Econometric Theory, **0**, 2026, 1–50.
doi:10.1017/S0266466626100620

ESTIMATION OF HETEROGENEOUS PANEL DATA MODELS WITH AN APPLICATION TO PROGRAM EVALUATION

XUN LU
Chinese University of Hong Kong

KE MIAO
Fudan University and Shanghai Institute of International Finance and Economics

LIANGJUN SU
Tsinghua University

This article studies panel data models with two-dimensional unobserved slope heterogeneity and interactive fixed effects. We propose a two-step approach to estimate the parameters in the model. We apply the proposed model to estimate heterogeneous treatment effects at both individual and aggregate levels. Monte Carlo simulations show that the proposed estimators perform remarkably well in finite samples in comparison with some existing methods such as the synthetic control ones. We apply our method to study the effect of economic liberalization on economic growth and find a positive and significant aggregate average treatment effect on the treated.

1. INTRODUCTION

Heterogeneity is a fundamental feature in many economic datasets and how to model heterogeneity has been a central issue in econometrics (see, e.g., Heckman, 2001). In this article, we propose a linear panel data model with a general form of heterogeneity. In our model, the outcome variable depends on both observable covariates and unobservable factors in a heterogeneous way. The slope coefficients of covariates are allowed to vary along both cross-sectional and time-series dimensions. We also allow multiple latent common factors (i.e., interactive fixed effects [IFEs hereafter]) to affect the outcome variable via an intercept term. Our model nests many commonly used panel data models as special cases, such as the homogeneous panel data model with IFEs, the random coefficient model, the time-varying coefficient model, and the panel data model with two-dimensional

The authors thank Editor-in-Chief Yixiao Sun, Co-Editor Martin Weidner, and two anonymous referees for their constructive comments. Address correspondence to Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, China, email: sulj@sem.tsinghua.edu.cn.

© The Author(s), 2026. Published by Cambridge University Press.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

heterogeneity, among others (for details, see the literature review below). We propose estimation and inference methods for our heterogeneous model. One important application of our model is to estimate the counterfactual outcomes for program evaluations.

An important feature of our model is that it allows the observable covariates, x_{it}, to influence the outcome heterogeneously, varying across both individuals and time. The importance of allowing heterogeneous effects has been well recognized in panel data models, and the homogeneous slope coefficients are often rejected empirically (see, e.g., Browning and Carro, 2007; Su and Chen, 2013). As we estimate all the slope coefficients, $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$, it is transparent to examine the role of covariates in explaining the outcome. To illustrate the importance of allowing for heterogeneous slopes, consider the famous Feldstein–Horioka puzzle in international economics (see, e.g., Feldstein and Horioka, 1980), where y_{it} and x_{it} are the investment rate and saving rate for country i in year t, respectively. The standard economic theory suggests that rational investors would invest in countries offering the highest return if there were no regulations in international financial markets. However, in practice, investors may face all kinds of regulations in investing in foreign countries. Therefore, the effect of domestic saving rates on investment rates depends on the constraints of capital mobility across countries. Based on this argument, the slope coefficient (β_{it}^0) of saving rates should be country-specific. It should also be time-varying, as the international capital market environments change over time. Lu and Su (2023) find strong evidence of heterogeneity in both dimensions using a simpler model that does not incorporate a factor structure. In our empirical study in Section 6, where we investigate the relationship between countries' economic growth and their consumption and investment shares, we also find substantial heterogeneity in the effects of these two variables on growth (see Figure 1 in Section 6). As Lu and Su (2023) argue, neglecting heterogeneity in slope coefficients can result in model misspecification and undermine the validity of subsequent estimation and inference in empirical research. A primary reason for this is that the true, heterogeneous coefficient β_{it}^0 may be correlated with x_{it}, potentially inducing endogeneity if β_{it}^0 is incorrectly specified as a constant.

Specifically, we assume that the outcome variable y_{it} follows a general panel data model with two-dimensional heterogeneity and a factor structure

$$y_{it} = \beta_{it}^{0\prime} x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it},\ i \in [N],\ t \in [T], \tag{1.1}$$

where x_{it} is a $K \times 1$ vector of observable covariates, f_t^0 is an $R^0 \times 1$ vector of unobservable factors, λ_i^0 is the corresponding factor loadings, R^0 is typically unknown, and u_{it} is the idiosyncratic error term. Here, $[a] \equiv \{1, 2, \ldots, a\}$ for any integer $a \geq 1$. Throughout this article, we use superscript "0" to denote true values. $\lambda_i^{0\prime} f_t^0$ is often referred to as the IFEs in the literature (see, e.g., Bai, 2009) and R^0 is the number of unobservable factors. We assume that the slope coefficient β_{it}^0 has an additive structure:

$$\beta_{it}^0 = \beta_{i\cdot}^0 + \beta_{\cdot t}^0, \tag{1.2}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

where $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ are two $K \times 1$ random vectors. Without loss of generality, we impose the normalization condition that $E(\beta_{i\cdot}^0) = 0$ and leave $E(\beta_{\cdot t}^0)$ unrestricted.[1] $E(\beta_{\cdot t}^0)$ can be interpreted as the average effects of x_{it} on y_{it}. $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ (or its centered version $\beta_{\cdot t}^0 - E(\beta_{\cdot t}^0)$) represent the effect heterogeneity along the individual and time dimensions, respectively. We allow $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ to be correlated with x_{it}, therefore, model (1.1) is a correlated random-coefficient (CRC)-type model with IFEs. In this model, we treat $\beta_{i\cdot}^0, \beta_{\cdot t}^0, \lambda_i^0, f_t^0$, and R^0 all as unknown parameters. Note that we do not allow dynamic models in this article, that is, x_{it} cannot be the lagged dependent variable.[2]

We propose a two-step estimation procedure for our model. In the first step, we obtain a preliminary consistent estimator of $\beta_{i\cdot}^0, \beta_{\cdot t}^0$, and Θ^0 via nuclear norm regularization (NNR), where Θ^0 is a $T \times N$ matrix with its (t,i)th element given by $\lambda_i^{0\prime} f_t^0$. Nevertheless, due to the use of NNR, we can only establish slow convergence rates for these estimators. In particular, we can only establish the weak consistency of the NNR estimator of the low-rank matrix Θ^0 in terms of Frobenius norm. But it is sufficient for us to obtain a consistent estimator of the number of factors, say, via a singular value thresholding procedure. In the second step, we adopt an iterative procedure to conduct post-NNR (PNNR) estimation of all parameters, including $\beta_{i\cdot}^0, \beta_{\cdot t}^0, \lambda_i^0$, and f_t^0. We establish the asymptotic properties of these estimators in the large panel framework where both N and T tend to infinity. One key challenge is to deal with the high dimensionality of the parameters. Note that the total dimension of $\{\beta_{i\cdot}^0, \beta_{\cdot t}^0, \lambda_i^0, f_t^0\}$ is $K(N+T-1)+R^0(N+T)$, which increases with both N and T. For example, in a reasonable setting where $N = 100, T = 100, K = 2$, and $R^0 = 2$, it is close to 800.

As argued above, our model in (1.1) and (1.2) can be of independent interest and applied directly to many empirical studies where $\{\beta_{i\cdot}^0, \beta_{\cdot t}^0\}$ are usually the parameters of interest. The model's utility is further demonstrated by its application to estimating counterfactual outcomes, which is central to program evaluation and treatment effect analysis. Following Rubin's potential outcome framework, let y_{it}^1 and y_{it}^0 be the potential treated (intervened) and untreated outcomes for individual i at time t, respectively. For example, consider the effect of economic liberalization (openness) on economic growth where the binary treatment indicator is whether a country is open and x_{it} is some observable covariates (e.g., investment share) that can explain economic growth. y_{it}^1 is the economic growth of country i at year t if it is open at year t, and y_{it}^0 is its economic growth if it is closed. The treatment effect is simply $y_{it}^1 - y_{it}^0$. The fundamental issue here is that we only observe either y_{it}^1 or y_{it}^0 but never both. We consider the treatment effect on the treated, that is, those countries which switch from "closed" to "open" in certain years. For those treated, we directly observe y_{it}^1, but y_{it}^0 is unobservable and counterfactual, which

[1] Alternatively, one can follow Lu and Su (2023) and write the slope vector as $\beta^0 + \beta_{i\cdot}^0 + \beta_{\cdot t}^0$, in which case we can impose $E(\beta_{i\cdot}^0) = E(\beta_{\cdot t}^0) = 0$ for identification.

[2] The main reason is that the two-dimensional heterogeneous coefficients create a complicated process for y_{it} if the x_{it} is the lagged y_{it}, say $y_{i,t-1}$. For a discussion on this issue for a simpler model where the factor structure is absent, see Lu and Su (2023). We leave this important extension of our model to future research.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

has to be estimated. We provide a general approach to estimate the counterfactual outcome in this framework. Specifically, we assume that the untreated potential outcome y_{it}^0 follows models (1.1) and (1.2) by taking into account the unobserved heterogeneity over both the countries and the years. Using the data of the control individuals (those who are never treated), we estimate $\beta_{\cdot t}^0$ and f_t^0. Then we estimate the individual-specific $\beta_{i\cdot}^0$ and λ_i^0 for the treated individuals and estimate the counterfactual untreated outcome for them. For details, see our Algorithm 2 in Section 4. We derive the asymptotic properties of various treatment effects on the treated based on our estimated counterfactual outcomes. Our approach can be viewed as an alternative to the popular synthetic control method (SCM) to estimate counterfactual outcomes. Relative to SCM, our approach can potentially exploit more information on covariates (x_{it}) when model (1.1) is correctly specified. In addition, we allow $\beta_{\cdot t}$ to have a nonstationary distribution. The nonstationarity of $\beta_{\cdot t}$ can invalidate many existing inference procedures applied to SCM, such as Chernozhukov, Wüthrich, and Zhu (2021), Masini and Medeiros (2021), and Masini and Medeiros (2022). This is because $\beta_{\cdot t}$ typically enters the error term of the SCM model, so nonstationary coefficients directly induce nonstationarity in the error term. Consequently, this violates a fundamental stationarity assumption underlying those inference methods. For a detailed discussion and evidence on these points, please refer to the simulations in Section 5 of the main text and Section S.7 of the Supplementary Material.

We apply our new method to evaluate the impact of economic liberalization on economic growth using panel data for 60 countries over 30 years (1965–1994). The outcome variable is annual economic growth, and the treatment indicator denotes whether a country is classified as "open." Two key covariates—the investment share and consumption share of GDP—are included in the analysis. Our main finding is that liberalization exerts a positive and statistically significant effect: the aggregate average treatment effects on the treated (ATT) is 1.24% annual growth. Country-level ATT estimates reveal substantial heterogeneity: 20 countries exhibit positive effects, while 15 show negative effects. Additionally, we find that the investment share is positively associated with the untreated potential outcome, whereas the consumption share exerts a small negative effect. Overall, the results support the view that openness fosters economic growth, though the magnitude and direction of effects vary across countries.

Our article is closely related to several strands of literature. First, our article pertains to the large literature on panel data models (see, e.g., Hsiao, 2022, for a review). Our model can be thought of as a generalization of many widely used panel data models. If we impose $\beta_{i\cdot}^0 + \beta_{\cdot t}^0$ to be constant, our model becomes the standard panel data model with IFEs (see, e.g., Bai, 2009; Moon and Weidner, 2015; Armstrong, Weidner, and Zeleneev, 2026). If we restrict that $\beta_{\cdot t}^0 = 0$ and leave $\beta_{i\cdot}^0$ unrestricted, this is the random coefficient model (see, e.g., Pesaran, 2006; Su, Shi, and Phillips, 2016; Li, Cui, and Lu, 2020). If we impose that $\beta_{i\cdot}^0 = 0$ while leaving $\beta_{\cdot t}^0$ unrestricted, our model becomes the time-varying coefficient model (see, e.g., Zhu, 2017). If the factor structure $\lambda_i^{0\prime} f_t^0$ is replaced by two-way additive

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

fixed effects, our model reduces to the panel data model with two-dimensional heterogeneity studied in Keane and Neal (2020) and Lu and Su (2023).

Second, our article is related to the literature on the estimation of counterfactual outcomes and treatment effects, which has been developing rapidly over the last two decades. Due to space limitations, here we only provide a brief review of the estimation methods with panel data. For cross-sectional data, see Imbens and Wooldridge (2009) for a review. In two seminal papers, Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010) propose the SCM to estimate counterfactual outcomes, which has been widely applied to many empirical contexts. SCM uses a weighted average of observed untreated outcomes of untreated individuals to estimate the counterfactual for those treated. As remarked in Athey and Imbens (2017), SCM is arguably the most important development in program evaluation in the last decade. The literature on SCM has been developing rapidly in the last two decades (see Abadie, 2021, for an excellent review). We have seen extensive development on the generalization or inference on SCM in the last decade. For example, Doudchenko and Imbens (2016) generalize SCM by allowing the weights to be negative and their sum to differ from one, but with an elastic-net type penalty. Xu (2017) proposes a generalized SCM based on panels with IFEs where the slope coefficients of observable covariates are assumed to be homogeneous. Athey et al. (2021) develop matrix completion estimators for counterfactuals. Arkhangelsky et al. (2021) propose a synthetic difference-in-differences method. Li (2020), Chernozhukov et al. (2021), Masini and Medeiros (2021), and Masini and Medeiros (2022) propose different inference methods for SCM, and Cattaneo, Feng, and Titiunik (2021) develop prediction intervals for SCM. Typically, a certain form of stationarity of the error term is required for the validity of inferences considered in these papers and inferences can be invalid when nonstationarity is not properly accounted for, say when it appears in the parameter. More recently, Chen (2023) shows that SCM can be represented as an online learning algorithm and its predictions of counterfactual outcomes perform almost as well as an oracle weighted average of control units' outcomes. Note that our estimator does not take a form of a weighted average of control units' outcomes, thus it is not in the class of estimators considered in Chen (2023). In addition, we allow for nonstationarity in $\{\beta_{\cdot t}\}$, in which case typical SCM inference procedures (e.g., Chernozhukov et al., 2021) that fail to take into account the covariate part are asymptotically invalid.

Third, in terms of methodology, our approach is closely related to the burgeoning literature on the use of NNR in econometrics, statistics, and machine learning (see Koltchinskii, Lounici, and Tsybakov, 2011; Negahban and Wainwright, 2011; Rohde and Tsybakov, 2011; Negahban et al., 2012; Bai and Ng, 2019; Fan, Gong, and Zhu, 2019; Athey et al., 2021; Ma, Su, and Zhang, 2022; Belloni et al., 2023; Miao, Phillips, and Su, 2023; Chernozhukov et al., 2026; Moon and Weidner, 2026). All these previous works focus on the error bounds (in Frobenius norm) for the NNR estimators, except Ma et al. (2022), Miao et al. (2023), Moon and Weidner (2026), and Chernozhukov et al. (2026), who study inference problems

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

in linear/nonlinear panel or VAR models with low-rank structures. Like these latter authors, we simply use the NNR to obtain consistent initial estimators and our final objective is to establish the usual inference theory for the complete model in (1.1) and (1.2).

The rest of the article is structured as follows. In Section 2, we propose our estimation method for models (1.1) and (1.2). Section 3 establishes the asymptotic properties of the estimators. In Section 4, we show how to apply our model to estimate various treatment effects and study their asymptotic properties. Section 5 reports Monte Carlo simulation results. In Section 6, we apply our method to study the effect of economic liberalization on economic growth. Section 7 concludes. All proofs of the main results are relegated to the Appendix. The Supplementary Material includes some detailed analyses and additional simulation and empirical results.

Notation. Let $\mathbb{I}_m$ denote an $m \times m$ identity matrix. For a real $m \times n$ matrix $A = (a_{ij})$, we use $\|A\|$, $\|A\|_{\text{sp}}$, and $||A||_*$ to denote its Frobenius norm, spectral norm, and nuclear norm, respectively: $||A||_{\text{sp}} \equiv \mu_1(A)$, $||A|| \equiv (\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2)^{1/2}$, and $||A||_* \equiv \sum_{k=1}^{m\wedge n} \mu_k(A)$, where $\mu_k(A)$ denotes the kth largest singular value of A for $k \leq m \wedge n \equiv \min(m,n)$ and $\equiv$ signifies a definitional relationship. Let $\|A\|_\infty$ and $\|A\|_1$ denote the L_∞ and L_1 norm of matrix A. That is, $\|A\|_\infty$ and $\|A\|_1$ are the maximum absolute row sum and column sum of A. Let $\mu_{\min}(A) \equiv \mu_{m\wedge n}(A)$ and $\mu_{\max}(A) \equiv \mu_1(A)$. Let A' denote the transpose of A. When A has full rank, let $\mathbb{P}_A = A(A'A)^{-1}A'$ and $\mathbb{M}_A = \mathbb{I}_m - \mathbb{P}_A$. For a positive integer n, let $[n] \equiv \{1, 2, \ldots, n\}$. Let $\text{vec}(\cdot)$ denote the usual vectorization operator. Let $\text{mat}(\cdot)$ be the reverse operator of $\text{vec}(\cdot)$: $\text{mat}(\text{vec}(A)) = A$. Let $\otimes$ denote the Kronecker product. Let $\mathbf{1}\{\cdot\}$ be the indicator function. Let $\text{blkdiag}(A_1, \ldots, A_k)$ represent a block-diagonal matrix with square diagonal blocks $A_1, \ldots, A_k$. The symbol $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and *plim* probability limit. We use $(N,T) \to \infty$ to signify that N and T pass to infinity jointly.

2. THE MODEL AND ESTIMATION METHOD

In this section, we first introduce the panel data model with two-dimensional unobserved slope heterogeneity and IFEs. Then we propose a two-step approach to estimate the parameters in the model.

2.1. The Model

The scalar form of our model is given by

$$y_{it} = (\beta_{i\cdot}^0 + \beta_{\cdot t}^0)' x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it}, \quad i \in [N],\ t \in [T], \tag{2.1}$$

where $x_{it}, f_t^0, \lambda_i^0$, and u_{it} are defined as above. Note that $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ are not separately identified. In other words, equation (2.1) cannot be directly employed to construct a least-squares estimator. To achieve identification, we assume $E(\beta_{i\cdot}^0) = 0$ and leave

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

$E(\beta_{\cdot t}^0)$ unrestricted. Define $\dot{\beta}_{i\cdot} = \beta_{i\cdot}^0 - \frac{1}{N}\sum_{j=1}^N \beta_{j\cdot}^0$ and $\dot{\beta}_{\cdot t} = \beta_{\cdot t}^0 + \frac{1}{N}\sum_{j=1}^N \beta_{j\cdot}^0$. By construction, $\sum_{j=1}^N \dot{\beta}_{j\cdot} = 0$. We can then rewrite the model as

$$y_{it} = (\dot{\beta}_{i\cdot} + \dot{\beta}_{\cdot t})' x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it}, \quad i \in [N-1],\ t \in [T], \tag{2.2}$$

$$y_{Nt} = \Big(-\sum_{j=1}^{N-1} \dot{\beta}_{j\cdot} + \dot{\beta}_{\cdot t} \Big)' x_{Nt} + \lambda_N^{0\prime} f_t^0 + u_{Nt}, \quad t \in [T]. \tag{2.3}$$

Let $\dot{B}^{(1)} \equiv [\dot{\beta}_{1\cdot}', \ldots, \dot{\beta}_{N-1\cdot}']' \in \mathbb{R}^{(N-1)K}$ and $\dot{B}^{(2)} \equiv [\dot{\beta}_{\cdot 1}', \ldots, \dot{\beta}_{\cdot T}']' \in \mathbb{R}^{TK}$. Equations (2.2) and (2.3) can be employed to construct a least squares estimator.

We establish two alternative representations of models (2.2) and (2.3). The vector representation facilitates the analysis of the least-squares estimators for elements in $\dot{B} \equiv [\dot{B}^{(1)\prime}, \dot{B}^{(2)\prime}]'$ in the following discussion, while the matrix representation is particularly useful for studying the estimator associated with the factor structure.

To proceed, we introduce some notation. Let $\mathbf{y}_{i\cdot} \equiv [y_{i1}, y_{i2}, \ldots, y_{iT}]' \in \mathbb{R}^T$, $\mathbf{y}_{\cdot t} \equiv [y_{1t}, y_{2t}, \ldots, y_{Nt}]' \in \mathbb{R}^N$, and $Y \equiv [\mathbf{y}_{1\cdot}', \mathbf{y}_{2\cdot}', \ldots, \mathbf{y}_{N\cdot}']' \in \mathbb{R}^{NT}$. Define $\mathbf{u}_{i\cdot}, \mathbf{u}_{\cdot t}$, and U analogously. Let $\mathbf{x}_{i\cdot} \equiv [x_{i1}, x_{i2}, \ldots, x_{iT}]' \in \mathbb{R}^{T\times K}$ and $\mathbf{x}_{\cdot t} \equiv [x_{1t}, x_{2t}, \ldots, x_{Nt}]' \in \mathbb{R}^{N\times K}$ be the covariate matrices that may appear in equations for individual i and time t, respectively. Define

$$X^{(1)} \equiv \underbrace{\begin{bmatrix} \mathbf{x}_{1\cdot} & & \\ & \ddots & \\ & & \mathbf{x}_{N-1\cdot} \\ -\mathbf{x}_{N\cdot} & \cdots & -\mathbf{x}_{N\cdot} \end{bmatrix}}_{NT\times(N-1)K},\ \mathbf{x}_{i\cdot}^{(2)} \equiv \underbrace{\begin{bmatrix} x_{i1}' & & \\ & \ddots & \\ & & x_{iT}' \end{bmatrix}}_{T\times TK},\ \text{and } X^{(2)} \equiv \underbrace{\begin{bmatrix} \mathbf{x}_{1\cdot}^{(2)} \\ \mathbf{x}_{2\cdot}^{(2)} \\ \vdots \\ \mathbf{x}_{N\cdot}^{(2)} \end{bmatrix}}_{NT\times TK}.$$

Let $X^{(1,2)} \equiv [X^{(1)}, X^{(2)}]$. Given the above notations, the model in (2.2) and (2.3) can be rewritten in an NT-vector form:

$$\textbf{Vector representation}: \quad Y = X^{(1)}\dot{B}^{(1)} + X^{(2)}\dot{B}^{(2)} + \theta^0 + U = X^{(1,2)}\dot{B} + \theta^0 + U, \tag{2.4}$$

where $\theta^0 \equiv \text{vec}\,(\Theta^0)$. The vector $\dot{B}$ stacks all free unit- and time-specific slope coefficients into a single $(N+T-1)K \times 1$ vector. Accordingly, for each pair (i,t), the corresponding row of $X^{(1,2)}$ consists of an expanded regressor vector, constructed by interacting x_{it} with the full set of unit and time dummies, while imposing the normalization constraint $\sum_{i=1}^N \dot{\beta}_{i\cdot} = 0$. Consequently, the product $X^{(1,2)}\dot{B}$ is equivalent to a regression specification that includes all unit–time interactive effects involving x_{it}. Next, we introduce the matrix representation. Denote the kth entries of $x_{it}, \dot{\beta}_{i\cdot}$, and $\dot{\beta}_{\cdot t}$ as $x_{k,it}$, $\dot{\beta}_{k,i\cdot}$, and $\dot{\beta}_{k,\cdot t}$, respectively. Let $\mathbf{X}_k$ be a $T \times N$ matrix with (t,i)th entry $x_{k,it}$, for $k \in [K]$. Similarly, we define $\mathbf{Y}$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

and $\mathbf{U}$. Then we have the matrix representation:

$$\textbf{Matrix representation}: \quad \mathbf{Y} = \sum_{k=1}^{K} (\dot{\mathbf{B}}_{k,T}\mathbf{X}_k + \mathbf{X}_k\dot{\mathbf{B}}_{k,N}) + \Theta^0 + \mathbf{U}, \tag{2.5}$$

where $\dot{\mathbf{B}}_{k,T} \equiv \text{diag}(\dot{\beta}_{k,\cdot 1}, \dot{\beta}_{k,\cdot 2}, \ldots, \dot{\beta}_{k,\cdot T})$ and $\dot{\mathbf{B}}_{k,N} \equiv \text{diag}(\dot{\beta}_{k,1\cdot}, \dot{\beta}_{k,2\cdot}, \ldots, \dot{\beta}_{k,N-1\cdot}, -\sum_{i=1}^{N-1} \dot{\beta}_{k,i\cdot})$ are $T \times T$ and $N \times N$ diagonal matrices, respectively.

We now outline the main steps for estimating model (2.1):

1. Solve the nuclear-norm-regularized problem defined in (2.7) below. This step yields a consistent estimator of the matrix Θ^0 in terms of Frobenius norm.
2. Determine the number of factors using (2.8) or other methods discussed below and obtain the initial estimators of the factors and factor loadings as in (2.9). This step provides consistent estimators of the number of factors, factors, and loadings.
3. Apply Algorithm 1, introduced in Section 2.3, to obtain the final estimators of the slope coefficients and latent factors and loadings. This step refines the estimators through iterations and produces the final estimators used for statistical inference.

2.2. Initial Estimation: Nuclear Norm Regularized Estimation

In this section, we produce initial estimators of the parameters in the model. Define the objective function as

$$\mathcal{L}_{NT}(B,\Theta) \equiv \frac{1}{2NT}||Y - X^{(1,2)}B - \theta||^2 + \gamma_{NT}\frac{||\Theta||_*}{\sqrt{NT}}, \tag{2.6}$$

where $B \in \mathbb{R}^{(N-1+T)K}$, $\theta \equiv$vec (Θ), $\Theta \in \mathbb{R}^{T\times N}$, and γ_{NT} is a regularization parameter such that $\gamma_{NT} \to 0$ as $(N,T) \to \infty$. Then the initial estimator is given by

$$(\hat{B}_o, \hat{\Theta}_o) \equiv \arg\min_{(B,\Theta)} \mathcal{L}_{NT}(B,\Theta). \tag{2.7}$$

At any fixed Θ, minimizing $\mathcal{L}_{NT}(B,\Theta)$ is a least-squares problem. Formally, we can write

$$\tilde{\mathcal{L}}_{NT}(\Theta) = \min_B \frac{1}{2NT}||Y - X^{(1,2)}B - \theta||^2 = \frac{1}{2NT}(Y-\theta)'\mathbb{M}_x(Y-\theta),$$

where $\mathbb{M}_x \equiv \mathbb{M}_{X^{(1,2)}}$ is the $NT \times NT$ projection matrix that projects out the columns of $X^{(1,2)}$. Hence, minimizing $\mathcal{L}_{NT}(B,\Theta)$ can be achieved by minimizing $\tilde{\mathcal{L}}_{NT}(\Theta) + \gamma_{NT}\frac{||\Theta||_*}{\sqrt{NT}}$. That is,

$$\begin{aligned}\hat{\Theta}_o &= \arg\min_{\Theta} \tilde{\mathcal{L}}_{NT}(\Theta) + \gamma_{NT}\frac{||\Theta||_*}{\sqrt{NT}} \\ &= \arg\min_{\Theta} \frac{1}{2NT}(Y-\theta)'\mathbb{M}_x(Y-\theta) + \gamma_{NT}\frac{||\Theta||_*}{\sqrt{NT}}.\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Given the initial estimators $\hat{B}_o$ and $\hat{\Theta}_o$, one can use various methods to determine the number of factors R^0. For example, we can estimate R^0 via the singular value thresholding as follows:

$$\hat{R} = \sum_{i=1}^{N\wedge T} \mathbf{1}\left\{\mu_i(\hat{\Theta}_o) \geq \left(\gamma_{NT}\sqrt{NT}\left\|\hat{\Theta}_o\right\|_{\text{sp}}\right)^{1/2}\right\}, \tag{2.8}$$

where $\mu_i(\hat{\Theta}_o)$ denotes the ith largest singular value of $\hat{\Theta}_o$. Alternatively, we can estimate R^0 by some other methods, including the information criteria and panel criteria methods of Bai and Ng (2002), the edge distribution estimator of Onatski (2010), the eigenvalue ratio and growth ratio method of Ahn and Horenstein (2013), and the cross-validation method of Jin, Miao, and Su (2021).

Given $\hat{R}$, we can obtain preliminary estimators of the factors and factor loadings. Consider the singular value decomposition (SVD) of $\hat{\Theta}_o$: $\hat{\Theta}_o = \hat{L}\hat{D}\hat{V}'$, where $\hat{D} = \text{diag}\,(\mu_1(\hat{\Theta}_o), \ldots, \mu_{N\wedge T}(\hat{\Theta}_o))$, and $\hat{L}$ and $\hat{V}$ stack the left and right singular vectors of $\hat{\Theta}_o$, respectively. We estimate F^0 and Λ^0, respectively, by

$$\hat{F}_o = \sqrt{T}\hat{L}_{*,[\hat{R}]} \text{ and } \hat{\Lambda}_o = \hat{V}_{*,[\hat{R}]}\hat{D}_{[\hat{R}],[\hat{R}]}/\sqrt{T}, \tag{2.9}$$

where $\hat{L}_{*,[\hat{R}]}$ denotes a submatrix of $\hat{L}$ that contains its first $\hat{R}$ columns and $\hat{D}_{[\hat{R}],[\hat{R}]} \equiv \text{diag}\,(\mu_1(\hat{\Theta}_o), \ldots, \mu_{\hat{R}}(\hat{\Theta}_o))$.

Remark 2.1. Here, we assume that β_{it}^0 follows an additive structure as in (1.2). Let $\beta_{k,it}^0$, $\beta_{k,i\cdot}^0$, and $\beta_{k,\cdot t}^0$ denote the k-th elements of β_{it}^0, $\beta_{i\cdot}^0$, and $\beta_{\cdot t}^0$, respectively. Since $\beta_{k,it}^0 = (\beta_{k,i\cdot}^0, 1)(1, \beta_{k,\cdot t}^0)'$, the $N \times T$ matrix $\boldsymbol{\beta}_k^0 \equiv \{\beta_{k,it}^0 : i \in [N],\ t \in [T]\}$ has rank two in general for each $k \in [K]$. It is possible to generalize $\beta_{k,it}^0$ to have a generic low-rank (factor) structure of rank R_k, namely, $\beta_{k,it}^0 = \gamma_{k,i}^{0\prime} g_{k,t}^0$, where $\gamma_{k,i}^0$ and $g_{k,t}^0$ are $R_k \times 1$ vectors for $k \in [K]$. In that case, the NNR needs to be applied to the slope coefficients too, as in Chernozhukov et al. (2026). However, inference in such settings typically requires stronger assumptions and a sample-splitting procedure, which we do not pursue in this article. See Remarks 3.6 and 3.7 in Section 3.3 below.

Remark 2.2. While both our study and Moon and Weidner (2026) utilize NNR, our contributions are distinct and address a fundamentally more challenging problem. Moon and Weidner (2026) apply NNR to estimate a homogeneous model, where the slope coefficient β_{it}^0 is assumed constant. In contrast, we consider a heterogeneous model, introducing high-dimensional parameters via the additive form $\beta_{it}^0 = \beta_{i\cdot}^0 + \beta_{\cdot t}^0$, varying across both individuals and time. This distinction is crucial. As noted in the literature (e.g., Jiang et al., 2021), the convergence of PCA estimators in panels with IFEs often hinges on a suitable initial estimator. The NNR method provides a consistent starting point that facilitates convergence. However, our article is one of the first to develop estimation and inference for a panel with two-dimensional slope heterogeneity and IFEs. Therefore, while NNR serves a similar initial-stage purpose, the subsequent analysis diverges completely.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

2.3. Post Nuclear Norm Regularized Estimation

In this section we propose an iterative procedure to obtain updated estimators of the regression coefficients, the factors, and the factor loadings. The algorithm goes as follows:

Algorithm 1 Iterated estimator for $\dot{B}$.

---

1. For $s = 0$, set $\hat{F}^{(s)} = \hat{F}_o$ and $\hat{\Lambda}^{(s)} = \hat{\Lambda}_o$.
2. Update $\hat{B}_s$ by

$$\hat{B}_s \equiv \arg\min_B \min_{g\in\mathbb{R}^{T\times\hat{R}}, h\in\mathbb{R}^{N\times\hat{R}}} \left\| \mathrm{mat}(Y - X^{(1,2)}B) - \hat{F}^{(s)}h' - g\hat{\Lambda}^{(s)\prime} \right\|^2$$

$$= \left[X^{(1,2)\prime}(\mathbb{M}_{\hat{\Lambda}^{(s)}} \otimes \mathbb{M}_{\hat{F}^{(s)}})X^{(1,2)}\right]^{-1} X^{(1,2)\prime}(\mathbb{M}_{\hat{\Lambda}^{(s)}} \otimes \mathbb{M}_{\hat{F}^{(s)}})Y. \quad \textbf{(2.10)}$$

   The corresponding sth step estimators for $\dot{\beta}_{j\cdot}$ and $\dot{\beta}_{\cdot t}$ are denoted as $\hat{\beta}_{j\cdot}^{(s)}$ and $\hat{\beta}_{\cdot t}^{(s)}$, respectively.
3. Update $\hat{F}^{(s+1)}$ and $\hat{\Lambda}^{(s+1)}$ based on $\mathrm{mat}[Y - X^{(1,2)}\hat{B}_s]$. Specifically, let SVD of $\mathrm{mat}[Y - X^{(1,2)}\hat{B}_s]$ be $\mathrm{mat}[Y - X^{(1,2)}\hat{B}_s] = \hat{L}^{(s+1)}\hat{D}^{(s+1)}\hat{V}^{(s+1)\prime}$. Estimate F^0 and Λ^0, respectively, by

$$\hat{F}^{(s+1)} = \sqrt{T}\hat{L}^{(s+1)}_{*,[\hat{R}]} \text{ and } \hat{\Lambda}^{(s+1)} = \hat{V}^{(s+1)}_{*,[\hat{R}]}\hat{D}^{(s+1)}_{[\hat{R}],[\hat{R}]}/\sqrt{T},$$

   where $\hat{L}^{(s+1)}_{*,[\hat{R}]}$ and $\hat{V}^{(s+1)}_{*,[\hat{R}]}$ denote submatrices of $\hat{L}^{(s+1)}$ and $\hat{V}^{(s+1)}$, respectively, consisting of their first $\hat{R}$ columns.
4. Iterate steps 2 and 3 for, $s^* \geq 2$ times.

   Set $\hat{B} \equiv \hat{B}_{s^*}$. Similar to the definitions $\dot{B} = [\dot{\beta}_{1\cdot}', \ldots, \dot{\beta}_{N-1\cdot}', \dot{\beta}_{\cdot 1}' \ldots, \dot{\beta}_{\cdot T}']'$, we set $\hat{B} \equiv [\hat{\beta}_{1\cdot}', \ldots, \hat{\beta}_{N-1\cdot}', \hat{\beta}_{\cdot 1}', \ldots, \hat{\beta}_{\cdot T}']'$. The estimator for $\dot{\beta}_{N\cdot}$ is given by $\hat{\beta}_{N\cdot} = -\sum_{i=1}^{N-1}\hat{\beta}_{i\cdot}$. Let $\hat{F} = \hat{F}^{(s^*)}$ and $\hat{\Lambda} = \hat{\Lambda}^{(s^*)}$. Denote tth row of $\hat{F}$ as $\hat{f}_t'$ and ith row of $\hat{\Lambda}$ as $\hat{\lambda}_i'$.

Note that the iterations in Algorithm 1 alternate between updating $\hat{B}_s$ and $(\hat{F}^{(s)}, \hat{\Lambda}^{(s)})$ for each iteration index s. However, the algorithm does not require numerical convergence. Instead, it can be terminated after any finite number of iterations $s^* \geq 2$. The advantages of this approach are twofold:

1. *Theoretical Perspective*: We do not require the solution to correspond to the global minimum of an optimization problem. As long as the initial estimators of the factors and factor loadings are consistent, we can show that a finite number of iterations yields the desired asymptotic properties. Moreover, our theoretical analysis indicates that stopping at $s^* = 2$ is sufficient, as additional iterations do not improve the asymptotic efficiency of the estimators in theory.
2. *Practical Perspective*: In implementation, it is unnecessary to verify numerical convergence, nor is there a concern that the algorithm may become trapped in a local minimum.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

3. ASYMPTOTIC THEORY

In this section, we study asymptotic properties of the initial estimator, the infeasible estimator that assumes F^0 and Λ^0 are observed as an intermediate step and the feasible estimator.

3.1. Asymptotic Properties of the Initial Estimators

To study the asymptotic properties of the initial estimators, we make the following assumptions.

Assumption 1. For $\gamma_{NT} = C_\gamma(N^{-1/2} + T^{-1/2})$, where C_γ is a sufficiently large positive constant, we have $(NT)^{-1/2}||\text{mat}\,(\mathbb{M}_x U)||_{sp} \leq \frac{\gamma_{NT}}{2}$ with probability approaching one (w.p.a.1).

Assumption 2. (Restricted Strong Convexity). Let $\mathbb{C}_{NT} \equiv \{\Theta \in \mathbb{R}^{T\times N} : \left\|\mathbb{M}_{F^0}\Theta\mathbb{M}_{\Lambda^0}\right\|_* \leq 3\left\|\Theta - \mathbb{M}_{F^0}\Theta\mathbb{M}_{\Lambda^0}\right\|_*\}$. There exists a $\mu > 0$, independent of N and T, such that for any $\theta \in \mathbb{R}^{NT}$ with $\text{mat}(\theta) \in \mathbb{C}_{NT}$, we have $\theta'\mathbb{M}_x\theta \geq \mu\theta'\theta$, w.p.a.1 conditional on (Λ^0, F^0).

Assumption 1 imposes some regularity conditions on $\mathbb{M}_x U$. It is close to the condition imposed on the tuning parameter in Lemma 2 of Moon and Weidner (2026), which studies the model with homogeneous slope coefficients. As one can argue, it suffices to choose $\gamma_{NT} = C_\gamma(N^{-1/2} + T^{-1/2})$ for some sufficiently large positive constant C_γ to ensure Assumption 1 to be satisfied under some primitive conditions on $\{x_{it}, u_{it}\}$. Such a rate ensures that the NNR estimator of Θ exhibits the optimal rate of convergence in terms of Frobenius norm (see Theorem 3.1 below). In our simulations and application, we follow the lead of Miao et al. (2023) and set $\gamma_{NT} = 0.5\max(N^{-1/2}, T^{-1/2})$ for simplicity, and leave the optimal choice of γ_{NT} for future research.

Assumption 2 imposes the celebrated restricted strong convexity (RSC) condition that was first introduced by Negahban and Wainwright (2012) and Negahban et al. (2012). We conjecture that this condition can be verified under more primitive assumptions by adapting the approach of Chernozhukov et al. (2026), accounting for the additional complexity introduced by the high-dimensional structure of our slope coefficients (see also Hong, Su, and Jiang, 2023; Moon and Weidner, 2026).

The following theorem studies the consistency of the nuclear-norm regularized estimator $\hat{\Theta}_o$ in terms of the Frobenius norm.

THEOREM 3.1. *Suppose Assumptions 1 and 2 hold. We have that* $\|\hat{\Theta}_o - \Theta^0\|/\sqrt{NT} \leq 3\sqrt{2R^0}\gamma_{NT}/\mu$, *w.p.a.1.*

Theorem 3.1 indicates that we can estimate the large-dimensional matrix Θ^0 consistently in terms of the Frobenius norm. Note that Θ^0 is a high-dimensional

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

matrix with NT entries and $\|\Theta^0\| = O_P(\sqrt{NT})$ in general. This explains why we need to normalize $\|\hat{\Theta}_o - \Theta^0\|$ by $\sqrt{NT}$ in Theorem 3.1.

Let $\max_i = \max_{1\le i\le N}$ and $\max_t = \max_{1\le t\le T}$. Define $\max_{i,j}$, $\max_{i,t}$, $\max_{t,s}$, $\min_i$, and $\min_t$ analogously. Next, we impose an assumption on the common factors and factor loadings.

Assumption 3. (Conditions on Factor Structure).

(i) $\left\|\frac{1}{N}\Lambda^{0\prime}\Lambda^0 - \Sigma_\Lambda\right\| = O_P(N^{-1/2})$ for some $\Sigma_\Lambda > 0$ and $\left\|\frac{1}{T}F^{0\prime}F^0 - \Sigma_F\right\| = O_P(T^{-1/2})$ for some $\Sigma_F > 0$.

(ii) The eigenvalues of $\Sigma_F^{1/2}\Sigma_\Lambda\Sigma_F^{1/2}$ are all distinct.

(iii) $\max_i ||\lambda_i^0|| = O_P(\log N)$ and $\max_t ||f_t^0|| = O_P(\log T)$.

Assumption 3(i) and (ii) posits a strong factor structure, a standard condition widely adopted in the literature (see, e.g., Bai and Ng, 2002; Bai, 2003, 2009; Su et al., 2016; Moon and Weidner, 2017; Miao et al., 2023, among others). Armstrong et al. (2026) propose a robust method for the estimation and inference of the slope coefficients in homogeneous panels involving weak factors. Assumption 3(iii) further imposes a uniform bound on the latent factors and factor loadings, which serves to facilitate the theoretical analysis in the subsequent sections. It holds when both λ_i^0 and f_t^0 are sub-exponentially distributed.

The following theorem reports the consistency of $\hat{R}, \hat{F}_o$, and $\hat{\Lambda}_o$.

THEOREM 3.2. *Suppose Assumptions 1–3 hold. Then we have*

(i) $P(\hat{R} = R^0) \to 1$ *as* $(N,T) \to \infty$.

(ii) $\frac{1}{\sqrt{T}}\left\|\hat{F}_o - F^0\tilde{H}\right\| = O_P(\gamma_{NT})$ *and* $\|\mathbb{M}_{\hat{F}_o} - \mathbb{M}_{F^0}\|_{sp} = O_P(\gamma_{NT})$, *where* $\tilde{H}$ *is an asymptotically nonsingular rotation matrix defined in the proof.*

(iii) $\frac{1}{\sqrt{N}}\left\|\hat{\Lambda}_o - \Lambda^0(\tilde{H}^{-1})'\right\| = O_P(\gamma_{NT})$ *and* $\|\mathbb{M}_{\hat{\Lambda}_o} - \mathbb{M}_{\Lambda^0}\|_{sp} = O_P(\gamma_{NT})$.

Theorem 3.2(i) indicates that $\hat{R}$ is a consistent estimator of R^0. Theorem 3.2(ii) indicates that the columns of $\hat{F}_o$ and F^0 span the same space asymptotically, so do the columns of $\hat{\Lambda}_o$ and Λ^0.

Remark 3.1. The result in Theorem 3.1 bears a conceptual resemblance to Lemma 2 of Moon and Weidner (2026), despite the significant increase in the number of regressors due to our two-dimensional heterogeneity. Theorem 3.2 follows from Theorem 3.1 under standard conditions on the factors and loadings in Assumption 3—a result also attainable for the homogeneous panel. The key mechanism that accommodates the high dimensionality of our slope parameters $\{\beta_{i\cdot}^0\}$ and $\{\beta_{\cdot t}^0\}$ is the high-level RSC condition in Assumption 2. It could be substantially more challenging to verify this RSC condition for our heterogeneous model than for a homogeneous one, which is the necessary statistical price for our more flexible specification.

Remark 3.2. Crucially, the results in Theorems 3.1 and 3.2 are preliminary. Despite the fact that we can establish the consistency of $\hat{\Theta}_o, \hat{F}_o$, and $\hat{\Lambda}_o$ in terms

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

of the Frobenius norm, we cannot obtain pointwise consistency or asymptotic distributions for the elements of these estimators. The primary role of Theorems 3.1 and 3.2 is to facilitate the analysis of the subsequent iterative (post-NNR) estimators. It is in the study of these post-NNR estimators where the substantive theoretical challenges and novelties of our article emerge. Due to the high-dimensional nature of the slope parameters, analyzing the asymptotic properties of the estimators for $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ becomes highly complex. As indicated in Section S.1 of the Supplementary Material, even the analysis of the infeasible estimators (assuming observed factors and loadings) is nontrivial. Furthermore, as detailed in Section S.2 of the Supplementary Material, the unobserved nature of the factors and loadings introduces additional layers of complication, substantially deepening the theoretical analysis beyond the homogeneous case.

3.2. Asymptotic Properties of the Infeasible Estimator

Despite the consistency of the initial estimators $\hat{F}^{(0)}$ and $\hat{\Lambda}^{(0)}$, it is challenging to study the asymptotic properties of $\hat{\beta}_{i\cdot}$ and $\hat{\beta}_{\cdot t}$ directly. Below we first study the asymptotic properties of their infeasible version as given in (3.1) and then study those of the feasible one. Our results show that the difference between the feasible and infeasible versions is asymptotically negligible. As an intermediate result, we first study the infeasible estimator in (3.1) in this section. Now we formally introduce the infeasible estimator. Given F^0 and Λ^0 are observed, we can estimate $\dot{B}$ by

$$\check{B} \equiv \left[X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})X^{(1,2)}\right]^{-1} X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})Y. \tag{3.1}$$

Similar to the definition $\dot{B} = [\dot{\beta}_{1\cdot}', \ldots, \dot{\beta}_{N-1\cdot}', \dot{\beta}_{\cdot 1}', \ldots, \dot{\beta}_{\cdot T}']'$, we set $\check{B} \equiv [\check{\beta}_{1\cdot}', \ldots, \check{\beta}_{N-1\cdot}', \check{\beta}_{\cdot 1}', \ldots, \check{\beta}_{\cdot T}']'$. The infeasible estimator for $\dot{\beta}_{N\cdot}$ is given by $\check{\beta}_{N\cdot} = -\sum_{i=1}^{N-1} \check{\beta}_{i\cdot}$. In Section S.1 of the Supplementary Material, we show that $X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})X^{(1,2)}$ is invertible asymptotically. Therefore, $\check{B}$ is well defined in large samples.

Remark 3.3. The infeasible estimator in equation (3.1) can be applied to estimate a heterogeneous panel data model that includes both observed time-invariant regressors λ_i^0 and factors f_t^0:

$$y_{it} = (\beta_{i\cdot}^0 + \beta_{\cdot t}^0)' x_{it} + \lambda_i^{0\prime} g_t + \gamma_i' f_t^0 + u_{it},$$

where g_t and γ_i are unknown coefficients.[3] In particular, the heterogeneous panel data model with two-way fixed effects (TWFE) model, as in Lu and Su (2023), arises as a special case. When $\lambda_i^0 = f_t^0 = 1$, the IFE structure simplifies to $\lambda_i^{0\prime} g_t + \gamma_i' f_t^0 = g_t + \gamma_i$, which represents the standard TWFE specification. An important distinction lies in the assumptions regarding the regressors x_{it}. Lu and Su (2023) impose the condition that x_{it}'s behave as idiosyncratic errors after

[3] We sincerely thank the co-editor and an anonymous referee for their helpful comments on this point.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

projecting out the individual and time effects. Our assumptions on x_{it} are more general and less restrictive. Consequently, the theoretical results developed in this section also apply directly to inference on the slope coefficients $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ within the specification described above.

To study the asymptotic properties of the infeasible estimator, we first introduce some notation and impose two regularity assumptions. Let $\dot{\mathbf{x}}_{i\cdot} \equiv \mathbb{M}_{F^0}\mathbf{x}_{i\cdot}$, and let $\dot{x}_{it}$ denote the transpose of the t'th row of $\dot{\mathbf{x}}_{i\cdot}$. Similarly, define $\ddot{\mathbf{x}}_{\cdot t} \equiv \mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t}$ and $\ddot{x}_{it}$. Define

$$\dot{\mathbf{s}}_{i\cdot} \equiv \frac{1}{T}\sum_{t=1}^{T}\dot{x}_{it}\dot{x}_{it}' = \frac{1}{T}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{x}_{i\cdot},\ \ddot{\mathbf{s}}_{\cdot t} \equiv \frac{1}{N}\sum_{i=1}^{N}\ddot{x}_{it}\ddot{x}_{it}' = \frac{1}{N}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t},$$
$$\dot{q}_i \equiv \mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot},\ \text{and}\ \ddot{q}_t \equiv \mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t}. \tag{3.2}$$

Let $\mathbf{Q}_{12}$ be an $(N-1)\times T$ block-partitioned matrix whose (i,t)-th block (of size $K\times K$) is given by $\dot{x}_{it}\ddot{x}_{it}' - \frac{1}{N-1}\sum_{j=1}^{N-1}\dot{x}_{jt}\ddot{x}_{jt}'$, for $i\in[N-1], t\in[T]$. Let $\delta_{NT} = \min(\sqrt{N},\sqrt{T})$.

Assumption 4. (Conditions on Regressors). For some universal positive constants $0<c_1<c_2<\infty$, we have

(i) $\min_i \mu_{\min}(\dot{\mathbf{s}}_{i\cdot}) \geq c_1$, $\min_t \mu_{\min}(\ddot{\mathbf{s}}_{\cdot t}) \geq c_1$, $\max_i \mu_{\max}(\dot{\mathbf{s}}_{i\cdot}) \leq c_2$, and $\max_t \mu_{\max}(\ddot{\mathbf{s}}_{\cdot t}) \leq c_2$ w.p.a.1;

(ii) $\max_i T^{-1}\sum_{t=1}^{T}\|x_{it}\|^\ell = O_P(1)$, $\max_t N^{-1}\sum_{i=1}^{N}\|x_{it}\|^\ell = O_P(1)$ for some $\ell>4$, $\max_{i,t}\|x_{it}\| = O_P[\log(NT)]$, $\max_i\|x_{it}\| = O_P[\log(N)]$, and $\max_t\|x_{it}\| = O_P[\log(T)]$;

(iii) there exists a sequence $\alpha_{NT}\to 0$. $||\mathbf{Q}_{12}||_{\text{sp}}/\sqrt{NT} = O_P(\alpha_{NT})$, $||\mathbb{M}_{F^0}\mathbf{X}_k||_{\text{sp}}/\sqrt{NT} = O_P(\alpha_{NT})$ and $||\mathbf{X}_k\mathbb{M}_{\Lambda^0}||_{\text{sp}}/\sqrt{NT} = O_P(\alpha_{NT})$ for all $k\in[K]$.

Assumption 4 imposes regularity conditions on $\dot{\mathbf{s}}_{i\cdot}$, $\ddot{\mathbf{s}}_{\cdot t}$, x_{it}, f_t^0, and λ_i^0. Assumption 4(i) introduces specific regularity requirements on $\dot{\mathbf{s}}_{i\cdot}$ and $\ddot{\mathbf{s}}_{\cdot t}$, which are constructed from the de-factored and de-factor-loading versions of x_{it}, denoted by $\dot{x}_{it}$ and $\ddot{x}_{it}$, respectively. These conditions can be verified under suitable primitive assumptions, although they do exclude certain cases. Assumption 4 (ii) holds under mild moment and weak dependence conditions on the sequence $\{x_{it}\}$. Assumption 4(iii) is central for ensuring the asymptotic non-singularity of the Gram matrix $X^{(1,2)\prime}(\mathbb{M}_\Lambda\otimes\mathbb{M}_F)X^{(1,2)}$. For discussions on low-rank covariates, see, for example, Moon and Weidner (2017) and Moon and Weidner (2026).

Remark 3.4. Assumption 4(iii) eliminates the possibility of low-rank regressors in x_{it} as in Bai (2009) and Armstrong et al. (2026). In particular, rank-one regressors, including x_{it} that vary only across i or only across t will certainly lead to an identification issue. We present a simple example here. Consider the model in Remark 3.3. Suppose a scalar x_{it} has rank one. Then it can be expressed as $x_{it} = a_i e_t$, where a_i and e_t are both scalars that take value 0 with zero probability. For this case, the model specification suffers from a fundamental identification

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

problem. For any vectors δ_1 and δ_2 of conformable dimensions, we have

$$\dot{\beta}_{\cdot t}x_{it}+\gamma_i'f_t^0=\dot{\beta}_{\cdot t}a_ie_t+\gamma_i'f_t^0=\left(\dot{\beta}_{\cdot t}+\frac{\delta_1'f_t^0}{e_t}\right)a_ie_t+(\gamma_i-a_i\delta_1)'f_t^0=\dot{\beta}_{\cdot t}^*x_{it}+\gamma_i^{*\prime}f_t^0,$$

and similarly,

$$\dot{\beta}_{i\cdot}x_{it}+\lambda_i^{0\prime}g_t=\dot{\beta}_{i\cdot}a_ie_t+\lambda_i^{0\prime}g_t=\left(\dot{\beta}_{i\cdot}+\frac{\delta_2'\lambda_i^0}{a_i}\right)a_ie_t+\lambda_i^{0\prime}(g_t-e_t\delta_2)=\dot{\beta}_{i\cdot}^*x_{it}+\lambda_i^{0\prime}g_t^*,$$

where $\dot{\beta}_{\cdot t}^*\equiv\dot{\beta}_{\cdot t}+\frac{\delta_1'f_t^0}{e_t},\gamma_i^*\equiv\gamma_i-a_i\delta_1,\dot{\beta}_{i\cdot}^*\equiv\dot{\beta}_{i\cdot}+\frac{\delta_2'\lambda_i^0}{a_i}$, and $g_t^*\equiv g_t-e_t\delta_2$. This algebraic equivalence demonstrates that the parameters in the model described in Remark 3.3 are not identified in the presence of rank-one regressors. This issue directly results in singular Gram matrix $X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0}\otimes\mathbb{M}_{F^0})X^{(1,2)}$. Further details are provided in Section S.6 of the Supplementary Material.

Assumption 5. (Conditions on Error Terms).

(i) $\max_t\|\mathbf{u}_{\cdot t}\|=O_P(\sqrt{N})$, $\max_i\|\mathbf{u}_{i\cdot}\|=O_P(\sqrt{T})$, $||\mathbf{U}||_{\text{sp}}=O_P(\sqrt{N+T})$, $||\mathbf{U}'\mathbf{U}||_{sp}=O_P(T)$ and $||\mathbf{U}\mathbf{U}'||_{sp}=O_P(N)$;

(ii) $||\sum_{i=1}^N\sum_{t=1}^T\lambda_i^0f_t^{0\prime}u_{it}||=O_P(\sqrt{NT})$, $\sum_{t=1}^Tx_{it}u_{it}=O_P(\sqrt{T})$, $\sum_{i=1}^Nx_{it}u_{it}=O_P(\sqrt{N})$, $\sum_{t=1}^Tf_t^0u_{it}=O_P(\sqrt{T})$, and $\sum_{i=1}^N\lambda_i^0u_{it}=O_P(\sqrt{N})$;

(iii) $\sum_{i=1}^N||\sum_{t=1}^Tx_{it}u_{it}||^2=O_P(NT)$ and $\sum_{t=1}^T||\sum_{i=1}^Nx_{it}u_{it}||^2=O_P(NT)$;

(iv) $\sum_{i=1}^{N-1}\sum_{t=1}^T\dot{\mathbf{s}}_{i\cdot}^{-1}\dot{x}_{it}u_{it}=O_P(\sqrt{NT})$ and $\sum_{i=1}^N\sum_{t=1}^T\dot{\mathbf{s}}_{i\cdot}^{-1}\dot{x}_{it}u_{it}=O_P(\sqrt{NT})$;

(v) $\max_i\left\|\sum_{t=1}^T\sum_{j=1}^N\varepsilon_{i,jt}\right\|=O_P(\sqrt{NT}\log N)$ for $\varepsilon_{i,jt}=\lambda_j^0u_{jt}u_{it},x_{it}\lambda_j^{0\prime}u_{jt}$, $\dot{x}_{it}\ddot{x}_{it}'\ddot{\mathbf{s}}_{\cdot t}^{-1}\ddot{x}_{jt}u_{jt}$, and $f_t^0\ddot{x}_{it}'\ddot{\mathbf{s}}_{\cdot t}^{-1}\ddot{x}_{jt}u_{jt}$;

(vi) $\max_t\left\|\sum_{s=1}^T\sum_{i=1}^N\varepsilon_{is,t}\right\|=O_P(\sqrt{NT}\log T)$ for $\varepsilon_{is,t}=f_t^0u_{is}u_{it},x_{is}f_s^{0\prime}u_{is}$, $\ddot{x}_{it}\dot{x}_{it}'\dot{\mathbf{s}}_{i\cdot}^{-1}\dot{x}_{is}u_{is}$, and $\lambda_i^0\dot{x}_{it}'\dot{\mathbf{s}}_{i\cdot}^{-1}\dot{x}_{is}u_{is}$;

(vii) $\max_i\left\|\sum_{t=1}^T\varepsilon_{it}\right\|=O_P(\sqrt{T}\log N)$ for $\varepsilon_{it}=x_{it}u_{it}$ and $f_t^0u_{it}$, and $\max_t\left\|\sum_{i=1}^N\varepsilon_{it}\right\|=O_P(\sqrt{N}\log T)$ for $\varepsilon_{it}=x_{it}u_{it}$ and $\lambda_i^0u_{it}$.

Assumption 5 imposes some high-level conditions on terms related to u_{it}'s. Assumption 5(i) is standard in the literature on factor models (see, e.g., Lu and Su, 2016; Moon and Weidner, 2017, 2026; Su and Wang, 2017; Chernozhukov et al., 2026). Moon and Weidner (2017) provide several examples to verify it. The rate conditions in Assumption 5(ii)–(vii) can be verified under some primitive conditions on $\{x_{it},f_t^0,u_{it}\}$ by calling up certain exponential inequalities. Assumption 5(ii) imposes pointwise asymptotic rate conditions on the relevant partial sums. In particular, $\sum_{t=1}^Tx_{it}u_{it}=O_P(\sqrt{T})$ is interpreted for each fixed i, and $\sum_{i=1}^Nx_{it}u_{it}=O_P(\sqrt{N})$ is interpreted for each fixed t; neither condition is intended to hold uniformly over its free index. The aggregate rate conditions required in our analysis are imposed separately in Assumption 5(iii) and (iv), whereas Assumption 5(v)–(vii) provides the required uniform bounds through maxima over the relevant indices.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

The following theorem establishes the asymptotic linear expansions of $\check{\beta}_{i\cdot}$ and $\check{\beta}_{\cdot t}$.

THEOREM 3.3. *Suppose that* $N^{-1/2}\log^2 T + T^{-1/2}\log^2 N \to 0$ *as* $(N,T) \to \infty$, *and Assumptions 3–5 hold. Then we have*

(i) $\check{\beta}_{i\cdot} - \dot{\beta}_{i\cdot} = (\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{x}_{i\cdot})^{-1}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} + o_P(1/\sqrt{T})$ *uniformly over* $i \in [N]$;

(ii) $\check{\beta}_{\cdot t} - \dot{\beta}_{\cdot t} = (\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t})^{-1}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t} + o_P(1/\sqrt{N})$ *uniformly over* $t \in [T]$.

Theorem 3.3 establishes the asymptotic leading term for the infeasible estimators. Under Assumptions 3–5, our theoretical analysis focuses on bounding the higher-order remainders in the expansions of $\check{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}$ and $\check{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}$. For the relative growth of N and T, it suffices that $N^{-1/2}\log^2 T + T^{-1/2}\log^2 N \to 0$, which contrasts with the stronger rate requirement in Theorem 3.4.

3.3. Asymptotic Properties of the Feasible Estimators

The next theorem reports the linear expansions of $\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}$ and $\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}$.

THEOREM 3.4. *Suppose that Assumptions 1–5 hold and* $(\sqrt{N}/T + \sqrt{T}/N)\log^2(NT) \to 0$. *Then we have*

(i) $\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot} = (\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{x}_{i\cdot})^{-1}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} + o_P(1/\sqrt{T})$ *uniformly over* $i \in [N]$;

(ii) $\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t} = (\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t})^{-1}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t} + o_P(1/\sqrt{N})$ *uniformly over* $t \in [T]$.

Theorem 3.4 presents the uniform asymptotic expansions of $\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}$ and $\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}$ under Assumptions 1–5, along with an additional rate condition on (N,T). The required rate condition is stronger than that in Theorem 3.3, reflecting the need to estimate the latent factors rather than treat them as observed. A standard rate condition in the literature on factor models is $(\sqrt{N}/T + \sqrt{T}/N) \to 0$. Given the greater complexity of our model and estimation procedure, we impose the slightly stronger condition $(\sqrt{N}/T + \sqrt{T}/N)\log^2(NT) \to 0$ to satisfy the technical demands of our asymptotic analysis.

We impose the following assumption to enable inference on the slope coefficients, factors, and factor loadings.

Assumption 6 (Conditions for Inference). (i) $E\left(\beta_{i\cdot}^0\right) = 0, \mathrm{Var}\left(\beta_{i\cdot}^0\right) = \Sigma_{1\beta} \geq 0$, and $E\left(u_{it}|x_{it}, \beta_{i\cdot}^0, \beta_{\cdot t}^0, F^0, \Lambda^0\right) = 0$;

(ii) $\max_i \left\|\dot{\mathbf{s}}_{i\cdot} - \mathbf{s}_{i\cdot}^0\right\| = O_P(T^{-1/2}\log N)$ and $\max_t \left\|\ddot{\mathbf{s}}_{\cdot t} - \mathbf{s}_{\cdot t}^0\right\| = O_P(N^{-1/2}\log T)$ for positive definite matrices $\mathbf{s}_{i\cdot}^0$ and $\mathbf{s}_{\cdot t}^0$;

(iii) $\frac{1}{\sqrt{N}}\sum_{i=1}^N \beta_{i\cdot}^0 \xrightarrow{d} N\left(0,\ \Sigma_{1\beta}\right), \frac{1}{\sqrt{T}}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} \xrightarrow{d} N\left(0, \Phi_{i\cdot}\right)$ for each i, and $\frac{1}{\sqrt{N}}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t} \xrightarrow{d} N\left(0, \Phi_{\cdot t}\right)$ for each t, where $\Phi_{i\cdot}$ and $\Phi_{\cdot t}$ are positive definite matrices;

(iv) $\frac{1}{N}\sum_{i=1}^N x_{it}\lambda_i^{0\prime} \xrightarrow{p} \mu_{\cdot t,x\lambda}$ and $\frac{1}{T}\sum_{t=1}^T x_{it}f_t^{0\prime} \xrightarrow{p} \mu_{i\cdot,xf}$ for each i;

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

(v) for each t, $\Xi_{\cdot t,N} \equiv \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [\lambda_i^0 - \mu'_{\cdot t,x\lambda}(\mathbf{s}_{\cdot t}^0)^{-1} \ddot{x}_{it}] u_{it} \xrightarrow{d} N(0, \Gamma_{\cdot t})$ with $\Gamma_{\cdot t} > 0$;

(vi) for each i, $\Xi_{i\cdot,T} \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [f_t^0 - \mu'_{i\cdot,xf}(\mathbf{s}_{i\cdot}^0)^{-1} \dot{x}_{it}] u_{it} \xrightarrow{d} N(0, \Gamma_{i\cdot})$ with $\Gamma_{i\cdot} > 0$.

Assumption 6(i) imposes moment and conditional mean restrictions on $\beta_{i\cdot}^0$ and u_{it}, while Assumption 6(ii) imposes uniform rate conditions on $\dot{\mathbf{s}}_{i\cdot}$ and $\ddot{\mathbf{s}}_{\cdot t}$ defined in (3.2). Assumption 6(iii) imposes central limit theorems for the aggregate slope heterogeneity and the score terms entering the estimators of the unit- and time-specific slope coefficients. Assumption 6(iv)–(vi) imposes probability limits and central limit theorems for terms entering the estimation of the factors and factor loadings. All convergence statements in Assumption 6(iii)–(vi) that contain a free index are pointwise. These conditions do not assert joint convergence over multiple units or time periods. Accordingly, Theorem 3.5 establishes pointwise asymptotic normality for each fixed unit-specific coefficient and each fixed time-specific coefficient, rather than joint or simultaneous inference over i or t.

The following theorem reports the asymptotic distributions of $\hat{\beta}_{i\cdot}$ and $\hat{\beta}_{\cdot t}$.

THEOREM 3.5. *Let* $V_{\beta,i\cdot} \equiv \Sigma_{1\beta}/N + (\mathbf{s}_{i\cdot}^0)^{-1}\Phi_{i\cdot}(\mathbf{s}_{i\cdot}^0)^{-1}/T$ *and* $V_{\beta,\cdot t} \equiv \Sigma_{1\beta}/N + (\mathbf{s}_{\cdot t}^0)^{-1}\Phi_{\cdot t}(\mathbf{s}_{\cdot t}^0)^{-1}/N$. *Suppose that Assumptions 1–6 hold and* $(\sqrt{N}/T + \sqrt{T}/N)\log^2(NT) \to 0$. *Then* $V_{\beta,i\cdot}^{-1/2}(\hat{\beta}_{i\cdot} - \beta_{i\cdot}^0) \xrightarrow{d} N(0, \mathbb{I}_K)$ *for each fixed* $i \in [N]$, *and* $V_{\beta,\cdot t}^{-1/2}(\hat{\beta}_{\cdot t} - \beta_{\cdot t}^0) \xrightarrow{d} N(0, \mathbb{I}_K)$ *for each fixed* $t \in [T]$.

Theorem 3.5 provides pointwise inference only for each fixed i or t; it does not establish joint or simultaneous inference across multiple units or time periods. Such inference, including inference after data-dependent selection, requires additional conditions. For the cell-level slope $\beta_{it}^0 = \beta_{i\cdot}^0 + \beta_{\cdot t}^0$, the natural rate of $\hat{\beta}_{it} - \beta_{it}^0$ is $\min\{\sqrt{N}, \sqrt{T}\}$. If $N/T \to \infty$ (resp. $N/T \to 0$), the unit-specific (resp. time-specific) error is the only leading term, and its marginal limit in Theorem 3.5 determines the limiting distribution. If $N/T \to \kappa \in (0, \infty)$, both terms contribute. Because Theorem 3.5 does not guarantee their joint asymptotic normality, in this case, the expansions in Theorem 3.4 must be supplemented by the additional condition that the two leading terms are asymptotically uncorrelated.

Uniform confidence bands for the collection $\{\beta_{i\cdot}^0 : 1 \le i \le N\}$ may also be of practical interest. However, the uniform asymptotic expansions in Theorem 3.4 and the marginal convergence results in Theorem 3.5 are not sufficient to justify such inference. Uniform inference requires additional conditions and theoretical results, including a uniform Gaussian approximation and the validity of an appropriate bootstrap procedure. Moreover, this problem is more challenging than that in Lu and Su (2023), as the latent factors and factor loadings must be estimated, and their estimation errors need to be controlled uniformly alongside those of the heterogeneous slope coefficients. We therefore leave the development of uniform inference to future research.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Given the results in Theorem 3.5, it is standard to construct pointwise confidence intervals (CIs) for the elements of $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$. To this end, we need to estimate the parameters that appear in the variance matrices $V_{\beta,i\cdot}$ and $V_{\beta,\cdot t}$:

- The quantities $s_{i\cdot}^0$ and $s_{\cdot t}^0$ can be estimated by $\hat{\mathbf{s}}_{i\cdot} \equiv x_{i\cdot}' M_{\hat{F}} x_{i\cdot}/T$ and $\hat{\mathbf{s}}_{\cdot t} \equiv x_{\cdot t}' M_{\hat{\Lambda}} x_{\cdot t}/N$, respectively.
- Under the additional assumption that the $\beta_{i\cdot}^0$'s are independent across i, $\Sigma_{1\beta}$ can be estimated by the sample variance of $\hat{\beta}_{i\cdot}$.
- Under the additional assumption that conditional on $(\Lambda^0, x_{\cdot t}), u_{it}$'s are uncorrelated across i, $\Phi_{\cdot t}$ equals the probability limit of $\frac{1}{N}\sum_{i=1}^N \ddot{x}_{it}\ddot{x}_{it}' u_{it}^2$. An estimator of $\Phi_{\cdot t}$ is given by $\hat{\Phi}_{\cdot t} \equiv \frac{1}{N}\sum_{i=1}^N \widehat{\ddot{x}}_{it}\widehat{\ddot{x}}_{it}' \hat{u}_{it}^2$, where $\widehat{\ddot{x}}_{it} \equiv x_{it} - x_{\cdot t}'\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_i$.
- Under the additional assumption that conditional on $(F^0, x_{i\cdot}), u_{it}$'s are uncorrelated across t, $\Phi_{i\cdot}$ equals the probability limit of $\frac{1}{T}\sum_{t=1}^T \dot{x}_{it}\dot{x}_{it}' u_{it}^2$. An estimator of $\Phi_{i\cdot}$ is given by $\hat{\Phi}_{i\cdot} \equiv \frac{1}{T}\sum_{t=1}^T \widehat{\dot{x}}_{it}\widehat{\dot{x}}_{it}' \hat{u}_{it}^2$, where $\widehat{\dot{x}}_{it} \equiv x_{it} - x_{i\cdot}'\hat{F}(\hat{F}'\hat{F})^{-1}\hat{f}_t$.

  If serial correlation is a concern, $\Phi_{i\cdot}$ represents the long-run variance of $\frac{1}{\sqrt{T}}\sum_{t=1}^T \dot{x}_{it}u_{it}$. In such cases, a heteroskedasticity-and-autocorrelation-consistent (HAC) estimator should be employed to estimate $\Phi_{i\cdot}$.

Remark 3.5. The variance matrices $\mathrm{Var}(\beta_{i\cdot}^0)$ and $\mathrm{Var}(\beta_{\cdot t}^0)$ can be interpreted as measures of heterogeneity in $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$, respectively. In practice, they can be consistently estimated using the sample variances of $\hat{\beta}_{i\cdot}$ and $\hat{\beta}_{\cdot t}$. Under some additional conditions to ensure uniform Gaussian approximation and bootstrap validity, one may consider a formal statistical test of heterogeneity or conduct uniform confidence bands on $\{\beta_{i\cdot}^0\}$ or $\{\beta_{\cdot t}^0\}$. But these certainly go beyond the scope of the article and are left for future research.

Remark 3.6. We now compare the results in Theorem 3.5 with those in Chernozhukov et al. (2026). Since we assume additive structure, $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ can be separately identified under the location normalization $E(\beta_{i\cdot}^0) = 0$. This makes it possible to obtain the linear expansions of $\hat{\beta}_{i\cdot} - \beta_{i\cdot}^0$ and $\hat{\beta}_{\cdot t} - \beta_{\cdot t}^0$ in Theorem 3.4 and to derive the asymptotic distributions of $\sqrt{T}(\hat{\beta}_{i\cdot} - \beta_{i\cdot}^0)$ and $\sqrt{N}(\hat{\beta}_{\cdot t} - \beta_{\cdot t}^0)$ in Theorem 3.5. Given these results, we can also, under additional assumptions, study the average effects by averaging $\beta_{i\cdot}^0$ over a group of individuals or $\beta_{\cdot t}^0$ over a group of time periods. In contrast, Chernozhukov et al. (2026) assume a low-rank structure for β_{it}^0 and focus on the estimation of certain group average effects at time t. For example, one object of interest is $\bar{\beta}_{\mathcal{G},t}^0 = \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \beta_{it}^0$, where $\mathcal{G}$ is a subset of $[N]$ and $|\mathcal{G}|$ denotes the cardinality of the set $\mathcal{G}$. Under the restrictive condition that $|\mathcal{G}| = o(N)$ along with some other restrictive conditions, they derive the asymptotic normality of

$$|\mathcal{G}|^{1/2}\left(\frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}}\hat{\beta}_{it} - \bar{\beta}_{\mathcal{G},t}^0\right).$$

When $\beta_{it}^0 = \beta_{i\cdot}^0 + \beta_{\cdot t}^0$, $\bar{\beta}_{\mathcal{G},t}^0 = \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \beta_{it}^0 = \frac{1}{|\mathcal{G}|}\sum_{i\in\mathcal{G}} \beta_{i\cdot}^0 + \beta_{\cdot t}^0$. In order to identify the time effect $\beta_{\cdot t}^0$, one needs to specify $\mathcal{G}$ to be $[N]$, which is unfortunately ruled

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

out by the condition that $|\mathcal{G}| = o(N)$ in Chernozhukov et al. (2026). So the results in our article and Chernozhukov et al. (2026) are quite different from each other and they are derived under different sets of assumptions. See the next remark.

For the group average $\bar{\beta}^0_{\mathcal{G},t}$, *define* $\widehat{\bar{\beta}}_{\mathcal{G},t} \equiv |\mathcal{G}|^{-1} \sum_{i\in\mathcal{G}} \hat{\beta}_{it}$. *Then*

$$\widehat{\bar{\beta}}_{\mathcal{G},t} - \bar{\beta}^0_{\mathcal{G},t} = \frac{1}{|\mathcal{G}|} \sum_{i\in\mathcal{G}} \left(\hat{\beta}_{i\cdot} - \beta^0_{i\cdot}\right) + \left(\hat{\beta}_{\cdot t} - \beta^0_{\cdot t}\right).$$

Conducting inference on this quantity thus requires an aggregate central limit theorem for the first term, as well as a characterization of its joint limiting behavior with the second. Under the conventional $\sqrt{T|\mathcal{G}|}$ rate for the first term, the relative magnitudes of $T|\mathcal{G}|$ *and* N determine which component dominates asymptotically. Furthermore, if the target of inference is a population—rather than realized—group mean, an additional central limit theorem is needed for the centered mean of the true effects $\beta^0_{i\cdot}$'s. These requirements are not implied by Theorems 3.4 and 3.5; rather, they play a role analogous to that of Assumption 8(vi).

Remark 3.7. The assumptions in our article are generally weaker than those in Chernozhukov et al. (2026), except for our assumption of an additive structure for β^0_{it}. First, Chernozhukov et al. (2026) assume that the regressor x_{it} has a strong factor structure ($x_{it} = l_i' w_t + e_{it}$), where the zero-mean error term e_{it} is serially independent and cross-sectionally weakly dependent. In contrast, we do not explicitly model x_{it} and only require that the projection residual of x_{it}—after partialling out the effect of the factors and loadings in the y-equation—is weakly dependent over i and t. In particular, we allow x_{it} to contain weak factors that are not present in the y-equation. Second, Chernozhukov et al. (2026) assume that (i) the error terms in both the y- and x-equations are independent across t; (ii) the error term in the x-equation is conditionally independent across t given the factors in the y- and x -equations, the factor component in β^0_{it}, and the error term in the y-equation, and (iii) the error term in the y-equation is conditionally independent across t given the factors in the y- and x -equations, the factor component in β^0_{it}, and the error term in the x-equation. Such (conditional) independence conditions across t are not realistic in static panels with large T, especially in empirical research. By contrast, we do not require these independence conditions; we only assume that the error terms are weakly dependent over t. Third, Chernozhukov et al. (2026) assume all loadings in the x- and y -equations and in β^0_{it} are nonrandom and uniformly bounded. They also assume sub-Gaussianity of $\{e_{it}u_{it}\}$. We impose no similar conditions. In particular, we allow the loadings to be random and require only that all variables (including the error terms) have certain finite polynomial-order moments.

To conduct inference on the factors and factor loadings, we first introduce some notation. Define $V = \text{diag}(v_1, \ldots, v_{R^0})$ and $Q = V^{1/2}\Upsilon'\Sigma_\Lambda^{-1/2}$, where $v_1 > v_2 > \cdots > v_{R^0} > 0$ are the eigenvalues of $\Sigma_\Lambda^{1/2}\Sigma_F\Sigma_\Lambda^{1/2}$, and Υ is the corresponding matrix of eigenvectors satisfying $\Upsilon'\Upsilon = \mathbb{I}_{R^0}$. Let $\tilde{\mathbf{Y}} = \mathbf{Y} -$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

$\text{mat}\big(X^{(1,2)}\hat{B}\big)$, and let V_{NT} denote the $R^0 \times R^0$ diagonal matrix containing the largest R^0 eigenvalues of $(NT)^{-1}\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}'$, arranged in decreasing order. Define $H = \big(N^{-1}\Lambda^{0\prime}\Lambda^0\big)\big(T^{-1}F^{0\prime}\hat{F}\big)V_{NT}^{-1}$. The following theorem establishes the asymptotic distributions of $\hat{f}_t$, $\hat{\lambda}_i$, and the product $\hat{\lambda}_i'\hat{f}_t$.

THEOREM 3.6. *Suppose that Assumptions 1–6 hold and* $(\sqrt{N}/T + \sqrt{T}/N)\log^2(NT) \to 0$. *Let* $V_{1,it} = \lambda_i^{0\prime}\Sigma_\Lambda^{-1}\Gamma_{\cdot t}\Sigma_\Lambda^{-1}\lambda_i^0/N$ *and* $V_{2,it} = f_t^{0\prime}\Sigma_F^{-1}\Gamma_{i\cdot}\Sigma_F^{-1}f_t^0/T$. *Then*

(i) $\sqrt{N}(\hat{f}_t - H'f_t^0) \xrightarrow{d} N\big(0, V^{-1}Q\Gamma_{\cdot t}Q'V^{-1}\big)$ *for each* $t \in [T]$;

(ii) $\sqrt{T}(\hat{\lambda}_i - H^{-1}\lambda_i^0) \xrightarrow{d} N\big(0, (Q')^{-1}\Gamma_{i\cdot}Q^{-1}\big)$ *for each* $i \in [N]$;

(iii) $(V_{1,it} + V_{2,it})^{-1/2}(\hat{\lambda}_i'\hat{f}_t - \lambda_i^{0\prime}f_t^0) \xrightarrow{d} N(0,1)$ *for each pair* $(i,t) \in [N] \times [T]$.

Theorem 3.6 provides pointwise distribution results. It does not establish joint or simultaneous inference across multiple time periods or units. In particular, uniform inference on the factor trajectory $\{f_t^0\}_{t\le T}$ under the adopted normalization (equivalently, on the rotated trajectory $\{H'f_t^0\}_{t\le T}$) requires additional conditions and a uniform asymptotic linear expansion for $\hat{f}_t - H'f_t^0$, analogous to the expansions in Theorem 3.4, together with a uniform Gaussian approximation and valid critical values for the corresponding maximum statistic. We leave the development of such uniform inference to future research.

The covariance matrices $\Gamma_{\cdot t}$ and $\Gamma_{\cdot i}$, as in Assumption 6(v) and (vi), are different from that often appear in the factor literature. This point indicates the slow convergences of $\hat{\beta}_{\cdot t}$ and $\hat{\beta}_{i\cdot}$ enter the asymptotics of $\hat{f}_t$ and $\hat{\lambda}_i$, respectively, and both enter the asymptotic distribution of the estimator of the common component $\lambda_i^{0\prime}f_t^0$.

4. ESTIMATION OF HETEROGENEOUS TREATMENT EFFECTS

In this section, we study the estimation of heterogeneous treatment effects by using the model proposed in the last section.

4.1. Basic Setup

One key application of our asymptotic results is to estimate various heterogeneous causal effects. To illustrate the main idea, we consider a simple setup where we have $N = N_1 + N_2$ individuals where the first N_1 individuals ($i \in [N_1]$) never receive any treatment, and the last N_2 individuals ($i \in [N] \setminus [N_1] \equiv \{N_1 + 1, \ldots, N\}$) receive treatment at some potentially different time points. We focus on the case where once a unit adopts the treatment, it stays treated afterward. We denote by $T_{0,i}$ and $T_{1,i}$ the number of control and treated time periods for unit i such that $T_{0,i} + T_{1,i} = T$. If individual i never receives treatment, then $T_{0,i} = T$ and $T_{1,i} = 0$. Following the literature of the potential outcome framework, we use y_{it}^1 and y_{it}^0 to denote the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

potential outcome for individual i who receives and does not receive the treatment at time t, respectively. Let y_{it} be the observed outcome: $y_{it} = y_{it}^1 d_{it} + y_{it}^0 (1 - d_{it})$, where $d_{it} = \mathbf{1}\{\text{individual } i \text{ receives treatment at time } t\}$ is the treatment indicator. Let $\mathcal{D}^0 \equiv \{(i,t) \mid d_{it} = 0, (i,t) \in [N] \times [T]\}$ and $\mathcal{D}^1 \equiv ([N] \times [T]) \backslash \mathcal{D}^0$. Thus, we observe y_{it}^0's for $(i,t) \in \mathcal{D}^0$ and y_{it}^1's for $(i,t) \in \mathcal{D}^1$, respectively.

The individual treatment effect measures the difference between the treated and control outcome:

$$\tau_{it} = y_{it}^1 - y_{it}^0.$$

We do not impose any assumptions on y_{it}^1 and therefore τ_{it}. In particular, y_{it}^1 does not need to follow a model like that for y_{it}^0. τ_{it} can be heterogeneous across both i and t and it does not necessarily have a factor structure as assumed in Xiong and Pelger (2023). We are only interested in the treatment effect on the treated, that is, τ_{it} for $(i,t) \in \mathcal{D}^1$. For them, we directly observe y_{it}^1, and therefore we only need to model y_{it}^0. We assume that y_{it}^0 is generated via the following data-generating process (DGP):

$$y_{it}^0 = \left(\beta_{i\cdot}^0 + \beta_{\cdot t}^0\right)' x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it} \text{ for } i \in [N] \text{ and } t \in [T], \tag{4.1}$$

where x_{it}, $\beta_{i\cdot}^0, \beta_{\cdot t}^0, \lambda_i^0, f_t^0$, and u_{it} are as defined in the last section. As we do not observe all y_{it}^0's for $(i,t) \in [N] \times [T]$, model (4.1) assists in imputing unobserved y_{it}^0's.

Nevertheless, identification of the counterfactual untreated outcomes relies on model (4.1), which is assumed to hold for all (i,t), including treated cells for which y_{it}^0 is counterfactual. In particular, conditional on x_{it}, the time-specific slope $\beta_{\cdot t}^0$ and common factor f_t^0 are shared across all units. Thus, rather than dispensing with a parallel-trends-type restriction, our conditional factor model for untreated outcomes imposes a factor-augmented *stability* condition in the same spirit as the conditional parallel-trends assumption in the staggered difference-in-differences literature.

Let $\mathcal{G}_t \equiv \{i \in [N] \backslash [N_1],\ d_{it} = 1\}$ and $\bar{T}_1 = \frac{1}{N_2} \sum_{i=N_1+1}^{N} T_{1,i}$. Let $|\mathcal{G}_t|$ denote the cardinality of the set $\mathcal{G}_t$, that is, the number of individuals who receive treatment at time t. In the following, we consider different levels of ATT:

1. Average treatment effect on the treated for i: $\tau_{i\cdot} = \frac{1}{T_{1,i}} \sum_{t=T_{0,i}+1}^{T} \tau_{it}$.
2. Average treatment effect on the treated at time t : $\tau_{\cdot t} = \frac{1}{|\mathcal{G}_t|} \sum_{i \in \mathcal{G}_t} \tau_{it}$.
3. Aggregate average treatment effect on the treated: $\bar{\tau} = \frac{1}{N_2 \bar{T}_1} \sum_{i=N_1+1}^{N} \sum_{t=T_{0,i}+1}^{T} \tau_{it}$.

Since we do not make any assumption on τ_{it}, it is impossible to eliminate the effect of u_{it} on the estimator of τ_{it} for a single cross-sectional unit at a time period. This motivates the above consideration of ATT. Of course, one can also consider other forms of treatment effects, say, weighted treatment effects, or treatment effects for certain selected groups of individuals or time periods. For each of the above three types of treatment effects, we derive the asymptotic distributions based on which we can test for the null hypothesis of no causal effects.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Nevertheless, if one would like to assume that u_{it} has the same distribution over either i or t, or both, we argue below that it is still possible to make inference on τ_{it} by relying on some other method.

4.2. Estimation of Heterogeneous Treatment Effects

Due to the presence of the two-dimensional unobserved slope heterogeneity in (4.1), we assume that both N_1 and T pass to infinity. This enables us to apply the asymptotic results in Section 3 to derive the asymptotic properties of certain estimators based on the data $\{(y_{it}, x_{it})\}_{i\in[N_1],t\in[T]}$. Then we can propose an estimator for the counterfactual outcome y_{it}^0 for $i = N_1+1,\ldots,N$ and $t = T_{0,i}+1,\ldots,T$, based on model (4.2) below.

Algorithm 2 (Estimation of the counterfactual outcome y_{it}^0 for $(i,t)\in\mathcal{D}^1$).

---

1. Run the nuclear-norm-regularized estimation based on the model

$$y_{it} = \left(\beta_{i\cdot}^0 + \beta_{\cdot}^0\right)' x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it} \quad \text{for } (i,t)\in[N_1]\times[T]. \tag{4.2}$$

   Obtain the preliminary estimator of the common component Θ, which is now a $T\times N_1$ matrix, based on which we obtain a consistent estimator of R^0. Obtain the post-nuclear norm iterated estimators of $\left\{\beta_{i\cdot}^0, \lambda_i^0\right\}_{i\in[N_1]}$ and $\left\{\beta_{\cdot t}^0, f_t^0\right\}_{t\in[T]}$, denoted as $\{\hat\beta_{i\cdot}, \hat\lambda_i\}_{i\in[N_1]}$ and $\{\hat\beta_{\cdot t}, \hat f_t\}_{t\in[T]}$.
2. Given $\{\hat\beta_{\cdot t}, \hat f_t\}_{t\in[T]}$, we estimate $\left\{\beta_{i\cdot}^0, \lambda_i^0\right\}_{i\in[N]\setminus[N_1]}$ by running a time-series regression for each $i\in[N]\setminus[N_1]$ using the data $\left\{y_{it}, x_{it}'\right\}_{t\in[T_{0,i}]}$:

$$y_{it} - \hat\beta_{\cdot t}' x_{it} = \beta_{i\cdot}^{0\prime} x_{it} + \lambda_i^{0\prime}\hat f_t + e_{it}, \quad \text{for } t\in[T_{0,i}],$$

   where e_{it} signifies the regression error that includes the parameter estimation error (PEE) from the early stage regression. Let $\hat\beta_{i\cdot}$ and $\hat\lambda_i$ denote the resulting estimators of $\beta_{i\cdot}^0$ and λ_i^0, respectively.
3. Estimate the counterfactual outcome y_{it}^0 as $\hat y_{it}^0 = (\hat\beta_{i\cdot} + \hat\beta_{\cdot t})' x_{it} + \hat\lambda_i'\hat f_t$ for $(i,t)\in\mathcal{D}^1$.

Let $\hat\tau_{it} = y_{it} - \hat y_{it}^0$ for $(i,t)\in\mathcal{D}^1$. $\hat\tau_{it}$ is an estimator of τ_{it} that is asymptotically unbiased but generally inconsistent. Then we can estimate $\tau_{i\cdot}$, $\tau_{\cdot t}$, and $\bar\tau$ as follows:

$$\hat\tau_{i\cdot} = \frac{1}{T_{1,i}}\sum_{t=T_{0,i}+1}^{T}\hat\tau_{it},\ \hat\tau_{\cdot t} = \frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t}\hat\tau_{it},\ \text{and } \hat\tau = \frac{1}{N_2\bar T_1}\sum_{i=N_1+1}^{N}\sum_{t=T_{0,i}+1}^{T}\hat\tau_{it}.$$

In the next section, we will study the asymptotic properties of $\hat\tau_{i\cdot}$, $\hat\tau_{\cdot t}$, and $\hat\tau$ under some regularity conditions, based on which one can make valid inferences on $\tau_{i\cdot}$, $\tau_{\cdot t}$, and $\bar\tau$.

For notational simplicity, here we consider the staggered treatment timing. Our approach can actually handle any irregular treatment pattern as long as the untreated period is reasonably long for the treated units. In this case, we simply

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

run a time-series regression using the data from the untreated period in Step 2 of Algorithm 2 and estimate the counterfactual outcome for the treated periods in Step 3. Neither step requires regularly spaced time-series data.

4.3. Asymptotic Distributions

In this section, we study the asymptotic distributions of $\hat{\tau}_{i\cdot}$, $\hat{\tau}_{\cdot t}$, and $\hat{\tau}$. We first state the potential outcome assumptions formally.

Assumption 7. (Conditions for Potential Outcomes). (i) The untreated and treated potential outcomes, y_{it}^0 and y_{it}^1, are given by

$$\begin{cases} y_{it}^0 = \left(\beta_{i\cdot}^0 + \beta_{\cdot t}^0\right)' x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it} \\ y_{it}^1 = y_{it}^0 + \tau_{it}, \end{cases}$$

where $i \in [N]$ and $t \in [T]$.

(ii) $E\left(u_{it} | x_{it}, \beta_{i\cdot}^0, \beta_{\cdot t}^0, F^0, \Lambda^0, \{d_{is}\}_{s \le t}\right) = 0$.

Assumption 7(i) is common in the synthetic control literature. For example, it is similar to Assumption 1 in Ferman and Pinto (2021) and Assumption 1 in Xu (2017) except that our model specification is more general with covariates and we allow for two-dimensional heterogeneity in the slope coefficients. Assumption 7(ii) strengthens the exogeneity Assumption 6(i) by requiring that u_{it} be mean-independent of the treatment history $\{d_{is}\}_{s \le t}$. This is similar to Assumption 3 in Ferman and Pinto (2021) and Assumption 2 in Xu (2017).

We now state some technical assumptions. With a little bit abuse of notation, we continue to use $\hat{\boldsymbol{B}}_{k,T} = \text{diag}(\hat{\beta}_{k,\cdot 1}, \ldots, \hat{\beta}_{k,\cdot T})$ and $\hat{\boldsymbol{B}}_{k,N_1} = \text{diag}(\hat{\beta}_{k,1\cdot}, \ldots, \hat{\beta}_{k,N_1\cdot})$ to denote estimators of $\dot{\boldsymbol{B}}_{k,T} = \text{diag}\,(\dot{\beta}_{k,\cdot 1}, \ldots, \dot{\beta}_{k,\cdot T})$ and $\dot{\boldsymbol{B}}_{k,N_1} = \text{diag}(\dot{\beta}_{k,1\cdot}, \ldots, \dot{\beta}_{k,N_1\cdot})$ based on the data $\{(y_{it}, x_{it})\}_{i \in [N_1], t \in [T]}$. The corresponding matrices $\mathbf{Y}$ (or $\mathbf{X}_k$) and B (or $\hat{B}$) are now of dimensions $T \times N_1$ and $(N_1 + T - 1)K \times 1$, respectively. Let $\tilde{\boldsymbol{Y}} = \mathbf{Y} - \sum_{k=1}^K (\hat{\boldsymbol{B}}_{k,T}\mathbf{X}_k + \mathbf{X}_k\hat{\boldsymbol{B}}_{k,N_1}) = \mathbf{Y} - \text{mat}\,(X^{(1,2)}\hat{B})$. Let V_{N_1T} be the $R^0 \times R^0$ diagonal matrix of the first R^0 largest eigenvalues of $(N_1T)^{-1}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}'$ in decreasing order along its main diagonal line. Let $H = (N_1^{-1}\Lambda^{0\prime}\Lambda^0)(T^{-1}F^{0\prime}\hat{F})V_{N_1T}^{-1}$, where $\Lambda^0 \equiv \Lambda_{N_1}^0 \equiv [\lambda_1^0, \ldots, \lambda_{N_1}^0]'$. Let $z_{it} = \left(x_{it}', f_t^{0\prime}H\right)'$ and $\bar{z}_{it} = \left(x_{it}', f_t^{0\prime}\right)'$. Similarly, $\mathbf{x}_{\cdot t} = [x_{1t}, \ldots, x_{N_1t}]'$ is now an $N_1 \times K$ matrix. Let $\mathbb{S}_1 \in \mathbb{R}^{K \times \left(K+R^0\right)}$ and $\mathbb{S}_2 \in \mathbb{R}^{R^0 \times \left(K+R^0\right)}$ denote selection matrices such that $\mathbb{S}_1 A$ and $\mathbb{S}_2 A$ select the first K rows and the last R^0 rows of A, respectively. Define

$$\mu_{i\cdot,x} = E(x_{it}),\ \mu_f = E\left(f_t^0\right),\ \mu_\lambda = E(\lambda_i^0),$$

$$\sigma_{i\cdot,u}^2 = \lim_{T_{1,i} \to \infty} \frac{1}{T_{1,i}} \sum_{t=T_{0,i}+1}^{T} \sum_{s=T_{0,i}+1}^{T} E\left(u_{it}u_{is}\right),$$

$$S_{i,T_{0,i}} = \frac{1}{T_{0,i}} \sum_{t=1}^{T_{0,i}} z_{it}z_{it}',\ \hat{\pi}_{t,x\lambda} = (\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}\mathbf{x}_{\cdot t},$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

$$S_i = \operatorname*{plim}_{T_{0,i}\to\infty} S_{i,T_{0,i}},$$

$$\pi_{t,x\lambda} = \text{plim}_{N_1\to\infty}\hat{\pi}_{t,x\lambda},\ \alpha_{i\cdot,T} = \frac{T_{1,i}}{\sqrt{T_{0,i}\bar{T}_1}}(\mu'_{i\cdot,x}\mathbb{S}_1 + \mu'_f H\mathbb{S}_2)S^{-1}_{i,T_{0,i}},$$

$$\mu_{\cdot t,x\lambda} = \frac{1}{N_1}\sum_{j=1}^{N_1} E(x_{jt}\lambda_j^{0\prime}),\ \nu_{\cdot t,x\lambda} = \text{plim}_{|\mathcal{G}_t|\to\infty}\frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t}[x_{it} - \mathbf{x}'_{\cdot t}\Lambda^0\left(\Lambda^{0\prime}\Lambda^0\right)^{-1}\lambda_i^0],$$

$$\Omega_{i,zu} = \lim_{T_{0,i}\to\infty}\frac{1}{T_{0,i}}\sum_{t=1}^{T_{0,i}}\sum_{s=1}^{T_{0,i}}\text{blkdiag}\left(\mathbb{I}_K, Q^{-1}\right)E\left[\bar{z}_{it}\bar{z}'_{is}u_{it}u_{is}\right]\text{blkdiag}\left(\mathbb{I}_K,(Q')^{-1}\right),$$

$$\Xi_{\cdot t} = \lim_{N_1\to\infty}\frac{1}{N_1}\sum_{j=1}^{N_1} E\left\{[\nu'_{\cdot t,x\lambda}(\mathbf{s}^0_{\cdot t})^{-1}(x_{jt} - \pi'_{t,x\lambda}\lambda_j^0) - \mu'_\lambda\Sigma_\Lambda^{-1}\lambda_j^0]u_{jt}\right\}^2,$$

$$\Xi = \operatorname*{plim}_{(N,T)\to\infty}\frac{1}{N_2}\sum_{i=N_1+1}^{N}\alpha_{i\cdot,T}\frac{1}{T_{0,i}}\sum_{t=1}^{T_{0,i}}\sum_{s=1}^{T_{0,i}}\text{blkdiag}\left(\mathbb{I}_K, Q^{-1}\right)$$
$$\times E\left[\bar{z}_{it}\bar{z}'_{is}u_{it}u_{is}\right]\text{blkdiag}\left(\mathbb{I}_K,(Q')^{-1}\right)\alpha'_{i\cdot,T},$$

where blkdiag $(\cdot)$ signifies a block diagonal matrix. Let $\underline{T}_0 = \min_{i\in[N]\setminus[N_1]} T_{0,i}$ and $\bar{T}_1 = \frac{1}{N_2}\sum_{i=N_1+1}^{N} T_{1,i}$. We add the following assumption.

Assumption 8.

(i) $\max_{i\in[N]\setminus[N_1]}\{\frac{1}{T_{0,i}}E||\sum_{t=1}^{T_{0,i}}(x_{it} - \mu_{i\cdot,x})||^2 + \frac{1}{T_{1,i}}E||\sum_{t=T_{0,i}+1}^{T}(x_{it} - \mu_{i\cdot,x})||^2 + \frac{1}{T_{1,i}}||\sum_{t=T_{0,i}+1}^{T}(f_t^0 - \mu_f)||^2 + \frac{1}{N_2\bar{T}_1}||\sum_{j,k=N_1+1}^{N}\sum_{t,s=T_{0,i}+1}^{T}E(\lambda_j^0\lambda_k^{0\prime}u_{jt}u_{ks})||\} \le C$;

(ii) $\max_{t\in[T]}\{||\frac{1}{N_1}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t} - \mathbf{s}^0_{\cdot t}|| + ||\frac{1}{N_1}\sum_{j=1}^{N_1}x_{jt}\lambda_j^{0\prime} - \mu_{\cdot t,x\lambda}|| + \left\|\hat{\pi}_{t,x\lambda} - \pi_{t,x\lambda}\right\|\} = O_P(N_1^{-1/2}\log T)$;

(iii) $\max_{i\in[N]\setminus[N_1]}||\frac{1}{T_{0,i}}\sum_{t=1}^{T_{0,i}}z_{it}z'_{it} - S_i|| = O_P[(\underline{T}_0/\log N)^{-1/2}]$ for some $S_i > 0$;

(iv) for each i, $\frac{1}{\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T}u_{it} \xrightarrow{d} N(0,\sigma^2_{i\cdot,u})$ and $\frac{1}{\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T}z_{it}u_{it} \xrightarrow{d} N(0,\Omega_{i,zu})$;

(v) for each $t \in [T]\setminus\left[\min_{i\in[N]}T_{0,i}\right]$, as $|\mathcal{G}_t|\to\infty$, $\frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}u_{it} \xrightarrow{d} N\left(0,\sigma^2_{\cdot t,u}\right)$ and $\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}[\nu'_{\cdot t,x\lambda}\left(\mathbf{s}^0_{\cdot t}\right)^{-1}(x_{jt} - \pi'_{t,x\lambda}\lambda_j^0) - \mu'_\lambda\Sigma_\Lambda^{-1}\lambda_j^0]u_{jt} \xrightarrow{d} N(0,\Xi_{\cdot t})$;

(vi) $\frac{1}{\sqrt{N_2\bar{T}_1}}\sum_{i=N_1+1}^{N}\sum_{t=T_{0,i}+1}^{T}u_{it} \xrightarrow{d} N\left(0,\sigma_u^2\right)$ and $\frac{1}{\sqrt{N_2}}\sum_{i=N_1+1}^{N}\alpha_{i\cdot,T}\frac{1}{\sqrt{T_{0,i}}}\sum_{t=1}^{T_{0,i}}z_{it}u_{it} \xrightarrow{d} N\left(0,\Xi\right)$;

(vii) $\max_{i\in[N]}\{||(N_1T)^{-1/2}\sum_{t=1}^{T}\sum_{j=1}^{N_1}\lambda_j^0u_{jt}x'_{it}|| + ||(N_1T)^{-1/2}\sum_{t=1}^{T}\zeta'_{it}(\mathbf{s}^0_{\cdot t})^{-1}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t}||\} = O_P(\log(N_1T))$ for $\zeta_{it} = \nu_{\cdot t,x\lambda}, \mu_{\cdot t,x\lambda}x_{k,it}$, and $x_{it}x_{k,it}$ with $k\in[K]$.

Assumption 8 imposes some conditions of $\left\{x_{it}, u_{it}, \lambda_i^0, f_t^0\right\}$ that can be verified under some primitive conditions. Assumption 8 contains three types of asymptotic

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

conditions. First, parts (iv) and (v) impose pointwise convergence conditions. These conditions support parts (i) and (ii), respectively, of Theorem 4.1. Second, part (vi) imposes aggregate convergence conditions jointly over treated units and post-treatment periods and is needed only for the overall average estimator in Theorem 4.1(iii). Third, parts (i)–(iii) and (vii) impose uniform bounds over the relevant unit and time indices. In establishing Theorem 4.1(iii), the aggregate limits in part (vi), together with these uniform bounds, control the accumulation of the leading terms and estimation remainders when averaging over both dimensions.

The following theorem states the asymptotic distributions of $\hat{\tau}_{i\cdot}$, $\hat{\tau}_{\cdot t}$, and $\hat{\tau}$.

THEOREM 4.1. *Suppose Assumptions 1–8 hold for* $\{(y_{it}, x_{it})\}_{i\in[N_1],t\in[T]}$ *with obvious changes to take into account the change in the sample size. Suppose that* $(\sqrt{N_1}/T + \sqrt{T}/N_1)\log^2(N_1T) \to 0$ *and* $\frac{T_{0,i}}{T} \to r_i \in (0,1]$ *as* $(N_1, T) \to \infty$. *Then*

(i) *For each fixed treated unit i, as* $T_{1,i} \to \infty$,

$$\sqrt{T_{1,i}}\left(\hat{\tau}_{i\cdot} - \tau_{i\cdot}\right) \xrightarrow{d} N\left(0, \sigma^2_{i\cdot,u} + \frac{1-r_i}{r_i}\Xi_{i\cdot}\right),$$

where $\Xi_{i\cdot} \equiv (\mu'_{i\cdot,x}\mathbb{S}_1 + \mu'_f Q^{-1}\mathbb{S}_2)S_i^{-1}\Omega_{i,zu}S_i^{-1}(\mathbb{S}'_1\mu_{i\cdot,x} + \mathbb{S}'_2(Q^{-1})'\mu_f)$.

(ii) *For each fixed time period* $t \in [T]\setminus\left[\min_{i\in[N]} T_{0,i}\right]$, *as* $|\mathcal{G}_t| \to \infty$ *and* $\frac{|\mathcal{G}_t|}{N_1} \to r_{\cdot t} \in [0,\infty)$,

$$\sqrt{|\mathcal{G}_t|}\left(\hat{\tau}_{\cdot t} - \tau_{\cdot t}\right) \xrightarrow{d} N\left(0, \sigma^2_{\cdot t,u} + r_{\cdot t}\Xi_{\cdot t}\right).$$

(iii) *As* $\bar{T}_1 \to \infty$ *and* N_2 *is either fixed or diverges to infinity such that* $N_2 = o(\min(N_1, \underline{T}_0/\log N, \delta^4_{N_1T}/\{\bar{T}_1[\log(N_1T)]^2\}))$,

$$\sqrt{N_2\bar{T}_1}\left(\hat{\tau} - \bar{\tau}\right) \xrightarrow{d} N\left(0, \sigma_u^2 + \Xi\right).$$

Theorem 4.1 studies the asymptotic distributions of $\hat{\tau}_{i\cdot}$, $\hat{\tau}_{\cdot t}$, and $\hat{\tau}$. To make inferences, we need to consistently estimate the asymptotic variances. Let $\hat{z}_{it} = (x'_{it}, \hat{f}'_t)'$ for $(i,t) \in [N]\times[T]$ and $\hat{u}_{it} = y_{it} - x'_{it}(\hat{\beta}_{i\cdot} + \hat{\beta}_{\cdot t}) - \hat{\lambda}'_i\hat{f}_t$ for $(i,t) \in \mathcal{D}^0$.

Case (1). Since $\sigma^2_{i\cdot,u}$ and $\Omega_{i,zu}$ denote the long-run variance of $\{u_{it}\}$ and $\{z_{it}u_{it}\}$, it is standard to use $\{\hat{u}_{it}\}$ and $\{\hat{z}_{it}\hat{u}_{it}\}$ to estimate them by HAC estimators. The main difficulty here lies in the fact that the above definition of $\hat{u}_{it}$ only works for the pre-treatment period when individual i receives the treatment in period t. When $(i,t) \in \mathcal{D}^1$, we cannot consistently estimate u_{it}. Under the covariance stationarity of $\{(u_{it}, \bar{z}_{it}u_{it}), t \geq 1\}$, we can estimate $\sigma^2_{i\cdot,u}$ and $\Omega_{i,zu}$, respectively, by

$$\hat{\sigma}^2_{i\cdot,u} = \frac{1}{T_{0,i}}\left[\sum_{t=1}^{T_{0,i}}\hat{u}^2_{it} + 2\sum_{r=1}^{T_{0,i}-1} w(r/m)\sum_{t=r+1}^{T_{0,i}}\hat{u}_{it}\hat{u}_{i,t-r}\right] \text{ and}$$

$$\hat{\Omega}_{i,zu} = \frac{1}{T_{0,i}}\sum_{t=1}^{T_{0,i}}\hat{z}_{it}\hat{z}'_{it}\hat{u}^2_{it} + \sum_{r=1}^{T_{0,i}-1} w(r/m)\sum_{t=r+1}^{T_{0,i}}\left(\hat{z}_{it}\hat{z}'_{i,t-r} + \hat{z}_{i,t-r}\hat{z}'_{it}\right)\hat{u}_{it}\hat{u}_{i,t-r},$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

where $w(\cdot)$ is a kernel function and $m = m_T$ is a truncation parameter. Of course, if $\{(u_{it}, \bar{z}_{it}u_{it})\}$ share the same variance–covariance matrix across $i \in [N]\backslash[N_1]$, one can pool over the treated cross-sectional units to estimate the above variances more accurately in finite samples. Then, we can estimate the asymptotic variance of $\sqrt{T_{1,i}}\left(\hat{\tau}_{i\cdot} - \tau_{i\cdot}\right)$ by

$$\widehat{\text{avar}}(\sqrt{T_{1,i}}\hat{\tau}_{i\cdot}) = \hat{\sigma}^2_{i\cdot,u} + \frac{T_{1,i}}{T_{0,i}}(\hat{\mu}'_{i\cdot,x}\mathbb{S}_1 + \hat{\mu}'_f\mathbb{S}_2)\hat{S}^{-1}_{i,T_{0,i}}\hat{\Omega}_{i,zu}\hat{S}^{-1}_{i,T_{0,i}}(\mathbb{S}'_1\hat{\mu}_{i\cdot,x} + \mathbb{S}'_2\hat{\mu}_f),$$

where $\hat{\mu}_{i\cdot,x} = \frac{1}{T_{1,i}}\sum_{t=T_{0,i}+1}^{T} x_{it}$, $\hat{\mu}_f = \frac{1}{T}\sum_{t=1}^{T}\hat{f}_t$, and $\hat{S}_{i,T_{0,i}} = \frac{1}{T_{0,i}}\sum_{t=1}^{T_{0,i}}\hat{z}_{it}\hat{z}'_{it}$. Alternatively, under the covariance stationarity of $\{\bar{z}_{it}, t \geq 1\}$, we can also use the following estimators $\tilde{\mu}_{i\cdot,x} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ and $\tilde{S}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{z}_{it}\hat{z}'_{it}$ to replace $\hat{\mu}_{i\cdot,x}$ and $\hat{S}_{i,T_{0,i}}$, respectively.

Case (2). Under the condition that $\left\{u_{it}, x_{it}, \lambda^0_i\right\}$ are independent across i, we can estimate $\sigma^2_{\cdot t,u}$ and $\Xi_{\cdot t}$, respectively, by

$$\hat{\sigma}^2_{\cdot t,u} = \frac{1}{|[N]\backslash\mathcal{G}_t|}\sum_{i\in[N]\backslash\mathcal{G}_t}\hat{u}^2_{it} \text{ and}$$

$$\hat{\Xi}_{\cdot t} = \frac{1}{N_1}\sum_{j=1}^{N_1}[\hat{\nu}'_{\cdot t,x\lambda}\left(\hat{\boldsymbol{s}}_{\cdot t}\right)^{-1}(x_{jt} - \mathbf{x}'_{\cdot t}\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_j) - \hat{\mu}'_\lambda(N_1^{-1}\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_j]^2\hat{u}^2_{jt},$$

where $\hat{\boldsymbol{s}}_{\cdot t} = \frac{1}{N_1}\mathbf{x}'_{\cdot t}\mathbb{M}_{\hat{\Lambda}}\mathbf{x}_{\cdot t}, \hat{\mu}_\lambda = \frac{1}{N_1}\sum_{j=1}^{N_1}\hat{\lambda}_j$, and $\hat{\nu}_{\cdot t,x\lambda} = \frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t}[x_{it} - \mathbf{x}'_{\cdot t}\hat{\Lambda}(\hat{\Lambda}'\hat{\Lambda})^{-1}\hat{\lambda}_i]$. Then we can estimate the asymptotic variance of $\sqrt{|\mathcal{G}_t|}\left(\hat{\tau}_{\cdot t} - \tau_{\cdot t}\right)$ by

$$\widehat{\text{avar}}(\sqrt{|\mathcal{G}_t|}\hat{\tau}_{\cdot t}) = \hat{\sigma}^2_{\cdot t,u} + \frac{|\mathcal{G}_t|}{N_1}\hat{\Xi}_{\cdot t}.$$

Case (3). We can estimate σ^2_u and Ξ by the panel HAC estimators $\hat{\sigma}^2_u$ and $\hat{\Xi}$, respectively, where

$$\hat{\sigma}^2_u = \frac{1}{N_2\bar{T}_0}\sum_{i=N_1+1}^{N} T_{0,i}\hat{\sigma}^2_{i\cdot,u},\ \hat{\Xi} = \frac{1}{N_2}\sum_{i=N_1+1}^{N}\hat{\alpha}_{i\cdot,T}\hat{\Omega}_{i,zu}\hat{\alpha}'_{i\cdot,T},$$

$$\bar{T}_0 = \frac{1}{N_2}\sum_{i=N_1+1}^{N} T_{0,i}, \text{ and } \hat{\alpha}_{i\cdot,T} = \frac{T_{1,i}}{\sqrt{T_{0,i}\bar{T}_1}}(\hat{\mu}'_{i\cdot,x}\mathbb{S}_1 + \hat{\mu}'_f\mathbb{S}_2)\hat{S}^{-1}_{i,T_{0,i}}.$$

As in Case (1), here we use the covariance stationarity of $\{(u_{it}, \bar{z}_{it}u_{it}), t \geq 1\}$ to obtain $\hat{\sigma}^2_u$. Then we can estimate the asymptotic variance of $\sqrt{N_2\bar{T}_1}\left(\hat{\tau} - \bar{\tau}\right)$ by

$$\widehat{\text{avar}}(\sqrt{N_2\bar{T}_1}\hat{\tau}) = \hat{\sigma}^2_u + \hat{\Xi}.$$

Despite the fact that $\hat{\tau}_{it}$ is generally inconsistent, we can still construct valid CIs for τ_{it} under certain stationarity conditions on $\{u_{it}\}$. This is because $\hat{\tau}_{it} = \tau_{it} +$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

$u_{it} + o_P(1)$ by Lemma A.7(i) in Section A.2 of the Supplementary Material. For example, if $\{u_{it}\}$ for the treated units are strictly stationary over both i and t, we can establish the $100(1-\alpha)\%$ CI for τ_{it} as $\left[\hat{\tau}_{it} + \hat{q}_{u,\alpha/2}, \hat{\tau}_{it} + \hat{q}_{u,1-\alpha/2}\right]$, where, for example, $\hat{q}_{u,\alpha/2}$ denotes the $\alpha/2$-sample quantile of $\{-\hat{u}_{it}, i \in [N]\setminus[N_1], t \in \left[T_{0,i}\right]\}$. Similarly, if we further impose strict stationarity for $\{u_{it}\}$ for both the control units and treated units, we can use the sample quantiles of the combination of $\{-\hat{u}_{it}, i \in [N]\setminus[N_1], t \in \left[T_{0,i}\right]\}$ and $\{-\hat{u}_{it}, i \in [N_1], t \in [T]\}$. Alternatively, we can apply the permutation test proposed by Chernozhukov et al. (2021) to conduct inference for τ_{it}.

5. MONTE CARLO SIMULATIONS

5.1. Data-Generating Processes and Implementations

To examine the finite sample performance of our estimation method, we conduct simulations with four DGPs. For all DGPs, the untreated outcome is generated as in models (1.1) and (1.2), and the treatment effect (TE), τ_{it} is randomly drawn from the $N(1,1)$ distribution.

We first consider simple DGPs 1 and 2 to illustrate the main idea of our approach. In both DGPs, x_{it} is a scalar, the number of factors (R^0) is 2, x_{it} and u_{it} are all generated from independent and identically distributed (IID) $N(0,1)$, and λ_i^0 and f_t^0 are drawn from IID $N(0, \mathbb{I}_2)$. We consider two different cases for $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$. In DGP 1 (heterogeneous case), $\beta_{i\cdot}^0$ and $\beta_{\cdot t}^0$ are IID $N(0,1)$ and $N(1,1)$ random variables. In DGP 2 (homogeneous case), $\beta_{i\cdot}^0 = 0$ and $\beta_{\cdot t}^0 = 1$ for all i and t. Note that for DGP 2, our proposed model is over-fitted, therefore we can examine the effect of over-fitting by comparing it with the traditional homogeneous model where $(\beta_{i\cdot}^0 + \beta_{\cdot t}^0)$ is restricted to be a constant.

DGPs 3 and 4 are more realistic. In both DGPs, there are two covariates $x_{it} = \left(x_{1,it}, x_{2,it}\right)'$ and two factors $f_t^0 = \left(f_{t,1}^0, f_{t,2}^0\right)'$. Correspondingly, define $\lambda_i^0 = \left(\lambda_{i,1}^0, \lambda_{i,2}^0\right)', \beta_{i\cdot}^0 = \left(\beta_{1,i\cdot}^0, \beta_{2,i\cdot}^0\right)'$, and $\beta_{\cdot t}^0 = \left(\beta_{1,\cdot t}^0, \beta_{2,\cdot t}^0\right)'$. The details of the specifications are as follows:

$$f_{t,1}^0 = \rho f_{t-1,1}^0 + \sqrt{1-\rho^2}v_{t,1}, \quad f_{t,2}^0 = \rho f_{t-1,2}^0 + \sqrt{1-\rho^2}v_{t,2},$$
$$x_{1,it} = 0.25 + 0.25\lambda_i^{0\prime} f_t^0 + \eta_{it}, \quad \eta_{it} = \rho\eta_{i,t-1} + \sqrt{1-\rho^2}v_{it,3}, \quad x_{2,it} = 0.5v_{it,4},$$
$$u_{it} = \sqrt{1 + 0.1x_{1,it}^2}\varepsilon_{it}, \quad \text{and } \varepsilon_{it} = \rho\varepsilon_{i,t-1} + \sqrt{1-\rho^2}v_{it,5},$$

$v_{t,1}, v_{t,2}, v_{it,3}, v_{it,4}, \lambda_{i,1}^0$, and $\lambda_{i,2}^0$ are drawn from IID $N(0, \mathbb{I}_6)$ and $v_{it,5}$ is IID random variable from the distribution of $\chi^2(1) - 1$. $\beta_{1,i\cdot}^0$ and $\beta_{1,\cdot t}^0$ are IID $N(0,1)$ and $N(1,1)$ random variables, and $\beta_{2,i\cdot}^0 = 0$ and $\beta_{2,\cdot t}^0 = 1$ for all i and t. Therefore, the slope coefficient of $x_{1,it}$ is heterogeneous and that of $x_{2,it}$ is homogeneous. The difference between DGPs 3 and 4 lies in ρ. In DGP 3, $\rho = 0$, therefore, there is no serial correlation in $x_{1,it}, f_{t,1}^0, f_{t,2}^0$, and u_{it}. In DGP 4, $\rho = 1/3$, so $x_{1,it}, f_{t,1}^0, f_{t,2}^0$, and u_{it} are all serially correlated. Note that DGP 4 is the most complicated where the error term is conditionally heteroskedastic and serially correlated.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

For DGPs 1–3, we use the White-type estimator for the variance terms to conduct inference. To improve the finite-sample performance, we impose the condition that the treated data are stationary along both cross-sectional and time dimensions. For DGP 4, we use an HAC-type estimator for the variance. It is well known that the performance of the standard HAC estimator can be poor when the time-series dimension is short (e.g., see the survey paper by Lazarus et al., 2018). To partially address this issue, we impose the stationarity of both treated and control data, and also apply a pre-whitening (PW) procedure as in Andrews and Monahan (1992) and Lu and Su (2023). We use the growth ratio method of Ahn and Horenstein (2013) to determine the number of factors in the initial nuclear norm regularized estimation procedure, as it outperforms singular value thresholding in our simulations and does not depend on any tuning parameter other than the maximum number of factors allowed. We set the tuning parameter as $\gamma_{NT} = 0.5\max(N_1^{-1/2}, T^{-1/2})$, following Miao et al. (2023). We consider different combinations of N_1 and T: $(N_1, T) = (40, 40)$, $(40, 80)$, $(80, 40)$, $(80, 80)$, and $(120, 120)$. We let $N_2 = 0.25N_1$ and $T_{0i} = T_{1i} = 0.5T$. The number of replications is 250.

5.2. Simulation Results

We first report the performance of our key Step 1 in Algorithm 2, where we use the data of control units to estimate $\{\beta_i^0, \lambda_i^0\}_{i\in[N_1]}$ and $\{\beta_t^0, f_t^0\}_{t\in[T]}$. We define the true slope coefficient as $\beta_{it}^0 = \beta_{i\cdot}^0 + \beta_{\cdot t}^0$ and the true factor component as $\theta_{it}^0 = \lambda_i^{0\prime} f_t^0$. Correspondingly, define their estimators as $\hat{\beta}_{it} = \hat{\beta}_{i\cdot} + \hat{\beta}_{\cdot t}$ and $\hat{\theta}_{it} = \hat{\lambda}_i' \hat{f}_t$, respectively. For ease of reporting, we only present the results for the slope coefficient of the first covariates for DGPs 3 and 4, where there are two covariates. Table 1 reports the mean squared errors (MSEs) of $\hat{\beta}_{1,it}$ (the first element of $\hat{\beta}_{it}$) and the correlation between $\hat{\beta}_{1,it}$ and $\beta_{1,it}^0$.[4] Specifically, we define

$$\text{MSE}_\beta = \frac{1}{N_1 T}\sum_{i=1}^{N_1}\sum_{t=1}^{T}\left(\hat{\beta}_{1,it} - \beta_{1,it}^0\right)^2 \text{ and } \rho_\beta = \widehat{\text{corr}}\left(\hat{\beta}_{1,it}, \beta_{1,it}^0\right),$$

and average both measures over 250 replications. Table 1 also reports the MSEs of $\hat{\theta}_{it}$ and the correlation between $\hat{\theta}_{it}$ and θ_{it}^0, which are defined as

$$\text{MSE}_\theta = \frac{1}{N_1 T}\sum_{i=1}^{N_1}\sum_{t=1}^{T}\left(\hat{\theta}_{it} - \theta_{it}^0\right)^2 \text{ and } \rho_\theta = \widehat{\text{corr}}\left(\hat{\theta}_{it}, \theta_{it}^0\right),$$

averaging over 250 replications. Table 1 shows that our estimation procedure in Step 1 works well. The correlations between $\hat{\beta}_{1,it}$ and $\beta_{1,it}^0$ and between $\hat{\theta}_{it}$ and θ_{it}^0 are all above 0.97 for DGPs 1 and 2, and all above 0.85 for DGPs 3 and 4. The MSEs of $\hat{\beta}_{1,it}$ and $\hat{\theta}_{it}$ decrease with N_1 and T, as expected. Our method can

[4] For brevity, we omit more detailed results, such as bias and variance, which are available upon request.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

TABLE 1. Performance of Step 1 in Algorithm 2

A: Performance of $\hat{\beta}_{1,it}$								
	DGP 1		DGP 2		DGP 3		DGP 4	
(N_1, T)	MSE_β $\times 100$	Corr: ρ_β	MSE_β $\times 100$	Corr: ρ_β	MSE_β $\times 100$	Corr: ρ_β	MSE_β $\times 100$	Corr: ρ_β
(40, 40)	6.02	0.985	6.02	–	16.33	0.962	17.53	0.959
(40, 80)	4.34	0.989	4.49	–	10.85	0.974	11.71	0.972
(80, 40)	4.42	0.989	4.39	–	10.73	0.974	12.20	0.971
(80, 80)	2.77	0.993	2.73	–	6.92	0.983	7.74	0.981
(120, 120)	1.77	0.996	1.79	–	4.42	0.989	4.98	0.988
B: Performance of $\hat{\theta}_{it}$								
	DGP 1		DGP 2		DGP 3		DGP 4	
(N_1, T)	MSE_θ $\times 100$	Corr: ρ_θ	MSE_θ $\times 100$	Corr: ρ_θ	MSE_θ $\times 100$	Corr: ρ_θ	MSE_θ $\times 100$	Corr: ρ_θ
(40, 40)	10.78	0.975	10.87	0.974	33.12	0.924	37.64	0.914
(40, 80)	7.89	0.981	7.89	0.981	21.13	0.950	22.87	0.946
(80, 40)	7.87	0.982	7.91	0.981	21.20	0.951	24.57	0.944
(80, 80)	5.18	0.988	5.14	0.988	13.04	0.968	14.70	0.965
(120, 120)	3.39	0.992	3.40	0.992	8.41	0.980	9.47	0.977

Note: Corr stands for correlation. For the homogeneous DGP 2, ρ_β (the correlation between $\hat{\beta}_{it}$ and β_{it}^0) is not defined.

determine the correct number of factors with a high probability. For simple DGPs 1 and 2, we determine the number of factors with a correct rate of 100% for all sample sizes considered. For more complicated DGPs 3 and 4, the correct rates are all above 90% and close to 100% when the sample size is large.

We next examine the performance of our estimator of various treatment effects, namely, $\hat{\tau}_{it}$, $\hat{\tau}_{i\cdot}$, $\hat{\tau}_{\cdot t}$, and $\hat{\tau}$. Table 2 reports their MSEs, which are defined as

$$\text{MSE}_{\hat{\tau}_{it}} = \frac{1}{N_2 T_1} \sum_{i=N_1+1}^{N} \sum_{t=T_{0,i}+1}^{T} \left(\hat{\tau}_{it} - \tau_{it}\right)^2, \ \text{MSE}_{\hat{\tau}_{i\cdot}} = \frac{1}{N_2} \sum_{i=N_1+1}^{N} \left(\hat{\tau}_{i\cdot} - \tau_{i\cdot}\right)^2,$$

$$\text{MSE}_{\hat{\tau}_{\cdot t}} = \frac{1}{\bar{T}_1} \sum_{t=T_0+1}^{T} \left(\hat{\tau}_{\cdot t} - \tau_{\cdot t}\right)^2 \text{ and } \text{MSE}_{\hat{\tau}} = \left(\hat{\tau} - \bar{\tau}\right)^2,$$

averaging over 250 replications. Note that $\hat{\tau}_{it}$ is not a consistent estimator of τ_{it}, and $\text{MSE}_{\hat{\tau}_{it}}$ converges to $\text{Var}(u_{it})$ in probability. $\text{MSE}_{\hat{\tau}_{i\cdot}}$ decreases at the rate $T_{1,i}^{-1}$, $\text{MSE}_{\hat{\tau}_{\cdot t}}$ decreases at the rate $|\mathcal{G}_t|^{-1}$, and $\text{MSE}_{\hat{\tau}}$ decreases at the rate of $(N_2 \bar{T}_1)^{-1}$. Our simulation results confirm all these theoretical results. We also compare the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

four MSE measures between our heterogeneous (hetero) and the conventional homogeneous (homo) models which assume that the slope coefficient $\beta_{i\cdot}^0 + \beta_{\cdot t}^0$ is a constant. Consider the simple DGPs 1 and 2 first. Note that the homo model is misspecified for DGP 1 and correctly specified for DGP 2. For DGP 1, our hetero model performs much better than the homo model, as expected. For example, MSE $_{\hat{\tau}_{it}}$ based on our hetero model and that based on the homo models are 1.08 and 3.17, respectively, when $(N_1, T) = (120, 120)$. The large difference is easy to understand. For DGP 1, the homo model effectively estimates

$$y_{it} = \beta_0^{0\prime} x_{it} + \lambda_i^{0\prime} f_t^0 + u_{it}^o, \text{ where } u_{it}^o = (\beta_{i\cdot}^0 + \beta_{\cdot t}^0 - \beta_0^0)' x_{it} + u_{it} \text{ and } \beta_0^0 = E(\beta_{\cdot t}^0).$$

That is, the homo model treats $\left(\beta_{i\cdot}^0 + \beta_{\cdot t}^0 - \beta_0^0\right)' x_{it}$ as part of the unpredictable error, which is actually predictable. In contrast, our hetero model can capture all of the predictable components. Note that for this DGP, Var $(u_{it}) = 1$ and Var $\left(u_{it}^o\right) = 3$, which explains the large difference in MSE $_{\hat{\tau}_{it}}$. For the same reason, MSE $_{\hat{\tau}_{i\cdot}}$, MSE $_{\hat{\tau}_{\cdot t}}$, and MSE $_{\hat{\tau}}$ based on our hetero models are much smaller than those based on the homo models. For DGP 2, where the true DGP is homogeneous, we find that the performance of our hetero model is similar to that of the homo model especially when the sample size is large, despite the fact that our hetero model is vastly over-fitted. For example, MSE $_{\hat{\tau}_{it}}$ based on our hetero model and the homo model are 1.08 and 1.05, respectively, when $(N_1, T) = (120, 120)$. This comparison in the simulations suggests that even when the true DGP is homogeneous (DGP 2), the estimation efficiency loss based on the overfitted hetero model is minimal. Conversely, when the true DGP is heterogeneous (DGP 1), our hetero model significantly outperforms the homo model. For DGPs 3 and 4, the performance of our hetero model is much better than the homo model. We also compare with the SCM and find that our hetero model outperforms SCM substantially for all DGPs and all sample sizes.[5] In our setup, SCM can only predict the factor components. Consider the simple DGPs 1 and 2 where the covariates x_{it} is cross-sectionally independent. SCM cannot predict $\left(\beta_{i\cdot}^0 + \beta_{\cdot t}^0\right)' x_{it}$ by taking the average of untreated outcomes. Therefore, SCM is essentially based on the model

$$y_{it} = c_0 + \lambda_i^{0\prime} f_t^0 + u_{it}^\dagger, \text{ where } u_{it}^\dagger = [(\beta_{i\cdot}^0 + \beta_{\cdot t}^0)' x_{it} - c_0] + u_{it} \text{ and}$$
$$c_0 = E[(\beta_{i\cdot}^0 + \beta_{\cdot t}^0)' x_{it}],$$

and treats $u_{it}^\dagger$ as an unpredictive error. MSE $_{\hat{\tau}_{it}}$ based on SCM is at least as large as Var $(u_{it}^\dagger)$, which is equals to 4 and 2 for DGPs 1 and 2, respectively. The main advantage of our approach is that we are able to utilize the information of covariates. In the Supplementary Material, we compare our approach with SCM for pure factor structures without any covariates and show that both approaches work well. We emphasize that the superior performance of our approach crucially

[5]The SCM is implemented using Synth MATLAB Code written by Alberto Abadie, Alexis Diamond, and Jens Hainmueller to whom we are indebted. One key issue in SCM is choosing predictors (see, e.g., Abadie, 2021). We have tried several different ways and find that it performs best in general for our DGPs when the predictors include the average of pre-treatment covariates (x_{it}) and all pre-treatment outcomes (y_{it}).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

TABLE 2. MSE of estimators of TE and ATT

	DGP 1			DGP 2			DGP 3			DGP 4		
$(N_1, T, N_2, T_{0i}, T_{1i})$	hetero	homo	synth	hetero	homo	synth	hetero	homo	synth	hetero	homo	synth
A: $\text{MSE}_{\hat{\tau}_{it}} \times 100$: TE for treated individuals at a given time												
(40, 40, 10, 20, 20)	129	364	592	128	117	304	329	531	825	337	541	835
(40, 80, 10, 40, 40)	118	337	527	117	110	270	279	491	729	283	496	733
(80, 40, 20, 20, 20)	124	352	558	124	115	288	302	513	801	311	526	805
(80, 80, 20, 40, 40)	112	325	511	113	108	262	266	475	730	271	482	738
(120, 120, 30, 60, 60)	108	317	485	108	105	246	250	461	690	253	464	695
B: $\text{MSE}_{\hat{\tau}_{i\cdot}} \times 100$: ATT for individuals												
(40, 40, 10, 20, 20)	6.42	17.9	29.0	6.59	5.90	15.5	17.3	32.0	48.6	31.6	52.1	81.5
(40, 80, 10, 40, 40)	2.79	7.99	13.0	2.92	2.78	6.49	7.78	18.7	24.2	14.4	28.0	38.8
(80, 40, 20, 20, 20)	6.29	17.5	27.5	6.31	5.79	14.4	15.8	30.7	47.3	30.2	49.5	76.5
(80, 80, 20, 40, 40)	2.85	8.40	12.6	2.72	2.65	6.47	7.01	18.4	24.9	13.5	27.3	38.8
(120, 120, 30, 60, 60)	1.81	5.23	7.85	1.77	1.72	4.16	4.51	13.7	16.4	8.68	19.6	25.5
C: $\text{MSE}_{\hat{\tau}_{\cdot t}} \times 100$: ATT for time periods												
(40, 40, 10, 20, 20)	13.6	38.6	78.1	13.0	11.9	37.6	32.7	58.5	105	33.6	59.7	110
(40, 80, 10, 40, 40)	11.7	33.7	65.6	11.9	11.2	34.3	28.3	54.5	90.4	28.9	55.6	91.7
(80, 40, 20, 20, 20)	6.45	17.3	38.4	6.19	5.73	18.3	15.0	31.5	54.7	15.5	32.1	55.6
(80, 80, 20, 40, 40)	5.64	16.7	35.9	5.44	5.24	16.6	13.4	30.1	48.1	13.6	30.4	48.6
(120, 120, 30, 60, 60)	3.60	10.5	21.8	3.53	3.44	11.0	8.53	21.5	31.5	8.58	21.4	31.8
D: $\text{MSE}_{\hat{\tau}} \times 100$: Aggregate ATT												
(40, 40, 10, 20, 20)	0.76	1.84	3.75	0.71	0.63	1.80	1.91	3.79	6.93	3.54	6.41	12.4
(40, 80, 10, 40, 40)	0.28	0.74	1.47	0.28	0.25	0.87	0.86	2.27	2.87	1.52	3.27	4.62
(80, 40, 20, 20, 20)	0.31	0.82	2.06	0.29	0.29	0.99	0.72	1.80	3.75	1.38	2.54	5.73
(80, 80, 20, 40, 40)	0.17	0.42	0.81	0.12	0.11	0.40	0.32	1.00	1.75	0.62	1.47	2.43
(120, 120, 30, 60, 60)	0.06	0.19	0.39	0.06	0.06	0.19	0.16	0.57	0.99	0.30	0.77	1.48

Note: hetero refers to the heterogeneous model proposed in this article, homo refers to the model where the slope coefficients are imposed to be constant, and synth refers to SCM (synthetic control method).

depends on the correct specification of our model. In the Supplementary Material, we design DGPs with grouped factor structure and show that SCM can outperform our method in this special case.

Table 3 presents the actual coverage of the 95% CI for the TE, τ_{it} and ATT, $\tau_{i\cdot}, \tau_{\cdot t}$, and τ. The CI of τ_{it} is based on the quantiles of the residuals and the CIs of $\tau_{i\cdot}, \tau_{\cdot t}$, and τ are based on our asymptotic normality result in Theorem 4.1. The actual coverages are slightly under-covered when the sample size is small, however they are all close to 95% when the sample size is relatively large. This suggests the validity of our inference procedure.

In Section S.7.3 of our Supplementary Material, we show that when $\beta_{\cdot t}$ is nonstationary, the permutation test proposed in Chernozhukov et al. (2021) applied to SCM may have severe size distortion. The nonstationarity of $\beta_{\cdot t}$ causes non-

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

TABLE 3. Empirical coverage of 95% confidence intervals

	DGP 1				DGP 2				DGP 3				DGP 4			
$(N_1, T, N_2, T_{0i}, T_{1i})$	τ_{it}	$\tau_{i\cdot}$	$\tau_{\cdot t}$	τ	τ_{it}	$\tau_{i\cdot}$	$\tau_{\cdot t}$	τ	τ_{it}	$\tau_{i\cdot}$	$\tau_{\cdot t}$	τ	τ_{it}	$\tau_{i\cdot}$	$\tau_{\cdot t}$	τ
(40, 40, 10, 20, 20)	0.90	0.93	0.89	0.91	0.90	0.92	0.90	0.92	0.88	0.92	0.87	0.88	0.87	0.87	0.86	0.85
(40, 80, 10, 40, 40)	0.93	0.95	0.92	0.95	0.93	0.94	0.92	0.94	0.92	0.93	0.91	0.92	0.91	0.89	0.90	0.89
(80, 40, 20, 20, 20)	0.90	0.92	0.90	0.92	0.90	0.93	0.90	0.95	0.89	0.92	0.89	0.93	0.88	0.89	0.88	0.91
(80, 80, 20, 40, 40)	0.93	0.94	0.93	0.92	0.93	0.94	0.93	0.96	0.92	0.94	0.92	0.96	0.91	0.92	0.92	0.93
(120, 120, 30, 60, 60)	0.94	0.94	0.93	0.95	0.94	0.95	0.94	0.94	0.93	0.94	0.93	0.94	0.93	0.92	0.93	0.94

stationarity of the error term in SCM, which invalidates many existing inference procedures, such as Chernozhukov et al. (2021), Masini and Medeiros (2021), and Masini and Medeiros (2022). Our approach estimates all $\beta_{\cdot t}$ as unknown parameters, and therefore is immune to this problem.

6. AN EMPIRICAL APPLICATION

In this section, we apply our new method to study the effect of economic liberalization on economic growth, which is an important question in international economics and has received considerable attention. Although it is intuitive to argue that the effect of liberalization should be positive, empirical studies remain challenging and the topic is still being heatedly debated; see, e.g., Sachs et al. (1995), Rodriguez and Rodrik (2000), and Billmeier and Nannicini (2013, BN hereafter).

We consider a sample of 60 countries over 30 years (1965–1994), that is, $N = 60$ and $T = 30$. The construction of the sample is mainly based on data availability. The outcome variable y_{it} is country i's economic growth in year t. The treatment indicator is whether a country is classified as "open" in year t. The treatment variable is constructed by Sachs et al. (1995) and updated by Wacziarg and Welch (2008). A country at one year is considered "close" if it satisfies certain conditions, such as imposing high tariffs, using large nontariff barriers, or having a socialist economic system, among others. For the detailed criterion for being "close," see, for example, BN (p. 985). There are many covariates that can potentially explain economic growth (see, e.g., Durlauf, Johnson, and Temple, 2005, for a comprehensive literature review). Considering the relatively small sample size here, we cannot include many covariates. We choose two important ones: investment share (share of gross capital formation) and consumption share (share of household consumption) out of a country's GDP measured at current purchasing power parities (PPPs). Investment and consumption are usually the two major components of a country's GDP. In our sample, the average investment share is about 17% and the average consumption share is about 67%. The economic growth data are constructed from the real GDP per capita data used in BN and the treatment indicator is directly from BN. The data source of investment and consumption share data is the Penn World Table.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

There are 25 control countries, meaning that they remained "close" for the whole sample period. Thirty-five countries are treated, meaning that they became open at certain years. We rule out countries that were always open or open for more than 20 years in our sample period.[6] The summary statistics are reported in Table S.4 in Section S.7 of the Supplementary Material. The list of treated and control countries can be found in Table S.5 in the Supplementary Material. In total, we have 241 treated outcomes in our sample.

Table 4 reports estimates of treatment effects based on various methods. We first consider our hetero model. We find that the estimate of the aggregate ATT is 1.24 with a 95% CI of [0.57, 1.92]. This suggests that being "open" can lead to 1.24% economic growth annually on average. The positive effect is largely consistent with the literature. For example, Sachs et al. (1995) state that "open economies grow, on average, by 2.45 percentage points more than close economies, with a highly statistically significant effect." Romalis (2007) concludes that "eliminating existing developed world tariffs would increase developing country growth rates by 0.6 to 1.6 percent per annum." We also try alternative methods, which are labeled as "simple," "synth," and "homo" in Table 4. The "simple" method refers to the regression of y_{it} on the binary treatment indicator and two covariates and the coefficient on the treatment indicator is interpreted as a treatment effect. For the aggregate ATT, TWFEs are included. The "synth" method refers to SCM where we use the average of pre-treatment covariates and all pre-treatment outcomes as predictors. The "homo" method refers to the same model as "hetero" except that the slope coefficient $\beta_{i\cdot}^0 + \beta_{\cdot t}^0$ is imposed to be constant. The estimates based on the three alternative methods are all positive, though the magnitudes (0.62 [simple], 1.05 [synth], and 0.22 [homo]) differ. This suggests the robustness of the "positive effects."

Table 4 also presents the ATT for each individual treated country. We find that 20 treated countries have a positive ATT, while 15 treated countries have a negative ATT. Table 4 also reports their 95% CIs. Note that in our sample, the time period is quite short, therefore, the CIs are usually wide. Only five countries' ATT are statistically significant at the 5% level (Argentina, Chile, Ghana, Guinea, and Uganda). Among the five countries, only one country (Guinea) has a negative effect. Table 4 also reports the estimates based on alternative methods. Although the exact estimates are different, there are some commonalities in the results. For example, if we compare the results between our hetero model and the synthetic control, 26 out of 35 ATT estimates share the same sign.

One important feature of our model is that we allow a general form of heterogeneity, which is supported by our empirical results. Figure 1 shows the histogram of the demeaned estimates of $\beta_{i\cdot}$ for investment share and consumption share. Figure 1 also plots the demeaned estimates of $\beta_{\cdot t}$ for those two variables over time. The detailed estimates are reported in Tables S.5 and S.6 in the Supplementary

[6]Note that we use the untreated data of the treated countries to estimate the counterfactual outcome. Therefore, we cannot consider a country with a too short untreated period.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Table 4. Treatment effects

	hetero			Estimates based on other methods		
	Estimates	95% CI		Simple	synth	homo
Aggregate ATT	1.24	[0.57	1.92]	0.62	1.05	0.22
ATT for individual treated countries						
Argentina	5.09	[0.62	9.57]	0.01	8.82	3.52
Benin	3.20	[−0.89	7.29]	2.61	1.67	−0.14
Botswana	2.61	[−0.54	5.76]	−0.88	6.91	4.80
Brazil	−2.13	[−6.61	2.34]	−5.37	−3.74	−2.68
Cameroon	−2.52	[−8.59	3.54]	−3.06	−5.51	−6.93
Cape Verde	−2.00	[−6.47	2.48]	−0.81	−0.44	−3.95
Chile	9.82	[6.53	13.10]	3.15	1.39	1.31
Colombia	0.46	[−2.90	3.83]	0.10	2.03	0.60
Costa Rica	0.38	[−2.99	3.74]	1.51	2.64	1.03
Gambia	−2.14	[−5.42	1.14]	−6.00	−2.19	−2.51
Ghana	3.76	[0.48	7.04]	4.43	2.00	0.18
Guatemala	2.23	[−1.39	5.86]	0.57	−0.14	0.16
Guinea	−3.91	[−7.28	−0.55]	0.95	0.22	0.57
Honduras	−2.24	[−6.72	2.23]	−2.68	2.67	−0.60
Israel	−1.34	[−4.62	1.94]	0.76	1.00	−0.08
Jamaica	1.15	[−1.98	4.29]	1.71	0.87	1.54
Kenya	−2.63	[−7.11	1.84]	−3.48	−3.64	−2.37
Mali	2.40	[−1.23	6.03]	2.60	0.41	1.16
Mexico	−0.53	[−3.90	2.83]	0.30	−1.13	−0.96
Morocco	−1.13	[−4.34	2.09]	−1.63	0.53	−0.24
Nepal	0.47	[−4.00	4.95]	1.40	4.30	0.87
Niger	0.96	[−7.44	9.36]	3.25	−1.30	−3.04
Paraguay	−1.31	[−5.14	2.51]	−0.18	0.85	−1.14
Peru	−1.39	[−5.22	2.43]	−2.07	−0.92	−1.09
Philippines	1.57	[−2.06	5.20]	0.49	0.06	−0.59
Romania	−2.78	[−7.83	2.28]	−8.31	−6.24	−7.97
South Africa	−1.17	[−5.65	3.30]	−0.91	−0.68	−3.75
Sri Lanka	2.19	[−1.09	5.47]	0.74	1.86	2.22
Trinidad & Tobago	0.51	[−4.55	5.56]	−1.23	−2.92	−1.66
Tunisia	−1.86	[−5.69	1.96]	−1.61	−1.18	0.83
Turkey	1.43	[−2.39	5.26]	0.49	2.11	1.09
Uganda	5.43	[1.80	9.05]	4.16	1.87	0.74
Uruguay	2.23	[−1.86	6.32]	1.91	2.96	1.12
Venezuela	2.23	[−2.24	6.71]	3.28	−0.48	−1.05
Zambia	2.77	[−3.29	8.84]	−6.64	3.16	−2.23

Note: "hetero" refers to the heterogeneous model proposed in this paper, "simple" refers to regressing the growth on the treatment indicator and covariates, "synth" refers to SCM (synthetic control method), and "homo" refers to the model where the slope coefficients are imposed to be constant. The CI is based on the pre-whitening HAC estimator of the variance-term.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

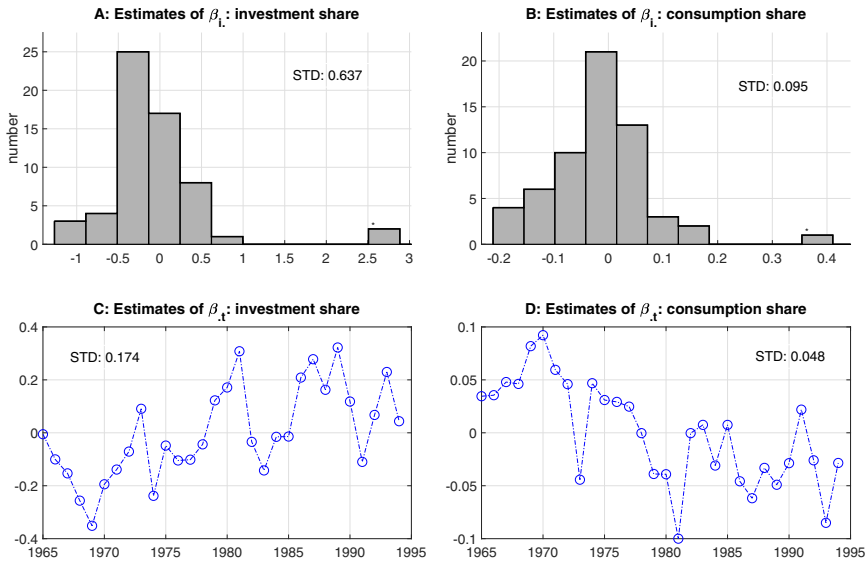


FIGURE 1. Estimates of $\beta_{i\cdot}$ and $\beta_{\cdot t}$ (demeaned).

Material. We observe a wide range of variations in the estimates, in particular the slope coefficient for investment share.[7] The average estimates of the slope coefficients of investment share and consumption share are 0.381 and −0.026, respectively. This shows that investment share has a large positive effect on a country's economic growth. We also find one unobservable factor, which suggests one common factor or shock that drives all countries' growth.

In sum, we find that the aggregate ATT of liberalization on countries' economic growth is positive and statistically significant. On average, "being open" leads to 1.24% economic growth annually. The ATT for individual countries is quite heterogeneous, though most of them are positive. We also find that investment share has a large positive effect on economic growth.

7. CONCLUSION

In this article, we consider the heterogeneous panel data models with IFEs where the slope coefficients of the covariates exhibit unobserved heterogeneity over both the individual and time dimensions. We propose a two-step procedure to estimate

[7]It is noticeable that two estimates of $\beta_{i\cdot}$ of investment share are larger than 2. These two countries are Ethiopia (2.67) and Guinea (2.69). One possible explanation is that the levels of the investment share of those two countries are low. The sample averages of investment share of these two countries are 5.1% and 5.6%, respectively, which are significantly lower than the average of the whole sample (17.25%). We leave the systematic study on this interesting question for future research.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

the parameters and study the asymptotic properties of the estimators. We apply the proposed model to study the consistent estimation of the average of heterogeneous treatment effects on the treated. As an empirical illustration, we investigate the effect of economic liberalization on economic growth and find a positive and significant aggregate ATT.

Our approach is based on a flexible outcome model with covariates and factor structures and can be thought of as an alternative to the popular SCM for estimating treatment effects. There are two main advantages of our approach. First, our estimator exploits the information of covariates and is typically more efficient than SCM if our model is correctly specified. Second, as emphasized above, some existing inference procedures such as that of Chernozhukov et al. (2021), Masini and Medeiros (2021), and Masini and Medeiros (2022) are typically invalid when $\{\beta_{\cdot t}^0\}$ exhibits some nonstationary behavior. Our estimator treats $\{\beta_{\cdot t}^0\}$ as unknown parameters to estimate, and thus does not suffer from this issue.

There are several interesting topics for further research. First, although we allow nonstationary $\{\beta_{\cdot t}\}$, we still impose stationarity on covariates and factors. It is helpful to relax those assumptions. Second, a fruitful direction of future research is to augment SCM with our outcome-based model, as in Ben-Michael, Feller, and Rothstein (2021). Third, we do not consider unbalanced panels in the current study. For unbalanced panels, we propose an algorithm to estimate the model parameters in Section S.5 of the Supplementary Material without a formal justification. We are exploring these topics in ongoing work.

This appendix establishes the main results of the article by drawing on several technical lemmas proved in Sections S.1.3, S.2.4, S.3, and S.4 of the Supplementary Material. We use $O_P(\cdot)$ and $o_P(\cdot)$ to denote the usual *big O* and *small o* in probability, respectively. For a random matrix A that is typically high-dimensional, we write $A = O_{\text{sp},P}(a_{NT})$ and $A = o_{\text{sp},P}(a_{NT})$ to indicate $\|A\|_{\text{sp}} = O_P(a_{NT})$ and $\|A\|_{\text{sp}} = o_P(a_{NT})$, respectively. For a matrix $A_{m\times n}$, denote the maximum absolute column-sum norm (ℓ_1-norm) and the maximum absolute row-sum norm (ℓ_∞-norm) as $\|A\|_1 \equiv \max_{1\le j\le n}\sum_{i=1}^m |a_{ij}|$ and $\|A\|_\infty \equiv \max_{1\le i\le m}\sum_{j=1}^n |a_{ij}|$. When A is an m-vector ($n = 1$), we have that $\|A\|_\infty$ represents the maximum of absolute value of its elements. We also define $\|A\|_{2,\infty} = \max_{1\le i\le m}\|A_{i,\cdot}\| = \max_{1\le i\le m}\sqrt{\sum_{j=1}^n a_{ij}^2}$, where $A_{i,\cdot}$ denotes the ith row of A. The following inequalities will be used frequently:

$$\|Ax\|_\infty \le \|A\|_{2,\infty}\|x\|, \quad \|AB\|_{2,\infty} \le \|A\|_\infty\|B\|_{2,\infty}, \quad \text{and } \|AB\|_{2,\infty} \le \|A\|_{2,\infty}\|B\|_{\text{sp}},$$

where x is a conformable vector and B is a matrix.

A. PROOFS OF THE MAIN RESULTS

A.1. Asymptotic Analyses for Section 3

To prove Theorem 3.1, we first introduce a lemma whose proof is similar to that of Lemma A.1 of Miao et al. (2023).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

LEMMA A.1. (i) *For any* $T \times N$ *matrix* Δ, $||\Theta^0 + \mathbb{M}_{F^0}\Delta\mathbb{M}_{\Lambda^0}||_* = ||\Theta^0||_* + ||\mathbb{M}_{F^0}\Delta\mathbb{M}_{\Lambda^0}||_*$.

(ii) *For any conformable matrices A and B,* $|\text{tr}(A'B)| \le ||A||_{sp}||B||_*$.

(iii) *rank* $(\Delta - \mathbb{M}_{F^0}\Delta\mathbb{M}_{\Lambda^0}) \le 2R^0$.

(iv) *For a matrix A with rank* $(A) = r$, $||A||_* \le \sqrt{r}||A||$.

Proof of Theorem 3.1. Let $\hat{\Delta}_\Theta = \hat{\Theta}_o - \Theta^0$ and $\hat{\delta}_\theta = \text{vec}(\hat{\Delta}_\Theta)$. Comparing the values of $\tilde{\mathcal{L}}_{NT}(\Theta) + \gamma_{NT}||\Theta||_*/\sqrt{NT}$ at $\hat{\Theta}_o$ and Θ^0, we have

$$
\begin{aligned}
0 &\ge \frac{1}{2NT}\hat{\delta}_\theta'\mathbb{M}_x\hat{\delta}_\theta - \frac{1}{NT}\hat{\delta}_\theta'\mathbb{M}_x U + \gamma_{NT}\frac{||\hat{\Theta}_o||_*}{\sqrt{NT}} - \gamma_{NT}\frac{||\Theta^0||_*}{\sqrt{NT}} \qquad \textbf{(A.1)}\\
&= \frac{1}{2NT}\hat{\delta}_\theta'\mathbb{M}_x\hat{\delta}_\theta - \frac{1}{NT}\hat{\delta}_\theta'\mathbb{M}_x U + \gamma_{NT}\frac{||\hat{\Delta}_\Theta + \Theta^0||_*}{\sqrt{NT}} - \gamma_{NT}\frac{||\Theta^0||_*}{\sqrt{NT}}\\
&\ge \frac{1}{2NT}\hat{\delta}_\theta'\mathbb{M}_x\hat{\delta}_\theta - \frac{1}{NT}\hat{\delta}_\theta'\mathbb{M}_x U + \gamma_{NT}\frac{||\Theta^0 + \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_* - ||\hat{\Delta}_\Theta - \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}}\\
&\quad - \gamma_{NT}\frac{||\Theta^0||_*}{\sqrt{NT}}\\
&= \frac{1}{2NT}\hat{\delta}_\theta'\mathbb{M}_x\hat{\delta}_\theta - \frac{1}{NT}\hat{\delta}_\theta'\mathbb{M}_x U + \gamma_{NT}\frac{||\mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_* - ||\hat{\Delta}_\Theta - \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}},
\end{aligned}
$$

where the second inequality holds by the triangle inequality and the last equality follows by Lemma A.1(i). Note that $\hat{\delta}_\theta'\mathbb{M}_x U = \text{tr}\,[\hat{\Delta}_\Theta'\,\text{mat}(\mathbb{M}_x U)]$. By Lemma A.1(ii), Assumption 1, and the triangle inequality, the following inequality holds w.p.a.1:

$$
\begin{aligned}
\frac{1}{NT}|\hat{\delta}_\theta'\mathbb{M}_x U| &\le \frac{||\text{mat}(\mathbb{M}_x U)||_{sp}}{\sqrt{NT}}\frac{||\hat{\Delta}_\Theta||_*}{\sqrt{NT}} \le \frac{\gamma_{NT}}{2}\frac{||\hat{\Delta}_\Theta||_*}{\sqrt{NT}}\\
&\le \frac{\gamma_{NT}}{2}\frac{||\mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}} + \frac{\gamma_{NT}}{2}\frac{||\hat{\Delta}_\Theta - \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}}. \qquad \textbf{(A.2)}
\end{aligned}
$$

Combining inequalities (A.1) and (A.2), we have that

$$
\frac{1}{NT}\hat{\delta}_\theta'\mathbb{M}_x\hat{\delta}_\theta + \gamma_{NT}\frac{||\mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}} \le 3\gamma_{NT}\frac{||\hat{\Delta}_\Theta - \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}}. \qquad \textbf{(A.3)}
$$

Inequality (A.3) implies that $\hat{\Delta}_\Theta \in \mathbb{C}_{NT}$, where $\mathbb{C}_{NT}$ is defined in Assumption 2. By Assumption 2, inequality (A.3), and Lemma A.1 (iii) and (iv), there exists a positive constant μ such that

$$
\begin{aligned}
\frac{\mu}{NT}\hat{\delta}_\theta'\hat{\delta}_\theta &\le 3\gamma_{NT}\frac{||\hat{\Delta}_\Theta - \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||_*}{\sqrt{NT}} \le 3\gamma_{NT}\sqrt{2R^0}\frac{||\hat{\Delta}_\Theta - \mathbb{M}_{F^0}\hat{\Delta}_\Theta\mathbb{M}_{\Lambda^0}||}{\sqrt{NT}}\\
&\le 3\gamma_{NT}\sqrt{2R^0}\frac{||\hat{\Delta}_\Theta||}{\sqrt{NT}}, \quad \text{w.p.a.1.}
\end{aligned}
$$

It follows that $\frac{||\hat{\Delta}_\Theta||}{\sqrt{NT}} \le \frac{3\sqrt{2R^0}}{\mu}\gamma_{NT}$. □

Proof of Theorem 3.2. (i) The proof follows from that of Theorem 3.2 in Hong et al. (2023) and thus omitted.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

(ii) Let $S_F = T^{-1}F^{0\prime}F^0, S_\Lambda = N^{-1}\Lambda^{0\prime}\Lambda^0$, and $\tilde{H} = S_F^{-1/2}v$, where v is the $R^0 \times R^0$ matrix whose columns are the eigenvectors of $S_F^{1/2}S_\Lambda S_F^{1/2}$. Conditional on $\hat{R} = R^0$, we can bound $T^{-1/2}\left\|\hat{F}_o - F^0\tilde{H}\right\|$. Specifically, we can follow Miao et al. (2023) and apply the Davis–Kahan sin (Θ) theorem (e.g., Yu, Wang, and Samworth, 2015) to obtain the following bound: $T^{-1/2}\left\|\hat{F}_o - F^0\tilde{H}\right\| = O_P(\gamma_{NT})$ under Assumptions 1–3, which implies that $\|\mathbb{M}_{\hat{F}_o} - \mathbb{M}_{F^0}\| = O_P(\gamma_{NT})$.

(iii) Analogous arguments of the last point show that $N^{-1/2}\left\|\hat{\Lambda}_o - \Lambda^0(\tilde{H}^{-1})'\right\| = O_P(\gamma_{NT})$ and $\|\mathbb{M}_{\hat{\Lambda}_o} - \mathbb{M}_{\Lambda^0}\| = O_P(\gamma_{NT})$. □

To prove Theorem 3.3, we introduce Lemmas A.2 and A.3 and some notations. Let

$$\tilde{Q} \equiv X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})X^{(1,2)} = \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix}, \tag{A.4}$$

$$\mathbf{q} \equiv X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})U = \begin{bmatrix} X^{(1)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})U \\ X^{(2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})U \end{bmatrix} \equiv \begin{bmatrix} \mathbf{q}^{(1)} \\ \mathbf{q}^{(2)} \end{bmatrix},$$

where $\tilde{Q}_{kl} = X^{(k)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})X^{(l)}$ for $k, l = 1, 2$. In Section S.1 of the Supplementary Material, we show that $\tilde{Q}$ is invertible *w.p.a.1* under Assumption 4. Define the inverse of $\tilde{Q}$, an $(N-1+T)K \times (N-1+T)K$ matrix, as

$$\tilde{\mathbb{Q}} \equiv \begin{bmatrix} \tilde{\mathbb{Q}}_{11} & \tilde{\mathbb{Q}}_{12} \\ \tilde{\mathbb{Q}}_{21} & \tilde{\mathbb{Q}}_{22} \end{bmatrix} \equiv \tilde{Q}^{-1} = \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{21} & \tilde{Q}_{22} \end{bmatrix}^{-1},$$

where $\tilde{\mathbb{Q}}_{kl}$ has the same size as $\tilde{Q}_{kl}$.

LEMMA A.2. *Let* $\mathbf{q}_a^{(1)} = \begin{pmatrix} \dot{q}_1' & \dot{q}_2' & \cdots & \dot{q}_{N-1}' \end{pmatrix}'$ *and* $\mathbf{q}_a^{(2)} = \begin{pmatrix} \ddot{q}_1' & \ddot{q}_2' & \cdots & \ddot{q}_T' \end{pmatrix}'$. *Suppose Assumptions 3–5 hold. Then we have:*

(i) $||\sqrt{T}\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)}||_\infty = O_P[(\log^2 N + \log^2 T)/\sqrt{N}]$;

(ii) $||\sqrt{N}\tilde{\mathbb{Q}}_{21}\mathbf{q}^{(1)}||_\infty = O_P[(\log^2 N + \log^2 T)/\sqrt{T}]$;

(iii) $||\sqrt{T}\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)} - \dot{\mathbf{S}}_{(1)}^{-1}\mathbf{q}_a^{(1)}/\sqrt{T}||_\infty = O_P[(\log^2 N + \log^2 T)/\sqrt{T}]$;

(iv) $||\sqrt{N}\tilde{\mathbb{Q}}_{22}\mathbf{q}^{(2)} - \ddot{\mathbf{S}}_{(2)}^{-1}\mathbf{q}_a^{(2)}/\sqrt{N}||_\infty = O_P[(\log^2 N + \log^2 T)/\sqrt{N}]$.

LEMMA A.3. *Suppose Assumptions 3–5 hold. Then we have:*

(i) $\sqrt{T}(\iota_{N-1}' \otimes \mathbb{I}_K)\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)} = \dot{\mathbf{s}}_N^{-1}\dot{q}_N/\sqrt{T} + O_P[(\log^2 N + \log^2 T)/\sqrt{N}]$;

(ii) $\sqrt{T}(\iota_{N-1}' \otimes \mathbb{I}_K)\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)} = O_P[(\log^2 N + \log^2 T)/\sqrt{N}]$.

The proofs of Lemmas A.1–A.3 are provided in Section S.1.3 of the Supplementary Material.

Proof of Theorem 3.3. Step 1. Obtain explicit expression of $\check{B} - \dot{B}$.

Recall the definition of $\check{B}$ as in (3.1), we have

$$\check{B} - \dot{B} = \left[X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})X^{(1,2)}\right]^{-1} X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})U.$$

As both $X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})X^{(1,2)}$ and $X^{(1,2)\prime}(\mathbb{M}_{\Lambda^0} \otimes \mathbb{M}_{F^0})U$ are high-dimensional objects, it is essential to carefully analyze their properties. After proper normalization, we

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

have

$$\mathbb{L}_{NT}(\check{B}-\dot{B})=\begin{bmatrix}\sqrt{T}\left(\check{B}^{(1)}-\dot{B}^{(1)}\right)\\ \sqrt{N}\left(\check{B}^{(2)}-\dot{B}^{(2)}\right)\end{bmatrix}=\begin{bmatrix}T\tilde{\mathbb{Q}}_{11} & \sqrt{NT}\tilde{\mathbb{Q}}_{12}\\ \sqrt{NT}\tilde{\mathbb{Q}}_{21} & N\tilde{\mathbb{Q}}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{q}^{(1)}/\sqrt{T}\\ \mathbf{q}^{(2)}/\sqrt{N}\end{bmatrix}$$
$$=\begin{bmatrix}\sqrt{T}\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)}+\sqrt{T}\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)}\\ \sqrt{N}\tilde{\mathbb{Q}}_{21}\mathbf{q}^{(1)}+\sqrt{N}\tilde{\mathbb{Q}}_{22}\mathbf{q}^{(2)}\end{bmatrix},$$

where $\mathbb{L}_{NT}\equiv \text{blkdiag}(\sqrt{T}\mathbb{I}_{K(N-1)},\sqrt{N}\mathbb{I}_{KT})$.

Step 2. Verify that $\max_{1\le i\le N-1}||\sqrt{T}(\check{\beta}_{i\cdot}-\dot{\beta}_{i\cdot})-\dot{\mathbf{s}}_{i\cdot}^{-1}\dot{q}_i/\sqrt{T}||=o_P(1)$ and $\max_{1\le t\le T}||\sqrt{N}(\check{\beta}_{\cdot t}-\dot{\beta}_{\cdot t})-\ddot{\mathbf{s}}_{\cdot t}^{-1}\ddot{q}_t/\sqrt{N}||=o_P(1)$.

Recall that $\dot{\mathbf{S}}_{(1)}\equiv\begin{pmatrix}\dot{\mathbf{s}}_{1\cdot} & & \\ & \ddots & \\ & & \dot{\mathbf{s}}_{N-1\cdot}\end{pmatrix}$, $\dot{\mathbf{S}}_{(2)}\equiv\begin{pmatrix}\ddot{\mathbf{s}}_{\cdot 1} & & \\ & \ddots & \\ & & \ddot{\mathbf{s}}_{\cdot T}\end{pmatrix}$, $\mathbf{q}_a^{(1)}\equiv\begin{pmatrix}\dot{q}_1\\ \vdots\\ \dot{q}_{N-1}\end{pmatrix}$, $\mathbf{q}_a^{(2)}\equiv\begin{pmatrix}\ddot{q}_1\\ \vdots\\ \ddot{q}_T\end{pmatrix}$, $\dot{\mathbf{S}}_{(1)}^{-1}\mathbf{q}_a^{(1)}=\left(\sum_{t=1}^T u_{1t}\dot{x}_{1t}'(\sum_{t=1}^T\dot{x}_{1t}\dot{x}_{1t}')^{-1},\ldots,\sum_{t=1}^T u_{N-1,t}\dot{x}_{N-1,t}'(\sum_{t=1}^T\dot{x}_{N-1,t}\dot{x}_{N-1,t}')^{-1}\right)'$ and $\ddot{\mathbf{S}}_{(2)}^{-1}\mathbf{q}_a^{(2)}=\left(\sum_{i=1}^N u_{i1}\dot{x}_{i1}'(\sum_{i=1}^N\dot{x}_{i1}\dot{x}_{i1}')^{-1},\ldots,\sum_{i=1}^N u_{iT}\dot{x}_{iT}'(\sum_{i=1}^N\dot{x}_{iT}\dot{x}_{iT}')^{-1}\right)'$.

Verifying $\max_{1\le i\le N-1}||\sqrt{T}(\check{\beta}_{i\cdot}-\dot{\beta}_{i\cdot})-\dot{\mathbf{s}}_{i\cdot}^{-1}\dot{q}_i/\sqrt{T}||=o_P(1)$, and $\max_{1\le t\le T}||\sqrt{N}(\check{\beta}_{\cdot t}-\dot{\beta}_{\cdot t})-\ddot{\mathbf{s}}_{\cdot t}^{-1}\ddot{q}_t/\sqrt{N}||=o_P(1)$ is equivalent to showing

$$\left\|\begin{matrix}\sqrt{T}\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)}+\sqrt{T}\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)}-\dot{\mathbf{S}}_{(1)}^{-1}\mathbf{q}_a^{(1)}/\sqrt{T}\\ \sqrt{N}\tilde{\mathbb{Q}}_{21}\mathbf{q}^{(1)}+\sqrt{N}\tilde{\mathbb{Q}}_{22}\mathbf{q}^{(2)}-\ddot{\mathbf{S}}_{(2)}^{-1}\mathbf{q}_a^{(2)}/\sqrt{N}\end{matrix}\right\|_\infty=o_P(1).$$

This result is a direct consequence of Lemma A.2: (1) $||\sqrt{T}\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)}||_\infty=o_P(1)$, (2) $||\sqrt{N}\tilde{\mathbb{Q}}_{21}\mathbf{q}^{(1)}||_\infty=o_P(1)$, (3) $||\sqrt{T}\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)}-\dot{\mathbf{S}}_{(1)}^{-1}\mathbf{q}_a^{(1)}/\sqrt{T}||_\infty=o_P(1)$, and (4) $||\sqrt{N}\tilde{\mathbb{Q}}_{22}\mathbf{q}^{(2)}-\ddot{\mathbf{S}}_{(2)}^{-1}\mathbf{q}_a^{(2)}/\sqrt{N}||_\infty=o_P(1)$.

Step 3. Verify $\sqrt{T}(\check{\beta}_{N\cdot}-\dot{\beta}_{N\cdot})=\dot{\mathbf{s}}_{N\cdot}^{-1}\dot{q}_N/\sqrt{T}+o_P(1)$.

It remains to study $\check{\beta}_{N\cdot}-\dot{\beta}_{N\cdot}$, where $\dot{\beta}_{N\cdot}=-\sum_{i=1}^{N-1}\dot{\beta}_{i\cdot}=-(\iota_{N-1}'\otimes\mathbb{I}_K)\dot{B}^{(1)}$ and $\check{\beta}_{N\cdot}=-\sum_{i=1}^{N-1}\check{\beta}_{i\cdot}=-(\iota_{N-1}'\otimes\mathbb{I}_K)\check{B}^{(1)}$. Then we have

$$\check{\beta}_{N\cdot}-\dot{\beta}_{N\cdot}=-(\iota_{N-1}'\otimes\mathbb{I}_K)(\check{B}-\dot{B}^{(1)})=-(\iota_{N-1}'\otimes\mathbb{I}_K)\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)}-(\iota_{N-1}'\otimes\mathbb{I}_K)\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)},$$

where $-\sqrt{T}(\iota_{N-1}'\otimes\mathbb{I}_K)\tilde{\mathbb{Q}}_{11}\mathbf{q}^{(1)}=-\dot{\mathbf{s}}_{N\cdot}^{-1}\dot{q}_N/\sqrt{T}+O_P(\log^2 N/\sqrt{N})$ and $\sqrt{T}(\iota_{N-1}'\otimes\mathbb{I}_K)\tilde{\mathbb{Q}}_{12}\mathbf{q}^{(2)}=O_P(\log N/\sqrt{N})$, by Lemma A.3. It follows that $\sqrt{T}(\check{\beta}_{N\cdot}-\dot{\beta}_{N\cdot})=-\dot{\mathbf{s}}_{N\cdot}^{-1}\dot{q}_N/\sqrt{T}+O_P(\log^2 N/\sqrt{N})$. □

To prove Theorem 3.4, we introduce Lemmas A.4 and A.5 and some notations. Let $\Lambda\in\mathbb{R}^{N\times R^0}$ and $F\in\mathbb{R}^{T\times R^0}$. Define

$$\tilde{Q}(\Lambda,F)\equiv X^{(1,2)\prime}(\mathbb{M}_\Lambda\otimes\mathbb{M}_F)X^{(1,2)}\text{ and }\tilde{Q}_{kl}(\Lambda,F)$$
$$\equiv X^{(k)\prime}(\mathbb{M}_\Lambda\otimes\mathbb{M}_F)X^{(l)}\text{ for }\quad k,l\in\{1,2\}.$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Next, define the projected score vectors

$$\mathbf{q}^{(1)}(\Lambda,F) \equiv X^{(1)\prime}(\mathbb{M}_\Lambda \otimes \mathbb{M}_F)U \text{ and } \mathbf{q}^{(2)}(\Lambda,F) \equiv X^{(2)\prime}(\mathbb{M}_\Lambda \otimes \mathbb{M}_F)U,$$

so that the vectors $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ from the previous section are obtained by evaluating at the true factors and loadings (Λ^0, F^0). Further, define

$$\mathbf{q}_\theta^{(1)}(\Lambda,F) \equiv X^{(1)\prime}(\mathbb{M}_\Lambda \otimes \mathbb{M}_F)\text{vec}(F^0\Lambda^{0\prime}) \text{ and } \mathbf{q}_\theta^{(2)}(\Lambda,F) \equiv X^{(2)\prime}(\mathbb{M}_\Lambda \otimes \mathbb{M}_F)\text{vec}(F^0\Lambda^{0\prime}),$$

which equal zero when $(\Lambda,F) = (\Lambda^0, F^0)$.

LEMMA A.4. *Under the conditions for Theorem 3.4, we have*
(i) $\tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})$ *is asymptotically non-singular, with expansion*

$$\tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})^{-1} = \tilde{\mathbb{Q}} \sum_{j=0}^{\infty} [(\tilde{Q} - \tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}))\tilde{\mathbb{Q}}]^j,$$

where $||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbb{L}_{NT}||_{sp} = O_P(1)$, *and* $||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\sum_{j=1}^{\infty}[(\tilde{Q} - \tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}))\tilde{\mathbb{Q}}]^j\mathbb{L}_{NT}||_{sp} = o_P(1)$;
(ii) $||\mathbb{L}_{NT}^{-1}(\tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) - \tilde{Q})\mathbb{L}_{NT}^{-1}||_{2,\infty} = O_P(\delta_{NT}^{-2}\log(NT))$, *and* $||\mathbb{L}_{NT}^{-1}(\tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) - \tilde{Q})\mathbb{L}_{NT}^{-1}||_{sp} = O_P(\sqrt{N+T}\delta_{NT}^{-2}\log(NT))$;
(iii) $||\mathbf{q}_\theta^{(1)}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})|| = o_P(\sqrt{T})$, *and* $||\mathbf{q}_\theta^{(2)}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})|| = o_P(\sqrt{N})$;
(iv) $||\mathbb{L}_{NT}^{-1}\mathbf{q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})|| = O_P(\sqrt{N+T})$.

LEMMA A.5. *Under the conditions for Theorem 3.4, we have*
(i) $\left\|\mathbb{L}_{NT}\tilde{\mathbb{Q}}(\mathbf{q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) - \mathbf{q})\right\|_\infty = o_P(1)$, *and* $||\mathbb{L}_{NT}\tilde{\mathbb{Q}}(\mathbf{q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) - \mathbf{q})|| = O_P(\sqrt{N+T}\delta_{NT}^{-1}\log^2(NT))$;
(ii) $\left\|\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbf{q}_\theta(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})\right\| = o_P(1)$;
(iii) $||\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\mathbb{L}_{NT}^{-1}||_{2,\infty} = O_P(\log(NT)/\sqrt{NT})$ *and* $||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\tilde{Q}^{(2,rem)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}||_{sp} = O_P(\log^2(NT)\delta_{NT}^{-2})$;
(iv) $\left\|\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbb{L}_{NT} - \text{blkdiag}(\dot{\mathbf{S}}_{(1)}^{-1}, \ddot{\mathbf{S}}_{(2)}^{-1})\right\|_{2,\infty} = O_P(\delta_{NT}^{-1}\log(NT))$,
where expressions of $\tilde{Q}^{(\infty)}$ *and* $\tilde{Q}^{(2,rem)}$ *are provided in the proof and* $\tilde{Q} - \tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) = \tilde{Q}^{(\infty)} + \tilde{Q}^{(2,rem)}$.

The proofs of the above two lemmas are given in Section S.2.4 of the Supplementary Material.

Proof of Theorem 3.4. We show that $\hat{B}_2$ is asymptotically equivalent to the infeasible estimator, that is, $||\mathbb{L}_{NT}(\hat{B}_2 - \dot{B})||_\infty = o_P(1)$. Then Theorem 3.3 leads to the desired result.

Recall that $\hat{B}_2 - \dot{B} = \tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})^{-1}\left[\mathbf{q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) + \mathbf{q}_\theta(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})\right]$. By Lemma A.4 (i), we have

$$\begin{aligned}\mathbb{L}_{NT}(\hat{B}_2 - \dot{B}) &= \mathbb{L}_{NT}\tilde{\mathbb{Q}}\left[\mathbf{q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)}) + \mathbf{q}_\theta(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})\right] \\ &\quad + \mathbb{L}_{NT}\tilde{\mathbb{Q}}[\tilde{Q} - \tilde{Q}(\hat{\Lambda}^{(2)}, \hat{F}^{(2)})]\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q} \\ &\equiv A_1 + A_2 + A_3,\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

where $A_1 \equiv \mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbf{q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)})$, $A_2 \equiv \mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbf{q}_\theta(\hat{\Lambda}^{(2)},\hat{F}^{(2)})$, $A_3 \equiv \mathbb{L}_{NT}\tilde{\mathbb{Q}}[\tilde{Q}-\tilde{Q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)})]\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}$, and $\delta\mathbf{q} \equiv \sum_{j=0}^{\infty}(\mathbb{L}_{NT}^{-1}(\tilde{Q}-\tilde{Q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)})\tilde{\mathbb{Q}}\mathbb{L}_{NT})^j\mathbb{L}_{NT}^{-1}[\mathbf{q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)}) + \mathbf{q}_\theta(\hat{\Lambda}^{(2)},\hat{F}^{(2)})]$.It suffices to study A_1, A_2, and A_3:

1. $||A_1 - \mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbf{q}||_\infty = o_P(1)$. This result is implied by Lemma A.5(i).

2. $||A_2||_\infty = o_P(1)$. This is implied by Lemma A.5(ii).

3. Finally, we show $||A_3||_\infty = o_P(1)$ below.

By Lemma A.4, we have that $||\mathbb{L}_{NT}^{-1}(\mathbf{q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)}) + \mathbf{q}_\theta(\hat{\Lambda}^{(2)},\hat{F}^{(2)}))|| = O_P(\sqrt{N+T})$. By the fact that

$$\left\|\sum_{j=0}^{\infty}\left(\mathbb{L}_{NT}^{-1}(\tilde{Q}-\tilde{Q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)})\tilde{\mathbb{Q}}\mathbb{L}_{NT}\right)^j\right\|_{sp} = O_P(1), \text{ and } ||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbb{L}_{NT}||_{sp} = O_P(1),$$

we have $||\delta\mathbf{q}|| = O_P(\sqrt{N+T})$. By the identity that $\tilde{Q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)}) - \tilde{Q} = \tilde{Q}^{(\infty)} + \tilde{Q}^{(2,rem)}$ as in Lemma A.5(iii), we have

$$\mathbb{L}_{NT}\tilde{\mathbb{Q}}[\tilde{Q}-\tilde{Q}(\hat{\Lambda}^{(2)},\hat{F}^{(2)})]\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q} = \mathbb{L}_{NT}\tilde{\mathbb{Q}}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q} + \mathbb{L}_{NT}\tilde{\mathbb{Q}}\tilde{Q}^{(2,rem)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}.$$

The spectrum norm bound of $\tilde{Q}^{(2,rem)}$ in Lemma A.5(iii) further implies $||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\tilde{Q}^{(2,rem)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}|| = O_P(\sqrt{N+T}\delta_{NT}^{-2}\log^2(NT)) = o_P(1)$. For $||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}$, we have

$$\begin{aligned}
&||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}||_\infty \\
&\quad\le \left\|\left(\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbb{L}_{NT} - \text{blkdiag}(\dot{\mathbf{S}}_{(1)}^{-1},\ddot{\mathbf{S}}_{(2)}^{-1})\right)\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\|_\infty \\
&\qquad + \left\|\text{blkdiag}(\dot{\mathbf{S}}_{(1)}^{-1},\ddot{\mathbf{S}}_{(2)}^{-1})\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\|_\infty \\
&\quad\le \left\|\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbb{L}_{NT} - \text{blkdiag}(\dot{\mathbf{S}}_{(1)}^{-1},\ddot{\mathbf{S}}_{(2)}^{-1})\right\|_{2,\infty}\left\|\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\| \\
&\qquad + ||\text{blkdiag}(\dot{\mathbf{S}}_{(1)}^{-1},\ddot{\mathbf{S}}_{(2)}^{-1})||_\infty\left\|\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\|_\infty \equiv A_{3,1} + A_{3,2}.
\end{aligned}$$

To bound $A_{3,1}$ and $A_{3,2}$, we first bound $\left\|\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\|$ and $\left\|\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\|_\infty$. By Lemma A.5(iii), we have

$$||\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}||_\infty \le ||\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\mathbb{L}_{NT}^{-1}||_{2,\infty}\cdot||\mathbb{L}_{NT}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}|| = O_P(\delta_{NT}^{-1}\log(NT)),$$

and $||\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}|| = O_P(\sqrt{N+T}\delta_{NT}^{-1}\log(NT))$. It follows that $A_{3,1} = O_P(\sqrt{N+T}\delta_{NT}^{-2}\log^2(NT)) = o_P(1)$ where we also use Lemma A.5(iv). Similarly, $A_{3,2} = O_P(\delta_{NT}^{-1}\log(NT)) = o_P(1)$ as $||\text{blkdiag}(\dot{\mathbf{S}}_{(1)}^{-1},\ddot{\mathbf{S}}_{(2)}^{-1})||_\infty = O_P(1)$ and $\left\|\mathbb{L}_{NT}^{-1}\tilde{Q}^{(\infty)}\tilde{\mathbb{Q}}\mathbb{L}_{NT}\delta\mathbf{q}\right\|_\infty = O_P(\delta_{NT}^{-1}\log(NT))$. Then $||A_3||_\infty = o_P(1)$. □

Proof of Theorem 3.5. (i) Note that $\sqrt{T}(\hat{\beta}_{i\cdot} - \beta_{i\cdot}^0) = -\sqrt{\frac{T}{N}}\frac{1}{\sqrt{N}}\sum_{j=1}^N\beta_j^0 + (\frac{1}{T}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{x}_{i\cdot})^{-1}\frac{1}{\sqrt{T}}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} + o_P(1)$ by Theorem 3.4(i). Then by Assumption 6(i)–(iii) and the Slutsky theorem, we have $\sqrt{T}(\hat{\beta}_{i\cdot} - \beta_{i\cdot}^0) \xrightarrow{d} N(0, \frac{1}{\kappa}\Sigma_{1\beta} + (\mathbf{s}_{i\cdot}^0)^{-1}\Phi_{i\cdot}(\mathbf{s}_{i\cdot}^0)^{-1})$ provided $\frac{N}{T} \to \kappa \in (0,\infty]$ as $(N,T) \to \infty$. When $\frac{N}{T} \to 0$ as $(N,T) \to \infty$, we have instead

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

that

$$\sqrt{N}(\hat{\beta}_{i\cdot} - \beta^0_{i\cdot}) = -\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\beta^0_{j\cdot} + \sqrt{\frac{N}{T}}\left(\frac{1}{T}\mathbf{x}'_{i\cdot}\mathbb{M}_{F^0}\mathbf{x}_{i\cdot}\right)^{-1}\frac{1}{\sqrt{T}}\mathbf{x}'_{i\cdot}\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} + o_P(1)$$
$$= -\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\beta^0_{j\cdot} + o_P(1) \xrightarrow{d} N(0, \Sigma_{1\beta}).$$

Summarizing the two cases, we have that $V^{-1/2}_{\beta,i\cdot}(\hat{\beta}_{i\cdot} - \beta^0_{i\cdot}) \xrightarrow{d} N(0, \mathbb{I}_K)$, for $V_{\beta,i\cdot} \equiv \Sigma_{1\beta}/N + (\mathbf{s}^0_{i\cdot})^{-1}\Phi_{i\cdot}(\mathbf{s}^0_{i\cdot})^{-1}/T$.

(ii) $\sqrt{N}(\hat{\beta}_{\cdot t} - \beta^0_{\cdot t}) = \frac{1}{\sqrt{N}}\sum_{j=1}^{N}\beta^0_{j\cdot} + (\frac{1}{N}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t})^{-1}\frac{1}{\sqrt{N}}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t} + o_P(1)$ by Theorem 3.4(ii). Then by Assumption 6 (i)–(iii) and the Slutsky theorem, we have $\sqrt{N}(\hat{\beta}_{\cdot t} - \beta^0_{\cdot t}) \xrightarrow{d} N(0, \Sigma_{1\beta} + (\mathbf{s}^0_{\cdot t})^{-1}\Phi_{\cdot t}(\mathbf{s}^0_{\cdot t})^{-1})$. □

To prove Theorem 3.6, we need the following lemma whose proof is given in Section S.3 of the Supplementary Material.

LEMMA A.6. *Under conditions for Theorem 3.6, we have*

(i) $\hat{f}_t - H'f^0_t = V^{-1}_{NT}T^{-1}\hat{F}'F^0\frac{1}{N}\sum_{i=1}^{N}[\lambda^0_i u_{it} - \lambda^0_i x'_{it}(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})] + O_P[\delta^{-2}_{NT}\log^2(NT)]$ *uniformly in* t;

(ii) $\hat{\lambda}_i - H^{-1}\lambda^0_i = H'\frac{1}{T}\sum_{t=1}^{T}[f^0_t u_{it} - f^0_t x'_{it}(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot})] + O_P[\delta^{-2}_{NT}\log^2(NT)]$ *uniformly in* i;

(iii) $\hat{\lambda}'_i\hat{f}_t - \lambda^{0\prime}_i f^0_t = \lambda^{0\prime}_i\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\frac{1}{N}\sum_{j=1}^{N}[\lambda^0_j u_{jt} - \lambda^0_j x'_{jt}(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})] + f^{0\prime}_t(T^{-1}F^{0\prime}F^0)^{-1}\frac{1}{T}\sum_{s=1}^{T}[f^0_s u_{is} - f^0_s x'_{is}(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot})] + O_P[\delta^{-2}_{NT}\log^2(NT)]$ *uniformly in* (i,t).

Proof of Theorem 3.6. (i) By Lemma A.6(i) and Theorem 3.4(ii),

$$\sqrt{N}(\hat{f}_t - H'f^0_t)$$
$$= V^{-1}_{NT}T^{-1}\hat{F}'F^0\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\lambda^0_i u_{it} - \lambda^0_i x'_{it}(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})\right] + O_P\left(\sqrt{N}\delta^{-2}_{NT}\log(NT)\right)$$
$$= V^{-1}_{NT}T^{-1}\hat{F}'F^0\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda^0_i u_{it} - \frac{1}{N}\sum_{i=1}^{N}\lambda^0_i x'_{it}\left(\frac{1}{N}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t}\right)^{-1}\frac{1}{\sqrt{N}}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t}\right\}$$
$$+ o_P(1)$$
$$= V^{-1}Q\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda^0_i u_{it} - \mu'_{\cdot t,x\lambda}(\mathbf{s}^0_{\cdot t})^{-1}\frac{1}{\sqrt{N}}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t}\right\} + o_P(1)$$
$$\xrightarrow{d} N\left(0, V^{-1}Q\Gamma_{\cdot t}Q'V^{-1}\right),$$

where the third line follows from the fact that $V_{NT} \xrightarrow{p} V > 0$ and $T^{-1}\hat{F}'F^0 \xrightarrow{p} Q$ as in Bai (2003) and Assumption 6.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

(ii) By Lemma A.6(ii) and Theorem 3.4(i),

$$\begin{aligned}
&\sqrt{T}(\hat{\lambda}_i - H^{-1}\lambda_i^0) \\
&= H' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[f_t^0 u_{it} - f_t^0 x_{it}'(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}) \right] + O_P\left(\sqrt{T}\delta_{NT}^{-2}\log(NT)\right) \\
&= H' \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t^0 u_{it} - \frac{1}{T}\sum_{t=1}^{T} f_t^0 x_{it}' \left(\frac{1}{T}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{x}_{i\cdot}\right)^{-1} \frac{1}{\sqrt{T}}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} \right] + o_P(1) \\
&= \left(Q'\right)^{-1} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t^0 u_{it} - \mu_{i\cdot,xf}'(\mathbf{s}_{i\cdot}^0)^{-1} \frac{1}{\sqrt{T}}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot} \right\} + o_P(1) \\
&\xrightarrow{d} N\left(0, (Q')^{-1}\Gamma_{i\cdot}Q^{-1}\right),
\end{aligned}$$

where the third line follows from the fact that $H \xrightarrow{p} Q^{-1}$ as in Bai (2003) and Assumption 6.

(iii) By Lemma A.6(iii) and Theorem 3.4(i) and (ii),

$$\begin{aligned}
\hat{\lambda}_i'\hat{f}_t - \lambda_i^{0\prime} f_t^0 = {} & \lambda_i^{0\prime}\left(N^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1} \frac{1}{N}\sum_{j=1}^{N}\left[\lambda_j^0 u_{jt} - \lambda_j^0 x_{jt}'(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})\right] \\
& + f_t^{0\prime}\left(T^{-1}F^{0\prime}F^0\right)^{-1} \frac{1}{T}\sum_{s=1}^{T}\left[f_s^0 u_{is} - f_s^0 x_{is}'(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot})\right] + O_P\left(\delta_{NT}^{-2}\log(NT)\right) \\
= {} & \lambda_i^{0\prime}\Sigma_\Lambda^{-1}\left[\frac{1}{N}\sum_{j=1}^{N}\lambda_j^0 u_{jt} - \mu_{\cdot t,x\lambda}'(\mathbf{s}_{\cdot t}^0)^{-1}\frac{1}{N}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t}\right] \\
& + f_t^{0\prime}\Sigma_F^{-1}\left[\frac{1}{T}\sum_{s=1}^{T} f_s^0 u_{is} - \mu_{i\cdot,xf}'(\mathbf{s}_{i\cdot}^0)^{-1}\frac{1}{T}\mathbf{x}_{i\cdot}'\mathbb{M}_{F^0}\mathbf{u}_{i\cdot}\right] + o_P(\delta_{NT}^{-1}).
\end{aligned}$$

Following the arguments as used in the proof of Theorem 3 in Bai (2003), we can show that $\left(V_{1,it} + V_{2,it}\right)^{-1/2}(\hat{\lambda}_i'\hat{f}_t - \lambda_i^{0\prime}f_t^0) \xrightarrow{d} N(0,1)$. □

A.2. Asymptotic Analysis for Section 4

To prove Theorem 4.1, we need the following two lemmas whose proofs are given in Section S.4 of the Supplementary Material.

LEMMA A.7. *Suppose Assumptions 1–8 hold and* $(\sqrt{N_1}/T + \sqrt{T}/N_1)\log^2(N_1T) \to 0$. *Then*

(i) $\hat{\tau}_{it} - \tau_{it} = u_{it} - x_{it}'(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}) - [x_{it}' - \lambda_i^{0\prime}(\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}\mathbf{x}_{\cdot t}](\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}) - \lambda_i^{0\prime}(N_1^{-1}\Lambda^{0\prime}\Lambda^0)^{-1}\frac{1}{N_1}\sum_{j=1}^{N_1}\lambda_j^0 u_{jt} - f_t^{0\prime}H\mathbb{S}_2 S_{i,T_{0,i}}^{-1}\frac{1}{T_{0,i}}\sum_{s=1}^{T_{0,i}} z_{is}u_{is} + O_p(\delta_{N_1T}^{-2}\log^2(N_1T)) \equiv \sum_{l=1}^{5}\xi_{it,l} + O_p(\delta_{N_1T}^{-2}\log^2(N_1T))$ *uniformly in* $i \in [N]\setminus[N_1]$ *and* $t \in [T]\setminus[T_{0,i}]$;

(ii) $\frac{1}{T_{0,i}}\sum_{t=1}^{T_{0,i}}\xi_{it}x_{it}'(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}) = O_P(\delta_{N_1T}^{-2}\log^2(N_1T))$ *uniformly in* i *for* $\xi_{it} = x_{it}, f_t^0$ *and* 1;

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

(iii) $\frac{1}{N_1 T_{0,i}} \sum_{j=1}^{N_1} \sum_{t=1}^{T_{0,i}} \lambda_j^0 x_{jt}'(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}) = O_P(\delta_{N_1 T}^{-2} \log^2(N_1 T))$ *uniformly in* i;
(iv) $\frac{1}{\sqrt{|\mathcal{G}_t|}} \sum_{i \in \mathcal{G}_t} x_{it}'(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}) = o_P(1)$ *as* $|\mathcal{G}_t| \to \infty$;
(v) $\frac{1}{\sqrt{|\mathcal{G}_t|}} \sum_{i \in \mathcal{G}_t} S_{i,T_{0,i}}^{-1} \frac{1}{T_{0,i}} \sum_{s=1}^{T_{0,i}} z_{is} u_{is} = o_P(1)$ *as* $|\mathcal{G}_t| \to \infty$.

LEMMA A.8. *Let* $\xi_{it,l}, l = 2, \ldots, 5$, *be as defined in Lemma A.7(i). Suppose Assumptions 1–8 hold and* $(\sqrt{N_1}/T + \sqrt{T}/N_1) \log^2(N_1 T) \to 0$. *Then*

(i) $\bar{\xi}_{2NT} \equiv \frac{1}{\sqrt{N_2 \bar{T}_1}} \sum_{i=N_1+1}^{N} \sum_{t=T_{0,i}+1}^{T} \xi_{it,2} = -\frac{1}{\sqrt{N_2}} \sum_{i=N_1+1}^{N} \frac{T_{1,i}}{\sqrt{T_{0,i} \bar{T}_1}} \mu_{i\cdot,x}' \mathbb{S}_1 S_i^{-1} \frac{1}{\sqrt{T_{0,i}}} \sum_{t=1}^{T_{0,i}} z_{it} u_{it} + o_P(1)$;

(ii) $\bar{\xi}_{3NT} \equiv \frac{1}{\sqrt{N_2 \bar{T}_1}} \sum_{i=N_1+1}^{N} \sum_{t=T_{0,i}+1}^{T} \xi_{it,3} = o_P(1)$;

(iii) $\bar{\xi}_{4NT} \equiv \frac{1}{\sqrt{N_2 \bar{T}_1}} \sum_{i=N_1+1}^{N} \sum_{t=T_{0,i}+1}^{T} \xi_{it,4} = o_P(1)$;

(iv) $\bar{\xi}_{5NT} \equiv \frac{1}{\sqrt{N_2 \bar{T}_1}} \sum_{i=N_1+1}^{N} \sum_{t=T_{0,i}+1}^{T} \xi_{it,5} = -\frac{1}{\sqrt{N_2}} \sum_{i=N_1+1}^{N} \frac{T_{1,i}}{\sqrt{T_{0,i} \bar{T}_1}} \mu_f' H \mathbb{S}_2 S_{i,T_{0,i}}^{-1} \frac{1}{\sqrt{T_{0,i}}} \sum_{s=1}^{T_{0,i}} z_{is} u_{is} + o_P(1)$.

Proof of Theorem 4.1. (i) By Lemma A.7(i), we have that uniformly in $i \in [N] \setminus [N_1]$ and $t \in [T] \setminus [T_{0,i}]$,

$$\begin{aligned}
\hat{\tau}_{it} - \tau_{it} &= u_{it} - x_{it}'(\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}) - [x_{it}' - \lambda_i^{0\prime} \left(\Lambda^{0\prime} \Lambda^0\right)^{-1} \Lambda^{0\prime} \mathbf{x}_{\cdot t}](\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t}) \\
&\quad - \lambda_i^{0\prime} \left(N_1^{-1} \Lambda^{0\prime} \Lambda^0\right)^{-1} \frac{1}{N_1} \sum_{j=1}^{N_1} \lambda_j^0 u_{jt} - f_t^{0\prime} H \mathbb{S}_2 S_{i,T_{0,i}}^{-1} \frac{1}{T_{0,i}} \sum_{s=1}^{T_{0,i}} z_{is} u_{is} \\
&\quad + O_p[\delta_{N_1 T}^{-2} \log^2(N_1 T)] \\
&\equiv \xi_{it,1} + \xi_{it,2} + \xi_{it,3} + \xi_{it,4} + \xi_{it,5} + O_p[\delta_{N_1 T}^{-2} \log^2(N_1 T)].
\end{aligned}$$

Then

$$\frac{1}{\sqrt{T_{1,i}}} \sum_{t=T_{0,i}+1}^{T} (\hat{\tau}_{it} - \tau_{it}) = \sum_{l=1}^{5} \frac{1}{\sqrt{T_{1,i}}} \sum_{t=T_{0,i}+1}^{T} \xi_{it,l} + o_P(1) \equiv \sum_{l=1}^{5} \bar{\xi}_{i\cdot,lT} + o_P(1).$$

It suffices to study each term in the last summation. First, $\bar{\xi}_{i\cdot,1T} = \frac{1}{\sqrt{T_{1,i}}} \sum_{t=T_{0,i}+1}^{T} u_{it} \xrightarrow{d} N(0, \sigma_{i\cdot,u}^2)$ by Assumption 8(iv). For $\bar{\xi}_{i\cdot,2T}$ and $\bar{\xi}_{i\cdot,5T}$, we have, by Theorem 3.4(i),

$$\begin{aligned}
\bar{\xi}_{i\cdot,2T} &= -\sqrt{\frac{T_{1,i}}{T_{0,i}}} \frac{1}{T_{1,i}} \sum_{t=T_{0,i}+1}^{T} x_{it}' \sqrt{T_{0,i}} (\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}) \\
&= -\sqrt{\frac{1 - r_i}{r_i}} \mu_{i\cdot,x}' \sqrt{T_{0,i}} (\hat{\beta}_{i\cdot} - \dot{\beta}_{i\cdot}) + o_P(1)
\end{aligned}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

$$= -\sqrt{\frac{1-r_i}{r_i}}\mu'_{i\cdot,x}\mathbb{S}_1 S_{i,T_{0,i}}^{-1}\frac{1}{\sqrt{T_{0,i}}}\sum_{t=1}^{T_{0,i}} z_{it}u_{it} + o_P(1) \text{ and}$$

$$\bar{\xi}_{i\cdot,5T} = -\sqrt{\frac{T_{1,i}}{T_{0,i}}}\frac{1}{T_{1,i}}\sum_{t=T_{0,i}+1}^{T} f_t^{0\prime} H\mathbb{S}_2 S_{i,T_{0,i}}^{-1}\frac{1}{\sqrt{T_{0,i}}}\sum_{s=1}^{T_{0,i}} z_{is}u_{is}$$

$$= -\sqrt{\frac{1-r_i}{r_i}}\mu_f H\mathbb{S}_2 S_{i,T_{0,i}}^{-1}\frac{1}{\sqrt{T_{0,i}}}\sum_{s=1}^{T_{0,i}} z_{is}u_{is} + o_P(1).$$

By Lemma A.7(ii) and (iii),

$$\bar{\xi}_{i\cdot,3T} = -\frac{1}{\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T} x'_{it}(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})$$

$$+\lambda_i^{0\prime}\left(N_1^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\frac{1}{N_1\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T}\Lambda^{0\prime}\mathbf{x}_{\cdot t}(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})$$

$$= \sqrt{T_{1,i}}O_p[\delta_{N_1T}^{-2}\log^2(N_1T)] = o_P(1).$$

In addition, $\bar{\xi}_{i\cdot,4T} = -\lambda_i^{0\prime}\left(N_1^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\frac{1}{N_1\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T}\sum_{j=1}^{N_1}\lambda_j^0 u_{jt} = o_P(1)$ under our assumptions. It follows that

$$\frac{1}{\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T}(\hat{\tau}_{it} - \tau_{it}) = \bar{\xi}_{i\cdot,1T} + \bar{\xi}_{i\cdot,2T} + \bar{\xi}_{i\cdot,5T} + o_P(1)$$

$$= \frac{1}{\sqrt{T_{1,i}}}\sum_{t=T_{0,i}+1}^{T} u_{it} - \sqrt{\frac{1-r_i}{r_i}}\left(\mu'_{i\cdot,x}\mathbb{S}_1 + \mu'_f H\mathbb{S}_2\right)S_{i,T_{0,i}}^{-1}\frac{1}{\sqrt{T_{0,i}}}\sum_{s=1}^{T_{0,i}} z_{is}u_{is} + o_P(1)$$

$$\xrightarrow{d} N\left(0, \sigma_{i\cdot,u}^2 + \frac{1-r_i}{r_i}[\mu'_{i\cdot,x}\mathbb{S}_1 + \mu'_f Q^{-1}\mathbb{S}_2]S_i^{-1}\Omega_{i,zu}S_i^{-1}[\mathbb{S}'_1\mu_{i\cdot,x} + \mathbb{S}'_2(Q')^{-1}\mu_f]\right),$$

where we recall that $S_i = \text{plim}_{T_{0,i}\to\infty} S_{i,T_{0,i}}$.

(ii) For any $t \geq \min_i T_{0,i} + 1$, we have, by Lemma A.7(i),

$$\frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}(\hat{\tau}_{it} - \tau_{it}) = \sum_{l=1}^{5}\frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}\xi_{it,l} + o_P(1) \equiv \sum_{l=1}^{5}\bar{\xi}_{\cdot t,lN} + o_P(1).$$

Note that $\bar{\xi}_{\cdot t,1N} = \frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t} u_{it} \xrightarrow{d} N\left(0, \sigma_{\cdot t,u}^2\right)$ by Assumption 8(v). For $\bar{\xi}_{\cdot t,3N}$ and $\bar{\xi}_{\cdot t,4N}$, we have, by Theorem 3.4(ii),

$$\bar{\xi}_{\cdot t,3N} = -\sqrt{\frac{|\mathcal{G}_t|}{N_1}}\frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t}[x'_{it} - \lambda_i^{0\prime}\left(\Lambda^{0\prime}\Lambda^0\right)^{-1}\Lambda^{0\prime}\mathbf{x}_{\cdot t}]\sqrt{N_1}(\hat{\beta}_{\cdot t} - \dot{\beta}_{\cdot t})$$

$$= -\sqrt{r_{\cdot t}}'\left(\mathbf{s}_{\cdot t}^0\right)^{-1}\frac{1}{\sqrt{N_1}}\mathbf{x}'_{\cdot t}\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t} + o_P(1), \text{ and}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

$$\bar{\xi}_{\cdot t,4N} = -\sqrt{\frac{|\mathcal{G}_t|}{N_1}}\frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t}\lambda_i^{0\prime}\left(N_1^{-1}\Lambda^{0\prime}\Lambda^0\right)^{-1}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\lambda_j^0 u_{jt}$$
$$= -\sqrt{r_{\cdot t}}\mu_\lambda'\Sigma_\Lambda^{-1}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\lambda_j^0 u_{jt} + o_P(1),$$

where $r_{\cdot t} = \lim_{(|\mathcal{G}_t|,N_1)\to\infty}|\mathcal{G}_t|/N_1$, $\nu_{\cdot t,x\lambda}' = \text{plim}_{|\mathcal{G}_t|\to\infty}\frac{1}{|\mathcal{G}_t|}\sum_{i\in\mathcal{G}_t}[x_{it}' - \lambda_i^{0\prime}(\Lambda^{0\prime}\Lambda^0)^{-1}\Lambda^{0\prime}\mathbf{x}_{\cdot t}]$, and $\mathbf{s}_{\cdot t}^0 = \text{plim}_{N_1\to\infty}\frac{1}{N_1}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{x}_{\cdot t}$. By Lemma A.7 (iv) and (v), $\bar{\xi}_{\cdot t,2N} = -\frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}x_{it}'(\hat{\beta}_{i\cdot} - \beta_{i\cdot}) = o_P(1)$, and $\bar{\xi}_{\cdot t,5N} = -f_t^{0\prime}H\mathbb{S}_2\frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}S_{i,T_{0,i}}^{-1}\frac{1}{T_{0,i}}\sum_{s=1}^{T_{0,i}}z_{is}u_{is} = o_P(1)$. In sum, we have

$$\frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}\left(\hat{\tau}_{it} - \tau_{it}\right)$$
$$= \bar{\xi}_{\cdot t,1N} + \bar{\xi}_{\cdot t,3N} + \bar{\xi}_{\cdot t,4N} + o_P(1)$$
$$= \frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}u_{it}$$
$$- \sqrt{r_{\cdot t}}\left\{\nu_{\cdot t,x\lambda}'\left(\mathbf{s}_{\cdot t}^0\right)^{-1}\frac{1}{\sqrt{N_1}}\mathbf{x}_{\cdot t}'\mathbb{M}_{\Lambda^0}\mathbf{u}_{\cdot t} - \mu_\lambda'\Sigma_\Lambda^{-1}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\lambda_j^0 u_{jt}\right\} + o_P(1)$$
$$= \frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}u_{it}$$
$$- \sqrt{r_{\cdot t}}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\left\{\nu_{\cdot t,x\lambda}'\left(\mathbf{s}_{\cdot t}^0\right)^{-1}[x_{jt} - \mathbf{x}_{\cdot t}'\Lambda^0\left(\Lambda^{0\prime}\Lambda^0\right)^{-1}\lambda_j^0]u_{jt} - \mu_\lambda'\Sigma_\Lambda^{-1}\lambda_j^0 u_{jt}\right\} + o_P(1)$$
$$= \frac{1}{\sqrt{|\mathcal{G}_t|}}\sum_{i\in\mathcal{G}_t}u_{it} - \sqrt{r_{\cdot t}}\frac{1}{\sqrt{N_1}}\sum_{j=1}^{N_1}\left[\nu_{\cdot t,x\lambda}'\left(\mathbf{s}_{\cdot t}^0\right)^{-1}(x_{jt} - \pi_{t,x\lambda}'\lambda_j^0) - \mu_\lambda'\Sigma_\Lambda^{-1}\lambda_j^0\right]u_{jt} + o_P(1)$$
$$\xrightarrow{d} N\left(0, \sigma_{\cdot t,u}^2 + r_{\cdot t}\Xi_{\cdot t}\right),$$

where we recall that $\pi_{t,x\lambda}' = \text{plim}_{N_1\to\infty}\mathbf{x}_{\cdot t}'\Lambda^0(\Lambda^{0\prime}\Lambda^0)^{-1}$ and $\Xi_{\cdot t} = \lim_{N_1\to\infty}\frac{1}{N_1}\sum_{j=1}^{N_1}E\{[\nu_{\cdot t,x\lambda}'\left(\mathbf{s}_{\cdot t}^0\right)^{-1}(x_{jt} - \pi_{t,x\lambda}'\lambda_j^0) - \mu_\lambda'\Sigma_\Lambda^{-1}\lambda_j^0]u_{jt}\}^2$.

(iii) $\bar{T}_1 \to \infty$ and N_2 is either fixed or diverge to infinity such that $N_2\bar{T}_1 = o(\delta_{N_1T}^4[\log(N_1T)]^{-2})$ and $N_2\log N = o\left(\underline{T}_0\right)$,

$$\frac{1}{\sqrt{N_2\bar{T}_1}}\sum_{i=N_1+1}^{N}\sum_{t=T_{0,i}+1}^{T}\left(\hat{\tau}_{it} - \tau_{it}\right) = \frac{1}{\sqrt{N_2\bar{T}_1}}\sum_{l=1}^{5}\sum_{i=N_1+1}^{N}\sum_{t=T_{0,i}+1}^{T}\xi_{it,l} + o_P(1)$$
$$\equiv \sum_{l=1}^{5}\bar{\xi}_{lNT} + o_P(1).$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

By Lemma A.8(i)–(iv),

$$\begin{aligned}
&\frac{1}{\sqrt{N_2\bar{T}_1}}\sum_{i=N_1+1}^{N}\sum_{t=T_{0,i}+1}^{T}\left(\hat{\tau}_{it}-\tau_{it}\right)\\
&=\bar{\xi}_{1NT}+\bar{\xi}_{2NT}+\bar{\xi}_{5NT}+o_P(1)\\
&=\frac{1}{\sqrt{N_2\bar{T}_1}}\sum_{i=N_1+1}^{N}\sum_{t=T_{0,i}+1}^{T}u_{it}\\
&\quad-\frac{1}{\sqrt{N_2}}\sum_{i=N_1+1}^{N}\frac{T_{1,i}}{\sqrt{T_{0,i}\bar{T}_1}}(\mu'_{i\cdot x,}\mathbb{S}_1+\mu'_f H\mathbb{S}_2)S^{-1}_{i,T_{0,i}}\frac{1}{\sqrt{T_{0,i}}}\sum_{t=1}^{T_{0,i}}z_{it}u_{it}+o_P(1)\\
&\xrightarrow{d}N\left(0,\sigma_u^2+\Xi\right),
\end{aligned}$$

where $\Xi=\lim_{(N,T)\to\infty}\frac{1}{N_2}\sum_{i=N_1+1}^{N}\alpha_{i\cdot}\frac{1}{T_{0,i}}\sum_{t,s=1}^{T_{0,i}}E\left[z_{it}z'_{is}u_{it}u_{is}\right]\alpha'_{i\cdot}$, $\alpha_{i\cdot}=\lim_{T\to\infty}\alpha_{i,T}$, and $\alpha_{i,T}=\frac{T_{1,i}}{\sqrt{T_{0,i}\bar{T}_1}}(\mu'_{i\cdot x,}\mathbb{S}_1+\mu'_f H\mathbb{S}_2)S^{-1}_{i,T_{0,i}}$. □

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466626100620.

FUNDING STATEMENT

X.L. acknowledges support from the Hong Kong Research Grants Council (RGC) under Grant Number 14500022 and the Chinese University of Hong Kong for the start-up fund. K.M. acknowledges the financial support from the National Natural Science Foundation of China (NSFC) under Grant Numbers 72573042 and 72121002. L.S. thanks the NSFC for financial support under Grant Number 72133002.

COMPETING INTEREST STATEMENT

The authors declare that no competing interests exist.

REFERENCES

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

Abadie, A., Diamond, A., & 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.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

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.

Andrews, D. W., & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60, 953–966.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12), 4088–4118.

Armstrong, T. B., Weidner, M., & Zeleneev, A. (2026). Robust estimation and inference in panels with interactive fixed effects. *Journal of Political Economy*, 134(8), 2223–2261. DOI: http://doi.org/10.1086/741623

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.

Athey, S., & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), 3–32.

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. (2019). Rank regularized estimation of approximate factor models. *Journal of Econometrics*, 212(1), 78–96.

Belloni, A., Chen, M., Padilla, O. H. M., & Wang, Z. K. (2023). High-dimensional latent panel quantile regression with an application to asset pricing. *Annals of Statistics*, 51(1), 96–121.

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

Billmeier, A., & Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *Review of Economics and Statistics*, 95(3), 983–1001.

Browning, M., & Carro, J. (2007). Heterogeneity and microeconometrics modelling. *Econometric Society Monographs*, 43, 47.

Cattaneo, M. D., Feng, Y., & Titiunik, R. (2021). Prediction intervals for synthetic control methods. *Journal of the American Statistical Association*, 116(536), 1865–1880.

Chen, J. (2023). Synthetic control as online linear regression. *Econometrica*, 91(2), 465–491.

Chernozhukov, V., Hansen, C., Liao, Y., & Zhu, Y. (2026). Inference for heterogeneous effects using low-rank estimation of factor slopes. *Review of Economics and Statistics*. Forthcoming. DOI: http://doi.org/10.1162/REST.a.1735

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

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

Durlauf, S. N., Johnson, P. A., & Temple, J. R. (2005). Growth econometrics. *Handbook of Economic Growth*, 1, 555–677.

Fan, J., Gong, W., & Zhu, Z. (2019). Generalized high-dimensional trace regression via nuclear norm regularization. *Journal of Econometrics*, 212(1), 177–202.

Feldstein, M., & Horioka, C. (1980). Domestic saving and international capital flows. *The Economic Journal*, 90(358), 314–329.

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

Heckman, J. J. (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. *Journal of Political Economy*, 109(4), 673–748.

Hong, S., Su, L., & Jiang, T. (2023). Profile gmm estimation of panel data models with interactive fixed effects. *Journal of Econometrics*, 235(2), 927–948.

Hsiao, C. (2022). *Analysis of panel data*. Cambridge University Press.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1), 5–86.

Jiang, B., Yang, Y., Gao, J., & Hsiao, C. (2021). Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics*, 224(2), 439–465.

Jin, S., Miao, K., & Su, L. (2021). On factor models with random missing: EM estimation, inference, and cross validation. *Journal of Econometrics*, 222(1), 745–777.

Keane, M., & Neal, T. (2020). Climate change and us agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics*, 11(4), 1391–1429.

Koltchinskii, V., Lounici, K., & Tsybakov, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Annals of Statistics*, 39(5), 2302–2329.

Lazarus, E., Lewis, D. J., Stock, J. H., & Watson, M. W. (2018). HAR inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4), 541–559.

Li, K., Cui, G., & Lu, L. (2020). Efficient estimation of heterogeneous coefficients in panel data models with common shocks. *Journal of Econometrics*, 216(2), 327–353.

Li, K. T. (2020). Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, 115(532), 2068–2083.

Lu, X., & Su, L. (2016). Shrinkage estimation of dynamic panel data models with interactive fixed effects. *Journal of Econometrics*, 190(1), 148–175.

Lu, X., & Su, L. (2023). Uniform inference in linear panel data models with two-dimensional heterogeneity. *Journal of Econometrics*, 235(2), 694–719.

Ma, S., Su, L., & Zhang, Y. (2022). Detecting latent communities in network formation models. *Journal of Machine Learning Research*, 23(310), 1–61.

Masini, R., & Medeiros, M. C. (2021). Counterfactual analysis with artificial controls: Inference, high dimensions, and nonstationarity. *Journal of the American Statistical Association*, 116(536), 1773–1788.

Masini, R., & Medeiros, M. C. (2022). Counterfactual analysis and inference with nonstationary data. *Journal of Business & Economic Statistics*, 40(1), 227–239.

Miao, K., Phillips, P. C., & Su, L. (2023). High-dimensional vars with common factors. *Journal of Econometrics*, 233(1), 155–183.

Moon, H. R., & Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4), 1543–1579.

Moon, H. R., & Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 33(1), 158–195.

Moon, H. R., & Weidner, M. (2026). Nuclear norm regularized estimation of panel regression models. *Journal of Econometrics*, 255, 106246.

Negahban, S., & Wainwright, M. J. (2011). Estimation of (near) low-rank matrices with noise and high-dimensional scaling. *Annals of Statistics*, 39(2), 1069–1097.

Negahban, S., & Wainwright, M. J. (2012). Restricted strong convexity and weighted matrix completion: Optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1), 1665–1697.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., & Yu, B. (2012). A unified framework for high-dimensional analysis of *m*-estimators with decomposable regularizers. *Statistical Science*, 27(4), 538–557.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4), 1004–1016.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967–1012.

Rodriguez, F., & Rodrik, D. (2000). Trade policy and economic growth: A skeptic's guide to the cross-national evidence. *NBER Macroeconomics Annual*, 15, 261–325.

Rohde, A., & Tsybakov, A. B. (2011). Estimation of high-dimensional low-rank matrices. *Annals of Statistics*, 39(2), 887–930.

Romalis, J. (2007). Market access, openness and growth. NBER Working paper.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620

Sachs, J. D., Warner, A., Åslund, A., & Fischer, S. (1995). Economic reform and the process of global integration. *Brookings Papers on Economic Activity*, 1995(1), 1–118.

Su, L., & Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, 29(6), 1079–1135.

Su, L., Shi, Z., & Phillips, P. C. B. (2016). Identifying latent structures in panel data. *Econometrica*, 84(6), 2215–2264.

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

Wacziarg, R., & Welch, K. H. (2008). Trade liberalization and growth: New evidence. *The World Bank Economic Review*, 22(2), 187–231.

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.

Yu, Y., Wang, T., & Samworth, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2), 315–323.

Zhu, Y. (2017). High-dimensional panel data with time heterogeneity: Estimation and inference. Working paper.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 14 Nov 2026 at 03:53:24, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266466626100620