Econometric Theory, **0**, 2026, 1–33.
doi:10.1017/S026646662610067X

AVERAGE TREATMENT EFFECT LOCALIZATION: PROJECTION METHODS IN SYNTHETIC CONTROL

RUEI-CHI LEE
Rutgers University

Many real-world policies and business interventions require assessing short-term effects to inform timely decisions, even though most causal inference methods focus on long-term average treatment effects. In this article, we introduce average treatment effect localization (ATEL), which captures localized, short-term policy impacts in panel data settings with a single treated unit and provides early indicators of policy impact. To accommodate both time-varying and nonlinear effects of observed and unobserved covariates, we propose a nonparametric model for untreated outcome, interpreted as a time-varying factor model via sieve approximation. Estimating the time-varying factor model is challenging due to the boundary bias and identification. Our estimation method based on diversified projection can effectively address these issues. We develop an asymptotic distribution theory to facilitate inference for the ATEL estimator. In an empirical application, we apply our proposed methodology to assess the impact of right-to-carry laws on violent crime rate.

1. INTRODUCTION

Evaluating policy interventions is crucial, as they are commonly encountered across various fields, including economics, political science, medical science, and others. Empirical researchers often focus on the long-term effects of policy interventions (e.g., Stevens, 1997; Angrist et al., 2006; Heckman et al., 2013; Chetty et al., 2014). In contrast, short-term policy interventions have received limited attention, partly because short-term effects may lack statistical significance. However, in practice, policymakers and business leaders must assess program effectiveness based on short-term results to inform decision-making and justify continued investment.

In this article, we estimate the average treatment effect localization (ATEL) to capture the short-term causal effects localized around the time the treatment is first received. Formally, ATEL is defined as

$$\alpha = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \mathbb{E}(Y_{1t}^I - Y_{1t}^N) K_h \left(\frac{t - T_0}{T_1} \right),$$

Address correspondence to Ruei-Chi Lee, Rutgers University, USA, e-mail: rl824@rutgers.edu.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

where Y_{1t}^I and Y_{1t}^N denote the treated and untreated potential outcomes, respectively, for the treated unit at time t, and T_0 and T_1 are the pre-treatment and post-treatment periods. The kernel function $K_h(\cdot)$, with bandwidth parameter h, is used to focus on treatment locally near T_0. ATEL specifically captures the immediate policy impact occurring soon after the intervention, providing a more precise and practical approach to short-term causal inference.

Although its name resembles the well-known local average treatment effect (LATE), the two quantities are conceptually distinct. LATE is unit-local: it averages the causal effect for compliers, the sub-population whose treatment status changes in response to an instrumental variable (Imbens and Angrist, 1994). ATEL is time-local: it averages the causal effect for the treated unit within a narrow window of periods immediately after the intervention.

The relevance of ATEL is particularly evident in contexts, such as public policy and corporate strategy. For example, when policymakers implement stricter gun control measures, such as background checks or assault weapon bans, the full impact on crime rates and public safety may take years to materialize. However, legislators and law enforcement agencies often need short-term indicators—such as reductions in firearm-related homicides, gun sales, or illegal firearm trafficking—to evaluate the policy's effectiveness and justify its continuation. Similarly, in the pharmaceutical industry, when a company launches a new drug, regulatory agencies and investors expect early clinical trial results and real-world patient outcomes before long-term health benefits can be fully observed. Failure to demonstrate early efficacy may lead to funding withdrawals or stricter regulatory scrutiny, even if the long-term benefits are substantial.

This article studies the problem of making inference and estimation on ATEL in aggregate panel data with a single treated unit. To construct the counterfactual outcomes, we consider the following nonparametric time-varying model for untreated potential outcomes:

$$Y_{it}^N = h_t(\eta_{it}, X_{it}) + u_{it}, \tag{1}$$

where η_{it} and X_{it} represent unobserved and observed covariates, respectively, u_{it} is the idiosyncratic error, and $h_t(\cdot)$ is a time-varying nonlinear function. This framework allows both unobserved and observed covariates to influence outcomes in nonlinear and time-varying ways, accommodating rich form of heterogeneity than existing methods.

Many existing methods for estimating counterfactuals, such as difference-in-differences (DID), rely on strong assumptions like parallel trends, which are often violated due to unobserved time-varying confounders. The synthetic control method (SCM), introduced by Abadie and Gardeazabal (2003) and Abadie et al. (2010), addresses this by constructing weighted averages of control units to account for time-varying unobservables. Factor-based models (e.g., Gobillon and Magnac, 2016; Xu, 2017) improve on this by using interactive fixed effects (IFEs), but still impose linear structures on the influence of unobserved confounders. Some studies have introduced nonparametric alternatives (e.g., Ouyang and Peng,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

2015, Carvalho et al., 2018), but focus on the time-invariant nonlinear function of observed covariates. In contrast, our model provides a general framework that incorporates both observed and unobserved confounders with nonlinear and time-varying effects.

Estimating this model is challenging due to the presence of both unobserved and observed covariates in the time-varying function. We address this issue by assuming the model can be approximated as a low-rank factor structure, interpreted as a sieve approximation:

$$h_t(\eta_{it}, X_{it}) \approx \beta_{it}' F_t, \tag{2}$$

where β_{it} is a vector of sieve basis function $\phi(\eta_{it}, X_{it})$ and F_t is a vector of sieve coefficients.

Since we allow covariates to vary across both time and individuals in the sieve basis function, our model can be approximated as a factor model with time-varying factor loadings. Our work relates to a strand of literature that studies time-varying factor models. For instance, Su and Wang (2017) propose local PCA that combines PCA with the kernel method to estimate time-varying factor models, allowing factor loadings to change smoothly over time. Similarly, Pelger and Xiong (2022) use local PCA with covariate-dependent kernels to estimate time-varying factor models in which covariates vary only across time.

However, estimating time-varying factor models with local PCA has several drawbacks. First, like the kernel regression method, local PCA suffers from boundary bias. Extending it to a local linear framework is infeasible because PCA requires nonnegative kernel weights to maintain the positive semidefiniteness of the covariance matrix, whereas local linear estimators may introduce negative weights to correct boundary bias, particularly near the boundaries. Second, as Cheung (2024) points out, while local PCA estimators are only consistent up to a time- or covariate-dependent transformation matrix, H_t, they face identification challenges. Specifically, H_t may lack smoothness under some circumstances, resulting in estimation inconsistency. For instance, the order of factors may change over time because PCA sorts factors by eigenvalues, potentially causing abrupt shifts in H_t. Cheung (2024) provides an example showing that such jumps in H_t can occur over time.

To address the limitations of the local PCA method, we separately estimate the factors and factor loadings. First, to consistently estimate the factors, we adopt a non-PCA-based approach known as diversified projection (DP), proposed by Fan and Liao (2022) within the pure factor model framework. Unlike PCA, DP provides a more straightforward estimation process that weighted average of observations across exogenous covariates cross-sectionally, such that the transformation matrix, H_t, doesn't rely on eigenvalue/eigenvectors. Second, for estimating time-varying factor loadings, we employ the local linear method, which estimates both the factor loadings and their first-order Taylor expansion. This enables the local linear method to effectively mitigate the boundary bias issue inherent in local PCA.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

In summary, our contributions can be outlined as follows:

1. ATEL is introduced to capture short-term, localized causal effects in panel data with a single treated unit. To facilitate the inference, its asymptotic normality is derived. Unlike conventional methods that focus on long-term average treatment effects, ATEL provides decision-makers with timely indicators of policy intervention impacts.
2. We propose a general nonparametric model to study treatment effects, which captures the nonlinear and time-varying relationships between observed and unobserved covariates and outcomes. This approach relaxes the restrictive assumptions often imposed on covariates in panel data models for treatment effect estimation.
3. Our model can be interpreted as a time-varying factor model using sieve approximations. While conventional PCA-based methods face challenges in estimating time-varying factor models, including boundary bias and identification of the transformation matrix, our estimation method effectively addresses these issues.

There is a substantial body of literature studying treatment effects in nonparametric panel data that incorporates both observed and unobserved covariates in the unknown function, known as nonseparable models. Several papers address the time-invariant nonparametric function, including works by Graham and Powell (2012), Hoderlein and White (2012), and Chernozhukov et al. (2013). More recent studies have developed models where the unknown function varies over time, such as those by Freyberger (2017), Fernández-Val et al. (2021), and Botosaru et al. (2023). These studies estimate the partial effects of covariates in the nonparametric function to analyze treatment effects, often requiring identification assumptions, such as weak monotonicity or time-invariant structural functions. Our objective, however, is not to estimate the effect of covariates but to control for them. Therefore, our model does not require the identification conditions typically assumed in the nonseparable models.

Our article is also closely related to the literature on treatment effect estimation under linear factor models (see, e.g., Hsiao et al., 2012; Gobillon and Magnac, 2016; Xu, 2017; Li and Sonnier, 2023, among others). Recent developments in factor structure causal inference focus on general treatment assignment patterns rather than a single treated unit. Bai and Ng (2021) propose a two-step PCA estimation for block structure treatment assignments and establish asymptotic analysis for the treatment effects. Other works allow more general treatment/missing patterns by treating the assignment mechanism as random. Athey et al. (2021) consider a matrix completion estimator with nuclear norm regularization for causal panel data and derive convergence rates. Xiong and Pelger (2023) apply PCA to a reweighted covariance matrix to accommodate assignment patterns beyond block structures. Duan et al. (2024) further extend this framework to non-stationary data and allow the missing probability to depend on fixed effects and factor loadings. Su and Wang (2025) propose an EM algorithm with a nuclear-norm-regularized estimator as the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

initial value and provide an asymptotic theory for the estimators. Different from these methods, we propose to estimate ATEL, a new estimand that focuses on treatment effects local to the treatment period.

However, these PCA-based approaches rely on a factor structure with time-invariant factor loadings. Bai and Ng (2021) construct counterfactual outcomes through their "tall-wide" block estimators that depend on the time-invariant transformation matrix. Similarly, the reweighted covariance matrix in Xiong and Pelger (2023) and Duan et al. (2024) restricts factor loadings to be time-invariant. Ignoring potential time variation generally leads to inconsistent estimation of the treated unit's factor loading. It is also possible to handle time-varying loadings in their settings by adopting the local PCA as in Su and Wang (2017). On the other hand, because the DP approach constructs the factor estimator through the cross-sectional average of observations with exogenous covariates, it can be adapted straightforwardly to random or block structure treatment assignment. Nevertheless, as discussed by Athey et al. (2021), treatment may be given to a single unit (synthetic control) or staggered by experimental design. In such cases, the assumption of random treatment assignment may be inappropriate. Since our analysis focuses on a single treated unit, we leave extensions of the DP approach to random treatment assignment for future research.

The rest of the article is organized as follows. Section 2 describes the model and introduces the estimation method for the ATEL estimator. Section 3 provides the asymptotic analysis results of the ATEL estimator. In Section 4, we report the simulation results. Section 5 presents an empirical application that estimates the ATEL of right-to-carry (RTC) laws on violent crime rate. Section 6 concludes. All proofs and technical lemmas are provided in the Supplementary Material.

We use the following notation. We denote the maximum and minimum eigenvalues of a square matrix A as $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$. For any matrix A, we use $\|A\|$ and $\|A\|_F$ to denote the operator norm and the Frobenius norm, respectively. $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. For two sequences a_{NT} and b_{NT}, we denote $a_{NT} \ll b_{NT}$ (or $b_{NT} \gg a_{NT}$) if $a_{NT} = o(b_{NT})$, $a_{NT} \lesssim b_{NT}$ (or $b_{NT} \gtrsim a_{NT}$) if $a_{NT} = O(b_{NT})$, and $a_{NT} \asymp b_{NT}$ if $a_{NT} = O(b_{NT})$ and $b_{NT} = O(a_{NT})$.

2. SETUP AND ESTIMATION

This section provides the main setup for the ATEL and estimation methods for our model.

2.1. The Average Treatment Effect Localization

We consider a panel data setting with $N+1$ units observed over T periods. The treated unit (the first unit) is untreated for T_0 period and remains treated thereafter. The N control units remain untreated for all periods. Let Y_{it}^I and Y_{it}^N be the potential outcomes of unit i in period t with and without treatment, respectively. The individual treatment effect for the ith unit at time t is defined as

$$\alpha_{it} = Y_{it}^I - Y_{it}^N.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

The fundamental challenge for causal inference is that we cannot observe both outcomes Y_{it}^I and Y_{it}^N simultaneously. Thus, the observed outcomes Y_{it} are related to potential outcomes as $Y_{it} = D_{it}Y_{it}^I + (1-D_{it})Y_{it}^N$, where $D_{it} = 1$ if $i = 1$ and $t > T_0$ and $D_{it} = 0$ otherwise.

In the absence of treatment, outcomes are generated by the following nonparametric model:

$$Y_{it}^N = h_t(\eta_{it}, X_{it}) + u_{it}, \text{ for } i = 1, \ldots, N+1, \ t = 1, \ldots, T, \tag{3}$$

where the covariates (η_{it}, X_{it}) are only partially observed. Specifically, η_{it} is a vector of unobserved variables, and X_{it} is a vector of observed covariates. This model captures the nonlinearity and time-varying effects from partially observed covariates (η_{it}, X_{it}) through the time-varying unknown function $h_t(\cdot, \cdot)$. The term u_{it} represents the idiosyncratic error, which is independent of (η_{it}, X_{it}).

After the treated unit receives a policy intervention or treatment, the treated unit outcome for the post-treatment periods is

$$Y_{1t} = Y_{1t}^I = h_t(\eta_{1t}, X_{1t}) + \alpha_{1t} + u_{1t}, \text{ for } t = T_0 + 1, \ldots, T. \tag{4}$$

Given that the treated unit receives the treatment after T_0, the conventional causal analysis often focuses on the long-term average treatment effect on the treated. It is formally defined as

$$\alpha_{\text{long-term}} = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \mathbb{E}(Y_{1t}^I - Y_{1t}^N),$$

where $T_1 = T - T_0$ is the number of post-treatment periods.

In contrast to the long-term average treatment effect, this article focuses on the policy intervention that is localized at the time of treatment T_0, to facilitate the analysis of short-term policy impacts. The parameter of interest is the ATEL, defined as

$$\alpha = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \mathbb{E}(Y_{1t}^I - Y_{1t}^N) K_h\left(\frac{t - T_0}{T_1}\right), \tag{5}$$

where $K_h(\cdot)$ is a kernel function with bandwidth parameter h. The kernel function assigns weights to post-treatment periods, localizing the treatment effect around T_0, so α represents the short-term effect soon after T_0. In particular, if $K_h(\cdot)$ follows a uniform kernel, α simplifies to the average treatment effect over the $\lfloor T_1 h \rfloor$ post-treatment periods. Additional details about the kernel function are provided in (19).

To motivate why short-term policy evaluation is important, our empirical example examines how RTC laws affect violent crime shortly after they are implemented. These laws usually allow people to carry concealed firearms in public with a permit. Supporters believe that RTC laws can help citizens protect themselves and deter crime, whereas critics argue that more guns in public increase the risk of violence and accidental shootings. Although many empirical studies have investigated the long-term average treatment effects of RTC laws, the findings

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

remain inconclusive (e.g., Lott Jr. and Mustard, 1997; Aneja et al., 2011; Donohue et al., 2019). Given that the United States has one of the highest rates of firearm violence among developed countries, it is crucial to evaluate RTC laws promptly.

The short-term policy impact especially matters to policymakers, such as governors, legislators, and mayors, who often serve short terms of 2–4 years. If an RTC law is linked to a sudden rise in violent crime soon after it goes into effect, it may cause a political backlash, no matter what the long-term data eventually show. Public perception also plays a large role in decisions, sometimes more so than academic findings. A short-term spike in gun violence could lead to the repeal or tightening of RTC laws. Therefore, policymakers need to pay close attention to the policy's short-term impact when deciding whether it is effective.

In practice, α represents the short-term policy effect right after T_0, which is useful for policymakers and business leaders who must make prompt decisions. In contrast, researchers often focus on the long-term average treatment effect on the treated. While $\alpha_{\text{long-term}}$ considers the treatment effect across all post-treatment periods, it can be diminished over time as treated units adapt or respond strategically, making it harder to measure the policy's direct impact. For example, in February 2025, the Trump administration reduced the USAID budget, forcing multiple labs to shut down and causing an immediate drop in academic output. Over time, academic institutions responded with various mitigation strategies, such as fundraising and layoffs, making it difficult to capture the policy's original impact.

As reported by *The New York Times*, these cuts at Johns Hopkins University, funded largely by USAID grants for public health and agricultural research, led to the elimination of research projects on tuberculosis, AIDS, and cervical cancer. In response, the university laid off more than 2,000 employees to cope with the reduced budget (Saul, 2025). For researchers examining how USAID budget reductions influence academic output, these institutional measures and other external adjustments can distort the long-term average treatment effect, obscuring the policy's direct influence in a long run. By contrast, our main objective, α, measures the policy intervention shortly after T_0, which remains largely unaffected by these subsequent actions and countermeasures.

2.2. Low-Rank Time-Varying Factor Structure: Approximation and Estimation

In this section, we discuss how the nonparametric model (3) can be approximated by a low-rank time-varying factor model with a slowly growing rank, and we outline its estimation methods.

We motivate the sieve approximation for the nonlinear time-varying function $h_t(\eta_{it}, X_{it})$ with two considerations. First, standard nonparametric regression is infeasible due to the presence of unobserved covariates η_{it}. Unobserved heterogeneity is typically handled via linear transformations (e.g., first-differencing) in linear panel data models to remove unit- or time-invariant components. However,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

because η_{it} varies across both units and time and $h_t(\cdot)$ imposes a nonlinear structure, there is no general rule of transformation to remove η_{it}. Second, the sieve expansion with truncation allows us to approximately represent the matrix formed by the elements $h_t(\eta_{it}, X_{it})$ using a low-rank matrix. This low-rank structure is fundamental to our procedure, as it enables us to recover the relevant low-dimensional components while avoiding fully nonparametric estimation of $h_t(\cdot)$.

Formally, let Θ be the $(N+1) \times T$ matrix where the (i,t) element of Θ is $h_t(\eta_{it}, X_{it})$. Suppose the unknown function $h_t(\eta_{it}, X_{it})$ has the following sieve approximation:

$$h_t(\eta_{it}, X_{it}) = \sum_{j=1}^{J} \phi_j(\eta_{it}, X_{it}) f_{jt} + r_{it} = \beta_{it}' F_t + r_{it}, \tag{6}$$

where $\beta_{it} = (\phi_1(\eta_{it}, X_{it}), \ldots, \phi_J(\eta_{it}, X_{it}))' \in \mathbb{R}^J$ is a set of sieve transformation of η_{it}, X_{it} using ϕ_j as the basis function (e.g., B-spline, Fourier series, wavelets, and polynomial series), $F_t = (f_{1t}, \ldots, f_{Jt})'$ is the vector of sieve coefficients, and r_{it} is the sieve approximation error. J denotes the number of sieve terms which grow slowly as $N, T \to \infty$. Let Θ_0 be the $(N+1) \times T$ matrix with elements $\beta_{it}' F_t$, and R be the $(N+1) \times T$ matrix of r_{it}. In matrix notation, we represent (6) as

$$\Theta = \Theta_0 + R.$$

For sufficiently smooth functions $\{h_t(\cdot)\}$, the sieve approximation error $r_{it} \to 0$ as $J \to \infty$, so the matrix Θ can be well approximated by the low-rank matrix Θ_0. For example, suppose h_t belongs to a Hölder class with parameter $b, C, d > 0$:

$$\{h : \|D^b h(x) - D^b h(y)\| \leq C\|x - y\|^d, \text{for all } x, y\}.$$

Then, for common bases such as B-splines, trigonometric polynomials, or polynomial series, the approximation error satisfies

$$\max_{it} |r_{it}| < CJ^{-a}, \quad a = (b+d)/(dim(\eta_{it}) + dim(X_{it})),$$

which can be made arbitrarily small for sufficiently smooth functions even if J grow slowly. The smoothness condition justifies the low-rank approximation of $h_t(\eta_{it}, X_{it})$ because it guarantees that the approximation error vanishes as J increases, allowing the truncation of the sieve expansion to the first J terms. In many empirical applications, the function $h_t(\cdot)$ represents the average outcome conditional on the characteristics η_{it} and X_{it}. It is plausible to assume that such function is sufficiently smooth so that it belongs to the Hölder class. Then it is well-known that such functions can be well-approximated by sieve functions with a small number of sieve bases, yielding a low-rank structure.

Combining (3) and (6), the untreated potential outcomes follow an approximate time-varying factor structure:

$$Y_{it}^N = \beta_{it}' F_t + r_{it} + u_{it}. \tag{7}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Several studies have considered low-rank factor structure to represent nonparametric function using sieve approximation (see, e.g., Fernández-Val et al., 2021; Chernozhukov et al., 2023; Choi et al., 2024). However, instead of a time-varying factor structure, these analyses approximate functions as pure factor structure, which restrict the unobserved covariates either be unit-invariant or time-invariant, while reasonable in certain applications, may be unrealistic in many practical settings.

For simplicity, we assume that the rank of the low-rank factor structure, J, is known. Since the time-varying function is approximated through a low-rank factor structure via a sieve representation, the rank equals the sieve dimension. When the rank is unknown in an empirical study, consistent estimators of rank J can be obtained using a conditioning argument (see, e.g., Bai and Ng, 2002; Ahn and Horenstein, 2013; Su and Wang, 2017).

Building on the factor representation, we separately discuss the estimation of factors and factor loadings. To estimate factors, we adopt the DP approach proposed by Fan and Liao (2022), which reduces dimensionality by projecting data onto low-dimensional subspaces using a diversified weight matrix W. Unlike conventional PCA, which requires a strong signal-to-noise ratio to identify factor estimators, the DP method estimates factors through weighted averages of observations rather than relying on eigenvectors. This projection is conducted across individuals, making the estimator robust to serial conditions. The factor estimator is given by

$$\hat{F}_t = \frac{1}{N}\sum_{i=2}^{N+1} W_{it} Y_{it}^N, \text{ for } t = 1, \ldots, T, \tag{8}$$

where $W_{it} = (w_{it1}, \ldots, w_{itJ})' \in \mathbb{R}^J$ that should satisfy:

(i) W_{it} is independent of u_{it}.
(ii) W_{it} is correlated with β_{it}.

Conditions (i) and (ii) are diversified weight's analogous to the exclusion restriction and relevance restriction in instrument variables. They are necessary to consistently estimate F_t up to affine transformation. Substituting (3) and (6) into $\hat{F}_t$ gives

$$\begin{aligned}\hat{F}_t &= \left(\frac{1}{N}\sum_{i=2}^{N+1} W_{it}\beta_{it}'\right) F_t + \frac{1}{N}\sum_{i=2}^{N+1} W_{it} r_{it} + \frac{1}{N}\sum_{i=2}^{N+1} W_{it} u_{it} \\ &= H_t F_t + e_t,\end{aligned} \tag{9}$$

where H_t is the $J \times J$ time-varying transformation matrix, and e_t is the estimation error that includes the sieve approximation error r_{it} and the idiosyncratic error u_{it}. Under the mild conditions of W, $e_t \xrightarrow{p} 0$ as $N \to \infty$.

A natural question is how to acquire the diversified weight matrix that satisfies these two conditions. A compelling choice is through observed characteristics X_{it},

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

which have been extensively explored in factor model. For instance, Connor et al. (2012) introduced a characteristic-based factor model to analyze stock returns. This model incorporates factor loadings influenced by firm-specific attributes, such as market capitalization and book-to-price ratio. Fan et al. (2016) propose a more general form in the factor loading that can be explained by both observed covariates and unobserved components.

We can then define the diversified weight W_{it} as the transformation of the observed characteristics X_{it} with a set of functions $\varphi_j(\cdot)$:

$$w_{itj} = \varphi_j(X_{it}), \text{ for } i \leq N+1,\ t \leq T,\ j \leq J,$$

where $\varphi_j(\cdot)$ is a sieve basis function, such as B-spline, trigonometric series, and polynomial series. Given that the factor loading is expressed as the sieve transformation of covariates (i.e., $\beta_{it} = \phi(\eta_{it}, X_{it})$) and u_{it} is independent of X_{it}, both conditions required for W_{it} are satisfied. However, certain choices of W_{it} may violate these conditions. For instance, Fan and Liao (2022) suggest constructing diversified weights from the initial outcome Y_{i1}, which has the advantage of not requiring additional observed characteristics. Since Y_{i1} contains the information of u_{i1}, if u_{it} is serially correlated, the resulting weights can be correlated with u_{it}, and the exogeneity condition fails.

The DP approach used here differs from that in Fan and Liao (2022) in several ways. First, their DP method is robust to over-estimating the number of factors, it does not require exact knowledge of this number for consistent factor estimation, as long as the rank of the diversified weight, W_{it}, is at least as large as the true number of factors. Intuitively, since DP constructs factors as cross-sectional weighted averages rather than eigenvectors, over-estimating the number of factors typically adds only redundant projections onto the same factor space. This differs from PCA, where additional eigenvectors correspond to non-spiked eigenvalues and are often driven by idiosyncratic noise. In contrast, our framework assumes that the number of factors, J, is known, allowing us to construct a diversified weight matrix with rank equal to J. Consequently, our DP method can be viewed as an exactly identified IV approach. It is interesting to extend this framework to overidentified IV settings. We examine robustness to over-estimating the number of factors in the simulation and empirical sections and leave the associated theoretical development for future work.

Second, Fan and Liao's (2022) diversified weights are restricted to being time-invariant, with the recommended choice for W_i based on individual-specific characteristics. Similar restrictions can also be found in some semi-parametric factor models (e.g., Connor et al., 2012; Fan et al., 2016). However, such restrictions may be unrealistic in practice. As discussed in Connor et al. (2012), financial applications often require certain asset characteristics to vary over time. Their work demonstrates the time-series evolution of cross-sectional correlations among asset characteristics, indicating the need for time variation in asset attributes. Similarly, in macro-level data, characteristics often change over time, as they are aggregated from individual or household-level data. For instance, characteristics

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

TABLE 1. Simulated coverage probability of 95% confidence interval for the standardized ATEL estimates and MSEs for different sieve basis functions $\varphi_j(\cdot)$ (B-splines, trigonometric series, and polynomial series) under DGP1–DGP3.

	DGP1		DGP2		DGP3	
	Coverage	MSE	Coverage	MSE	Coverage	MSE
B-spline	0.9350	0.2023	0.9323	0.0675	0.9400	0.0653
Trigonometric	0.9330	0.2269	0.9300	0.0709	0.9353	0.0691
Polynomial	0.9327	0.2055	0.9373	0.0681	0.9407	0.0655

Note: The sample size is $N = 100$, $T_0 = 50$, $T_1 = 50$, and the number of factors $J = 2$. Based on simulations with 3,000 repetitions.

such as gender, race/ethnicity, or educational attainment are time-invariant at the individual level but vary over time after aggregation. Thus, the DP approach we consider is more adaptable to practical applications.

For practical implementation of W_{it}, since the number of factors, J, is equal to the dimension of the diversified weight, we recommend a data-driven rank estimation procedure for choosing J. Specifically, we adopt the information criteria suggested by Su and Wang (2017), which are tailored for time-varying factor structures. For the diversified weight functions $\varphi_j(\cdot)$, we recommend using B-spline bases.[1] Although standard sieve bases such as polynomial series, trigonometric series, and B-splines can all generate smooth functions, B-splines offer better numerical stability and more suitable boundary behavior. High-order polynomial series tend to induce severe multicollinearity as the sieve dimension J increases. Trigonometric bases impose a periodic boundary condition on the covariate support, which is typically unnatural for economic covariates. In contrast, B-spline basis functions only affect a local region of the covariate space, which provides stable boundary behavior and well-conditioned weight matrices. For an excellent review of the sieve basis function, we refer the readers to Chen (2007).

Table 1 examines the sensitivity of the ATEL estimator to the choice of sieve basis functions, $\varphi_j(\cdot)$, used in the construction of W_{it}. We evaluate this across different data-generating processes (DGPs) with the number of factors $J = 2$. Specifically, DGP1 specifies $h_t(\cdot)$ as a linear function, DGP2 uses a polynomial-series specification, and DGP3 uses a trigonometric-series specification. More details on these DGPs are provided in Section 4. Table 1 reports the simulation coverage probabilities of the 95% confidence intervals for the standardized ATEL estimates and the mean squared errors (MSEs). Overall, the simulation results are satisfactory. Across all DGPs, empirical coverage is close to the nominal level, and MSEs remain uniformly small. These findings suggest that our estimator

[1] Specifically, for $J = 2$ and $J = 3$, we use linear and quadratic splines (orders $k = 2$ and $k = 3$) without knots. For $J \geq 4$, we use cubic B-splines (order $k = 4$) with $J - k$ knots, so that the spline space has dimension J.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

is insensitive to the specific choice of sieve basis functions. Additionally, we present further simulations in Section 4 across different number of factors J for the construction of W_{it}. Those results indicate that our ATEL estimates remain robust to the choice of J as well.

After estimating the factors, the next step is to estimate the time-varying factor loadings. The conventional PCA may fail under time-varying factor loadings. To see this, let $\tilde{F}_t$ be the PCA estimator of F_t. Under regular conditions, the PCA estimator for the factor loading is

$$\tilde{\beta}_i = \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t Y_{it} = \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t F_t'\beta_{it} + o_P(1).$$

This representation shows that $\tilde{\beta}_i$ is a weighted average (over time) of β_{it}. As a result, the PCA estimator captures only a global average factor loading rather than the spontaneous, time-varying parameter β_{it}, and thereby failing to capture the variation in factor loadings over time.

To estimate time-varying factor loadings, we use local linear regression, a widely used nonparametric method that mitigates boundary bias (see, e.g., Fan, 1996; Li and Racine, 2007). By assuming factor loading is a smooth function of t/T, we have

$$\beta_{it} = \beta_i\left(\frac{t}{T}\right).$$

Since $\beta_i : [0,1] \to \mathbb{R}$ is a smooth function of t/T, we can approximate $\beta_i(\frac{t}{T})$ by $\beta_i(\frac{r}{T})$ if r/T is close to t/T enough for $r \in \{1,\ldots,T\}$. Specifically, through Taylor expansion, we have

$$\beta_i(\frac{t}{T}) = \beta_i\left(\frac{r}{T}\right) + \beta_i^{(1)}\left(\frac{r}{T}\right)\left(\frac{t-r}{T}\right) + d_i(t,r),$$

where $\beta_i^{(1)}(\cdot)$ is the first-order derivative of $\beta_i(\cdot)$ and $d_i(t,r)$ is the Taylor approximation error.

Based on this local linear approximation, the estimator for the time-varying factor loading is obtained by minimizing the locally weighted sum of squared residuals:

$$\min_{\beta_{1r},\beta_{1r}^{(1)}} \frac{1}{T_0}\sum_{t=1}^{T_0}\left[Y_{1t}^N - \beta_{1r}'\hat{F}_t - \beta_{1r}^{(1)'}\hat{F}_t\left(\frac{t-r}{T_0}\right)\right]^2 K_h\left(\frac{t-r}{T_0}\right), \text{ for } r = 1,\ldots,T_0, \tag{10}$$

where $K_h(\cdot)$ is a kernel function and $h = h_{T_0}$ is the bandwidth parameter. To handle boundary issues and ensure uniform results, we adopt the kernel function as defined in Su and Wang (2017).

Let $q_{tr} = \left(\hat{F}_t', (\frac{t-r}{T_0})\hat{F}_t'\right)'$ be a $2J \times 1$ vector. Stacking these over T_0 periods, we define $Q_r = (q_{1r},\ldots,q_{T_0 r})'$ as a $T_0 \times 2J$ matrix. Additionally, let

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

$M_r = diag\left(K_h(\frac{1-r}{T_0}), \ldots, K_h(\frac{T_0-r}{T_0})\right)$ be a $T_0 \times T_0$ matrix, and $Y_1 = (Y_{11}^N, \ldots, Y_{1T_0}^N)'$ be a $T_0 \times 1$ vector. Then the minimizer of (10) is given by

$$\begin{pmatrix} \hat{\beta}_{1r} \\ \hat{\beta}_{1r}^{(1)} \end{pmatrix} = (Q_r' M_r Q_r)^{-1} Q_r' M_r Y_1. \tag{11}$$

The local linear estimator at the boundary point where $r = T_0$ is used for post-treatment periods. To better illustrate the idea, the Taylor expansion of β_{1t} is

$$\beta_1\left(\frac{t}{T_1}\right) = \beta_1\left(\frac{T_0}{T_1}\right) + \beta_1^{(1)}\left(\frac{T_0}{T_1}\right)\left(\frac{t-T_0}{T_1}\right) + d_1(t, T_0).$$

This motivates our estimator of time-varying factor loading in the post-treatment periods is

$$\hat{\beta}_{1t} = \hat{\beta}_{1T_0} + \hat{\beta}_{1T_0}^{(1)}\left(\frac{t-T_0}{T_1}\right), \text{ for } t = T_0+1, \ldots, T. \tag{12}$$

2.3. Comparison with Synthetic Control Method

The SCM, proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010), is a widely used approach for estimating the treatment effect. Given the structural similarities in how SCM and our DP approach construct estimators, it is important to compare these methods. SCM also conducts a cross-sectional weighted average of untreated outcomes to construct the estimator:

$$\hat{Y}_{1t}^{SC} = W^{SC'} Y_t,$$

where $W^{SC} = (W_2^{SC}, \ldots, W_{N+1}^{SC})$ and $Y_t = (Y_{2t}^N, \ldots, Y_{N+1,t}^N)$ are $N \times 1$ vectors. The weights are selected by best fitting the outcome of the treated unit using pre-treatment data, and the weights are nonnegative and sum to one. Specifically, we focus on the case where all pre-treatment outcomes are included as predictors:

$$\min_{W \in \Delta} \sum_{t=1}^{T_0} \left(Y_{1t}^N - W' Y_t^N\right)^2, \tag{13}$$

where $\Delta = \{W \in \mathbb{R}^N : W_i \geq 0 \ \forall i, \sum_{i=2}^{N+1} W_i = 1\}$.

There are two fundamental differences between our DP approach and SCM. First, SCM uses only one weighted vector while we use the J weighting vectors. This difference arises because we use DP to estimate multiple factors, whereas SCM uses a weight to directly estimate the counterfactual outcomes. Second, SCM is primarily motivated by a time-invariant factor structure. To make the comparison transparent, we focus on the case where the untreated outcome follows a factor model with time-invariant factor loading with just one factor:

$$Y_{it}^N = \beta_i F_t + u_{it}, \quad i = 1, \ldots, N+1.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Suppose we just use the SCM weight W_i^{SC} to form our DP:

$$\hat{F}_t = \sum_{i=2}^{N+1} W_i Y_{it}^N.$$

SCM uses $\hat{F}_t$ directly as the estimated counterfactual outcome:

$$\hat{F}_t = \hat{Y}_{1t}^{SC},$$

while our DP approach uses $\hat{F}_t$ as the estimated factor.

Consequently, our approach requires an additional step to estimate the factor loadings of the treated unit by regressing the treated unit's outcome on the estimated factors over the pre-treatment periods:

$$\hat{\beta}_1 = \left(\sum_t \hat{F}_t^2\right)^{-1} \sum_t \hat{F}_t Y_{1t}^N.$$

Abadie et al. (2010) provide conditions under which the existence of W^{SC} requires the factor loadings of the treated unit to be perfectly reconstructed via a convex combination of the factor loadings in the control group; that is,

$$\beta_1 = \sum_{i=2}^{N+1} W_i^{SC} \beta_i. \tag{14}$$

In this case,

$$\begin{aligned}\hat{F}_t &= \sum_{i=2}^{N+1} W_i \beta_i F_t + \sum_{i=2}^{N+1} W_i u_{it} \\ &= \beta_1 F_t + o_P(1).\end{aligned}$$

This shows that, under the SCM weight condition, the DP estimated factors $\hat{F}_t$ indeed can be used as the counterfactual estimator, which is consistent for $\beta_1 F_t$.

Meanwhile, we use $\hat{F}_t$ as the factor estimator and apply one further step to estimate β_1. Under the SCM condition (14), our approach is also consistent. To see this, note that

$$\begin{aligned}\hat{\beta}_1 &= \left(\sum_t \hat{F}_t^2\right)^{-1} \sum_t \hat{F}_t (F_t \beta_1 + u_{1t}) \\ &= \left(\sum_t \hat{F}_t^2\right)^{-1} \sum_t \sum_{i=2}^{N+1} \hat{F}_t F_t \beta_i W_i + \left(\sum_t \hat{F}_t^2\right)^{-1} \sum_t \hat{F}_t u_{1t} \\ &= \left(\sum_t \hat{F}_t^2\right)^{-1} \sum_t \sum_{i=2}^{N+1} \hat{F}_t (Y_{it}^N - u_{it}) W_i + \left(\sum_t \hat{F}_t^2\right)^{-1} \sum_t \hat{F}_t u_{1t} \\ &= 1 + \zeta = 1 + o_P(1),\end{aligned}$$

where $\zeta = (\sum_t \hat{F}_t^2)^{-1} \sum_t \hat{F}_t u_{1t} - (\sum_t \hat{F}_t^2)^{-1} \sum_t \sum_{i=2}^{N+1} \hat{F}_t u_{it} W_i$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

As a result,

$$\hat{\beta}_1 \hat{F}_t = \hat{F}_t + o_P(1) = \beta_1 F_t + o_P(1).$$

Therefore, if the treated unit's factor loading lies within the convex hull of the control group's factor loadings, both methods are consistent. However, when this condition fails, SCM is not consistent. SCM requires the weights to be well-chosen to satisfy the exact matching condition of factor loadings. In contrast, because we flexibly estimate β_1, our framework does not require the weights to be precisely chosen. Indeed, our requirement for W is weaker than SCM.

Finally, we acknowledge that SCM is simpler to implement which only requires one set of weights. The DP, however, requires choosing the number of factors and choosing multiple weights to satisfy our conditions.

2.4. Estimation Procedure

With low-rank approximations, untreated potential outcomes can be modeled as a factor model. Following the factor-based approach proposed by Gobillon and Magnac (2016) and Li and Sonnier (2023), we separately estimate the factors and factor loadings (the treated unit). The detailed estimation steps are outlined below:

Step 1: (Initialization) Pre-determine the diversified weights W_{it} as described in Section 2.2.

Step 2: (Factors F_t) Using the DP method, the factor F_t is estimated by projecting the untreated outcomes Y_{it}^N of the control group onto the space defined by the diversified weights W_{it}:

$$\hat{F}_t = \frac{1}{N} W_t' Y_t, \text{ for } t = 1, \ldots, T, \tag{15}$$

where $W_t = (W_{2t}, \ldots, W_{N+1,t})'$ is the $N \times J$ matrix.

Step 3: (Factor Loading β_{1t}) Run local linear estimator Y_{1t}^N on $[\hat{F}_t, \hat{F}_t(\frac{t-T_0}{T_0})]$ with pre-treatment periods, obtain $\hat{\beta}_{1T_0}$ and $\hat{\beta}_{1T_0}^{(1)}$, then

$$\hat{\beta}_{1t} = \hat{\beta}_{1T_0} + \hat{\beta}_{1T_0}^{(1)} \left(\frac{t - T_0}{T_1} \right), \text{ for } t = T_0 + 1, \ldots, T.$$

Step 4: (Counterfactual Outcome) With the estimated factors and factor loading, the counterfactual outcomes for the treated unit $i = 1$ during the post-treatment period $t = T_0 + 1, \ldots, T$ are constructed as

$$\hat{Y}_{1t}^N = \hat{\beta}_{1t}' \hat{F}_t, \text{ for } t = T_0 + 1, \ldots, T. \tag{16}$$

Step 5: (ATEL Estimator) Using the estimated counterfactual outcomes, the treatment effect for each post-treatment period ($t = T_0 + 1, \ldots, T$) is given by

$$\hat{\alpha}_{1t} = Y_{1t} - \hat{Y}_{1t}^N. \tag{17}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

The ATEL estimator is then computed as

$$\hat{\alpha} = \frac{1}{T_1}\hat{\alpha}_1' \mathbf{K}_{T_1}, \tag{18}$$

where $\hat{\alpha}_1 = (\hat{\alpha}_{1,T_0+1}, \ldots, \hat{\alpha}_{1T})'$ and $\mathbf{K}_{T_1} = \left(K_h(\frac{(T_0+1)-T_0}{T_1}), \ldots, K_h(\frac{T-T_0}{T_1})\right)'$ are $T_1 \times 1$ vector.

Our estimation procedure is computationally efficient and scales well to large panels. It consists of two main steps: (i) factor estimation via DP, a cross-sectional weighted average of untreated outcomes, which can be computationally less costly than PCA because it avoids eigen-decomposition for large matrices and (ii) factor-loading estimation via local linear smoothing (weighted linear regression), which is a standard procedure and remains easy to compute even if T_0 is large, since it uses only the $\lfloor T_0 h \rfloor$ observations in a neighborhood of the boundary point T_0. In practice, both steps can be implemented efficiently via parallelization over t.[2]

3. ASYMPTOTIC RESULTS

3.1. Inferential Theory

This section presents the inferential theory. We provide the asymptotic normality of the ATEL. Remind the following notation:

$$h_t(\eta_{it}, X_{it}) = \sum_{j=1}^{J} \phi_j(\eta_{it}, X_{it}) f_{jt} + r_{it} = \beta_{it}' F_t + r_{it}.$$

Let $W_t = (W_{2t}, \ldots, W_{N+1,t})'$ be the $N \times J$ matrix. Each of its rows $W_{it} = (w_{it1}, \ldots, w_{itJ})'$ is a $J \times 1$ vector for $i = 2, \ldots, N+1$ and $t = 1, \ldots, T$. $\beta_t = (\beta_{2t}, \ldots, \beta_{N+1,t})'$ is the $N \times J$ matrix. $F_t = (f_{t1}, \ldots, f_{tJ})'$ is a $J \times 1$ vector. $u_t = (u_{1t}, \ldots, u_{N+1,t})'$ is $(N+1) \times 1$ vectors for $t = 1, \ldots, T$.

We now state the conditions and asymptotic results of estimators.

Assumption 1 (Sieve Approximation).

(i) $\{h_t(\cdot)\}_{t \le T}$ belong to ball $H(U, \|\cdot\|_{L_2}, C)$ inside a Hilbert space spanned by the basis $\{\phi_j : j = 1, \ldots\}$ with a uniform L_2-bound C:

$$\sup_{h \in H(U, \|\cdot\|_{L_2})} \|h\| \le C,$$

where U is the support of (η_{it}, X_{it}).

(ii) The sieve approximation satisfies $\max_{it} |r_{it}| < CJ^{-a}$ for some $a, C > 0$.

(iii) For some $C > 0$, $\max_{j \le J} \sup_{\eta, X} |\phi_j(\eta, X)| < C$.

[2] To facilitate implementation for applied users, we provide the MATLAB toolbox (ATEL), which is available at https://github.com/rueichilee/ATEL.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Assumption 1 presents some conditions for sieve approximation. Assumption 1(i) restricts $\{h_t(\cdot)\}_{t\leq T}$ to a Hilbert space with a uniform L_2 bound such that $\max_{t\leq T}\sum_{j=1}^{\infty} f_{tj}^2 < \infty$. Moreover, F_t is nonrandom since it's sieve coefficients. Then we illustrate that, under some sensible conditions on the functional space and the sieve bases, Assumption 1(ii) is well satisfied. Suppose h_t belongs to a Hölder class: For some $b, C, d > 0$, $\{h : \|D^b h(x) - D^b h(y)\| \leq C\|x-y\|^d$, for all $x, y\}$. Moreover, suppose the sieve bases follow some common basis functions, such as B-spline, trigonometric polynomials, or polynomial series. Then, with the sufficiently smooth functions $\{h_t(\cdot)\}$, sieve approximation error will satisfy $\max_{it}|r_{it}| < CJ^{-a}$, $a = (b+d)/(dim(\eta_{it}) + dim(X_{it}))$. Assumption 1(iii) can be satisfied if the basis functions are bounded such as trigonometric basis or the support of unobserved and observed covariates η_{it}, X_{it} is bounded. It implies that $\|\beta_{it}\|^2 = O_P(J)$ since factor loading is a vector of sieve basis function and the number of sieve terms grows slowly as $N, T \to \infty$.

Assumption 2 (Diversified Weights). There exist constants $c, C > 0$ such that as probability approaching one ($N \to \infty$),

(i) $\max_{i,t}|w_{itj}| < C$.
(ii) For each t, the $J \times J$ matrix $\frac{1}{N}W_t'W_t$ satisfies

$$c < \lambda_{\min}\left(\frac{1}{N}W_t'W_t\right) < \lambda_{\max}\left(\frac{1}{N}W_t'W_t\right) < C.$$

(iii) W_{it} is independent of u_{it}.

Assumption 2 ensures that the diversified idiosyncratic error, $\frac{1}{N}\sum_{i=2}^{N+1} W_{it}u_{it}$, satisfies the conditions for a cross-sectional central limit theorem when the idiosyncratic errors $(u_{2t}, \ldots, u_{N+1,t})$ are cross-sectionally weakly dependence. Furthermore, according to Assumption 1(ii), the diversified sieve approximation error $\frac{1}{N}\sum_{i=2}^{N+1} W_{it}r_{it}$ is negligible than the diversified idiosyncratic error such that

$$\hat{F}_t - H_t F_t = \frac{1}{N}\sum_{i=2}^{N+1} W_{it}u_{it} + o_p(1).$$

Therefore, the $\sqrt{N}(\hat{F}_t - H_t F_t)$ is asymptotically normal when the idiosyncratic errors $(u_{2t}, \ldots, u_{N+1,t})$ are cross-sectionally weakly dependence for each $t \leq T$. In particular, this asymptotic normality result still holds even in the finite T, which differs from conventional PCA requiring both $N, T \to \infty$.

Assumption 3 (Factors and Factor Loadings).

(i) $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \to \Sigma_F$ for some $J \times J$ positive definite matrix Σ_F.
(ii) For each t, there are some constants $c, C > 0$ such that as probability approaching one ($N \to \infty$),

$$c < \lambda_{\min}\left(\frac{\beta_t'\beta_t}{N}\right) \leq \lambda_{\max}\left(\frac{\beta_t'\beta_t}{N}\right) < C.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Assumption 3 imposes conditions on factors and factor loadings, similar to those in Su and Wang (2017). The key difference is that they assume random factors and nonrandom factor loadings, whereas we assume nonrandom factors and random factor loadings, derived from the sieve approximation. Assumption 3(i) ensures the nonsingularity of $\frac{1}{T}\sum_{t=1}^{T} F_t F_t'$ for sufficiently large T. Assumption 3(ii) imposes pervasive condition on factor loadings.

Assumption 4 (Smoothness over Time). $\beta_{it} = \beta_i(\frac{t}{T})$ and $W_{it} = W_i(\frac{t}{T})$ are smooth functions with continuous derivatives up to second order on the compact interval $[0, 1]$.

Assumption 4 imposes the common smoothness restriction on time-varying factor loadings typically assumed in the existing literature. The main difference is that, while many studies directly assume that factor loadings are functions of t/T, we approximate the unknown function $h_t(\eta_{it}, X_{it})$ via a sieve approach, where the factor loadings are represented by sieve basis functions of the covariates η_{it} and X_{it}. The assumption is plausible because aggregate variables often evolve slowly. In macroeconomic applications, variables such as GDP, inflation, and interest rates are commonly assumed to have trends or low-frequency components that evolve smoothly (see, e.g., Nelson and Plosser, 1982; Hodrick and Prescott, 1997).

When η_{it} and X_{it} follow smooth trajectories over time, that is, $\eta_{it} = \eta_i(\frac{t}{T})$ and $X_{it} = X_i(\frac{t}{T})$, we have

$$\beta_{it} = \phi(\eta_{it}, X_{it}) = \phi\left(\eta_i\left(\frac{t}{T}\right), X_i\left(\frac{t}{T}\right)\right) = \beta_i\left(\frac{t}{T}\right).$$

Likewise, the diversified weight W_{it} is also a sieve basis function of X_{it}, it follows that W_{it} is a smooth function of time. Accordingly, Assumption 4 implies that $H_t = \frac{1}{N}\sum_{i=2}^{N+1} W_{it}\beta_{it}$ is continuous with respect to t.

We impose the α-mixing condition. Let $\mathcal{F}_{-\infty}^{0}$ and $\mathcal{F}_{T}^{\infty}$ denote the σ-algebra generated by $\{u_t : t \le 0\}$ and $\{u_t : t \ge T\}$, respectively. Define the mixing coefficient

$$\alpha(T) = sup_{A\in\mathcal{F}_{\infty}^{0}, B\in\mathcal{F}_{T}^{\infty}} |P(A)P(B) - P(AB)|.$$

Assumption 5 (Weakly Dependence).

(i) $\{u_t : t \le T\}$ is a stationary process, satisfying $E(u_{it}) = 0$ and for some $\delta > 0$, there exists $C > 0$ such that $\mathbb{E}(|u_{it}|^{2+\delta}) < C$ for all $i \le N+1, t \le T$. Furthermore, u_{it} is independent of η_{it} and X_{it} for all $i \le N+1, t \le T$.

(ii) Strong Mixing: There exist $C_1 > 0$ such that for all $T > 0, \alpha(T) < exp(-C_1 T)$.

(iii) There is a constant $C > 0$, so that, $\max_{j\le N+1}\sum_{i=2}^{N+1} |\mathbb{E}(u_{it}u_{jt})| < C$, $\max_{t\le T}\frac{1}{N}\sum_{i,j=2}^{N+1}\sum_{s=1}^{T} |Cov(u_{it}, u_{js})| < C$, $\max_{t\le T}\frac{1}{N}\sum_{i,j=2}^{N+1}\sum_{s=1}^{T} |Cov(u_{it}, u_{js})| < C$, $\max_{t\le T}\max_{i,j\le N+1}\frac{1}{N}\sum_{q,v=2}^{N+1}\sum_{s=1}^{T} |Cov(u_{it}u_{qt}, u_{js}u_{vs})| < C$.

Assumption 5(i) and (ii) is a standard requirement for moments and the mixing coefficient for α-mixing time series. In particular, since diversified weight W_{it} is chosen based on X_{it}, the independence condition in Assumption 5(i) implies Assumption 2(iii). Assumption 5(iii) allows the idiosyncratic error u_{it} to be weak

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

dependence on both time series and cross section, which are in the same spirit of approximate factor analysis (see, e.g., Bai and Ng, 2002; Bai, 2003).

Assumption 6 (Signal-to-Noise). The rank J satisfies:

(i) $J^a \gg \sqrt{NT_1h}$.
(ii) $\sqrt{N} \gg J$.

Assumption 6(i) is employed to bound the approximation error of the sieve approximation, requiring $\{h_t(\cdot)\}_{t\leq T}$ to be sufficiently smooth to satisfy this condition.

Local linear regression is applied to estimate $\hat{\beta}_{it}, \hat{\beta}_{it}^{(1)}$. Standard practice involves imposing mild conditions on the kernel functions and the convergence properties of the bandwidth, h.

Assumption 7 (Kernel Estimation).

(i) The kernel function $K^* : [-1, 1] \to \mathbb{R}^+$ is a symmetric bounded probability density function that satisfies Lipschitz continuity.
(ii) As $N, T_0, T_1 \to \infty$, $h \to 0$, $T_0h \to \infty$, $T_1h \to \infty$, $N/T_0h \to \infty$, $T_0h^3J^2 \to 0$.

Assumption 7 is the standard assumption for local linear regression. Several types of kernel function satisfy Assumption 7(i), such as Epanechnikov, uniform, and quartic kernels. To handle boundary issues, we define the kernel function $K(\cdot)$ based on $K^*(\cdot)$ following Su and Wang (2017):

$$K_h(\frac{t-r}{T}) = \frac{1}{h}K(\frac{t-r}{Th}) = \begin{cases} \frac{1}{h}K^*(\frac{t-r}{Th}) / \int_{-r/(Th)}^{1} K^*(u)du, & \text{if } r \in [1, \lfloor Th \rfloor) \\ \frac{1}{h}K^*(\frac{t-r}{Th}), & \text{if } r \in [\lfloor Th \rfloor, T - \lfloor Th \rfloor] \\ \frac{1}{h}K^*(\frac{t-r}{Th}) / \int_{-1}^{(1-r/T)/h} K^*(u)du, & \text{if } r \in (T - \lfloor Th \rfloor, T]. \end{cases} \tag{19}$$

To establish an asymptotic result, define $\mu_j = \int u^j K(u)du$ and $\nu_j = \int u^j K^2(u)du$. Specifically, $\mu_0 = 1$ uniformly.

Formally, the following theorem shows that $\hat{\alpha}$ is asymptotic normal.

THEOREM 1. *With Assumptions 1–7, as* $N, T_0, T_1 \to \infty$,

$$(\frac{1}{T_0h}\Sigma_1 + \frac{1}{T_1h}\Sigma_2)^{\frac{-1}{2}}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, 1),$$

where $\Sigma_1 = \nu_0\lambda'\xi^{-1}\Omega\xi^{-1}\lambda$, $\lambda = \lim_{T_1\to\infty}\frac{1}{T_1}\sum_{t=T_0+1}^{T} F_t$, $\xi = \lim_{T_0\to\infty}\frac{1}{T_0}\sum_{t=1}^{T_0} F_tF_t'$, $\Omega = \lim_{T_0\to\infty}\left[\frac{1}{T_0}\sum_{t=1}^{T_0} F_tF_t'\mathbb{E}(u_{1t}^2) + \frac{2}{T_0}\sum_{t=1}^{T_0}\sum_{s=t+1}^{T_0} F_tF_s'\mathbb{E}(u_{1t}u_{1s})\right]$, $\Sigma_2 = \nu_0\sigma^2$, $\sigma^2 = \lim_{T_1\to\infty}\left[\frac{1}{T_1}\sum_{t=T_0+1}^{T}\mathbb{E}(u_{1t}^2) + \frac{2}{T_1}\sum_{t=T_0+1}^{T}\sum_{s=t+1}^{T}\mathbb{E}(u_{1t}u_{1s})\right]$.

The asymptotic result holds when N, T_0, and T_1 are large, corresponding to the factor-based estimation of the average treatment effect (e.g., Bai and Ng, 2021; Li and Sonnier, 2023). The asymptotic variance comprises two terms. Because

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

the two variance terms vanish at different rates, the rate of convergence of $\hat{\alpha}$ is $1/\delta_{T_0,T_1,h}$, where $\delta_{T_0,T_1,h} = \min\{\sqrt{T_0 h}, \sqrt{T_1 h}\}$. It is important to note that the ATEL estimator is always asymptotically normal regardless of the limit relationship between T_0 and T_1. To see this, the asymptotic distribution can be rewritten as

$$\left(\frac{\delta_{T_0,T_1,h}^2}{T_0 h}\Sigma_1 + \frac{\delta_{T_0,T_1,h}^2}{T_1 h}\Sigma_2\right)^{-\frac{1}{2}} (\delta_{T_0,T_1,h}(\hat{\alpha}-\alpha)) \xrightarrow{d} N(0,1).$$

In particular, if $T_0/T_1 \to 0$, $\sqrt{T_0 h}(\hat{\alpha}-\alpha) \xrightarrow{d} N(0,\Sigma_1)$; if $T_1/T_0 \to 0$, $\sqrt{T_1 h}(\hat{\alpha}-\alpha) \xrightarrow{d} N(0,\Sigma_2)$.

The validity of the asymptotic inference does not require the consistent estimation of H_t because, similar to the standard time-invariant factor model, the rotation matrix is not identifiable. This issue is analogous to the rotation problem in the usual PCA setting, where factors are estimated up to an unknown rotation H that itself cannot be consistently estimated. However, this lack of identification does not affect the validity of asymptotic inference, because the asymptotic variance would depend on H only through either HF_t or $H^{-1}\beta_i$. In either case, we can replace these terms with their consistent estimators, $\hat{F}_t$ and $\hat{\beta}_i$, to consistently estimate the variance. This analogy carries over to our setting where the rotation matrix H_t varies over time. In other words, the asymptotic variance is rotation-free and can be consistently estimated, even though the rotation matrix H_t itself is not.

We now turn to the consistent estimation of the asymptotic covariance in Theorem 1. Given the weak dependence of the time series on the idiosyncratic error, an estimator of Σ_1 is given by

$$\hat{\Sigma}_1 = \hat{\lambda}'\hat{\xi}^{-1}\hat{\Omega}\hat{\xi}^{-1}\hat{\lambda},$$

where $\hat{\lambda} = \frac{1}{T_1\mu_0}\sum_{t=T_0+1}^{T}\hat{F}_t K_h(\frac{t-T_0}{T_1})$, $\hat{\xi} = \frac{1}{T_0\mu_0}\sum_{t=1}^{T_0}\hat{F}_t\hat{F}_t' K_h(\frac{t-T_0}{T_0})$, and

$$\hat{\Omega} = \frac{h}{T_0}\sum_{t=1}^{T_0}\hat{u}_{1t}^2\hat{F}_t\hat{F}_t' K_h^2\left(\frac{t-T_0}{T_0}\right) + \frac{2h}{T_0}\sum_{j=1}^{T_0-1}\sum_{t=j+1}^{T_0}\hat{u}_{1t}\hat{u}_{1j}\hat{F}_t\hat{F}_j' K_h\left(\frac{t-T_0}{T_0}\right)K_h\left(\frac{j-T_0}{T_0}\right),$$

where $\hat{u}_{1t} = Y_{1t} - \hat{\beta}_{1t}'\hat{F}_t$.

An estimator of Σ_2 is given by

$$\hat{\Sigma}_2 = \frac{h}{T_0}\sum_{t=1}^{T_0}\hat{u}_{1t}^2 K_h^2\left(\frac{t-T_0}{T_0}\right) + \frac{2h}{T_0}\sum_{j=1}^{T_0-1}\sum_{t=j+1}^{T_0}\hat{u}_{1t}\hat{u}_{1j}K_h\left(\frac{t-T_0}{T_0}\right)K_h\left(\frac{j-T_0}{T_0}\right).$$

If u_{1t} is serially uncorrelated, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ can be simplified:

$$\hat{\Sigma}_1 = \hat{\lambda}'\hat{\xi}^{-1}\hat{V}\hat{\xi}^{-1}\hat{\lambda}, \quad \hat{\Sigma}_2 = \frac{h}{T_0}\sum_{t=1}^{T_0}\hat{u}_{1t}^2 K_h^2(\frac{t-T_0}{T_0}),$$

where $\hat{V} = \frac{h}{T_0}\sum_{t=1}^{T_0}\hat{u}_{1t}^2\hat{F}_t\hat{F}_t' K_h^2(\frac{t-T_0}{T_0})$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

In practice, a data-driven approach is recommended for bandwidth selection. One effective method is leave-one-out cross-validation, which determines the optimal bandwidth, h_{cv}, by minimizing the following objective function:

$$h_{cv} = \arg\min_{h} \sum_{t=1}^{T_0} (y_{1t} - \hat{\beta}_{1t}^{(-t)} \hat{F}_t)^2, \tag{20}$$

where $\hat{\beta}_{1t}^{(-t)}$ is the leave-one-out estimator. Specifically, $\hat{\beta}_{1t}^{(-t)}$ is the counterpart of $\hat{\beta}_{1t} = \hat{\beta}_{1T_0} + \hat{\beta}_{1T_0}^{(1)}(\frac{t-T_0}{T_0})$, but excludes the t-th observation in the local linear estimation process.

While our main focus is inference for the ATEL, practitioners may also be interested in inference on the estimated counterfactual at each post-treatment period by constructing pointwise confidence intervals. This requires further assumption on β_{it}. Consider the estimated counterfactual outcomes at $t > T_0$:

$$\hat{Y}_{1t}^N = -\underbrace{\left(\hat{\beta}_{1T_0} - H_{T_0}^{-1}\beta_{1T_0}\right)' H_t F_t}_{O_P\left(\frac{1}{\sqrt{T_0 h}}\right)} - \underbrace{\left(\hat{\beta}_{1T_0}^{(1)} - H_{T_0}^{-1}\beta_{1T_0}^{(1)}\right)' H_t F_t \left(\frac{t-T_0}{T_1}\right)}_{O_P\left(\frac{1}{\sqrt{T_0 h}}\right)} + \underbrace{\beta_{1t}' F_t}_{O_P(1)} + o_P(1).$$

The first two terms, estimation error in the factor loading for the pre-treatment periods, are the source of Σ_1 (before averaging over post-treatment periods), while the third term, $\beta_{1t}' F_t$, is the common components of the factor model. Since we are not considering the treatment effect, the common component $\beta_{1t}' F_t$ is dominant here. If we further assume that β_{1t} are i.i.d. and follow a normal distribution, the statistical inference can still be implemented. For example, suppose $\beta_{1t} \sim N(\bar{\beta}, \Sigma_\beta)$ with $\Sigma_\beta = diag\left(\sigma_1^2, \ldots, \sigma_J^2\right)$. Then, for each post-treatment period, the variance of the estimated counterfactual outcome is

$$\tilde{\Sigma}_t = \tilde{\Sigma}_{1,t} + \tilde{\Sigma}_{2,t},$$

where $\tilde{\Sigma}_{1,t} = \frac{1}{T_0 h} v_0 F_t' \xi^{-1} \Omega \xi^{-1} F_t$ and $\tilde{\Sigma}_{2,t} = F_t' \Sigma_\beta F_t$.

The estimators for $\tilde{\Sigma}_{1,t}$ and $\tilde{\Sigma}_{2,t}$ are given by

$$\hat{\tilde{\Sigma}}_{1,t} = \frac{1}{T_0 h} \hat{F}_t \hat{\xi}^{-1} \hat{\Omega} \hat{\xi}^{-1} \hat{F}_t, \quad \hat{\tilde{\Sigma}}_{2,t} = \hat{F}_t' \hat{\Sigma}_\beta \hat{F}_t,$$

where $\hat{\Sigma}_\beta = \frac{1}{T_1} \sum_{t=T_0+1}^{T} (\hat{\beta}_{1t} - \hat{\bar{\beta}})(\hat{\beta}_{1t} - \hat{\bar{\beta}})' K_h(\frac{t-T_0}{T_1})$ and $\hat{\bar{\beta}} = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \hat{\beta}_{1t} K_h(\frac{t-T_0}{T_1})$. Accordingly, the pointwise confidence intervals for the estimated counterfactual outcome can be constructed as

$$\hat{Y}_{1t}^N \pm c_\alpha \hat{\tilde{\Sigma}}_t^{1/2}, \text{ for } t = T_0 + 1, \ldots, T, \tag{21}$$

where c_α is the critical value of a standard normal or t-distribution for the given confidence interval, and $\hat{\tilde{\Sigma}}_t$ is the estimator for $\tilde{\Sigma}_t$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

3.2. Extensions

We conclude this section with a few extensions of our estimator that highlight the flexibility of our approach.

3.2.1. *Dynamic ATEL.* The original ATEL parameter focuses on the short-term effect localized around the intervention date. However, researchers are often also interested in the dynamic treatment effect, that is, how the policy impact evolves after the initial intervention (e.g., Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). In particular, one may ask: How does the effect of treatment vary with the length of exposure? For example, does ATEL increase or decrease as more time elapses after treatment begins?

The ATEL framework can be naturally generalized to a dynamic version that localizes the treatment effect at later post-treatment periods. Specifically, for a given post-treatment horizon m, define

$$\alpha_m = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \mathbb{E}(Y_{1t}^I - Y_{1t}^N) K_h \left(\frac{t-(T_0+m)}{T_1} \right),$$

which denotes the dynamic ATEL, capturing the short-term effect localized around the post-treatment $T_0 + m$. The corresponding estimator is constructed analogously as

$$\hat{\alpha}_m = \frac{1}{T_1} \sum_{t=T_0+1}^{T} \hat{\alpha}_{1t} K_h \left(\frac{t-(T_0+m)}{T_1} \right).$$

The asymptotic normality result in Theorem 1 remains valid for $\hat{\alpha}_m$, provided that m is fixed or $m/T_1 h \to 0$.

3.2.2. *ATEL with Carryover Effects.* In many applications, treatment effects may carry over in time (Imai et al., 2023). Treatment received in earlier periods can continue to influence outcomes in later periods, even if the current treatment status is the same. In other words, the outcome at time t may depend not only on whether the unit is treated at t, but also on how long it has been treated and on its past treatment history.

We extend the ATEL framework to accommodate such carryover effects. Let the binary treatment path up to time t be denoted by $D_{1:t} \in \{0,1\}^t$. In the potential outcomes framework, the potential outcome is written as $Y_{1t}(D_{1:t})$, representing the outcome that the treated unit would experience at time t under the treatment path $D_{1:t}$

Following Imai et al. (2023), we impose a limited (finite) carryover assumption that the realization of the potential outcome depends only on the most recent L lags of the treatment path.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Assumption 8 (Finite Carryover). For all treatment paths $D_{1:t} \in \{0,1\}^t$,

$$Y_{1t}(D_{1:t}) = Y_{1t}(D_{t-L+1}, \ldots, D_t).$$

This assumption explicitly restricts carryover effects to at most L periods. Since ATEL measures a local treatment effect using a narrow window in the post-treatment period, it can be generalized to this finite carryover setting. The estimand of interest is now defined as

$$\alpha(L) = \frac{1}{T_1} \sum_{t=T_0+L+1}^{T} \mathbb{E}(Y_{1t}(\mathbf{1}) - Y_{1t}(\mathbf{0}))K_h\left(\frac{t-(T_0+L)}{T_1}\right),$$

where $\mathbf{1}$ and $\mathbf{0}$ denote L-dimensional vectors of ones and zeros, respectively. Thus, $Y_{1t}(\mathbf{1})$ and $Y_{1t}(\mathbf{0})$ are the potential outcomes for the treated unit that has been continuously treated or untreated over the last L periods prior to t.

The key point is that we exclude the first L post-treatment periods, $t = T_0 + 1, \ldots, T_0 + L$, because the corresponding potential outcomes $Y_{1t}(\mathbf{1})$ involve a counterfactual history in which the unit would have already been treated for L periods, which is incompatible with the observed treatment path immediately after T_0. We therefore construct the ATEL estimator with carryover effects as

$$\hat{\alpha}(L) = \frac{1}{T_1} \sum_{t=T_0+L+1}^{T} \hat{\alpha}_{1t} K_h\left(\frac{t-(T_0+L)}{T_1}\right).$$

This estimator is a kernel-weighted average of the post-treatment effects after the policy has been implemented for at least L periods. The asymptotic normality result in Theorem 1 remains valid for $\hat{\alpha}(L)$, since trimming a fixed number L of post-treatment periods is asymptotically negligible.

4. SIMULATIONS

In this section, we conduct a set of Monte Carlo simulations to evaluate our theoretical results. We study the asymptotic results in approximating the finite sample distribution of ATEL. To evaluate the performance of our methods, we consider three DGPs for the potential outcomes in the absence of the treatment:

- *DGP1*: $Y_{it}^N = X_{it}\theta_t + \lambda_{i1}f_{1t} + \lambda_{i2}f_{2t} + u_{it}$, where $\theta_t \sim N(0,1)$ and $\lambda_{i1}, f_{1t}, \lambda_{i2}, f_{2t} \sim N(\frac{1}{\sqrt{2}}, 1)$,
- *DGP2*: $Y_{it}^N = h_t(\eta_{it}, X_{it}) + u_{it}$, where $h_t(\eta_{it}, X_{it}) = \sum_{r=1}^{\infty} \frac{|S_{tr}|}{r^3}\eta_{it}^r + \sum_{r=1}^{\infty} \frac{|V_{tr}|}{r^3}X_{it}^r$,
- *DGP3*: $Y_{it}^N = h_t(\eta_{it}, X_{it}) + u_{it}$, where $h_t(\eta_{it}, X_{it}) = \sum_{r=1}^{\infty} \frac{|S_{tr}|}{r^3}\sin(r\eta_{it}) + \sum_{r=1}^{\infty} \frac{|V_{tr}|}{r^3}\sin(rX_{it})$,

where X_{it}, η_{it} are generated from $U(0,1)$, S_{tr}, V_{tr} are generated from $N(2,1)$, and u_{it} is generated from $N(0,1)$.

Followed from (3) and (4), the treated potential outcomes are then generated from $Y_{1t}^I = Y_{1t}^N + \alpha_{1t}$ for $t = T_0 + 1, \ldots, T$ with $\alpha_{1t} = 0.5$. We consider the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

several combination of sample size $N = \{30, 100, 200\}$, $T_0 = \{15, 60, 100\}$, $T_1 = \{15, 40, 100\}$. Since the rank J is not inferred from the data, we consider $J \in \{2, 3, 4\}$.

We construct the diversified weight, W_{it}, followed from Section 2.2, using observed covariates with B-spline series expansion. For the local linear estimator, the Epanechnikov kernel function $K(u) = \frac{3}{4}(1-u^2)I(|u| <= 1)$ is used. To select the optimal bandwidth, the leave-one-out cross-validation method is employed.

Table 2 reports the simulation coverage probability and average length of 95% confidence intervals for the standardized estimate α based on 3,000 simulation replications. Overall, the simulation results are consistent with our asymptotic theory. Across designs, the empirical coverage is close to the nominal level and remains stable to the choices of J in the simulation. In addition, we find that the average length of the confidence intervals under DGP3 is shorter than under DGP1 and DGP2.

Figure 1 presents the histograms of the standardized estimates of ATEL with the standard normal density. All DGPs exhibit a good approximation to the standardized normal distribution, irrespective of the choice of J. This indicates that our inferential procedure is robust to the sieve dimension.

To evaluate the performance of our estimation method, we consider the following DGPs for partially linear models, where the observed covariates X_{it} linearly affect outcome variables and unobserved covariates η_{it} in the nonlinear, time-varying function $h_t(\cdot)$.

- *DGP4*: $Y_{it}^N = h_t(\eta_{it}) + X_{it}\theta + u_{it}$, where $h_t(\eta_{it}) = \sum_{r=1}^{\infty} \frac{|S_{tr}|}{r^3}\eta_{it}^r$,
- *DGP5*: $Y_{it}^N = h_t(\eta_{it}) + X_{it}\theta + u_{it}$, where $h_t(\eta_{it}) = \sum_{r=1}^{\infty} \frac{|S_{tr}|}{r^3}\sin(r\eta_{it})$,

where X_{it}, η_{it} is generated from $U(0,1)$, S_{tr} is generated from $N(2,1)$, θ is generated from $N(\frac{1}{\sqrt{2}}, 1)$, and u_{it} is generated from $N(0,1)$.

Since $h_t(\eta_{it})$ is a time-varying function dependent on unobserved covariates η_{it}, it can also be approximated by a low-rank time-varying factor structure, as we did previously, such that $h_t(\eta_{it}) \approx \beta_{it}' F_t$. Therefore, the time-varying factor model can be seen as a special case of the model in DGPs 4 and 5, where this general nonlinear function is approximated using a time-varying factor structure, and observed covariates affecting the outcome linearly through θ.

For comparison, in addition to the DP method, we also employ an alternative estimation approach based on the local PCA method of Su and Wang (2017), who restrict their analysis to a pure time-varying factor structure without observed covariates. In order to estimate θ simultaneously, we adapt their method by iteratively running local PCA to estimate the time-varying factor structure with observed covariates. This procedure follows the same principle as Bai (2009), who propose an IFE model and run PCA iteratively until the coefficient estimates converge. Further details of our estimation scheme are provided in the Supplementary Material.

To estimate the treatment effect using the local PCA method, we proceed as follows:

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

TABLE 2. Simulated coverage probability of 95% confidence interval and average length of the CI for the standardized estimate α.

Coverage									
	DGP1			DGP2			DGP3		
	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$
$N=30, T_0=15, T_1=15$	0.9147	0.9360	0.9417	0.9247	0.9340	0.9450	0.9217	0.9317	0.9430
$N=100, T_0=15, T_1=15$	0.9113	0.9310	0.9357	0.9287	0.9297	0.9453	0.9223	0.9330	0.9450
$N=100, T_0=60, T_1=40$	0.9350	0.9357	0.9440	0.9527	0.9357	0.9390	0.9360	0.9357	0.9507
$N=200, T_0=60, T_1=40$	0.9360	0.9450	0.9447	0.9420	0.9320	0.9463	0.9453	0.9447	0.9417
$N=200, T_0=100, T_1=100$	0.9367	0.9433	0.9423	0.9430	0.9397	0.9402	0.9487	0.9407	0.9367
Average length									
	DGP1			DGP2			DGP3		
	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$	$J=2$	$J=3$	$J=4$
$N=30, T_0=15, T_1=15$	3.4371	5.7315	10.2042	2.0062	3.9684	5.6613	1.9124	2.9438	6.8371
$N=100, T_0=15, T_1=15$	3.3668	5.8876	9.0321	1.9586	2.9892	5.6640	1.8970	2.9984	5.6781
$N=100, T_0=60, T_1=40$	1.7794	1.7959	1.8156	0.9974	1.0008	1.0319	0.9737	0.9785	0.9996
$N=200, T_0=60, T_1=40$	1.7995	1.7874	1.8205	0.9931	1.0006	1.0208	0.9741	0.9818	0.9949
$N=200, T_0=100, T_1=100$	1.2271	1.2090	1.2196	0.6833	0.6753	0.6804	0.6730	0.6661	0.6719

Note: Based on simulations with 3,000 repetitions.

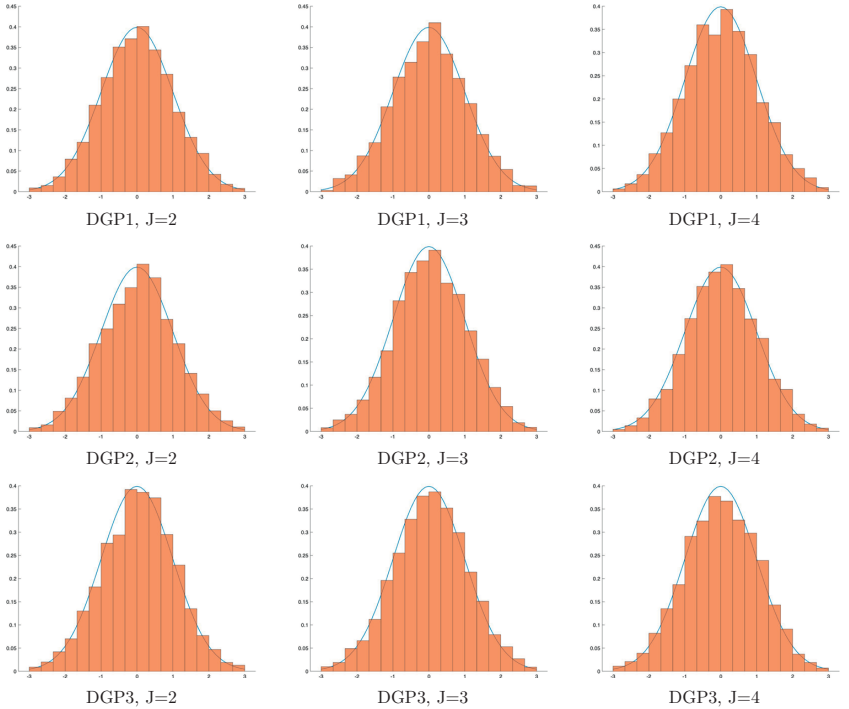


FIGURE 1. Histogram of estimated average treatment effect localization. These histograms are for the standardized estimates α (i.e., $\hat{\alpha}-\alpha$ divided by the estimated asymptotic standard deviation). The standard normal density is superimposed on the histogram. Note: The sample size is $N=200, T_0=100, T_1=100$. Based on simulations with 3,000 repetitions.

Step 1: Using local PCA method iteratively, we estimate F_t and θ using observations in the control group ($i=2,\ldots,N+1$, $t=1,\ldots,T$).

Step 2: Using local PCA method iteratively, we estimate β_{it} using observations in the pre-treatment periods ($i=1,\ldots,N+1$, $t=1,\ldots,T_0$).

Step 3: Construct counterfactual outcome: $\hat{Y}_{1t}^N=\hat{\beta}_{1T_0}'\hat{F}_t+X_{1t}\hat{\theta}$ and then compute the treatment effect: $\hat{\alpha}_{1t}=Y_{1t}-\hat{Y}_{1t}^N$ for $t=T_0+1,\ldots,T$.

Because the untreated outcome Y_{1t}^N is unobserved in the post-treatment periods, we use the estimation of β_{1T_0} to approximate β_{1t} when constructing the counterfactual in Step 3. To assess the accuracy of the estimated treatment effect, we compute a weighted MSE that localizes our estimator around T_0 via a kernel function. We refer to this measure as MSEL (MSE localization), defined by

$$MSEL=\frac{1}{T_1}\sum_{t=T_0+1}^{T}(\alpha_{1t}-\hat{\alpha}_{1t})^2K_h\left(\frac{t-T_0}{T_1}\right).$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

TABLE 3. Comparison of MSEL between DP and local PCA with $J = 2$.

	DGP4		DGP5	
	DP	Local PCA	DP	Local PCA
$N = 30, T_0 = 15, T_1 = 15$	0.9655	2.3463	0.9577	2.3093
$N = 100, T_0 = 15, T_1 = 15$	0.9680	2.1321	1.0361	2.1156
$N = 100, T_0 = 60, T_1 = 40$	1.3245	1.4523	1.2871	1.3776
$N = 200, T_0 = 60, T_1 = 40$	1.2979	1.3966	1.2667	1.3301
$N = 200, T_0 = 100, T_1 = 100$	1.3024	1.3674	1.2451	1.2977

Note: Based on simulations with 1,000 repetitions.

Table 3 reports the MSEL results with the number of factors $J = 2$. We can observe that our DP method has smaller MSELs compared to the local PCA method across all sample sizes considered. Moreover, the DP method provides more stable MSELs as the sample size decreases for both DGP 4 and DGP 5. By contrast, the local PCA method's MSELs deteriorate with smaller sample sizes. These results indicate that our method performs well even when the model is partially linear.

5. EMPIRICAL APPLICATION: ATEL OF RIGHT-TO-CARRY LAWS ON VIOLENT CRIME

We illustrate the usefulness of the DP inference with an application about the short-term effects of RTC law on violent crime. Determining whether a policy achieves its intended effect in a timely manner is crucial for making future policy decisions, especially in the realm of firearm regulation. Among developed countries, the United States has the highest rate of firearm violence, with gun homicides occurring at a rate roughly 25% higher than in other high-income countries (Bangalore and Messerli, 2013). Over the past several decades, lawmakers have pursued various strategies to address gun violence by regulating the purchase, possession, and use of firearms. Although the long-term average treatment effects of firearm policies have been studied extensively, relatively little attention has been given to their short-term impacts.

RTC laws, which allow individuals to carry concealed firearms in public, aim to regulate the possession and use of concealed firearms, balancing personal safety with public security. The primary debate of RTC laws revolves around their impact on public safety. On the one hand, these laws could reduce violent crime by deterring potential offenders who fear encountering armed victims. On the other hand, they could increase violent crime if permit holders engage in criminal activities or if the presence of concealed firearms exacerbates criminal behavior among others.

Empirical research on RTC laws has struggled to find consensus. For example, Lott Jr. and Mustard (1997) and LaTulip Jr. (2000) propose the "More Gun, Less Crime" hypothesis, using panel data analysis of county-level data to support their

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

theory that RTC laws reduce crime rate in every violent crime category between 5% and 8%. However, more recent studies, employing updated data and different statistical methods, suggest that RTC laws have minimal impact or may increase violent crime rates (e.g., Doucette et al., 2019; Siegel et al., 2019; Donohue et al., 2022). For instance, Aneja et al. (2011) found that RTC laws have negligible effects on murder rates but increase other violent crime rates by 20%–30%. Similarly, Donohue et al. (2019), using SCMs, observed that RTC laws are associated with higher violent crime rates after adoption.

We investigate the impact of RTC law implementation in Arizona, where RTC legislation took effect in 1994. Prior to 1994, Arizona prohibited concealed firearms for the general public, categorizing it as a may-issue state, with only law enforcement personnel or those with strong ties to law enforcement allowed to carry concealed firearms. In 1994, Arizona transitioned to a shall-issue state by passing legislation allowing state-permitted gun owners to carry concealed weapons.

We use state-level data on annual violent crime rates (per 100,000 residents) from 1977 to 2006 from the FBI's Uniform Crime Reports (UCRs).[3] The analysis distinguishes between the pre-treatment period (1977–1993) and the post-treatment period (1994–2006). The potential control group comprises 14 states: nine states with no RTC legislation before 2014 and five states that enacted RTC laws after 2006, such as Nebraska and Kansas. The explanatory variables include the police rate and the poverty rate. Using these explanatory variables, we construct diversified weights based on a J-dimensional B-spline series expansion.

Using the data-driven factor selection procedure of Su and Wang (2017), the information criteria select $J = 2$. As a robustness check, we also consider alternative specifications with $J = 2, 3, 4,$ and 5. The ATEL estimation results are presented in Figure 2, which displays the actual violent crime rate and estimated counterfactual outcomes during the period 1977–2006. We also report the corresponding pointwise 95% confidence intervals for the post-treatment periods, constructed under the normality assumption on the factor loadings in (21). We observe that before the intervention occurs, the fitted counterfactual lines closely follow the actual violent crime rate, indicating a good in-sample fit. After the intervention, the counterfactual crime rates decline more sharply than the actual rates, suggesting a positive treatment effect of the RTC law. This pattern holds across all values of J. The numerical analysis in Table 4 reports the ATEL estimates for different values of J. We can observe a consistent result that the ATEL is statistically significant at the 5% level for all choices of J. These results demonstrate that over-estimating the number of factors maintains satisfactory estimation accuracy, suggesting that our method does not rely on accurate estimation of J.

For comparison, we evaluate two alternative methods, including IFE methods proposed by Bai (2009) and SCM proposed by Abadie et al. (2010). Our analysis

[3]The data are available from http://works.bepress.com/john_donohue/89/.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

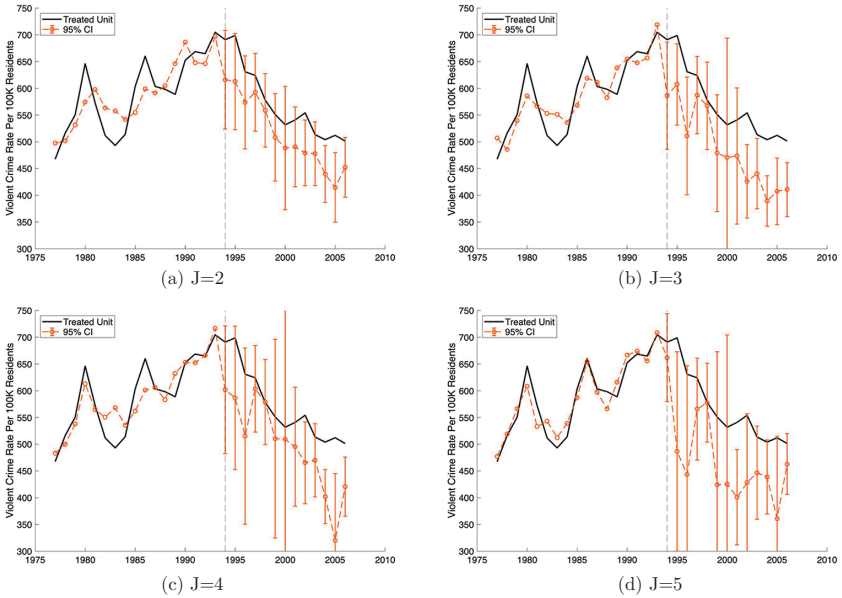


FIGURE 2. Treatment effect of the RTC law on Arizona with different J. The solid line is the actual violent crime rate. The dashed lines are the counterfactual outcomes with different J. The vertical lines are the pointwise 95% confidence intervals. The vertical dashed line denotes the beginning of the treatment.

TABLE 4. Average treatment effect localization of RTC law of Arizona with different J.

J	ATEL	p-value
2	49.7665(15.0055)	0.0069
3	71.9248(12.3949)	0.0001
4	57.8420(13.3495)	0.0012
5	99.0401(13.7880)	0.0000

Note: The number in the parentheses is standard error.

focuses on factor models with $J = 2$. Since SCM requires that all predictors are time-invariant, all explanatory covariates are averaged over pre-treatment periods, and all pre-treatment outcomes are included.[4] As shown in Figure 3, all methods follow a similar trend during the post-treatment periods. We can observe that the

[4]Ferman et al. (2020) highlighted a key limitation of SCM: the absence of clear guidelines for selecting predictor variables. They demonstrated that selectively including certain predictors—especially pre-treatment outcomes—to improve pre-treatment fit may lead to misleading inferences.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

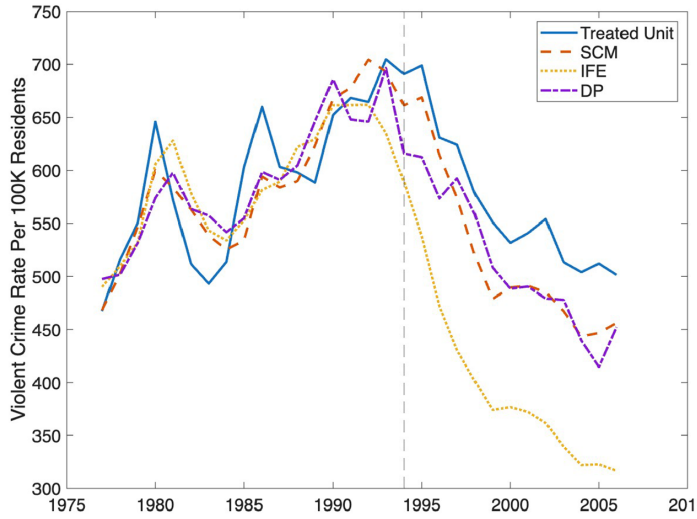


FIGURE 3. Treatment effect of the RTC law of Arizona with different methods. The solid line is the actual violent crime rate. The dashed lines are the counterfactual outcomes with different methods. The vertical line denotes the beginning of the treatment effect.

TABLE 5. Average treatment effect localization of RTC law on Arizona with DP and IFE.

Method	ATEL	*p*-value
DP	49.7665(15.0055)	0.0069
IFE	153.5546(40.1636)	0.0024

Note: The number in the parentheses is standard error.

DP method closely aligns with SCM results, while the IFE method consistently has larger treatment effect estimates, diverging from the other methods. Lastly, since SCM in Abadie et al. (2010) relies on placebo tests for inference, our comparison in Table 5 focuses exclusively on the DP and IFE methods, evaluating the ATEL, standard error, and *p*-value. It is noteworthy that the ATEL estimated by the IFE method is substantially larger than that of the DP method, as observed in Figure 3, resulting in a smaller *p*-value.

6. CONCLUSION

This article introduced ATEL to capture localized, short-term impacts of policy interventions. We model untreated potential outcomes to flexibly account for both observed and unobserved covariates in a time-varying, nonparametric framework, which can be represented as a low-rank, time-varying factor structure via sieve

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

approximation. By combining DP for factor estimation with local linear smoothing for time-varying loadings, our method overcomes the limitations of conventional PCA-based approaches. We also develop the asymptotic distribution to facilitate inference on the ATEL. An empirical application on RTC laws demonstrates the practical value of ATEL for timely policy evaluation.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S026646662610067X.

FUNDING STATEMENT

The author declares that no specific funding has been received for this article.

COMPETING INTEREST

The author declares that no competing interests exist.

REFERENCES

Abadie, A., Diamond, A., & 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., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque Country. *American Economic Review*, 93(1), 113–132.

Ahn, S. C., & Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3), 1203–1227.

Aneja, A., Donohue, J., & Zhang, A. (2011). The impact of right-to-carry laws and the NRC report: Lessons for the empirical evaluation of law and policy. *American Law and Economics Review*, 13(2), 565–631.

Angrist, J., Bettinger, E., & Kremer, M. (2006). Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia. *American Economic Review*, 96(3), 847–862.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, 116(536), 1716–1730.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1), 135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4), 1229–1279.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

Bai, J., & Ng, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *Journal of the American Statistical Association*, 116(536), 1746–1763.

Bangalore, S., & Messerli, F. H. (2013). Gun ownership and firearm-related deaths. *The American Journal of Medicine*, 126(10), 873–876. https://doi.org/10.1016/j.amjmed.2013.04.012

Botosaru, I., Muris, C., & Pendakur, K. (2023). Identification of time-varying transformation models with fixed effects, with an application to unobserved heterogeneity in resource shares. *Journal of Econometrics*, 232(2), 576–597.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Carvalho, C., Masini, R., & Medeiros, M. C. (2018). ArCo: An artificial counterfactual approach for high-dimensional panel time-series data. *Journal of Econometrics*, 207(2), 352–380.

Chen, X. (2007). Chapter 76 large sample sieve estimation of semi-nonparametric models. In J. J. Heckman & E. E. Leamer (Eds.), Handbook of Econometrics (Vol. 6B, pp. 5549–5632). Elsevier.

Chernozhukov, V., Fernández-Val, I., Hahn, J., & Newey, W. (2013). Average and quantile effects in nonseparable panel models. *Econometrica*, 81(2), 535–580.

Chernozhukov, V., Hansen, C., Liao, Y., & Zhu, Y. (2023). Inference for low-rank models. *Annals of Statistics*, 51(3), 1309–1330.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.

Cheung, Y. L. (2024). Identification of time-varying factor models. *Journal of Business & Economic Statistics*, 42(1), 76–94.

Choi, J., Kwon, H., & Liao, Y. (2024). Inference for low-rank completion without sample splitting with application to treatment effect estimation. *Journal of Econometrics*, 240(1), 105682.

Connor, G., Hagmann, M., & Linton, O. (2012). Efficient semiparametric estimation of the Fama-French model and extensions. *Econometrica*, 80(2), 713–754.

Donohue, J. J., Aneja, A., & Weber, K. D. (2019). Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies*, 16(2), 198–247.

Donohue, J. J., Cai, S. V., Bondy, M. V., & Cook, P. J. (2022). Why does right-to-carry cause violent crime to increase? (Tech. rep.), National Bureau of Economic Research.

Doucette, M. L., Crifasi, C. K., & Frattaroli, S. (2019). Right-to-carry laws and firearm workplace homicides: A longitudinal analysis (19922017). *American Journal of Public Health*, 109(12), 1747–1753.

Duan, J., Pelger, M., & Xiong, R. (2024). Factor analysis for causal inference on large non-stationary panels with endogenous treatment. SSRN 4823360.

Fan, J. (1996). *Local polynomial modelling and its applications* (Monographs on Statistics and Applied Probability, 66). Routledge.

Fan, J., & Liao, Y. (2022). Learning latent factors from diversified projections and its applications to over-estimated and weak factors. *Journal of the American Statistical Association*, 117(538), 909–924.

Fan, J., Liao, Y., & Wang, W. (2016). Projected principal component analysis in factor models. *Annals of Statistics*, 44(1), 219.

Ferman, B., Pinto, C., & Possebom, V. (2020). Cherry picking with synthetic controls. *Journal of Policy Analysis and Management*, 39(2), 510–532.

Fernández-Val, I., Freeman, H., & Weidner, M. (2021). Low-rank approximations of nonseparable panel models. *The Econometrics Journal*, 24(2), C40–C77.

Freyberger, J. (2017). Non-parametric panel data models with interactive fixed effects. *The Review of Economic Studies*, 85(3), 1824–1851.

Gobillon, L., & Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics*, 98(3), 535–551.

Graham, B. S., & Powell, J. L. (2012). Identification and estimation of average partial effects in irregular correlated random coefficient panel data models. *Econometrica*, 80(5), 2105–2152.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6), 2052–2086.

Hoderlein, S., & White, H. (2012). Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics*, 168(2), 300–314.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.

Hsiao, C., Steve Ching, H., & Ki Wan, S. (2012). A panel data approach for program evaluation: Measuring the benefits of political and economic integration of Hong Kong with mainland China. *Journal of Applied Econometrics*, 27(5), 705–740.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X

Imai, K., Kim, I. S., & Wang, E. H. (2023). Matching methods for causal inference with time-series cross-sectional data. *American Journal of Political Science*, 67(3), 587–605.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467–475.

LaTulip, K. P. , Jr. (2000). More guns, less crime, by john R. Lott, Jr. *Journal of Health Care Law and Policy*, 4(1), 147.

Li, K. T., & Sonnier, G. P. (2023). Statistical inference for the factor model approach to estimate causal effects in quasi-experimental settings. *Journal of Marketing Research*, 60(3), 449–472.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Lott, J. R. , Jr., & Mustard, D. B. (1997). Crime, deterrence, and right-to-carry concealed handguns. *The Journal of Legal Studies*, 26(1), 1–68.

Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconmic time series: Some evidence and implications. *Journal of Monetary Economics*, 10(2), 139–162.

Ouyang, M., & Peng, Y. (2015). The treatment-effect estimation: A case study of the 2008 economic stimulus package of China. *Journal of Econometrics*, 188(2), 545–557.

Pelger, M., & Xiong, R. (2022). State-varying factor models of large dimensions. *Journal of Business & Economic Statistics*, 40(3), 1315–1333.

Saul, S. (2025). Federal cuts prompt Johns Hopkins to cut more than 2,000 workers. *The New York Times*.

Siegel, M. B., Pahn, M., Xuan, Z., Fleegler, E. W., & Hemenway, D. (2019). The impact of state firearm laws on homicide and suicide deaths in the USA, 1991–2016: A panel study. *Journal of General Internal Medicine*, 34(10), 2021–2028.

Stevens, A. H. (1997). Persistent effects of job displacement: The importance of multiple job losses. *Journal of Labor Economics*, 15(1, Part 1), 165–188.

Su, L., & Wang, F. (2025). Inference for large dimensional factor models under general missing data patterns *Journal of Econometrics*, 250, 106022.

Su, L., & Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1), 84–101.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199.

Xiong, R., & Pelger, M. (2023). Large dimensional latent factor modeling with missing observations and applications to causal inference. *Journal of Econometrics*, 233(1), 271–301.

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

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 13 Nov 2026 at 22:52:34, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S026646662610067X