{1,t}$. In Section 4.2, we present a real-world application of our method, serving as a motivating example to help readers understand the concepts of donors, surrogates, proxies, and factors.

3 MAIN RESULTS

In this section, we present our main results on identification and estimation. We apply the off-the-shelf GMM approach for identifying model parameters and estimating the average treatment effect on the treated (ATT). To begin with, under assumptions provided in the previous section, we have the following identification results:

Theorem 3.1 (Identification) *Under Assumption 2.1-2.2, we have $E[Y_t(0)] = E[W_t'\alpha]$ and $E[Y(1)] =$*

$E[W_t'\alpha + X_t'\gamma]$ *for any* $1 \le t \le T$. *Under Assumption 2.1-2.4, (α, γ) is uniquely identified by solving*

$$E[Y_t - W_t'\alpha \mid Z_{0,t}, t \le T_0] = 0 \text{ for any } t \le T_0 ,$$
$$E[Y_t - W_t'\alpha - X_t'\gamma \mid Z_{0,t}, Z_{1,t}, t > T_0] = 0 \quad (5)$$
$$\text{for any } t > T_0 .$$

Using the identification results, we formulate the following moment conditions:

$$0 = E[U_t(\alpha, \gamma)]$$
$$= E\left[\begin{pmatrix} g_0(Z_{0,t})(Y_t - W_t'\alpha)I\{t \le T_0\} \\ g_1(Z_{0,t}, Z_{1,t})(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \end{pmatrix}\right],$$

where g_0, g_1 are d_0- and d_1-dimensional ($d_0 + d_1 \ge N + H$) vector of user-specified functions. The above moment equation motivates the use of the generalized method of moments (GMM) method with $U_t(\alpha, \gamma)$ providing identifying moment restrictions (Hansen (1982)). The GMM solves

$$(\hat{\alpha}, \hat{\gamma}) = \arg\min_{\alpha, \gamma} m(\alpha, \gamma)' \Omega m(\alpha, \gamma),$$

where $m(\alpha, \gamma) = 1/T \sum_{t=1}^{T} U_t(\alpha, \gamma)$ is the sample moments evaluated at an arbitrary value α, γ and Ω is a $(d_0 + d_1) \times (d_0 + d_1)$ user-specified symmetric and positive-definite weight matrix.

An interesting observation is that by forming the unconditional moments $E[g_1(Z_{0,t}, Z_{1,t})(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\}] = 0$ using equation (5), we can identify α and γ, as long as the dimension of g_1 exceeds the number of donors and surrogates. This finding suggests the possibility of using only post-treatment data for estimation. In Section 4.1, we validate the practical relevance of this idea through a simulation study.

Then, we make the following assumptions on the error terms:

Assumption 3.1 (Stationary error)
$\epsilon_{Y,t}, \epsilon_{W,t}, \delta_t, \epsilon_{X,t}$ *are stationary and weakly dependent processes.*

We adopt the weakly dependent assumption as a substitute for the independent and identically distributed errors assumption, while still ensuring the applicability of ergodic laws of large numbers and central limit theorems. The consistency and asymptotic normality of GMM estimators have been well-documented under such conditions, as demonstrated in studies by Hansen (1982), Newey and West (1986), and Hall (2005b). The theorem presented below is derived from Theorem 3.2 in Hall (2005b) and its proof can be found in Section B of the Supplementary Material.

Theorem 3.2 *Under Assumptions 2.1-2.4 and 3.1 and regularity conditions B.1-B.9 listed in Section B.2, as $T \to \infty$, we have*

$$\sqrt{T}(\hat{\theta} - \theta)$$
$$\xrightarrow{d} N\left(0, (G^\top \Omega G)^{-1} G^\top \Omega S \Omega^\top G (G^\top \Omega^\top G)^{-1}\right)$$

where $\theta = (\alpha', \gamma')'$, $S = \lim_{T \to \infty} \text{Var}\{\sqrt{T}m(\alpha, \gamma)\}$ is the variance-covariance matrix of the limiting distribution of $\sqrt{T}m(\alpha, \gamma)$ and the formula for G is provided as equation (8) in Section B.2.

Remark 3.1 *When utilizing the general method of moments framework, over-identification tests on moment restrictions are applicable when the number of moment conditions surpasses the number of parameters to be estimated (i.e., $d_0 + d_1 > N + H$). A testing procedure for this scenario was proposed by Hansen (1982), which involves calculating the J-statistic as follows:*

$$J_T = m(\hat{\theta})' \hat{\Omega} m(\hat{\theta}). \quad (6)$$

The validity of this testing procedure is supported by its asymptotic distribution, which is a chi-square distribution with $d_0 + d_1 - (N + H)$ degrees of freedom, as demonstrated in Lemma 4.2 of Hansen (1982). ∎

Remark 3.2 *It is worth noting that Assumption 2.4 is a stronger condition than what is strictly necessary for estimating the parameters of interest, θ. In the context of the IFEM model, we only require the invertibility of $G^\top \Omega G$. In other words, the rank conditions ensuring the relevance of $Z_{0,t}$ and W_t (as well as $Z_{0,t}, Z_{1,t}$ and W_t, X_t) are sufficient for the validity of the result.* ∎

Theorem 3.2 provides the asymptotic results required for estimation and statistical inference of the parameters (α, γ) identified by the structural model. The synthetic weights α potentially provide meaningful interpretation for the synthetic control, while γ can heuristically be used to gauge the potential predictive ability of surrogates for the target treatment effect over time. We aim to make inferences about the ATT, for which we provide two distinct estimation strategies that we incorporate in the GMM framework. Note that, we have

$$\tau = E\left[\frac{1}{T - T_0} \sum_{t=T_0}^{T} Y_t(1) - Y_t(0)\right]$$
$$= E\left[\frac{1}{T - T_0} \sum_{t=T_0}^{T} \rho_t'\theta\right] = E\left[\frac{1}{T - T_0} \sum_{t=T_0}^{T} X_t'\gamma\right],$$

which motivates the following estimator that plugs in the estimated surrogate coefficients and replaces pop-

ulation quantities with their sample counterparts:

$$\hat{\tau} = \frac{1}{T - T_0} \sum_{t=T_0}^{T} X_t' \hat{\gamma} \ .$$

Alternatively, we can use the difference between outcomes of the target unit and the synthetic control to estimate the average treatment effect as follows:

$$\tau = E\left[\frac{1}{T - T_0} \sum_{t=T_0}^{T} Y_t(1) - W_t'\alpha\right] \ ,$$

which motivates the following plug-in estimator:

$$\hat{\tau} = \frac{1}{T - T_0} \sum_{t=T_0}^{T} Y_t - W_t'\hat{\alpha} \ .$$

By incorporating both estimation strategies, we form the moment conditions $E[\tilde{U}_t(\alpha, \gamma, \tau)] = 0$ where

$$\tilde{U}_t(\alpha, \gamma, \tau)$$
$$= \begin{pmatrix} g_0(Z_{0,t})(Y_t - W_t'\alpha)I\{t \leq T_0\} \\ g_1(Z_{0,t}, Z_{1,t})(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \\ (Y_t - \tau - W_t'\alpha)I\{t > T_0\} \\ (X_t'\gamma - \tau)I\{t > T_0\} \end{pmatrix} \ .$$

Then, we can estimate τ and structural paramters (α, γ) simultaneously through the GMM approach that solves

$$(\hat{\alpha}, \hat{\gamma}, \tau) = \arg\min_{\alpha, \gamma, \tau} \tilde{m}(\alpha, \gamma, \tau)' \Omega \tilde{m}(\alpha, \gamma, \tau)$$

where $\tilde{m}(\alpha, \gamma, \tau) = 1/T \sum_{t=1}^{T} \tilde{U}_t(\alpha, \gamma, \tau)$. Similar to that of Theorem 3.2, we then present the following theorem for estimation and inference of ATT.

Theorem 3.3 *Under Assumptions 2.1-2.4 and 3.1 and regularity conditions B.1-B.9 listed in Section B.2, as $T \to \infty$, we have*

$$\sqrt{T}(\hat{\theta} - \theta)$$
$$\xrightarrow{d} N\left(0, \left(G^\top \Omega G\right)^{-1} G^\top \Omega S \Omega^\top G \left(G^\top \Omega^\top G\right)^{-1}\right)$$

where $\theta = (\alpha', \gamma', \tau)'$ includes τ as an extra parameter, and $S = \lim_{T \to \infty} \text{Var}\{\sqrt{T}\tilde{m}(\alpha, \gamma, \tau)\}$ is the variance-covariance matrix of the limiting distribution of $\sqrt{T}\tilde{m}(\alpha, \gamma, \tau)$ and the formula for G is provided as equation (9) in Section B.2.

Remark 3.3 *The GMM approach provided in Theorem 3.3 offers a flexible framework for conducting estimation and inference for various types of causal effects. For example, the moment equations $\tilde{U}_t(\alpha, \gamma, \tau)$ can be modified for estimating ATT of any specified period of time with $E[(X_t'\gamma - \tau)I\{t_1 < t < t_2\}] = 0$, percentage lift with $E[(X_t'\gamma - W_t'\alpha \cdot \tau)I\{t_1 < t < t_2\}]$, and so on.* ∎

4 EXPERIMENTS

4.1 Experiments on Synthetic Data

In this section, we examine the finite sample performance of our proposed method under various conditions. We simulate time series data for N control units and one treated unit with fixed pre-treatment period $T_0 = 100$ across different lengths of total periods $T = 200, 400, 800$. We assume the following data generating mechanism:

$$Y_t = I\{t > T_0\}(\rho_t'\theta + \delta_t) + \lambda_t'\beta + \epsilon_{Y,t}$$
$$W_{i,t} = \lambda_t'\Gamma_i + \epsilon_{i,t}$$
$$X_{i,t} = \rho_t'\Phi_i + \epsilon_{X,t}'$$

where $\epsilon_{Y,t}, \epsilon_{i,t}, \epsilon_{X,i,t} \overset{i.i.d}{\sim} N(0,1)$ denotes random errors. We simulate a vector of latent factors $\lambda_t = (\lambda_{1,t}, \ldots, \lambda_{F,t})'$ for donor model, where $\lambda_{k,t} \overset{i.i.d}{\sim} N(1,1)$ or $\lambda_{k,t} \overset{i.i.d}{\sim} N(\log(t), 1)$, $k = 1, \ldots, F$, and $F = 1, 5, 10$, and a vector of latent factors $\rho_t = \{\rho_{1,t}, \ldots, \rho_{K,t}\}$ for surrogate model, where $\rho_{i,t} \overset{i.i.d}{\sim} N(\mu_i, 1)$, $j = 1, \ldots, K$, and $K = 1, 5$. We set $\mu_1 = 1$ and $\mu_i = 0$ for $2 \leq i \leq K$. In this case, the average treatment effect is equal to one, i.e. $\tau = 1$. For each setting, we assume the number of control units $N = 2F$ and the number of surrogates $H = 2K$, and the first half of the control units and surrogates constitute the donor and surrogate pool. We specify factor loadings as follows: $\beta = (1, \ldots, 1)'$ and $\theta = (1, \ldots, 1)'$ are two vectors of ones, and $\Gamma = (I_F, I_F)\Phi = (I_K, I_K)$ stack two identity matrices, which correspondingly generate donors/surrogates and proxies. Respectively, Φ_i and Γ_i correspond to i-th column of Φ and Γ. This specification leads to exact identified linear systems with solutions $\beta = \alpha$ and $\gamma = \theta$. Notably, the SC weights α do not sum to one.

To estimate τ, we implement our proposed method taking the first half of control units as donors $\{W_{i,t} : i = 1, \ldots, F\}$ and second half of control units as supplemental proxies $\{W_{i,t} : i = F+1, \ldots, 2F\}$. Similarly, we use the first half of surrogates $\{X_{i,t} : i = 1, \ldots, K\}$ to predict causal effects and the second half of surrogates as proxies $\{X_{i,t} : i = K+1, \ldots, 2K\}$. As a comparison, we implement the following methods:

1. **(SC)** The unconstrained OLS under following linear regression model: $Y_t = \alpha_0 + I\{t > T_0\}\tau + W_t'\alpha + \nu_t$, $t = 1, \ldots, T$.

2. **(SC-S)** The unconstrained OLS with surrogates under following linear regression model: $Y_t = \alpha_0 + I\{t > T_0\}X_t'\gamma + W_t'\alpha + \nu_t$, $t = 1, \ldots, T$. We perform inference through the following moment

equations:

$$\tilde{U}_t(\alpha, \gamma, \tau) =$$
$$\begin{pmatrix} I\{t > T_0\}X_t(Y_t - (I\{t > T_0\}X_t'\gamma + W_t'\alpha)) \\ W_t(Y_t - (I\{t > T_0\}X_t'\gamma + W_t'\alpha)) \\ (X_t'\gamma - \tau)I\{t > T_0\} \end{pmatrix} .$$

3. **(PI)** The proximal inference method without using surrogates in Shi et al. (2021), which is based on the following moment conditions:

$$\tilde{U}_t(\alpha, \gamma, \tau) = \begin{pmatrix} Z_{0,t}(Y_t - W_t'\alpha)I\{t \le T_0\} \\ (Y_t - \tau - W_t'\alpha)I\{t > T_0\} \end{pmatrix} .$$

4. **(PI-P)** A proximal inference method using post-treatment period only. Specifically, we use the following moment conditions:

$$\tilde{U}_t(\alpha, \gamma, \tau) = \begin{pmatrix} Z_{0,t}(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \\ Z_{1,t}(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \\ (X_t'\gamma - \tau)I\{t > T_0\} \end{pmatrix} .$$

5. **(PI-S)** Our proposed proximal inference with surrogates from Section 3. Specifically, we use the following moment conditions:

$$\tilde{U}_t(\alpha, \gamma, \tau) = \begin{pmatrix} Z_{0,t}(Y_t - W_t'\alpha)I\{t \le T_0\} \\ Z_{1,t}(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \\ (X_t'\gamma - \tau)I\{t > T_0\} \end{pmatrix} .$$

Table 1 reports the Mean Squared Errors (MSE) of each estimation method for both independent and AR(1) errors, computed across 2,000 Monte Carlo iterations. We compare our proposed method, **PI-S**, with other methods under two data generating processes, one with an i.i.d normal latent factor λ_t with a mean and variance equal to one, and another with an i.i.d normal λ_t with a time trend $\log(t)$. We find that **PI-S** outperforms all other methods under almost all model specifications, having the lowest MSE. However, under $\lambda_t \sim N(1,1)$, **SC** and **SC-S** perform better than **PI-S** when the number of factors is large relative to the number of total time periods, i.e. $K = 5, T = 200$, possibly due to the "curse of dimensionality". Additionally, we observe that the performance of **PI-P** and **PI-S** improves as the post-treatment period increases, while the performance of other methods deteriorates, highlighting the advantage of our approach in the case of long post-treatment periods. Finally, we also find that as the dimension of latent factors increases, the performance of all proximal inference methods declines, particularly for **PI-P**.

Table 2 reports coverage rates at a 5% significance level for different estimation methods. Under the stationary model with $\lambda_t \sim N(1,1)$, all five methods achieve

coverage rates of approximately 95%. However, under the model with a time trend, i.e., $\lambda_t \sim N(\log(t), 1)$, regression-based methods (**SC** and **SC-S**) produce a large bias and fail to cover the ground truth, while all three proximal inference methods (**PI**, **PI-P**, and **PI-S**) exhibit good coverage rates. The comparison of coverage rates in the "robust" section versus the "HAC" section demonstrates the validity of our HAC robust inference approach for models with time series autocorrelation. When the dimension of latent factors is large relative to the number of observations, the proximal inference methods with surrogates (**PI-P** and **PI-S**) tend to become conservative. Importantly, alongside Table 1, we demonstrate the potential of applying proximal inference to estimate causal effects with post-intervention data only by highlighting relatively low MSEs and good coverage rates.

4.2 Experiments on Real-World Data

In this section, we examine and compare the five synthetic control methods discussed in Section 4.1 (**SC**, **SC-S**, **PI**, **PI-S**, and **PI-P**) using data collected from the study of the Panic of 1907, one of the most severe financial crises in US history. Fohlin and Lu (2021) re-analyzed the impact of this event on NYC trusts using a new high-frequency dataset on market valuations and a generalized synthetic control method proposed by Xu (2017). Their analysis employs a dataset covering 59 trust companies from January 5, 1906, to December 30, 1908, with a triweekly frequency, derived from the "Trust and Surety Company" table published in the New York Tribune. These trust companies are divided into three groups based on their connections and status during the crisis: the "troubled" group, the "connected" group, and the "independent" group. The troubled group comprises three trusts that experienced severe runs during the panic, while the connected group consists of seven trusts linked to four major "money trusts". The remaining 49 trusts are classified as the independent group.

In the original analysis by Fohlin and Lu (2021), the primary focus was on assessing the impact of the panic on both troubled and connected trusts collectively, taking into account the average performance of firms within each respective group. However, our study aims to highlight the disparities in estimation outcomes arising from the application of various methods explored in this paper. To achieve this, we consider the logarithm of the stock mid-price of Knickerbocker Trust Company as the target outcome, with the three troubled trusts serving as surrogates. Specifically, we use the bid price (logarithm) of these trusts, including the bid price (logarithm) of Knickerbocker as surrogates, with their corresponding ask prices acting as

Std Error	(K,T)	$\lambda_t \sim N(1,1)$					$\lambda_t \sim N(\log(t),1)$				
		SC	**SC-S**	**PI**	**PI-P**	**PI-S**	**SC**	**SC-S**	**PI**	**PI-P**	**PI-S**
Robust	$(1,200)$	0.057	0.066	0.060	0.073	0.048	0.684	0.690	0.078	0.115	0.051
	$(1,800)$	0.022	0.024	0.023	0.009	0.010	1.961	1.963	0.061	0.014	0.020
	$(5,200)$	0.150	0.158	0.215	2.617	0.201	6.886	6.749	0.290	1.744	0.231
	$(5,800)$	0.057	0.058	0.112	0.044	0.048	23.770	23.742	0.247	0.050	0.081
HAC	$(1,200)$	0.062	0.071	0.065	0.076	0.049	0.674	0.684	0.086	0.122	0.057
	$(1,800)$	0.024	0.026	0.024	0.010	0.010	2.004	2.004	0.072	0.015	0.022
	$(5,200)$	0.162	0.168	0.258	6.369	0.214	6.931	6.807	0.360	7.122	0.262
	$(5,800)$	0.062	0.063	0.120	0.045	0.050	24.065	24.064	0.279	0.058	0.090

Table 1: Mean Squared Errors under various methods

Std Error	(K,T)	$\lambda_t \sim N(1,1)$					$\lambda_t \sim N(\log(t),1)$				
		SC	**SC-S**	**PI**	**PI-P**	**PI-S**	**SC**	**SC-S**	**PI**	**PI-P**	**PI-S**
Robust	$(1,200)$	94.00%	91.85%	95.05%	93.25%	94.25%	11.40%	12.75%	94.55%	94.40%	94.25%
	$(1,800)$	94.60%	94.00%	94.45%	94.75%	94.75%	0.00%	0.00%	94.30%	94.35%	95.00%
	$(5,200)$	94.35%	94.30%	96.70%	97.65%	97.05%	0.15%	0.15%	95.95%	97.60%	97.60%
	$(5,800)$	95.05%	94.65%	94.35%	95.35%	94.60%	0.00%	0.00%	94.30%	95.55%	94.80%
HAC	$(1,200)$	93.30%	91.05%	93.85%	92.90%	93.80%	14.70%	15.35%	94.10%	93.10%	93.25%
	$(1,800)$	94.05%	93.00%	94.65%	94.25%	95.30%	0.00%	0.00%	92.45%	94.95%	93.50%
	$(5,200)$	93.60%	92.80%	95.45%	97.95%	96.40%	0.45%	0.45%	93.75%	97.30%	96.20%
	$(5,800)$	94.85%	94.85%	94.40%	95.65%	95.40%	0.00%	0.00%	92.95%	94.45%	94.35%

Table 2: Coverage rate under various methods

proxies. Similarly, the mid-prices (logarithm) of the 48 independent trusts are used as donors.[1] The relevance of our IFEM model to this application is supported by the common assumption of factor models in studying the stock market within the asset pricing literature (see, for example, discussions of factors and loadings in Fama and French (1993)).

Figure 1 displays the trends of the logarithmic stock price of Knickerbocker Trust Company alongside the synthetic control estimates derived from the five methods presented in the Section 4.1. The plot reveals that the classic synthetic control regression generates a counterfactual series closely aligned with the observed outcome in the post-intervention period, resulting in a relatively small causal effect. In contrast, the proximal inference methods (**PI** and **PI-S**) yield series that visually extend the pre-intervention price trend more consistently, suggesting that these approaches might be better suited to this context. Notably, the post-intervention predictions, particularly

with **PI** and **PI-S**, exhibit higher volatility compared to the pre-intervention series, reflecting the inherent uncertainty associated with the estimates. Meanwhile, **SC-S** remains below the pre-intervention average price but also appears to perform relatively well.

Table 4.2 displays the estimated average treatment effect on the treated (ATT) obtained using the five methods, along with their corresponding relative efficiency and standard errors enclosed in parentheses. In the **PI-S** column, the standard errors of the estimator are included, while in the other columns, we show the relative efficiency calculated as the ratio of variances of **PI-S** against the other methods. To estimate standard errors and relative efficiency, we use heteroskedasticity- and autocorrelation-consistent (HAC) variance estimators. Each row section corresponds to different lengths of post-intervention periods used for estimation, with the total length of the post-treatment period available in the data being 182 (triweekly frequency). The results suggest that the **SC** method performs poorly, as its point estimates differ considerably from those of the other methods. On the other hand, the **PI-S** method demonstrates smaller standard errors than **PI** when em-

---

[1]The original analysis consists of 49 independent trusts, one of which has a missing date and thus was dropped in our analysis. We adopt the contaminated surrogate approach from Section A.3, as it is arguably more suitable for this dataset.
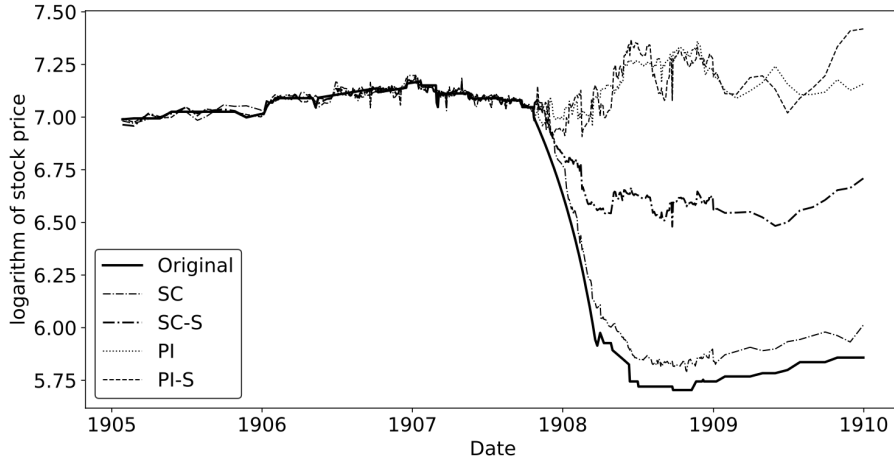
Figure 1: Stock price of Knickerbocker and its synthetic controls using various methods

Table 3: Estimated ATT under different lengths of post-intervention period

#time steps	SC†	SC-S†	PI†	PI-S*	PI-P†
80	-0.020	-0.370	-0.600	-0.593	-0.739
	(120.843)	(12.750)	(0.999)	(**0.154**)	(0.022)
100	-0.013	-0.495	-0.771	-0.769	-0.361
	(182.137)	(9.725)	(0.972)	(**0.161**)	(0.012)
120	-0.061	-0.566	-0.909	-0.920	-0.590
	(113.319)	(10.030)	(0.998)	(**0.165**)	(0.245)
160	-0.069	-0.626	-1.086	-1.086	-0.531
	(107.123)	(10.882)	(0.959)	(**0.166**)	(0.082)
182	-0.089	-0.643	-1.138	-1.134	-1.220
	(56.464)	(11.036)	(0.962)	(**0.158**)	(0.092)

* Numbers in parentheses represent the estimated standard errors.
† Numbers in parentheses represent the relative efficiencies, calculated as the ratio of variances of PI-S against the other methods.

ploying different post-intervention periods for training. This finding highlights the benefits of augmenting pre-intervention estimation with post-intervention data. Nonetheless, the difference between **PI** and **PI-S** is not substantial. Interestingly, despite relying solely on post-intervention data, **PI-P** is capable of producing reasonably accurate estimates. This observation underscores the potential advantages of using proximal inference methods, such as **PI**, **PI-P** and **PI-S**, in estimating the impact of financial crises on stock prices.

5 CONCLUSION

Our paper introduces a novel framework that enhances the synthetic control method (SCM) for estimating the long-term impact of interventions. By integrating surrogates and leveraging the proximal causal inference framework, our approach effectively incorporates post-intervention information, improving the es-

timation of causal effects. Importantly, our method demonstrates the feasibility of estimating causal effects using only post-intervention period data, which is particularly valuable in situations with limited pre-intervention data. Moving forward, several avenues for future research arise from our work. One direction involves the development of prediction intervals for our proposed estimator, as seen in the work of Chernozhukov et al. (2021). This would enable the quantification of uncertainty and provide a more comprehensive assessment of the estimated treatment effects. Additionally, exploring weighting approaches to achieve double robustness properties, as discussed in Qiu et al. (2023), could further enhance the robustness of our estimates. Furthermore, in Section A.3, we present a solution for surrogates contaminated by the outcomes of the control group. Extending this approach to address interference problems in the synthetic control setup holds promise for future research. Overall, our framework opens up new possibilities for

refining and expanding the synthetic control method, offering valuable insights for researchers in the field of impact evaluation.

Acknowledgements

References

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

Abadie, A., Diamond, A., and Hainmueller, J. (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.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132.

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

Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *Journal of Machine Learning Research*, 19(22):1–51.

Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803.

Chernozhukov, V., Wüthrich, K., and Zhu, Y. (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 116(536):1849–1864.

Cui, Y., Pu, H., Shi, X., Miao, W., and Tchetgen, E. T. (2023). Semiparametric proximal causal inference. *Journal of the American Statistical Association*, 0(0):1–12.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Working Paper 22791, National Bureau of Economic Research.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Ferman, B. (2021). On the properties of the synthetic control estimator with many periods and many controls. *Journal of the American Statistical Association*, 116(536):1764–1772.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221.

Fohlin, C. and Lu, Z. (2021). How contagious was the panic of 1907? new evidence from trust company stocks. *AEA Papers and Proceedings*, 111:514–19.

Hall, A. R. (2005a). *Generalized method of moments*. Oxford University Press.

Hall, P. (2005b). *Generalized Method of Moments*. Oxford University Press, Oxford, UK.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Kellogg, M., Mogstad, M., Pouliot, G. A., and Torgovitsky, A. (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association*, 116(536):1804–1816. PMID: 35706442.

Miao, W., Geng, Z., and Tchetgen Tchetgen, E. J. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993.

Miao, W., Shi, X., and Tchetgen, E. T. (2020). A confounding bridge approach for double negative control inference on causal effects.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Powell, D. (2018). *Imperfect Synthetic Controls: Did the Massachusetts Health Care Reform Save Lives?* RAND Corporation, Santa Monica, CA.

Qiu, H., Shi, X., Miao, W., Dobriban, E., and Tchetgen, E. T. (2023). Doubly robust proximal synthetic controls.

Shi, X., Li, K., Miao, W., Hu, M., and Tchetgen, E. T. (2021). Theory for identification and inference with synthetic controls: A proximal causal inference framework.

Shi, X., Miao, W., Nelson, J. C., and Tchetgen, E. J. T. (2020). Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 82(2):521–540.

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

Checklist

1. For all models and algorithms presented, check if you include:

   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes. Please see in Section 2 and 3.]

   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Not Applicable.]

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable.]

2. For any theoretical claim, check if you include:

   (a) Statements of the full set of assumptions of all theoretical results. [Yes. Please see in Section 2 and B.2.]

   (b) Complete proofs of all theoretical results. [Yes. Please see in Section B.1 and B.2]

   (c) Clear explanations of any assumptions. [Yes. Please see in Section 2.]

3. For all figures and tables that present empirical results, check if you include:

   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes.]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes. Details are explained in Section 4.1 and 4.2]

   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes. We report MSEs and coverage rates.]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [No. Our method is not computationally expensive.]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Yes. The original real-world data application is cited in Section 4.2.]

   (b) The license information of the assets, if applicable. [Not Applicable.]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable.]

   (d) Information about consent from data providers/curators. [Not Applicable. Data is open sourced.]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable.]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable.]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable.]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable.]

The appendix is organized as follows: in Appendix A we provide three useful extensions to our primary framework. In Appendix B.1, we present the proof of Theorem 3.1-3.3.

A EXTENSIONS

In this section, we explore three extensions to our primary framework. First, we incorporate adjustments for measured covariates, which can be beneficial in practice to mitigate the issue of unmeasured confounders. Second, we aim to relax the linear structure in the IFEM model. We outline the necessary assumptions for nonparametric identification, which may be useful for practitioners when designing and implementing nonparametric methods in practice. Finally, we include a useful extension of the surrogacy model to consider "contaminated" surrogates. Such surrogates can potentially be alternative outcomes of the target unit or outcomes of other units affected by the intervention.

A.1 Adjustment for Measured Covariates

In practice, one may wish to incorporate available covariate data measured across units and over time, either to account for endogeneity or to improve efficiency. As such, we generalize the IFEM model as follows:

$$
\begin{aligned}
Y_t(1) &= \rho_t'\theta + \delta_t + \lambda_t'\beta + C_{Y,t}'\xi_Y + \epsilon_{Y,t} \\
Y_t(0) &= \lambda_t'\beta + C_{Y,t}'\xi_Y + \epsilon_{Y,t} \\
W_t' &= \lambda_t'\Gamma + \xi_W'C_{W,t} + \epsilon_{W,t}' \\
X_t' &= \rho_t'\Phi + \xi_X'C_{X,t} + \epsilon_{X,t}' \ ,
\end{aligned}
$$

where $C_{Y,t}, C_{W,t}, C_{X,t}$ are vectors/matrices of measured covariates that are not causally impacted by the intervention, i.e. $C_{Y,t}(1) = C_{Y,t}(0) = C_{Y,t}$. We modify Assumption 2.1 and 2.3 to include measured covariates as follows:

Assumption A.1 (Conditional independent errors) *The error terms are conditional mean independent and independent of each other.* $E\left(\epsilon_{Y,t} \mid \lambda_t, C_{Y,t}\right) = E\left(\epsilon_{W,t} \mid \lambda_t, C_{W,t}\right) = E\left(\delta_t \mid \rho_t\right) = E\left(\epsilon_{X,t} \mid \rho_t, C_{X,t}\right) = 0$ *for* $t = 1, \ldots, T$.

Assumption A.2 (Existence of proxies) *We observe* $\{Z_{0,t}, Z_{1,t}\}$ *such that*

$$
Z_{0,t} \perp\!\!\!\perp \{Y_i(0), W_t\} \mid \lambda_t, C_{Y,t}, C_{W,t} \text{ for any } t \le T_0 \ ,
$$
$$
\{Z_{0,t}, Z_{1,t}\} \perp\!\!\!\perp \{Y_i(1), W_t, X_t\} \mid \lambda_t, \rho_t, C_{Y,t}, C_{W,t}, C_{X,t} \text{ for any } t > T_0 \ .
$$

Let $\tilde{Y}_t = Y_t - C_{Y,t}'\xi_Y$, $\tilde{W}_t = W_t - C_{W,t}'\xi_W$ and $\tilde{X}_t = X_t - C_{X,t}'\xi_X$. Under the modified assumptions above, we have the following identification results:

Theorem A.1 (Identification) *Under Assumption A.1 and 2.2, we have* $E[Y_t(0)] = E[C_{Y,t}'\xi_Y + \tilde{W}_t'\alpha]$ *and* $E[Y(1)] = E[C_{Y,t}'\xi_Y + \tilde{W}_t'\alpha + \tilde{X}_t'\gamma]$ *for any* $1 \le t \le T$. *Under Assumption A.1-A.2, 2.2 and 2.4,* (α, γ) *is uniquely identified by solving*

$$
E[\tilde{Y}_t - \tilde{W}_t'\alpha \mid Z_{0,t}, C_{Y,t}, C_{W,t}, t \le T_0] = 0 \text{ for any } t \le T_0 \ ,
$$
$$
E[\tilde{Y}_t - \tilde{W}_t'\alpha - \tilde{X}_t'\gamma \mid Z_{0,t}, Z_{1,t}, C_{Y,t}, C_{W,t}, C_{X,t}, t > T_0] = 0 \text{ for any } t > T_0 \ .
$$

The proof is similar to that of Theorem 3.1 in Appendix B.1.

Next, we present an example of applying Theorem A.1 to treatment effect estimations. First, let $\xi = (\xi_Y, \xi_W, \xi_X)$, $\tilde{Z}_{0,t} = (Z_{0,t}, C_{Y,t}, C_{W,t})$. We have moment conditions $E[\tilde{U}_t(\alpha, \gamma, \tau, \xi)] = 0$ where

$$
\tilde{U}_t(\alpha, \gamma, \tau, \xi) = \begin{pmatrix} g_0(\tilde{Z}_{0,t})(Y_t - W_t'\alpha - C_{Y,t}'\xi_Y - C_{W,t}'\xi_W\alpha)I\{t \le T_0\} \\ g_1(\tilde{Z}_{0,t}, Z_{1,t}, C_{X,t})(Y_t - W_t'\alpha - X_t'\gamma - C_{Y,t}'\xi_Y - C_{W,t}'\xi_W\alpha - C_{X,t}'\xi_X\gamma)I\{t > T_0\} \\ (Y_t - \tau - W_t'\alpha - C_{Y,t}'\xi_Y - C_{W,t}'\xi_W\alpha)I\{t > T_0\} \\ (X_t'\gamma - C_{X,t}'\xi_X\gamma - \tau)I\{t > T_0\} \end{pmatrix} .
$$

Lastly, we can employ the standard GMM approach as outlined in Section 3 for estimation and inference.

A.2 Nonparametric Identification and Estimation

The given approach in section 3 is heavily dependent on the linearity assumptions in equations (3) and (4). In practice, this linear structure could potentially be violated. Thus, in this section, we extend our proposed framework to obtain nonparametric identification and estimation. To begin, we replace Assumption 2.3 with the following assumption:

Assumption A.3 (Existence of confounding bridge) *There exists a function $h(W_t)$ such that the outcome model for $Y_t(0)$ is equivalent to a model for $h(W_t)$:*

$$E[Y_t(0) \mid \lambda_t] = E[h(W_t) \mid \lambda_t], \quad \forall t \geq 1.$$

Meanwhile, there exists a function $g(X_t)$ such that the causal effect model for $Y_t(1) - Y_t(0)$ is equivalent to a model for $g(X_t)$:

$$E[Y_t(1) - Y_t(0) \mid \rho_t] = E[g(X_t) \mid \rho_t], \quad \forall t \geq 1.$$

From Assumption A.3, we can immediately deduce that nonparametric identification of ATT can be expressed as follows:

$$\tau = E[Y_t(1) - Y_t(0) \mid t > T_0] = E[g(X_t) \mid t > T_0].$$

Subsequently, the following theorem presents the nonparametric identification results for $h(\cdot), g(\cdot)$:

Theorem A.2 (Moment condition for $h(\cdot), g(\cdot)$) *Under Assumption 2.1, A.3 and 2.3, the confounding bridge function $h(\cdot), g(\cdot)$ satisfies the moment condition*

$$E[Y_t - h(W_t) \mid Z_{0,t}, t \leq T_0] = 0 \text{ for any } t \leq T_0 ,$$
$$E[Y_t - h(W_t) - g(X_t) \mid Z_{0,t}, Z_{1,t}, t > T_0] = 0 \text{ for any } t > T_0 .$$

The proof follows the same procedure as in Appendix B.1. By Assumption 2.3, we have $E[Y_t - h(W_t) \mid Z_{0,t}, \lambda_t, t \leq T_0] = E[Y_t - h(W_t) \mid \lambda_t, t \leq T_0] = 0$. By integrating over λ_t, we obtain $E[Y_t - h(W_t) \mid Z_{0,t}, t \leq T_0] = 0$. The same argument applies to the second moment condition, and thus the result follows.

A.3 Contaminated Surrogates

In this section, we extend our IFEM model to accommodate situations where "pure" surrogates are not available. Instead, we only observe surrogates that are "contaminated" by latent factors that drive the outcomes of the control. In many applications, such contaminated surrogates are essentially alternative outcomes of the target units and are expected to be driven by λ_t in (3) in the pre-intervention periods, and driven by both λ_t and ρ_t in (4) in the post-intervention periods. For example, if the target outcome of interest is the GDP growth of California, contaminated surrogates can potentially be inflation and the unemployment rate of California, or the GDP growth, inflation, and unemployment rate of other states affected by the treatment.

To formalize this setup, we consider the following latent factor model for the surrogates as a replacement of X_t from equation (4):

$$\begin{aligned} X_t'(0) &= \lambda_t' \Theta + \epsilon_{X,t}' \\ X_t'(1) - X_t'(0) &= \rho_t' \Phi + \delta_{X,t} , \end{aligned} \tag{7}$$

where Θ is a $(F \times H)$ matrix of factor loadings. In addition to Assumption A.2, we assume the following:

Assumption A.4 (Existence of synthetic control for contaminated surrogates) *There exist a $(N \times H)$ matrix of weights Ψ such that $\Gamma \Psi = \Theta$.*

Remark A.1 *Similar to Remark 2.1, such an assumption holds when the number of donors exceeds the number of latent factors λ_t, i.e. $N \geq F$.* ∎

Under the model of contaminated surrogates and the additional Assumption A.4, we have the following results for identification:

Theorem A.3 (Identification) *Under Assumption 2.1-2.2 and A.4, we have $E[Y_t(0)] = E[W_t'\alpha]$, $E[X_t(0)'] = E[W_t'\Psi]$ and $E[Y(1)] = E[W_t'\alpha + (X_t'(1) - W_t'\Psi)\gamma]$ for any $1 \leq t \leq T$. Under Assumption 2.1-2.4, (α, γ) is uniquely identified by solving*

$$E[Y_t - W_t'\alpha \mid Z_{0,t}, t \leq T_0] = 0 \text{ for any } t \leq T_0 ,$$
$$E[X_t - W_t'\Psi \mid Z_{0,t}, t \leq T_0] = 0 \text{ for any } t \leq T_0 ,$$
$$E[Y_t - W_t'\alpha - (X_t' - W_t'\Psi)\gamma \mid Z_{0,t}, Z_{1,t}, t > T_0] = 0 \text{ for any } t > T_0 .$$

The proof is similar to that of Theorem 3.1 in Appendix B.1. Next, we use GMM approach for estimation and inference about the average treatment effects on the treated (ATT). The moment equations defining the ATT can be given as $E[(Y_t - \tau - W_t'\alpha)I\{t > T_0\}] = 0$ and $E[((X_t' - W_t'\Psi)\gamma - \tau)I\{t > T_0\}] = 0$.

B PROOFS

B.1 Proof of Theorem 3.1

By Assumption 2.1 and 2.2, we have

$$E[Y_t(0)] = E[\lambda_t'\beta] = E[\lambda_t'\Gamma\alpha] = E[(W_t - \epsilon_{W,t})'\alpha] = E[W_t'\alpha]$$

and

$$E[Y_t(1)] = E[Y_t(0)] + E[\rho_t'\theta] = E[W_t'\alpha] + E[\rho_t'\Phi\gamma] = E[W_t'\alpha] + E[(X_t - \epsilon_{X,t})'\gamma] = E[W_t'\alpha + X_t'\gamma] .$$

Again, by Assumption 2.1 and 2.2, we have

$$E[Y_t(0) - W_t'\alpha \mid \lambda_t, t \leq T_0] = E[\epsilon_{Y,t} - \epsilon_{W,t}'\alpha \mid \lambda_t, t < T_0] = 0$$

and

$$E[Y_t(1) - W_t'\alpha - X_t'\gamma \mid \lambda_t, \rho_t, t > T_0] = E[Y_t(1) - Y_t(0) - X_t'\gamma \mid \lambda_t, \rho_t, t > T_0]$$
$$= E[\delta_t - \epsilon_{X,t}'\gamma \mid \lambda_t, \rho_t, t > T_0] = 0 .$$

By Assumption 2.3, we have

$$E[Y_t - W_t'\alpha \mid Z_{0,t}, t \leq T_0] = E[E[Y_t(0) - W_t'\alpha \mid \lambda_t, Z_{0,t}, t \leq T_0] \mid Z_{0,t}, t \leq T_0]$$
$$= E[E[Y_t(0) - W_t'\alpha \mid \lambda_t, t \leq T_0] \mid Z_{0,t}, t \leq T_0] = 0 ,$$

and

$$E[Y_t - W_t'\alpha - X_t'\gamma \mid Z_{0,t}, Z_{1,t}, t > T_0] = E[E[Y_t(1) - W_t'\alpha - X_t'\gamma \mid \lambda_t, \rho_t, Z_{1,t}, t > T_0] \mid Z_{0,t}, Z_{1,t}, t > T_0]$$
$$= E[E[Y_t(1) - W_t'\alpha - X_t'\gamma \mid \lambda_t, \rho_t, t > T_0] \mid Z_{0,t}, Z_{1,t}, t > T_0] = 0 .$$

To show uniqueness, assume there exists two sets of vectors of weights (α, γ), $(\tilde{\alpha}, \tilde{\gamma})$ such that $E[Y_t - W_t'\alpha - X_t'\gamma \mid Z_{0,t}, Z_{1,t}, t > T_0] = E[Y_t - W_t'\tilde{\alpha} - X_t'\tilde{\gamma} \mid Z_{0,t}, Z_{1,t}, t > T_0] = 0$. Thus, $E[W_t'(\alpha - \tilde{\alpha}) + X_t'(\gamma - \tilde{\gamma}) \mid Z_{0,t}, Z_{1,t}, t > T_0] = 0$, which by Assumption 2.4 implies $\alpha = \tilde{\alpha}$ and $\gamma = \tilde{\gamma}$. ∎

B.2 Details and Proof for Theorem 3.2 and 3.3

B.2.1 Details for the Theorem Statements

In Thereom 3.2, the formula for G is given as follows:

$$G = E\left[\begin{pmatrix} g_0(Z_{0,t})W_t'I\{t \leq T_0\} & 0 \\ g_1(Z_{0,t}, Z_{1,t})W_t'I\{t > T_0\} & g_1(Z_{0,t}, Z_{1,t})X_t'I\{t > T_0\} \end{pmatrix}\right] . \tag{8}$$

In Theorem 3.3, the formula for G is given as follows:

$$G = E\left[\begin{pmatrix} g_0(Z_{0,t})W_t'I\{t \leq T_0\} & 0 & 0 \\ g_1(Z_{0,t}, Z_{1,t})W_t'I\{t > T_0\} & g_1(Z_{0,t}, Z_{1,t})X_t'I\{t > T_0\} & 0 \\ W_t'I\{t > T_0\} & 0 & I\{t > T_0\} \\ 0 & -X_t'I\{t > T_0\} & I\{t > T_0\} \end{pmatrix}\right] . \tag{9}$$

B.2.2 Proof for Theorem 3.2 and 3.3

We first show that $E[U_t(\alpha, \gamma)] = E[\tilde{U}_t(\alpha, \gamma, \tau)] = 0$ at the true value under Assumptions 2.1-2.4 and 3.1. For brevity, we only show $E[U_t(\alpha, \gamma)] = 0$. We have

$$
\begin{aligned}
E[U_t(\alpha, \gamma)] &= E\left[\begin{pmatrix} g_0(Z_{0,t})(Y_t - W_t'\alpha)I\{t \le T_0\} \\ g_1(Z_{0,t}, Z_{1,t})(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \end{pmatrix}\right] \\
&= E\left[\begin{pmatrix} E\left[g_0(Z_{0,t})(Y_t - W_t'\alpha)I\{t \le T_0\} \mid Z_{0,t}, t \le T_0\right] \\ E\left[g_1(Z_{0,t}, Z_{1,t})(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \mid Z_{0,t}, Z_{1,t}, t > T_0\right] \end{pmatrix}\right] \\
&= E\left[\begin{pmatrix} g_0(Z_{0,t})E\left[(Y_t - W_t'\alpha)I\{t \le T_0\} \mid Z_{0,t}, t \le T_0\right] \\ g_1(Z_{0,t}, Z_{1,t})E\left[(Y_t - W_t'\alpha - X_t'\gamma)I\{t > T_0\} \mid Z_{0,t}, Z_{1,t}, t > T_0\right] \end{pmatrix}\right] \quad \text{(by Theorem 3.1)} \\
&= 0 .
\end{aligned}
$$

Now we focus on proving Theorem 3.2. Similar arguments apply to Theorem 3.3. We follow Chapter 3 of Hall (2005a) to prove the asymptotic normality of our estimator. To summarize, we have a population moment condition $E[U_t(\theta)] = 0$ with a moment function $U_t(\alpha, \gamma) = \mathcal{V}_t(Y_t - \mathcal{D}_t'\theta)$, $\theta = (\alpha, \gamma) \in \mathcal{R}^{N+H}$, $\mathcal{V}_t = \{g_0(Z_{0,t})'I\{t \le T_0\}, g_1(Z_{0,t}, Z_{1,t})'I\{t > T_0\}\}' \in \mathcal{R}^{d_1+d_2}$, $\mathcal{D}_t = \{W_t', X_t'I\{t > T_0\}\}' \in \mathcal{R}^{N+H}$. Let $\mathcal{O}_t = \{Y_t, W_t, X_t, T_0, Z_{0,t}, Z_{1,t}\} \in \mathcal{R}^{N+H+\dim(\{Z_{0,t}, Z_{1,t}\})+2}$, denote the observable vector of random variables, where $\dim(\{Z_{0,t}, Z_{1,t}\})$ is the dimension of $\{Z_{0,t}, Z_{1,t}\}$. Let $\Theta \subseteq \mathcal{R}^{N+H}$ denote the parameter space of θ, and let $O \subseteq \mathcal{R}^{N+H+\dim(\{Z_{0,t}, Z_{1,t}\})+2}$ denote the sample space of $\mathcal{O}_t$. Then $U_t = U_t(\mathcal{O}_t; \theta)$ is a mapping from $O \times \Theta$ to $\mathcal{R}^{d_1+d_2}$. We impose the following regularity conditions.

Assumption B.1 (Strict stationarity) *The observable vector of random variables $\mathcal{O}_t$ form a strictly stationary process, such that all expectations of functions of $\mathcal{O}_t$ do not depend on time.*

Assumption B.2 (Regularity conditions for U_t) *The function $U_t : O \times \Theta \to \mathcal{R}^{d_1+d_2}$ where $d_1 + d_2 < \infty$ satisfies: (i) it is continuous on Θ for each $\mathcal{O}_t \in O$; (ii) $E[U_t(\mathcal{O}_t; \theta)]$ exists and is finite for every $\theta \in \Theta$; (iii) $E[U_t(\mathcal{O}_t; \theta)]$ is continuous on Θ.*

Assumption B.3 (Regularity conditions on $\partial U_t(\mathcal{O}_t; \theta)/\partial \theta'$) *(i) The derivative matrix $\partial U_t(\mathcal{O}_t; \theta)/\partial \theta'$ exists and is continuous on Θ for each $\mathcal{O}_t \in O$; (ii) The true value of θ does not lie on the boundary of Θ; (iii) $E[\partial U_t(\mathcal{O}_t; \theta)/\partial \theta']$ exists and is finite.*

Assumption B.4 (Properties of the Weighting Matrix) *The user-specified weight matrix Ω is a positive semi-definite matrix, possibly depends on data, and converges in probability to the positive definite matrix of constants.*

Assumption B.5 (Ergodicity) *The random process $\{\mathcal{O}_t; -\infty < t < \infty\}$ is ergodic.*

Assumption B.6 (Compactness of Θ) *Θ is a compact set.*

Assumption B.7 (Domination of $U_t(\mathcal{O}_t; \theta)$) *$E[\sup_{\theta \in \Theta} \|U_t(\mathcal{O}_t; \theta)\|] < \infty$.*

Assumption B.8 (Properties of the variance of the sample moment) *Let $\theta^* = (\alpha^*, \gamma^*)$ denote the true value of θ. (i) $E[U_t(\mathcal{O}_t; \theta)U_t(\mathcal{O}_t; \theta)']$ exists and is finite; (ii) S exists and is a finite valued positive definite matrix.*

Assumption B.9 (Properties of $G_T(\theta) = T^{-1}\sum_{t=1}^{T} \partial U_t(\mathcal{O}_t; \theta)/\partial \theta'$) *(i) $E[\partial U_t(\mathcal{O}_t; \theta)/\partial \theta']$ is continuous on some neighbourhood N_ϵ of the true value θ^* in Θ; (ii) Uniform convergence of $G_T(\theta)$: $\sum_{\theta \in N_\epsilon} \|G_T(\theta) - E[\partial U_t(\mathcal{O}_t; \theta)/\partial \theta']\| \xrightarrow{p} 0$.*

Under Assumptions 2.1-2.4 and 3.1 and B.1-B.9, we have that Theorem 3.2 holds by Theorem 3.2 of Hall (2005a).