Large Dimensional Latent Factor Modeling with Missing Observations and Applications to Causal Inference

Ruoxuan Xiong * and Markus Pelger †

First version: October 16, 2019

This version: November 19, 2019

Abstract

This paper develops the inferential theory for latent factor models estimated from large dimensional panel data with missing observations. We estimate a latent factor model by applying principal component analysis to an adjusted covariance matrix estimated from partially observed panel data. We derive the asymptotic distribution for the estimated factors, loadings and the imputed values under a general approximate factor model. The key application is to estimate counterfactual outcomes in causal inference from panel data. The unobserved control group is modeled as missing values, which are inferred from the latent factor model. The inferential theory for the imputed values allows us to test for individual treatment effects at any time. We apply our method to portfolio investment strategies and find that around 14% of their average returns are significantly reduced by the academic publication of these strategies.

Keywords: Factor Analysis, Principal Components, Synthetic Control, Causal Inference, Treatment Effect, Missing Entry, Large-Dimensional Panel Data, Large N and T, Matrix Completion

JEL classification: C14, C38, C55, G12

*Stanford University, Department of Management Science & Engineering, Email: rxiong@stanford.edu.

†Stanford University, Department of Management Science & Engineering, Email: mpelger@stanford.edu.

We thank Susan Athey, Mohsen Bayati, Guillaume Basse, Jianqing Fan, Kay Giesecke, Peter Glynn, Lisa Goldberg, Serena Ng, Sam Pimentel, seminar and conference participants at Stanford, INFORMS and the Marketplace Innovation Workshop for helpful comments.

1 Introduction

Large dimensional panel data with missing entries are prevalent. In causal panel data, the main focus is to estimate the unobserved potential outcomes. In financial data, stock returns are missing before a company is listed, after its bankruptcy or because of illiquidity. In macroeconomic datasets, panel data might be collected at different frequencies or not for all geographical locations resulting in missing entries. In the famous Netflix challenge, a majority of users' ratings for films are missing. Estimating missing entries in panel data is a fundamental problem with applications in social science, statistics, and computer science.

This paper presents an inferential theory for latent factor models estimated from large dimensional panel data with missing observations. We propose a novel approach to estimate a latent factor model by applying principal component analysis (PCA) to an adjusted covariance matrix, which is estimated from the partially observed panel data. We derive the asymptotic normal distribution for the estimated factors, loadings, and imputed values.

The key application is to estimate counterfactual outcomes for causal inference. The unobserved control group is modeled as missing values, which are inferred from the latent factor model. The inferential theory for the imputed values allows us to test for individual treatment effects at a particular time. This granular test is of practical importance because we learn not only for whom but also when the treatment is effective which allows us to optimally allocate treatments to units over time.

1.1 Main Contribution

Our work contributes to three distinct fields: large dimensional factor modeling, matrix completion and causal inference. First, we extend the inferential theory of latent factors to large dimensional data with general patterns in missing entries. Second, matrix completion methods impute missing entries under the assumption of a low-rank structure which is corrupted with noise. We provide confidence intervals for the imputed values. Lastly, the key question in causal inference is the estimation of counter-factual outcomes, i.e. what would have been the outcome if a unit had not been treated or if a unit had been treated. The unobserved counter-factual outcome can naturally be formulated as a missing observation problem. We are the first to provide a test for the point-wise treatment effect that can be heterogeneous and time dependent under general adoption patterns where the units can be affected by unobserved factors.

The inferential theory for latent factor models with missing data is important for a number of reasons. First, we show how to consistently impute the missing observations in a large dimensional data set, which can then be used as an input for other applications. Second, we provide confidence intervals for the imputed values, which serve as a decision criterion if the imputed data should be used. Third, we provide the conditions under which missing values can be inferred. Fourth, the distribution of the missing observations can actually be the object of interest itself. For example, the imputed values serve as the synthetic control for which we need an asymptotic distribution theory. The inferential theory is key for deriving a test statistic for a treatment effect.

Our method is very simple to adopt and but works under general assumptions. Conventional factor estimation in large dimensional panel data applies PCA to a sample covariance matrix, which requires a fully observed balanced panel. To tackle the missing entries in the panel, our estimator replaces them with zeros and re-weights the observed entries. The next step is to simply apply PCA to the covariance matrix of this transformed panel. The missing entries are estimated by the common components of the factor model. We only need to make the standard assumptions of an approximate factor model.

Our framework allows for very general patterns of missing observations. The patterns are modeled as general functions of the unobserved loadings and unit specific features. In this case, the re-weighting of the observed entries is based on a propensity score for which we provide a consistent estimator. Allowing the missing pattern to be a function of unit-specific characteristics is relevant for the causal inference application as the treatment of units is typically not random. Furthermore, we cover the common scenario of a simultaneous and staggered treatment adoption where the treatment cannot be removed once implemented. Our framework also allows for the common case studied in the matrix completion literature[1] that the data is missing independently of the underlying factor model. In this case, the re-weighting of the observed entries is simply based on the proportion of missing to observed entries.

Deriving the inferential theory under these general conditions is a challenging problem. The missing observations have a complex effect on the asymptotic covariance matrix of the imputed

[1] See (Candès and Recht, 2009; Negahban and Wainwright, 2012).

entries. In particular, the asymptotic variance has an additional variance correction term compared with the fully observed panel. This term results in a larger asymptotic variance than in the fully observed case.

In our empirical analysis, we study the effect of academic publications on the return of anomaly portfolios. There is an ongoing debate in asset pricing on whether academic publications result in the disappearance, reversion or attenuation of anomalies in equity returns. An anomaly describes a pattern in average returns that cannot be explained by a benchmark asset pricing model as for example the important Capital Asset Pricing Model (CAPM). Schwert (2003), McLean and Pontiff (2016) and Chen and Zimmermann (2018) suggest that the return of anomalies are reduced after their publication, mainly because investors become aware of the effect and correct the mispricing. Our novel methodology allows us to test if the average return or pricing error of an anomaly portfolio is significantly reduced by its publication. At a 5% confidence level, merely 14% of the anomalies are significantly reduced by publication. Importantly, a naive estimation of the publication effect which simply compares time-series means before or after the publication date is more likely to find an effect as the sample mean returns are in general lower in the latter part of the data set. Our approach correctly accounts for time effects and the uncertainty in the estimation showing that the risk premium of most "classical" anomalies have not been affected by publication.

1.2 Related Literature

We show the inferential theory for large dimensional factor models from incomplete panel data with general missing patterns. This paper works under the framework of an approximate factor structure where both the cross-section dimension and time-series dimension are large. When the data is fully observed, Bai and Ng (2002) show that the factor model can be estimated with PCA applied to the covariance matrix of the data. Bai (2003) derives the consistency and asymptotic normality of the estimated factors, loadings and common components, which are the product of factors and loadings. Bai (2009) extends the inferential theory to a model with observed covariates and latent factors.[2] When a panel has missing entries, a common approach is to estimate the factor

[2]A current active research topic is to extend the constant loading factor model to a time-varying loading model by using a projection in the cross-section dimension (Fan et al., 2016; Kelly et al., 2018), a local window or high frequency approach (Pelger, 2019; Aït-Sahalia and Xiu, 2018) or a kernel projection in the time dimension (Pelger and Xiong, 2018). However, the current literature relies on a fully observed panel data set of the projected data.

model from a subset of the data for which a balanced panel is available. This approach has two drawbacks: First, it is in general less efficient as our approach makes use of all the data. Second, it can lead to a biased estimate if the data is not missing at random. For example, a complete panel of stock returns suffers from survivorship bias as only companies can be included that did not go bankrupt.

The inferential theory of large dimensional factor models with missing observations is an active area of research. Our paper is most closely related to the recent papers by Su et al. (2019), Bai and Ng (2019) and Chen et al. (2019). The papers differ in the algorithms to impute the missing observations, the generality of the missing patterns and the proportion of required observed entries relative to the missing entries. There is a trade-off in terms of generality of the model and required observations, where our work allows the most general pattern in missing observations with a general approximate factor structure at the cost of observing entries at the same rate as missing entries. Importantly, in contrast to the other papers our framework allows the missing pattern to depend on unit specific features and to test for an individual treatment effect at any time for any cross-section unit or a weighted treatment effect. This is exactly what we need for the main application in causal inference. Su et al. (2019) estimate the latent factor model with the expectation–maximization (EM) algorithm under the assumption of randomly missing values.[3] Independently and simultaneously, Bai and Ng (2019) provide the inferential theory for the factor-based imputed values based on the innovative idea of shuffling rows and columns such that there exist fully observed TALL and WIDE blocks for estimating the factor model.[4] Chen et al. (2019) approach the problem from a matrix completion perspective which can also be mapped into a factor model framework. They solve a nuclear norm regularized optimization problem to estimate

[3]Stock and Watson (2002b); Bańbura and Modugno (2014); Negahban and Wainwright (2012) propose to use EM algorithms to estimate the factor model from the panel data with missing observations. Giannone et al. (2008); Doz et al. (2011); Jungbacker et al. (2011); Stock and Watson (2016) propose to use the state space framework and Kalman Filtering to estimate the factor model with missing observations. Gagliardini et al. (2019) propose a simple diagnostic criterion for approximate factor structure in large (unbalanced) panel datasets. Other work to impute missing values using EM algorithms includes Rubin (1976); Dempster et al. (1977); Meng and Rubin (1993) that study the problem under a different framework, i.e., on cross-sectional data (but not panel data).

[4]Our paper differs from Bai and Ng (2019) in three aspects: 1. We allow the observational pattern to depend on the loadings or observed covariates; 2. We provide general tests for treatment effects, such as an individual treatment effect at any time or a weighted treatment effect. 3. Their re-shuffling of rows and columns imposes some restrictions on the missing patterns and might result in using less observations for estimating missing entries. Our first point requires a re-weighting of the observed entries by a generalized propensity score which we assume to be positive. The second point, requires the number of observed and missing entries to grow at a similar rate. The third point complicates our derivations of the inferential theory as we have to deal with many local rotation matrices of the latent factors.

the missing entries and develop an inferential theory under the assumption of random sampling and i.i.d. noise. The last two papers require less observed entries than our framework, which is relevant for problems such as the Netflix challenge, but have restrictive assumptions on missing patterns or the factor model, which limits the application to causal inference in the social science, which is our main objective.

Our imputed values are point-wise consistent and have asymptotic normal distributions which is relevant for the matrix completion literature that studies a similar problem. Both our paper and the matrix completion literature assume a low-rank structure in the panel data. In the matrix completion literature, the most popular method is to estimate the low-rank matrix from a convex optimization problem.[5] The main results in the matrix completion literature are upper bounds for the mean-squared estimation error for the estimated matrix. However, point-wise consistency does not hold in general because the typically used nuclear norm regularization results in a bias in the estimated matrix. In their path-breaking work, Chen et al. (2019) propose de-biased estimators and provide an inferential theory under the assumption of i.i.d. sampling and i.i.d. noise. Our paper contributes to the matrix completion literature by allowing general observation patterns and dependent error structures, which is particularly relevant for applications in the social science.

Our paper allows for heterogeneous and time dependent treatment effects of an intervention and general intervention adoption patterns compared with the synthetic control methods in causal inference. Furthermore, our paper provides a flexible test for the treatment effects. In comparative case studies, a key question is to estimate the counter-factual outcomes for the treated units. A valid control unit is "close" to the treatment unit except for the treatment effect. Typically synthetic controls are weighted averages of untreated units where the weights depend on unit specific features. A popular model assumption is that the potential outcome is linear in observed covariates and unobserved common factors (Abadie et al., 2010, 2015). Abadie et al. (2010, 2015), Hsiao et al. (2012), Doudchenko and Imbens (2016), Li and Bell (2017), Li (2017), Carvalho et al. (2018), and Masini and Medeiros (2018) propose to match each treated unit by weighted averages of all control units using the pretreatment observations. Li and Bell (2017) and Li (2017) further

[5]The conventional optimization problem is to minimize the mean squared error between the observations and the corresponding entries in the estimated matrix while regularizing the nuclear norm of the estimated matrix (Mazumder et al., 2010; Negahban and Wainwright, 2011, 2012). The nuclear norm of a matrix is similar to the ℓ_1 norm of a vector. The optimal solution has a lower rank if the nuclear norm has more weight in the objective function.

show the inferential theory for the average treatment effect over time. Li and Bell (2017) propose using the LASSO method to select control units and Carvalho et al. (2018) show the inferential theory for the LASSO method. Masini and Medeiros (2018) focus on the high-dimensional, non-stationary data. These methods rely on the assumption that there is only one treated unit and the treatment effects are either constant or stationary. Another method is to regress the post-treatment outcomes for the control units on the pre-treatment outcomes and covariates and use the coefficients to predict the counter-factual outcome for the treated/control units. Athey et al. (2018) proposes to use matrix completion methods to complete the control panel data and allow for more general treatment adoption patterns: multiple treated units and staggered treatment adoption. However, the point-wise guarantee for the imputed values is not provided in Athey et al. (2018). In this paper, we do not only allow for general treatment adoption patterns, but also provide the point-wise inferential theory for the imputed counter-factual outcomes. Furthermore, we can test for treatment effects even if they are heterogeneous and time dependent. Our approach does not require a priori knowledge on which covariates describe if a treated and control units are a good match. Instead, our latent loadings capture all unit-specific information in a data-driven way. The synthetic control that we impute is a weighted average of the untreated units that takes all unit-specific information into account.

The rest of the paper is organized as follows. Section 2 introduces the model and provides the estimator for factors, loadings, and common components. Section 3 states the necessary assumptions for our theoretical results. Section 4 shows the asymptotic results and the tests for the point-wise treatment effect. Section 5 provides a feasible estimator for the propensity score which is needed as a weight to construct our estimator. Section 6 demonstrates simulation results. In our empirical analysis in section 7 we study the effect of academic publications on investment strategies. Section 8 concludes the paper. Additional results and the proofs are collected in the Appendix.

2 Model and Estimation

2.1 Model

Assume we partially observe a panel data set with T time periods and N cross-sectional units. This panel data has a factor structure with r common factors. Denote X_{it} as the cross-sectional observation i at time t, $F_t \in \mathbb{R}^{r\times 1}$ as the latent factors at time t, $\lambda_i \in \mathbb{R}^{r\times 1}$ as the factor loadings of the cross-sectional unit i and e_{it} as the idiosyncratic error:

$$X_{it} = \lambda_i^\top F_t + e_{it} \quad i = 1, 2, \cdots, N \text{ and } t = 1, 2, \cdots, T$$

or in vector notation,

$$\underbrace{X_t}_{N\times 1} = \underbrace{\Lambda}_{N\times r} \underbrace{F_t}_{r\times 1} + \underbrace{e_t}_{N\times 1} \qquad \text{for } t = 1, 2, \cdots, T.$$

In an asymptotic setup where N and T are both large, we randomly observe some entries in $X = [X_1, X_2, \cdots, X_T] \in \mathbb{R}^{N\times T}$. Let $W_{it} \in \{0, 1\}$ be the indicator variable, where $W_{it} = 1$ indicates that the (i,t)-th entry is observed and $W_{it} = 0$ otherwise. In this paper we will estimate the latent factors F and loadings Λ from the partially observed X, impute the missing values and provide the inferential theory for all estimators.

2.2 Estimation

There are two steps to estimate the latent factor model from the partially observed panel data: First, we need to estimate the covariance matrix of the data and second we estimate the latent factors and loadings based on the eigenvectors of the estimated covariance matrix. The conventional latent factor estimator without missing values applies principal component analysis to the sample covariance matrix. A natural way to deal with the missing values is to set these entries to zero. However, the conventional PCA estimator will then be biased. Our estimator correctly reweights the entries in the covariance matrix before applying PCA.

We first impute the missing entries by 0 and denote the imputed matrix as $\widetilde{X}$:

$$\widetilde{X}_{it} = X_{it} W_{it}, \quad \text{for } i = 1, 2, \cdots, N \text{ and } t = 1, 2, \cdots, T$$

In matrix notation, we have $\widetilde{X} = X \odot W$, where $\odot$ denotes the Hadamard product.

When some entries are missing in X, the conventional sample covariance estimator $\frac{1}{T}\widetilde{X}\widetilde{X}^\top$ is biased because the actual realization of the missing values is not equal to zero. We propose the natural estimator of the covariance matrix where for each entry we only use the time periods when both units are observed. This is equivalent to estimating the sample covariance matrix with

$\widetilde{X}$, but reweighting the entries. Figure 1 is a simple example to illustrate the covariance matrix estimation if for a part of the cross-section the entries are missing in the second half of the data. More generally, our sample covariance matrix estimator equals

$$\widetilde{\Sigma}_{ij} = \frac{1}{|\mathcal{Q}_{ij}|} \sum_{t \in \mathcal{Q}_{ij}} X_{it} X_{jt}, \tag{1}$$

where $\mathcal{Q}_{ij}$ is the set of time periods t when both units i and j are observed. Under the assumptions imposed in this paper, $\frac{1}{|\mathcal{Q}_{ij}|} \sum_{t \in \mathcal{Q}_{ij}} X_{it} X_{jt}$ is a consistent estimator for the covariance between unit i and j.

$\mathbf{X}_{1,1} \quad \cdots \quad \mathbf{X}_{1,T_0}$	$\mathbf{X}_{1,T_0+1} \quad \cdots \quad \mathbf{X}_{1,T}$
$\mathbf{X}_{2,1} \quad \cdots \quad \mathbf{X}_{2,T_0}$	$\mathbf{X}_{2,T_0+1} \quad \cdots \quad \mathbf{X}_{2,T}$

(a) Observation pattern for X: Shaded entries are missing.

$\frac{1}{T_0} \sum_{t=1}^{T_0} \mathbf{X}_{1,t} \mathbf{X}_{1,t}^\top$	$\frac{1}{T_0} \sum_{t=1}^{T_0} \mathbf{X}_{1,t} \mathbf{X}_{2,t}$
$\frac{1}{T_0} \sum_{t=1}^{T_0} \mathbf{X}_{2,t} \mathbf{X}_{1,t}^\top$	$\frac{1}{T} \sum_{t=1}^{T} \mathbf{X}_{2,t} \mathbf{X}_{2,t}^\top$

(b) Sample covariance matrix $\widetilde{\Sigma}$: Shaded entries are estimated using observations up to time T_0

Table 1: Covariance matrix estimation for X with missing entries. For $t = T_0 + 1, ..., T$ the first N_0 cross section units are missing, while the elements $N_0 + 1, ..., N$ are observed for all t, i.e. $\mathbf{X}_{1,t} = \begin{pmatrix} X_{1,t} & \cdots & X_{N_0,t} \end{pmatrix}$ and $\mathbf{X}_{2,t} = \begin{pmatrix} X_{N_0+1,t} & \cdots & X_{N,t} \end{pmatrix}$.

When the data is fully observed, we can apply Principal Component Analysis (PCA) to $\frac{1}{NT} X X^\top$ to estimate the loadings.[6] Up to rescaling the eigenvectors of the largest eigenvalues estimate the loadings. Then, we regress X on the estimated loadings to get the estimated factors.[7]

Similarly, for the partially observed data we apply PCA to $\frac{1}{N}\widetilde{\Sigma}$ to estimate the loadings.[8] We first estimate loadings and impose the identification assumption $\widetilde{\Lambda}^\top \widetilde{\Lambda}/N = I_r$ to uniquely identify the loadings.[9] Estimated loadings $\widetilde{\Lambda}$ are $\sqrt{N}$ times the r eigenvectors corresponding to the largest eigenvalues of the sample covariance matrix, that is

$$\frac{1}{N} \widetilde{\Sigma} \widetilde{\Lambda} = \widetilde{\Lambda} \widetilde{V}. \tag{2}$$

[6]Alternatively, we can apply PCA to $\frac{1}{NT} X X^\top$ to estimate the loadings and then regress $X^\top$ on the estimated loadings to get the estimated factors. The estimators are also consistent and asymptotic normal. Assume we have demeaned X_t for every t so in $\left(\frac{1}{N} X X^\top\right)_{ij}$ is a root-$\sqrt{N}$ consistent estimate for the covariance $cov(X_{it}, X_{jt})$.

[7]Bai and Ng (2002) and Bai (2003) develop the inferential theory, i.e., the consistency and asymptotic normality, for the factors and loadings estimated from PCA.

[8]We divide $\widetilde{\Sigma}$ by N such that the eigenvalues of $\frac{1}{N}\widetilde{\Sigma}$ do not scale with N and T.

[9]We assume the true number of factors is r and has been consistently estimated as in Bai (2003).

The next step is to estimate the factors. When the data is fully observed, we can regress X_{it} on $\widetilde{\lambda}_i$ to estimate the factors at time t, $\widetilde{F}_t$. However, when $X_{:,t}$ is only partially observed, we propose to regress only the observed X_{it} on $\widetilde{\lambda}_i$. This is a consistent estimator of the factors if the missing pattern does not depend on the λ_i, i.e. the unit-specific attributes.

However, we allow for more general observation patterns, that is, the probability of whether X_{it} is observed can depend on some observed covariates $S \in \mathbb{R}^{N\times k}$. In particular, the covariates S can depend on the unit-specific attributes Λ. We use $P(W_{it} = 1|S)$ to denote the probability of X_{it} being observed, which is an extension of the propensity score used in causal inference[10] and introduces the concept of time into the propensity score. We will have a more detailed discussion on the observation pattern in Assumption 1 in Section 3.

Given the observation probability $P(W_{it} = 1|S)$, we estimate $\widetilde{F}$ from a weighted average[11]

$$\widetilde{F}_t = \frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it} = 1|S)} X_{it}\widetilde{\lambda}_i = \frac{1}{N}\sum_{i\in\mathcal{O}_t} \frac{1}{P(W_{it} = 1|S)} X_{it}\widetilde{\lambda}_i, \tag{3}$$

where $\mathcal{O}_t = \{i : W_{it} = 1\}$ is the set of units observed at time t.[12] The weight $\frac{1}{P(W_{it}=1|S)}$ is always at least 1, which compensates for the missing entries at time t and removes the biases in $\widetilde{F}_t$. The estimator for $\widetilde{F}_t$ is closely related to the inverse propensity score estimator in causal inference[13]. In the special case when all entries at time t are missing at random with equal probability and independent of the covariates S, that is $P(W_{it} = 1|S) = p_t = \lim_{N\to\infty} |\mathcal{O}_t|/N$ for all i, then equation (3) simplifies to

$$\widetilde{F}_t = \frac{1}{|\mathcal{O}_t|}\sum_{i\in\mathcal{O}_t} X_{it}\widetilde{\lambda}_i \tag{4}$$

The last step is to estimate the common component $C_{it} = \lambda_i^\top F_t$. We use the plug-in estimator, $\widetilde{C}_{it} = \widetilde{\lambda}_i^\top \widetilde{F}_t$. If X_{it} is not observed, we estimate X_{it} by $\widetilde{C}_{it}$.

[10] See (Rosenbaum and Rubin, 1983)

[11] The standard form of weighted least squares is $\frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it}=1|S)} X_{it}\widetilde{\lambda}_i \left(\frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it}=1|S)}\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\right)^{-1}$. Since $\frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it}=1|S)}\lambda_i\lambda_i^\top \xrightarrow{P} \Sigma_\Lambda$ from Assumption 2.2, we have $\frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it}=1|S)} X_{it}\widetilde{\lambda}_i - \frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it}=1|S)} X_{it}\widetilde{\lambda}_i \left(\frac{1}{N}\sum_{i=1}^{N} \frac{W_{it}}{P(W_{it}=1|S)}\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\right)^{-1} = O_p\left(\frac{1}{\sqrt{N}}\right)$ and we use Equation (3) for notational simplicity.

[12] We assume $\frac{0}{0} = 0$ in this paper. If $P(W_{it} = 1|S) = 0$, we have $\frac{W_{it}}{P(W_{it}=1|S)} = 0$.

[13] Compare with (Hahn, 1998; Hirano et al., 2003)

2.3 An Illustrative Toy Example

We illustrate in a simple example how our estimator differs from the conventional PCA estimator. Assume that we have only one factor and the factor, loadings and residual components are i.i.d. normally distributed:

$$X_{it} = \lambda_i F_t + e_{it} \qquad F_t \overset{i.i.d.}{\sim} N(0, \sigma_F^2) \qquad \lambda_i \overset{i.i.d.}{\sim} N(0,1) \qquad e_{it} \overset{i.i.d.}{\sim} N(0, \sigma_e^2).$$

As in Table 1 the cross sectional observations $1, ..., N_0$ are missing for $T_0 + 1, ..., T$. We separate the vector of factor realizations into its first $\mathbf{F}_1 = \begin{pmatrix} F_1 & \cdots & F_{T_0} \end{pmatrix}^\top$ and second part $\mathbf{F}_2 = \begin{pmatrix} F_{T_0+1} & \cdots & F_T \end{pmatrix}^\top$ and similarly for the loadings $\mathbf{\Lambda}_1 = \begin{pmatrix} \lambda_1 & \cdots & \lambda_{N_0} \end{pmatrix}^\top$ and $\mathbf{\Lambda}_2 = \begin{pmatrix} \lambda_{N+0+1} & \cdots & \lambda_N \end{pmatrix}^\top$.

We start with the simplest case without error terms e_t to illustrate the logic of reweighting the entries. In this case the conventional covariance matrix equals

$$\frac{\widetilde{X}\widetilde{X}^\top}{T} = \frac{1}{T}\begin{pmatrix} \mathbf{\Lambda}_1 \mathbf{F}_1^\top & 0 \\ \mathbf{\Lambda}_2 \mathbf{F}_1^\top & \mathbf{\Lambda}_2 \mathbf{F}_2^\top \end{pmatrix} \begin{pmatrix} \mathbf{F}_1 \mathbf{\Lambda}_1^\top & \mathbf{F}_1 \mathbf{\Lambda}_2^\top \\ 0 & \mathbf{F}_2 \mathbf{\Lambda}_2^\top \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{T_0}{T}} \mathbf{\Lambda}_1 \\ \mathbf{\Lambda}_2 \end{pmatrix} \left(\sigma_F^2 + o_p(1)\right) \begin{pmatrix} \sqrt{\frac{T_0}{T}} \mathbf{\Lambda}_1^\top & \mathbf{\Lambda}_2^\top \end{pmatrix}.$$

Obviously, the eigenvector of this matrix is a biased estimate of the loadings. In contrast, the eigenvector of the correctly weighted sample covariance matrix consistently estimates the loadings:

$$\widetilde{\Sigma} = \begin{pmatrix} \mathbf{\Lambda}_1 \frac{\mathbf{F}_1^\top \mathbf{F}_1}{T_0} \mathbf{\Lambda}_1^\top & \mathbf{\Lambda}_1 \frac{\mathbf{F}_1^\top \mathbf{F}_1}{T_0} \mathbf{\Lambda}_2^\top \\ \mathbf{\Lambda}_2 \frac{\mathbf{F}_1^\top \mathbf{F}_1}{T_0} \mathbf{\Lambda}_1^\top & \mathbf{\Lambda}_2 \frac{\mathbf{F}_1^\top \mathbf{F}_1 + \mathbf{F}_2^\top \mathbf{F}_2}{T} \mathbf{\Lambda}_2^\top \end{pmatrix} = \begin{pmatrix} \mathbf{\Lambda}_1 \\ \mathbf{\Lambda}_2 \end{pmatrix} \left(\sigma_F^2 + o_p(1)\right) \begin{pmatrix} \mathbf{\Lambda}_1^\top & \mathbf{\Lambda}_2^\top \end{pmatrix}.$$

The same logic carries over to the estimator of the factors. Assume that we know the population loadings. Then, the estimator of the factor from the regression on the loadings equals

$$\frac{1}{N} \widetilde{X}^\top \Lambda = \frac{1}{N} \begin{pmatrix} \mathbf{F}_1 \mathbf{\Lambda}_1^\top & \mathbf{F}_2 \mathbf{\Lambda}_1^\top \\ 0 & \mathbf{F}_2 \mathbf{\Lambda}_2^\top \end{pmatrix} \begin{pmatrix} \mathbf{\Lambda}_1 \\ \mathbf{\Lambda}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \frac{N_0}{N} \end{pmatrix} + o_p(1).$$

which is a biased estimator for the second time period of the factor. The weighted least square regression provides a correct estimator

$$\tilde{F} = \begin{pmatrix} \mathbf{F}_1 \frac{\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2^\top \mathbf{\Lambda}_2}{N} \\ \mathbf{F}_2 \frac{\mathbf{\Lambda}_1^\top \mathbf{\Lambda}_1}{N_0} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{pmatrix} + o_p(1),$$

Note that in this special case the probability of observing an entry equals $P(W_{it} = 1|S) = \frac{N_0}{N}$ which is independent of any covariates S.

The proper reweighting in the loading and factor estimation leads to an additional correction term in the asymptotic variance of the estimator. As an illustration of this additional challenge, we add the i.i.d. error term e_{it} to our example. In our simplified setup our consistent estimator for the loadings $\tilde{\Lambda}$ has the following expansion for $i = 1, ..., N_0$:

$$\sqrt{T}\left(\tilde{\lambda}_i - \lambda_i\right) = \sqrt{\frac{T}{T_0}} \left(\frac{\tilde{F}^\top \tilde{F}}{T}\right)^{-1} \frac{1}{\sqrt{T_0}} \sum_{t=1}^{T_0} F_t e_{it} + \sqrt{T} \left(\frac{\tilde{F}^\top \tilde{F}}{T}\right)^{-1} \left(\frac{\mathbf{F}_1^\top \mathbf{F}_1}{T_0} - \frac{F^\top F}{T}\right) \lambda_i + o_p(1)$$

which results in the asymptotic normal distribution

$$\sqrt{T}\left(\tilde{\lambda}_i - \lambda_i\right) \xrightarrow{d} \begin{cases} N\left(0, \frac{T}{T_0}\frac{\sigma_e^2}{\sigma_F^2} + 2\frac{T-T_0}{T_0}\right) & \text{for } i = 1, ..., N_0 \\ N\left(0, \frac{\sigma_e^2}{\sigma_F^2}\right) & \text{for } i = N_0 + 1, ..., N. \end{cases} \tag{5}$$

The second term in the asymptotic expansion is due to averaging over different number of units for different elements of the loadings. This additional variance correction term vanishes for $T_0 \to T$. Similar terms appear in the distribution of the estimators of the factors and common components. We show under general conditions how these correction terms arise in the asymptotic distribution and how to take them into account for the inferential theory.

2.4 Application to Causal Inference

One of the most important applications of our inferential theory is to test for a treatment effect in the panel data setting. The fundamental problem in causal inference is that we only observe the outcome under treatment and would like to compare it with the unobserved outcome without treatment. We will model the counter-factual outcome as the missing observation. Our estimator allows us to impute the missing observations which serve as the counter-factual control outcome.

The treatment effect is the difference between the treatment and control outcomes. Our inferential theory is key to provide feasible test statistics for the treatment effect.

A valid control unit is "close" to the treatment unit except for the treatment effect. Typically synthetic controls are weighted averages of untreated units where the weights depend on observed covariates. Our approach is more general. We do not need to take a stand a priori on which covariates describe if a treated and control units are a good match. Instead, our latent loadings capture all unit-specific information in a data-driven way. The common component that we impute is a weighted average of the untreated units that takes all unit-specific information into account.

Our object of interest is the common component of the units after treatment adoption C_{it}^{treat} and the common component of the synthetic control C_{it}^{ctrl}. The treatment effect for unit i at time t is $\tau_{it} = C_{it}^{treat} - C_{it}^{ctrl}$. Previous literature in causal inference in the panel data setting focusses on the average treatment effect over time[14]. Importantly, our novel approach allows us to test an entry-wise effect:

$$\mathcal{H}_0 : \tau_{it} = 0 \quad \mathcal{H}_1 : \tau_{it} \neq 0$$

as we can provide the asymptotic distributions for $\widetilde{C}_{it}^{ctrl}$ and $\widetilde{C}_{it}^{treat}$.

Obviously, we can also accommodate an average treatment effect $\frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\tau_{it}$. More generally, we allow for regressions of the observed treatment units X_{it}^{treat} and unobserved control units X_{it}^{ctrl} on observed covariates Z and can test for a treatment effect in the regression coefficients. The time-series average treatment effect is just a special case. The time-series regression on covariates Z averages out the residual term e_{it} and hence allows us to extend the analysis beyond the common component. We provide a test for a treatment effect for the case where the treatment and control units have different loadings but share the same latent factors and the more general case where also the latent factors can be different after treatment.

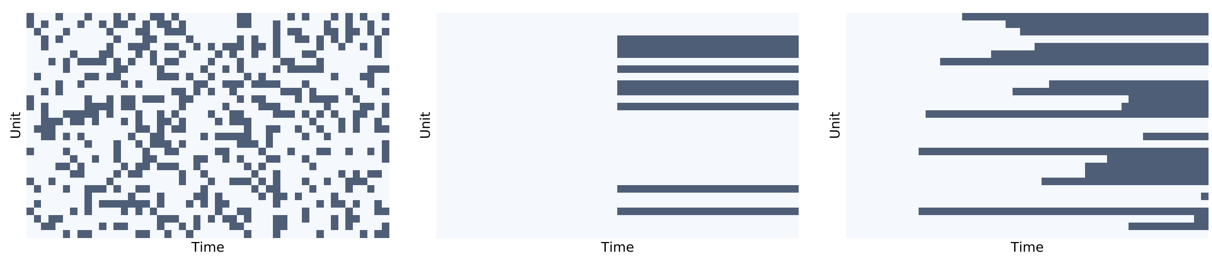
3 Assumptions

Notation. Let $M < \infty$ denote a generic constant. Let $\|v\|$ denote the vector norm and $\|A\| = trace(A^\top A)^{1/2}$ the Frobenius norm of matrix A. We denote the set of (time-series/cross-section)

[14] For example (Li, 2017; Li and Bell, 2017)

indices corresponding to the observed entries by $\mathcal{O}_t = \{i : W_{it} = 1\}$, $\mathcal{O}_{st} = \{i : W_{it} = 1 \text{ and } W_{is} = 1\}$, $\mathcal{Q}_i = \{t : W_{it} = 1\}$ and $\mathcal{Q}_{ij} = \{t : W_{it} = 1 \text{ and } W_{jt} = 1\}$. Thus, $\mathcal{O}_{ss} = \mathcal{O}_s$ and $\mathcal{Q}_{ii} = \mathcal{Q}_i$. The weighting matrices of the inverse probability are $\Pi^{(-1)} = [1/\pi_{st}] = [N/|\mathcal{O}_{st}|]$ and $Q^{(-1)} = [1/q_{ij}] = [T/|\mathcal{Q}_{ij}|]$, where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$.

We allow for very general patterns in the missing observations. Figure 1 shows three observation patterns that are allowed by Assumption 1. These three patterns are widely seen in empirical applications. The first one is the randomly missing pattern, that is, whether an entry is observed or not does not depend on other entries or observable covariates. For example, the observational pattern of the Netflix challenge is modeled as randomly missing entries.[15]. The second and third ones are the observation patterns for control panels in simultaneous and staggered treatment adoptions. Once a unit adopts the treatment, it stays treated afterwards. These two patterns are widely assumed in the literature of causal inference in panel data.[16].



(a) Randomly missing

(b) Missing pattern for the control panel in the simultaneous treatment adoption

(c) Missing pattern for the control panel in the staggered treatment adoption

Figure 1: Patterns of missing observations. The shaded entries indicate the missing entries.

Assumption 1. *Missing observations:*

1. $\lim_{N\to\infty} |\mathcal{O}_s|/N \geq \underline{\pi} > 0$, $\lim_{N\to\infty} |\mathcal{O}_{st}|/N \geq \underline{\pi} > 0$; *Similarly,* $\lim_{T\to\infty} |\mathcal{Q}_i|/T \geq \underline{q} > 0$, $\lim_{T\to\infty} |\mathcal{Q}_{ij}|/T \geq \underline{q} > 0$.

2. *W is independent of Λ conditional on S. W is independent of F and e.*

[15] (Candès and Recht, 2009; Zhou et al., 2008).

[16] See (Athey et al., 2018; Athey and Imbens, 2018)

3. *For $j \neq i$, W_{it} is independent of W_{js} conditional on S.*[17] *The probability of $W_{it} = 1$ depends on S, denoted as $P(W_{it} = 1|S)$. $0 < \underline{p} \leq P(W_{it} = 1|S)$.*

4. *$lim_{N\to\infty}|\mathcal{O}_t|/N = p_t$ where $0 < \underline{p} \leq p_t$.*

Under Assumption 1.1, the number of observed realizations for every cross-section unit i goes to infinity as $T \to \infty$; similarly, the number of observed outcomes for every time period t goes to infinity as $N \to \infty$. This assumption is necessary for the pointwise asymptotic results for the factors and loadings, which requires many observations for every time period and every cross-section unit. The observation pattern can depend on observable covariates S. All our results go through if these covariates are actually the loadings Λ themselves. However, using the consistent estimator $\widetilde{\Lambda}$ for S still provides consistent estimator of all quantities, but affects the asymptotic distribution due to the estimation error. By working with observable covariates we avoid this additional term in the asymptotic distribution.

Assumption 1.2 is closely connected to the **unconfoundedness** assumption in causal inference.[18] W_{it} can depend on the outcome X_{it}, i.e. observations with specific attributes can be more likely to be missing. Assumption 1.3 is related to the **propensity score** and **overlap assumption** in causal inference. We allow the observation probability to change over time, which generalizes the propensity score that is static. We condition on S instead of $S_{i,\cdot}$ to allow a general dependency structure of W on S. $P(W_{it} = 1|S) = P_t(W_{it} = 1|S_{i,:})$ would rule out network effects, which is usually assumed in the definition of the propensity score[19]. We assume $P(W_{it} = 1|S)$ is bounded away from 0, such that $\frac{1}{P(W_{it}=1|S)}$ does not diverge, which is equivalent to the overlap assumption in causal inference.

Assumption 2. *Factor Model:*

1. *Factors: $\forall\, t$, $\mathbb{E}[\|F_t\|^4] \leq \overline{F} < \infty$. There exists some positive definite $r \times r$ matrix Σ_F, such that $\frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top \xrightarrow{P} \Sigma_F$ and $\mathbb{E}\left\|\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} F_t F_t^\top - \Sigma_F\right)\right\| \leq M$. Furthermore, for any $\mathcal{Q}_{ij}$, $\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t F_t^\top \xrightarrow{P} \Sigma_F$ and $\mathbb{E}\left\|\sqrt{|\mathcal{Q}_{ij}|}\left(\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t F_t^\top - \Sigma_F\right)\right\| \leq M$.*

[17] t and s can be the same.
[18] Compare with (Rosenbaum and Rubin, 1983).
[19] E.g. (Rosenbaum and Rubin, 1983)

2. *Factor loadings: loadings are random, independent of factors and errors and have bounded fourth moments. There exists some positive definite* $r \times r$ *matrix* Σ_Λ *such that* $\frac{1}{N}\sum_{i=1}^N \lambda_i \lambda_i^\top \xrightarrow{P} \Sigma_\Lambda$ *and* $\mathbb{E}\left\|\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N \lambda_i\lambda_i^\top - \Sigma_\Lambda\right)\right\| \leq M$. *Moreover,* $\frac{1}{N}\sum_{i=1}^N \frac{W_{it}}{P(W_{it}=1|S)}\lambda_i\lambda_i^\top \xrightarrow{P} \Sigma_\Lambda$ *and* $E\left\|\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N \frac{1}{P(W_{it}=1|S)} W_{it}\lambda_i\lambda_i^\top - \Sigma_\Lambda\right)\right\| \leq M$.

3. *Time and cross-section dependence and heteroskedasticity of errors: There exists a positive constant* $M < \infty$, *such that for all* N *and* T:

   (a) $\mathbb{E}[e_{it}] = 0$, $\mathbb{E}|e_{it}|^8 \leq M$.

   (b) $\mathbb{E}[e_{is}e_{it}] = \gamma_{st,i}$ *with* $|\gamma_{st,i}| \leq \gamma_{st}$ *for some* γ_{st} *and all* i. *For all* t, $\sum_{s=1}^T \gamma_{st} \leq M$.

   (c) $\mathbb{E}[e_{it}e_{jt}] = \tau_{ij,t}$ *with* $|\tau_{ij,t}| \leq \tau_{ij}$ *for some* τ_{ij} *and all* t. *For all* i, $\sum_{j\in\mathcal{S}_{st}} \tau_{ij} \leq M$.

   (d) $\mathbb{E}[e_{it}e_{js}] = \tau_{ij,ts}$ *and* $\sum_{j=1}^N \sum_{s=1}^T |\tau_{ij,ts}| \leq M$ *for all* i *and* t.

   (e) *For all* i *and* j, $\mathbb{E}\left|\frac{1}{|\mathcal{Q}_{ij}|^{1/2}}\sum_{t\in\mathcal{Q}_{ij}}\left(e_{it}e_{jt} - \mathbb{E}[e_{it}e_{jt}]\right)\right|^4 \leq M$.

4. *Weak dependence between factor and idiosyncratic errors: for every* (i,j),

$$\mathbb{E}\left\|\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{i\in\mathcal{Q}_{ij}} F_t e_{it}\right\|^2 \leq M.$$

5. *Eigenvalues: The eigenvalues of* $\Sigma_\Lambda\Sigma_F$ *are distinct.*

Assumption 2 describes an approximate factor structure and is at a similar level of generality as Bai (2003): (1) Assumption 2.1 ensures that each factor has a nontrivial contribution to the variation in X. (2) We assume loadings are random but independent of factors and errors in Assumption 2.2. We could study a factor model conditioned on some particular realization of the loadings and the analysis would essentially be equivalent to that under the assumption that loadings are nonrandom. (3) Assumption 2.3 allows errors to be time-series and cross-sectionally weakly correlated. (4) Assumption 2.4 allows factors and idiosyncratic errors to be weakly correlated. (5) Assumption 2.5 guarantees that each loading and factor can be uniquely identified up to some rotation matrix.

Additionally, we assume that these aspects also hold if we look at a subset of all time periods (the subset is denoted as $\mathcal{Q}_{ij}$ in Assumption 2). Together with Assumption 1.2, our covariance matrix estimator (1) using incomplete observations has similar properties as the conventional covariance matrix estimator $\frac{1}{T}XX^\top$ using full observations. For example, both $\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} X_{it}X_{jt}$ and $\frac{1}{T}\sum_{t=1}^T X_{it}X_{jt}$ are consistent estimators for Σ_{ij}. Moreover, the eigenvalues and principal components estimated from both matrices are consistent, which we show in the next section and which is the foundation to develop the inferential theory of the factor model estimated from Equation (1).

We allow W to depend on S. As a result, the unweighted average $\mathcal{O}_{st}$, $\frac{1}{|\mathcal{O}_{st}|}\sum_{i\in\mathcal{O}_{st}}\lambda_i\lambda_i^\top$ does not necessarily converge to Σ_Λ but requires reweighting by the propensity score.[20] Assumption 2.2 arises naturally as illustrated by the following example. Assume that the probability of observing an entry depends on the unit-specific features captured by the loadings $P(W_{it}|S) = P(W_{it}|\lambda_i)$. For simplicity we assume that λ_i is i.i.d. with second moment Σ_Λ and rule out network effects. [21] By the Law of Large Numbers $\frac{1}{N}\sum_{i=1}^N \frac{1}{P(W_{it}|S)} W_{it}\lambda_i\lambda_i^\top \xrightarrow{P} \mathbb{E}\left[\frac{\mathbb{E}[W_{it}|\lambda_i]}{P(W_{it}|\lambda_i)}\lambda_i\lambda_i^\top\right] = \Sigma_\Lambda$ for all t.

Assumption 3. *Moments and Central Limit Theorems: $\exists M < \infty$ s. t. for all N and T*

1. $\mathbb{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}\left(e_{it}e_{jt} - \mathbb{E}[e_{it}e_{jt}]\right)\right\|^2 \leq M$ *for every j.*

2. $\mathbb{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N \frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}\lambda_i F_t^\top e_{it}\right\|^2 \leq M$ *for every t*

3. $\frac{\sqrt{T}}{N}\sum_{i=1}^N \lambda_i\lambda_i^\top \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{it} \xrightarrow{d} N(0,\Phi_j)$ *for every j, where*
$\Phi_j = \lim_{N,T\to\infty}\frac{T}{N^2}\sum_{i=1}^N\sum_{l=1}^N \mathbb{E}\left[\lambda_i\lambda_i^\top\left(\frac{1}{|\mathcal{Q}_{ij}||\mathcal{Q}_{lj}|}\sum_{s\in\mathcal{Q}_{ij},t\in\mathcal{Q}_{lj},(s,t)\in\Omega_{e_j}}\mathbb{E}\left[F_sF_t^\top e_{js}e_{jt}\right]\right)\lambda_l\lambda_l^\top\right]$.

4. $\frac{1}{\sqrt{N}}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\lambda_i e_{it} \xrightarrow{d} N(0,\Gamma_t)$ *for every t, where*
$\Gamma_t = \lim_{N\to\infty}\frac{1}{N}\sum_{i\in\mathcal{O}_t,l\in\mathcal{O}_t,(i,l)\in\Omega_{e_t}}\frac{1}{P(W_{it}=1|S)P(W_{lt}=1|S)}\mathbb{E}[\lambda_i\lambda_l^\top]\mathbb{E}[e_{it}e_{lt}]$.

5. $\frac{\sqrt{T}}{N}\sum_{i=1}^N\lambda_i\lambda_i^\top\left(\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_tF_t^\top - \frac{1}{T}\sum_{t=1}^T F_tF_t^\top\right) \xrightarrow{d} N(0,\Xi_{F,j})$[22] *for every j.*

6. $\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N\frac{W_{it}\lambda_i\lambda_i^\top}{P(W_{it}=1|S)} - \frac{1}{N}\sum_{i=1}^N\lambda_i\lambda_i^\top\right) \xrightarrow{d} N(0,\Theta_{\Lambda,t})$[23] *for any t.*

7. $\frac{1}{N^2}\sum_{i=1}^N\sum_{l=1}^N\lambda_l\lambda_l^\top\left(\frac{1}{|\mathcal{Q}_{li}|}\sum_{s\in\mathcal{Q}_{li}}F_sF_s^\top - \frac{1}{T}\sum_{s=1}^T F_sF_s^\top\right)\lambda_i\lambda_i = o_p\left(\frac{1}{\sqrt{T}}\right)$

[20] In the causal inference literature, the inverse propensity score weighted estimator is widely used to estimate the treatment effect, see for example (Hahn, 1998; Hirano et al., 2003).

[21] We can interpret this case as we first sample λ_i from some i.i.d. distribution and we estimate the latent model of X with missing entries conditional on Λ.

[22] This statement should be read as $\frac{\sqrt{T}}{N}vec\left(\sum_{i=1}^N\lambda_i\lambda_i^\top\left(\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_tF_t^\top - \frac{1}{T}\sum_{t=1}^T F_tF_t^\top\right)\right) \xrightarrow{d} N(0,\Xi_{F,j})$, where vec is the vectorization operator.

[23] This statement should be read as $\sqrt{N}vec\left(\frac{1}{N}\sum_{i=1}^N\frac{W_{it}\lambda_i\lambda_i^\top}{P(W_{it}=1|S)} - \frac{1}{N}\sum_{i=1}^N\lambda_i\lambda_i^\top\right) \xrightarrow{d} N(0,\Theta_{\Lambda,t}))$, where vec is the vectorization operator.

8. $\frac{1}{N^2}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\sum_{l=1}^{N}\lambda_l\lambda_l^\top\left(\frac{1}{|\mathcal{Q}_{li}|}\sum_{s\in\mathcal{Q}_{li}}F_sF_s^\top-\frac{1}{T}\sum_{s=1}^{T}F_sF_s^\top\right)\lambda_i e_{it}=o_p\left(\frac{1}{\sqrt{T}}\right)$

Assumption 3 is not required to show the consistency of loadings and factors but is only used to show the asymptotic normality of the estimators. Assumption 3.1-4 is closely related to the moment and CLT assumptions in Bai (2003). The first two parts in Assumptions 3 restrict the second moments of certain averages. The 3rd and 4th point state the necessary central limit theorems. $\frac{\sqrt{T}}{N}\sum_{i=1}^{N}\lambda_i\lambda_i^\top\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_te_{it}\xrightarrow{d}N(0,\Phi_j)$ is one of the leading terms in the asymptotic distribution of the estimated loadings $\widetilde{\lambda}_i$. However, $\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_te_{it}$ varies with i so we cannot separately average over the cross-sectional and time dimension as in the conventional framework. Point 5-8 are specific to the missing value problem and introduce the correction terms that appear in the asymptotic distribution. They are due to the fact that our estimator averages over different number of observations for different entries in the covariance matrix.

4 Asymptotic Results

4.1 Consistency

We first show the consistency of our estimators. Similar as Bai (2003), our analysis starts with plugging in $\widetilde{X}=(\Lambda^\top F+e)\odot W$ into Eq. (2). With some algebra, we have

$$\begin{aligned}\widetilde{\lambda}_j=&\underbrace{\frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^{N}\widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i\odot W_j)F\lambda_j/q_{ij}}_{H_j\lambda_j}+\underbrace{\frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^{N}\widetilde{\lambda}_i e_i^\top diag(W_i\odot W_j)F\lambda_j/q_{ij}}_{\mathrm{I}}\\&+\underbrace{\frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^{N}\widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i\odot W_j)e_j/q_{ij}}_{\mathrm{II}}+\underbrace{\frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^{N}\widetilde{\lambda}_i e_i^\top diag(W_i\odot W_j)e_j/q_{ij}}_{\mathrm{III}},\end{aligned}$$

where $H_j=\frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^{N}\widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i\odot W_j)F/q_{ij}$ is different for different units j. Let $H=\frac{1}{NT}\widetilde{V}^{-1}\widetilde{\Lambda}^\top\Lambda F^\top F$, which is defined similarly to the rotation matrix H in Bai and Ng (2002). Then we have

$$\widetilde{\lambda}_j-H\lambda_j=\widetilde{\lambda}_j-H_j\lambda_j+(H_j-H)\lambda_j=\mathrm{I}+\mathrm{II}+\mathrm{III}+(H_j-H)\lambda_j.$$

We show in the appendix that the time-series averages of the square of I, II and III converge to 0 at the rate $O_p\left(\min\left(\frac{1}{N},\frac{1}{T}\right)\right)$. Furthermore, Under Assumption 2.1, $\frac{1}{T}F^\top F\xrightarrow{P}\Sigma_F$ and

$\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t F_t^\top \xrightarrow{P} \Sigma_F$, we can show $H_j - H = O_p\left(\min\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right)\right)$.

Then we have the following theorem for the consistency of the estimated loadings.

Theorem 1. *Define* $\delta_{NT}^2 = \min(N,T)$. *Under Assumptions 1 and 2, we obtain*

$$\delta_{NT}^2\left(\frac{1}{N}\sum_{j=1}^{N}\left\|\widetilde{\lambda}_j - H\lambda_j\right\|^2\right) = O_p(1), \tag{6}$$

where $H = \frac{1}{NT}\widetilde{V}^{-1}\widetilde{\Lambda}^\top \Lambda F^\top F$.

Theorem 1 states that the whole loading matrix can be consistently estimated up to an appropriate rotation as $N, T \to \infty$ even if we only observe an incomplete panel matrix. The convergence rate is the same rate as for the fully observed panel in Bai and Ng (2002). Theorem 1 is based on the assumption that the observed entries are representative for the missing entries and hence provide a consistent estimation. Theorem 1 is a critical intermediate step to show the asymptotic normality of the estimated factor model in the next section.

4.2 Asymptotic Normality

The factors, loadings and common components are asymptotically normally distributed.

Theorem 2. *Under Assumptions 1-3 and if* $\sqrt{T}/N \to 0$, *then for each* i *as* $N, T \to \infty$:

$$\sqrt{T}(\widetilde{\lambda}_j - H\lambda_j) = \frac{\sqrt{T}}{N}\widetilde{V}^{-1}H\sum_{i=1}^{N}\lambda_i\lambda_i^\top \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} + \sqrt{T}\widetilde{V}^{-1}(H_j - H)\lambda_j + o_p(1) \tag{7}$$

$$\xrightarrow{d} N\left(0, V^{-1}(Q^{-1})^\top(\Phi_j + (\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j \otimes I))Q^{-1}V^{-1}\right), \tag{8}$$

where $Q = V^{1/2}\Upsilon\Sigma_F^{-1/2}$, V *is a diagonal matrix with the diagonal entries being the eigenvalues of* $\Sigma_F^{1/2}\Sigma_\Lambda\Sigma_F^{1/2}$, Υ *is the corresponding eigenvectors,* Φ_j *and* $\Xi_{F,j}$ *are defined in Assumption 3. Assume we know the auto-correlation structure in error terms that are only weakly serially dependent, then the plug-in estimator* $\widetilde{\Gamma}_{\lambda_j}$ *for the asymptotic variance in (8) is consistent and yields*

$$\sqrt{T}\widetilde{\Gamma}_{\lambda_j}^{-1/2}(\widetilde{\lambda}_j - H\lambda_j) \xrightarrow{d} N(0, I_r). \tag{9}$$

Theorem 2 states that the estimated loadings converge at the rate of $\sqrt{T}$, which is the same as the conventional PCA in Bai (2003). The asymptotic distribution of estimated loadings is determined by two terms: the time-series average of $F_t e_{jt}$, the first term in the right-hand side (RHS) of Equation (7), and the difference between the unit-specific rotation matrix H_j and the unified rotation matrix H, the second term in the RHS of Equation (7). In the conventional PCA, the asymptotic distribution of the loadings only depends on the first term, the time-series average of $F_t e_{jt}$. The difference between H_j and H has mean 0 but is of the order of $O_p\left(\frac{1}{\sqrt{T}}\right)$, and thus the difference contributes to the asymptotic distribution of the loadings. Compared with the estimated loadings from the fully observed data, the estimated loadings from the partially observed data have a larger variance. This finding makes intuitively sense as estimating loadings from the partially observed data is equivalent to estimating the loadings with less data, i.e., a smaller panel. The asymptotic normal distribution comes from Assumptions 3.3 and 3.5, which describe asymptotically independent distributions.

Theorem 3. *Under Assumptions 1-3 and if $\sqrt{N}/T \to 0$, then for each t as $N, T \to \infty$:*

$$\begin{aligned} \sqrt{N}(\widetilde{F}_t - (H^{-1})^\top F_t) &= \frac{1}{\sqrt{N}} \sum_{i \in \mathcal{O}_t} \frac{1}{P(W_{it}=1|S)} H\lambda_i e_{it} + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} v_{i,t} H \lambda_i \lambda_i^\top F_t + o_p(1) &(10)\\ &\xrightarrow{d} N(0, (Q^{-1})^\top (\Gamma_t + (F_t^\top \otimes I)\Theta_{\Lambda,t}(F_t \otimes I))Q^{-1}), &(11) \end{aligned}$$

where $v_{i,t} = \frac{1}{P(W_{it}=1|S)} - 1$ for $i \in \mathcal{O}_t$ and $v_{i,t} = -1$ for $i \notin \mathcal{O}_t$. Assume we know the cross-section correlation structure in the error terms and they are only weakly dependent [24]*, then the plug-in estimator $\widetilde{\Theta}_{F_t}$ for the asymptotic variance in (11) is consistent and yields*

$$\sqrt{T}\widetilde{\Theta}_{F_t}^{-1/2}(\widetilde{F}_t - (H^\top)^{-1} F_t) \xrightarrow{d} N(0, I_r). \tag{12}$$

Theorem 3 states that the convergence rate of the estimated factors is $\sqrt{N}$, which is the same as the conventional PCA. Similar to the estimated loadings, the asymptotic distribution of estimated factors is determined by two terms: the cross-section weighted average of $\lambda_i e_{it}$, the first term in the RHS of Equation (10), and the difference between the time-specific rotation matrix and the unified rotation matrix $(H^{-1})^\top$, the second term in the RHS of Equation (10). When the data is fully observed, the asymptotic distribution of F_t is driven by the first term and the second term

[24] This assumption can be replaced by an appropriate sparsity assumption with a corresponding threshold estimator.

vanishes as in Bai (2003). The second term has mean 0 and its variance is increasing in the number of missing observations. The asymptotic normal distribution comes from the two asymptotically independent terms in Assumptions 3.4 and 3.6.

Theorem 4. *Under Assumptions 1-3, then for each t and i as $N, T \to \infty$:*

$$\begin{aligned}
\delta_{NT}(\widetilde{C}_{it} - C_{it}) &= \delta_{NT}(\widetilde{\lambda}_i - H\lambda_i)^\top (H^\top)^{-1} F_t + \delta_{NT}(H\lambda_i)^\top (\widetilde{F}_t - (H^\top)^{-1} F_t) + o_p(1) \quad (13)\\
&\xrightarrow{d} N\left(0, \frac{\delta_{NT}^2}{N} \lambda_i^\top \Sigma_\Lambda^{-1} (\Gamma_t + (F_t^\top \otimes I)\Theta_{\Lambda,t}(F_t \otimes I))\Sigma_\Lambda^{-1}\lambda_i \right.\\
&\quad \left. + \frac{\delta_{NT}^2}{T} F_t^\top \Sigma_F^{-1} \Sigma_\Lambda^{-1} (\Phi_i + (\lambda_i^\top \otimes I)\Xi_{F,i}(\lambda_i \otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1} F_t \right) \quad (14)
\end{aligned}$$

where $\widetilde{C}_{it} = \widetilde{\lambda}_i^\top \widetilde{F}_t$ and $C_{it} = \lambda_i^\top F_t$. Furthermore, for consistent estimators $\widetilde{\Gamma}_{\lambda_i}$ as in Theorem 2 and $\widetilde{\Theta}_{F_t}$ as in Theorem 3 we have

$$\left(\frac{1}{T} \widetilde{\lambda}_i^\top \widetilde{\Theta}_{F_t} \widetilde{\lambda}_i + \frac{1}{N} \widetilde{F}_t^\top \widetilde{\Gamma}_{\lambda_i} \widetilde{F}_t \right)^{-1/2} (\widetilde{C}_{it} - C_{it}) \xrightarrow{d} N(0,1). \quad (15)$$

Theorem 4 states that the asymptotic distribution of the estimated common component is determined by the asymptotic distributions of both the estimated loadings and factors. It depends on the order of N and T which distribution dominates. If N is of a smaller order, the asymptotic distribution of the factors dominates; if T is of the smaller order, the asymptotic distribution of the loadings dominates; otherwise both distributions contribute.

4.3 Test of Treatment Effects

The asymptotic results in Section 4.2, in particular Theorem 4, can be used to answer the important question in causal inference, whether a unit's treatment effect at a particular time period, denoted as τ_{it}, is significant or not. We assume the potential outcome for both the control and treated have an approximate low-rank structure. The treatment effect for unit i at time t is $\tau_{it} = C_{it}^{treat} - C_{it}^{ctrl}$. In this paper, we want to test if the treatment effect τ_{it} is significantly different from 0 as in Equation (6). In the following, we discuss how to estimate and test τ_{it} for two cases:

1. Control factors F^{ctrl} and treated factors F^{treat} are *different*, that is, F^{ctrl} and F^{treat} span different vector spaces.

2. Control factors F^{ctrl} and treated factors F^{treat} are the *same*, that is, F^{ctrl} and F^{treat} span the same space.

4.3.1 Control and Treated Panel Have Different Factors

When F^{ctrl} and F^{treat} are different, we estimate a factor model from the incomplete control panel and another one from the incomplete treated panel using the estimation approach in Section 2. This means we apply our estimation approach twice where we either view the treated units as missing values to obtain the loadings and factors for the control or we view the untreated units as missing values to obtain the loadings and factors for the treatment. We can directly extend Theorem 4 to obtain the asymptotic distribution of the estimated treatment effect $\widetilde{\tau}_{it} = \widetilde{C}_{it}^{treat} - \widetilde{C}_{it}^{ctrl}$.

Theorem 5. *Assume the control panel Y^c and the treated panel Y^t both follow Assumptions 1-3. For each i and t, as $N, T \to \infty$, we have*

$$\begin{aligned} \delta_{NT}(\widetilde{\tau}_{it} - \tau_{it}) &= \delta_{NT}(\widetilde{C}_{it}^{treat} - C_{it}^{treat}) - \delta_{NT}(\widetilde{C}_{it}^{ctrl} - C_{it}^{ctrl}) \\ &\xrightarrow{d} N(0, M_{it}^{ctrl} + M_{it}^{treat}) \end{aligned} \tag{16}$$

where M_{it}^{ctrl} and M_{it}^{treat} are the asymptotic variances of $\delta_{NT}(\widetilde{C}_{it}^{ctrl} - C_{it}^{ctrl})$ and $\delta_{NT}(\widetilde{C}_{it}^{treat} - C_{it}^{treat})$ defined in (14). Furthermore, we have

$$\frac{\widetilde{\tau}_{it} - \tau_{it}}{\widetilde{\sigma}_{\tau_{it}}} \xrightarrow{d} N(0,1), \tag{17}$$

where $\widetilde{\sigma}_{\tau_{it}}^2 = \widetilde{M}_{it}^{ctrl} + \widetilde{M}_{it}^{treat}$ with $\widetilde{M}_{it}^{ctrl} = \frac{1}{T}(\widetilde{\lambda}_i^{ctrl})^\top \widetilde{\Theta}_{F_t^{ctrl}} \widetilde{\lambda}_i^{ctrl} + \frac{1}{N}(\widetilde{F}_t^{ctrl})^\top \widetilde{\Gamma}_{\lambda_i^{ctrl}} \widetilde{F}_t^{ctrl}$, $\widetilde{M}_{it}^{treat} = \frac{1}{T}(\widetilde{\lambda}_i^{treat})^\top \widetilde{\Theta}_{F_t^{treat}} \widetilde{\lambda}_i^{treat} + \frac{1}{N}(\widetilde{F}_t^{treat})^\top \widetilde{\Gamma}_{\lambda_i^{treat}} \widetilde{F}_t^{treat}$, and $\delta_{NT}^2 \widetilde{M}_{it}^{ctrl}$ and $\delta_{NT}^2 \widetilde{M}_{it}^{treat}$ are consistent estimators of M_{it}^{ctrl} and M_{it}^{treat}.

The asymptotic distribution of $\widetilde{\tau}_{it} - \tau_{it}$ depends on the asymptotic distributions of $\widetilde{C}_{it}^{treat} - C_{it}^{treat}$ and $\widetilde{C}_{it}^{ctrl} - C_{it}^{ctrl}$. The asymptotic distributions of $\widetilde{C}_{it}^{treat} - C_{it}^{treat}$ and $\widetilde{C}_{it}^{ctrl} - C_{it}^{ctrl}$ depend on e_{it}^{treat} and e_{it}^{ctrl} respectively. For each i and t, we observe at most one of X_{it}^{treat} and X_{it}^{ctrl}. Hence, the asymptotic distributions of $\widetilde{C}_{it}^{treat} - C_{it}^{treat}$ and $\widetilde{C}_{it}^{ctrl} - C_{it}^{ctrl}$ are asymptotically independent. As a result, $\widetilde{\tau}_{it} - \tau_{it}$ is asymptotically normal with the asymptotic variance being the sum of the asymptotic variances of $\widetilde{C}_{it}^{treat} - C_{it}^{treat}$ and $\widetilde{C}_{it}^{ctrl} - C_{it}^{ctrl}$.

Theorem 5 allows us to test individual treatment effects for each i and t which is novel in the literature on causal inference for panel data.[25] In many empirical applications, the object of interest is a unit's average treatment effect over time (Abadie et al., 2010, 2015; Doudchenko and Imbens, 2016; Li and Bell, 2017). That is,

$$\mathcal{H}_0 : \tfrac{1}{T-T_{0,i}} \textstyle\sum_{t=T_{0,i}+1}^{T} \tau_{it} = 0 \quad \mathcal{H}_1 : \tfrac{1}{T-T_{0,i}} \textstyle\sum_{t=T_{0,i}+1}^{T} \tau_{it} \neq 0, \tag{18}$$

where $T_{0,i}$ is the last time period with control observations for unit i.[26] We estimate $\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} \tau_{it}$ by the plug-in estimator $\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} \widetilde{\tau}_{it}$. We need an additional assumption for the CLT on the subset of the treated data which is closely related to Assumption 3:

Assumption 4. *For $T_{0,i} < T$ satisfying $T - T_{0,i} \to \infty$, it holds*

1. $\frac{1}{T-T_{0,i}} \sum_{T_{0,i}+1}^{T} F_t \xrightarrow{P} \mu_F$.
2. $\frac{1}{\sqrt{T-T_{0,i}}} \sum_{T_{0,i}+1}^{T} F_t e_{it} \xrightarrow{d} N(0, \Psi_i)$.
3. $\frac{1}{\sqrt{N}(T-T_{0,i})} \sum_{t=T_{0,i}+1}^{T} \sum_{j \in \mathcal{O}_t} \frac{1}{P(W_{jt}=1|S)} \lambda_j e_{jt} = o_p(1)$
4. $\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left(\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} \left(\frac{W_{jt}\lambda_j\lambda_j^\top}{P(W_{jt}=1|S)} - \lambda_j\lambda_j^\top \right) F_t \right) \xrightarrow{d} N(0, \Theta_{\Lambda,i})$.

Assumption 4 is an extension of Assumptions 3.4 and 3.6. It is used to show that the time-series weighted average of estimated common components and imputed values is asymptotically normal as in Lemma 1 and Theorem 6. Then, we have the following result about the asymptotic distribution of the estimated time-series average treatment effect $\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} \widetilde{\tau}_{it}$.

Lemma 1. *Assume the control panel Y^c and the treated panel Y^t both follow Assumptions 1-4. As $N, T, T - T_{0,i} \to \infty$, for each i, the average estimated common component for the control C_{it}^{ctrl} and treated C_{it}^{treat} has (for notation simplicity, we omit the superscript ctrl and treat in the following)*

$$\begin{aligned} &\tfrac{\delta_{NT}}{T-T_{0,i}} \textstyle\sum_{t=T_{0,i}+1}^{T} \left(\widetilde{C}_{it} - C_{it} \right) \\ \xrightarrow{d}\;\; & N\left(0, \lambda_i^\top \Sigma_\Lambda^{-1} \Theta_{\Lambda,i} \Sigma_\Lambda^{-1} \lambda_i + \mu_F^\top \Sigma_F^{-1} \Sigma_\Lambda^{-1} (\Phi_i + (\lambda_j^\top \otimes I) \Xi_{F,i} (\lambda_i \otimes I)) \Sigma_\Lambda^{-1} \Sigma_F^{-1} \mu_F \right) \end{aligned} \tag{19}$$

[25] For example, if we want to test $\mathcal{H}_0 : \tau_{it} = 0$ for some i and t, we reject $\mathcal{H}_0$ if $|\widetilde{\tau}_{it}/\widetilde{\sigma}_{\tau_{it}}|$ is larger than 1.96 for the two-sided test (or larger than 1.645 for the one-sided test) at 95% confidence level.

[26] For the one-sided test, $\mathcal{H}_1 : \frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} \tau_{it} > 0$ or $\mathcal{H}_1 : \frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} \tau_{it} < 0$.

where Θ_i and $\Xi_{F,i}$ are defined in Assumption 3, and $\Theta_{\Lambda,i}$ is defined in Assumption 5. Denote the asymptotic variance in Equation (19) as M_i. Then we have for the average treatment effect

$$\frac{1}{\widetilde{\sigma}_{\tau_i}}\left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^{T}\widetilde{\tau}_{it}-\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^{T}\tau_{it}\right)\xrightarrow{d}N(0,1), \tag{20}$$

where $\widetilde{\sigma}^2_{\tau_i}=\widetilde{M}_i^{ctrl}+\widetilde{M}_i^{treat}$, $\delta^2_{NT}\widetilde{M}_i^{ctrl}$ and $\delta^2_{NT}\widetilde{M}_i^{treat}$ are consistent estimators of M_i^{ctrl} and M_i^{treat}, $\widetilde{M}_i^{ctrl}=\frac{1}{T}(\widetilde{\lambda}_i^{ctrl})^\top\widetilde{\Theta}_{\Lambda^{ctrl},i}\widetilde{\lambda}_i^{ctrl}+\frac{1}{N}\left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t^{ctrl}\right)^\top\widetilde{\Gamma}_{\lambda_i^{ctrl}}\left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t^{ctrl}\right)$, $\widetilde{M}_i^{treat}=\frac{1}{T}(\widetilde{\lambda}_i^{treat})^\top\widetilde{\Theta}_{\Lambda^{treat},i}\widetilde{\lambda}_i^{treat}+\frac{1}{N}\left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t^{treat}\right)^\top\widetilde{\Gamma}_{\lambda_i^{treat}}\left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t^{treat}\right)$, $\widetilde{\Gamma}_{\lambda_i^{ctrl}}$ and $\widetilde{\Gamma}_{\lambda_i^{treat}}$ are defined in Theorem 2, and $\widetilde{\Theta}_{\Lambda^{ctrl},i}$ and $\widetilde{\Theta}_{\Lambda^{treat},i}$ are defined in Lemma 14 in the Appendix.

4.3.2 Control and Treated Panel Share the Same Factors

When the control and treated panels share the same set of factors, we can write the potential outcome for the control as $X_{it}^{ctrl}=C_{it}^{ctrl}+e_{it}^{ctrl}=(\lambda_i^{ctrl})^\top F_t+e_{it}^{ctrl}$ and the treated as $X_{it}^{treat}=C_{it}^{treat}+e_{it}^{treat}=(\lambda_i^{treat})^\top F_t+e_{it}^{treat}$. While the factors are the same, we allow the loadings λ_i^{ctrl} and λ_i^{treat} to be different. The implication of this setting is that the treatment does not affect the latent factors, but only affects the units' exposure to factors. For example, in our empirical study about the publication effect of anomalies, the no-arbitrage principle implies the existence of a stochastic discount factor (SDF) that is spanned by the same latent factors and can price all assets. Our SDF does not change after publications, but the anomaly's exposure to this SDF captured by the loadings may change after publications. Hence, the average returns or exposure to risk factors can be affected by the publication as the portfolios load on different parts of the SDF after investors become aware of an anomaly.

When the majority of the observations are control observations as in most causal inference applications such as our empirical study and Abadie et al. (2010, 2015), we estimate the factor model from the incomplete control panel X^{ctrl}. Assuming a permanent treatment effect, we can use ordinary least squares (OLS) to estimate the loadings for treated $\widetilde{\lambda}_i^{treat}$,[27]

$$\widetilde{\lambda}_i^{treat} = \left(\sum_{t=T_{0,i}+1}^{T} \widetilde{F}_t \widetilde{F}_t^{\top} \right)^{-1} \sum_{t=T_{0,i}+1}^{T} \widetilde{F}_t X_{it}^{treat}, \tag{21}$$

where T_i is the treatment adoption time for unit i and T is the total number of periods. The common components for the treated panel can be estimated by $\widetilde{C}_{it}^{treat} = (\widetilde{\lambda}_i^{treat})^{\top} \widetilde{F}_t$. We have the following lemma to show the asymptotic distribution for $\widetilde{C}_{it}^{treat}$.

Lemma 2. *Assume $T - T_{0,i} \to \infty$ for $T_{0,i} < T$ and define $\delta^2_{N,T-T_{0,i}} = \min(N, T - T_{0,i})$. Under Assumptions 1-4 we have*

$$\begin{aligned}\delta_{N,T-T_{0,i}}(\widetilde{C}_{it}^{treat} - C_{it}^{treat}) \quad \xrightarrow{d} \quad & N\Bigg(0, \tfrac{\delta^2_{N,T-T_{0,i}}}{N}(\lambda_i^{treat})^{\top}\Sigma_{\Lambda}^{-1}(\Gamma_t + (F_t^{\top} \otimes I)\Theta_{\Lambda,t}(F_t \otimes I))\Sigma_{\Lambda}^{-1}\lambda_i^{treat} \\ & + \tfrac{\delta^2_{N,T-T_{0,i}}}{T-T_{0,i}} F_t^{\top}\Sigma_F^{-1}\Psi_i\Sigma_F^{-1}F_t.\Bigg) \end{aligned} \tag{22}$$

Similar to Theorem 4, there are two terms in the asymptotic distribution. For $(T-T_{0,i})/N \to 0$ or $N/(T-T_{0,i}) \to 0$ only one term remains; otherwise both of them contribute. One question of interest is whether the average common component over time changes by the treatment. That is to test

$$\mathcal{H}_0 : \frac{1}{T - T_{0,i}} \sum_{t=T_{0,i}+1}^{T} (C_{it}^{ctrl} - C_{it}^{treat}) = 0, \quad \mathcal{H}_1 : \frac{1}{T - T_{0,i}} \sum_{t=T_{0,i}+1}^{T} (C_{it}^{ctrl} - C_{it}^{treat}) \neq 0.$$

Note, that $\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^{T} C_{it} = (\vec{1}^{\top}\vec{1})^{-1}\vec{1}^{\top} C_{i,(T_{0,i}+1):T}$ (with the superscript to be either *ctrl* or *treat*). We can generalize $\vec{1}$ to a generic Z and let

$$\beta_i^{ctrl} = (Z^{\top}Z)^{-1}Z^{\top}C^{ctrl}_{i,(T_{0,i}+1):T} \quad \text{and} \quad \beta_i^{treat} = (Z^{\top}Z)^{-1}Z^{\top}C^{treat}_{i,(T_{0,i}+1):T}$$

and test

$$\mathcal{H}_0 : \beta_i^{ctrl} = \beta_i^{treat} \quad \mathcal{H}_1 : \beta_i^{ctrl} \neq \beta_i^{treat}.$$

[27]If units switch between treatment and control, we can modify Equation (21) to $\widetilde{\lambda}_i^{treat} = \left(\sum_{t\in\mathcal{S}_i} \widetilde{F}_t\widetilde{F}_t^{\top}\right)^{-1} \sum_{t=\mathcal{S}_i} \widetilde{F}_t X_{it}^{treat}$, where $\mathcal{S}_i$ it the set of indices for the treated observations. Lemma 2 can be adapted accordingly.

We require the additional weak assumption that links the regressors with the approximate factor model.

Assumption 5. *For $T_{0,i} < T$ satisfying $T - T_{0,i} \to \infty$ it holds for $Z \in \mathbb{R}^{(T-T_{0,i})\times L}$*

$$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \left(\sum_{t=T_{0,i}+1}^{T} \left(\frac{W_{jt}\lambda_j\lambda_j^\top}{P\left(W_{jt}=1|\Lambda\right)} - \lambda_j\lambda_j^\top \right) F_t Z_{t-T_{0,i}}^\top (Z^\top Z)^{-1} \right) \xrightarrow{d} N(0, \Theta_{\Lambda,i,Z})$$

We obtain feasible estimates by regressing the estimated common components $\widetilde{C}_{it}^{ctrl}$ and $\widetilde{C}_{it}^{treat}$ on the observed covariates Z to estimate the coefficients

$$\widetilde{\beta}_i^{ctrl} = (Z^\top Z)^{-1} Z^\top \widetilde{C}_{i,(T_{0,i}+1):T}^{ctrl} \quad \text{and} \quad \widetilde{\beta}_i^{treat} = (Z^\top Z)^{-1} Z^\top \widetilde{C}_{i,(T_{0,i}+1):T}^{treat}.$$

Equipped with Lemma 2 and Theorem 4, we can show the asymptotic distributions of $\widetilde{\beta}_i^{ctrl}$, $\widetilde{\beta}_i^{treat}$ and $\widetilde{\beta}_i^{ctrl} - \widetilde{\beta}_i^{treat}$. Here, we present the distribution for $\widetilde{\beta}_i^{ctrl} - \widetilde{\beta}_i^{treat}$ and delegate the distributions for the individual $\widetilde{\beta}_i^{ctrl}$ and $\widetilde{\beta}_i^{treat}$ to the Appendix.

Theorem 6. *Suppose Assumptions 1-5 hold and $T - T_{0,i} \to \infty$:*

$$\delta_{N,T-T_{0,i}}((Z^\top Z)^{-1} Z^\top M Z (Z^\top Z)^{-1} + M_Z)^{-1/2} \left((\widetilde{\beta}_i^{ctrl} - \widetilde{\beta}_i^{treat}) - (\beta_i^{ctrl} - \beta_i^{treat}) \right) \xrightarrow{d} N(0, I) \quad (23)$$

where M is a $(T-T_{0,i}) \times (T-T_{0,i})$ matrix with

$$\begin{aligned}
M_{t-T_{0,i},t-T_{0,i}} =& \frac{\delta_{N,T-T_{0,i}}^2}{T} F_t^\top \Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_i + ((\lambda_i^{ctrl})^\top \otimes I)\Xi_{F,i}(\lambda_i^{ctrl} \otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_t \\
&+ \frac{\delta_{N,T-T_{0,i}}^2}{T-T_{0,i}} F_t^\top \Sigma_F^{-1}\Psi_i\Sigma_F^{-1}F_t \\
M_{t-T_{0,i},s-T_{0,i}} =& \frac{\delta_{N,T-T_{0,i}}^2}{T} F_t^\top \Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_i + ((\lambda_i^{ctrl})^\top \otimes I)\Xi_{F,i}(\lambda_i^{ctrl} \otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_s \\
&+ \frac{\delta_{N,T-T_{0,i}}^2}{T-T_{0,i}} F_t^\top \Sigma_F^{-1}\Psi_i\Sigma_F^{-1}F_s \\
M_{Z,lm} =& \frac{\delta_{N,T-T_{0,i}}^2}{N} (\lambda_i^{ctrl} - \lambda_i^{treat})^\top \Sigma_\Lambda^{-1}\Theta_{\Lambda,i,Z,lm}\Sigma_\Lambda^{-1}(\lambda_i^{ctrl} - \lambda_i^{treat})
\end{aligned}$$

Given the asymptotic distribution for $\widetilde{\beta}_i^{ctrl}$ and $\widetilde{\beta}_i^{treat}$ in Lemma 6, we can test if β_i is affected by the treatment, that is, if β_i^{ctrl} and β_i^{treat} are the same or different. As an example, for $Z = \vec{1}$, we can use (23) to test if the average treatment effect is 0, and then in Equation (23), we have $(Z^\top Z)(Z^\top MZ)^{-1/2} = \frac{T-T_{0,i}}{\sqrt{\sum_{t=1}^{T-T_{0,i}}\sum_{s=1}^{T-T_{0,i}} M_{ts}}}$.

5 Feasible Estimator of $P(W_{it}|S)$

We provide a feasible estimator for $P(W_{it}|S)$ which we need in (3) to estimate the factors. If the estimator satisfies $\widehat{P}(W_{it}|S) = P(W_{it}|S) + o_p\left(\min\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right)\right)$, then the estimation error of $\widehat{P}(W_{it}|S)$ can be neglected and does not affect the asymptotic distributions of estimated factors $\widetilde{F}_t$ and common components $\widetilde{C}_{it}$. Then, Theorems 3 and 4 continue to hold if we use $\widetilde{P}(W_{it}|S)$ in (3).

As previously mentioned, we could set $S = \Lambda$, i.e. consider $P(W_{it}|\Lambda)$. Then, the observation pattern can depend on all unobserved unit-specific attributes. However, the major challenge is to estimate the probability $P(W_{it}|\lambda_i)$ as we do not observe the latent features λ_i. If we use the estimated features $\widetilde{\lambda}_i$, whose estimation error is of the order $O_p\left(\frac{1}{\sqrt{N}}\right)$ based on Theorem 2, the estimation error for the observed probability is of an order of at least $O_p\left(\frac{1}{\sqrt{N}}\right)$. Hence, the estimation error of $\widehat{P}(W_{it}|\widetilde{\lambda}_i)$ would contribute to the asymptotic distributions of $\widetilde{F}_t$ and $\widetilde{C}_{it}$.

In order to avoid this issue, we condition on observed covariates S. These covariates can have the same span as the latent factors, but do not suffer from the estimation error. In the following we discuss three cases for a feasible estimator of $P(W_{it}|S)$.

In the first case, we assume that the probability of observing an entry does not depend on unit specific features, i.e. $P(W_{it}|S) = P(W_{it})$. Then, we can estimate $P(W_{it})$ by $\frac{|\mathcal{O}_t|}{N}$. The convergence rate of the estimated probability $\widehat{P}(W_{it})$ is $\frac{1}{N}$. If $\sqrt{T}/N \to 0$, which is also assumed in Theorem 2, then the estimation error of $\widehat{P}(W_{it})$ is of the order $o_p\left(\min\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right)\right)$.

In the second case, $P(W_{it}|S) = f(S)$ we allow for unit specific features where we can use the full observation matrix to estimate $f(S)$. Since we have NT observations, if we impose a parametric form on $f(S)$, for example a logit model, then $f(S)$ can be consistently estimated at the rate $\sqrt{NT}$. Alternatively, if we do not impose a functional form $f(S)$, we can use a nonparametric regression to estimate $f(S)$. This probability can be consistently estimated at the rate $\sqrt{NTh_1h_2\cdots h_p}$, where $h_1, \cdots, h_p$ are the bandwidths for 1st, 2nd, ..., p-th coordinate in S. When $NTh_1h_2\cdots h_p \gg \max(\sqrt{T}, \sqrt{N})$, the estimation error is of the order $o_p\left(\min\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right)\right)$.

In the third case, when S only takes finitely many values, then $\widehat{P}(W_{it}|S = s) = \frac{|\mathcal{O}_{\varsigma,t}|}{N_s}$, where

$N_s = \sum_{i=1}^{N} \mathbf{1}(S = s)$ and $\mathcal{O}_{S,t} = \{i : W_{it} = 1 \text{ and } S = s\}$. The convergence rate is $\frac{1}{N_s}$. When $N_s \gg \max(\sqrt{T}, \sqrt{N})$ for all s, the estimation error is of the order $o_p\left(\min\left(\frac{1}{\sqrt{N}}, \frac{1}{\sqrt{T}}\right)\right)$.

6 Simulation

In this section, we demonstrate the finite sample properties of our asymptotic results for both the observed entries and the missing entries. We confirm the theoretical distribution results for the factor, loadings and common components and show that without the proper reweighting and variance correction term the asymptotic distribution is severely biased.

We generate the data from a one-factor model for 2,000 Monte Carlo simulations:

$$X_{it} = \lambda_i^\top F_t + e_{it}$$

where $F_t \overset{\text{i.i.d.}}{\sim} N(0,1)$, $\lambda_i \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $e_{it} \overset{\text{i.i.d.}}{\sim} N(0,1)$. We study two missing patterns

1. Data is randomly missing.

2. Staggered adoption with irreversible treatment: The probability of missing observations depends on the unit specific features and once not observed the unit's observations stay missing.

6.1 Asymptotic Distribution

For both observation patterns, the finite sample distribution results of our estimators work well. The main text presents the results for the randomly missing observations while the Appendix collects the results for the staggered adoption case.

In the randomly missing case, the observation matrix W is generated from $W_{it} \sim \text{Bernoulli}(p)$, where $p = 0.5$ or 0.9. Figure 2 shows the histogram of the standardized common components for randomly selected observed entries and missing entries. We present the corresponding histograms of the standardized estimated factors and loadings in Figures 14 and 15 in the Appendix. The estimates are centered and standardized using consistent estimates of the theoretical mean and standard deviation. [28]

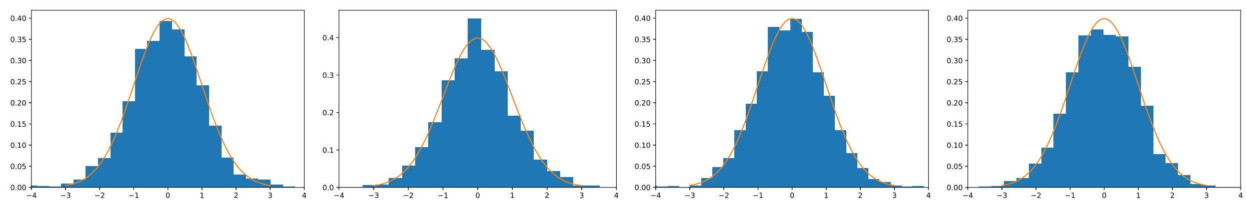
[28]We use C_{it} ($\lambda_i^\top F_t$) as the theoretical mean in Figure 2. Moreover, because we know H in the simulation, we use HF_t and $(H^\top)^{-1}\lambda_i$ as the theoretical means in Figures 14 and 15 in the Appendix. The theoretical standard deviations are calculated based on the plug in estimators for F_t, λ_i and e_{it}. If $(i,t) \in \mathcal{O}$, $\widetilde{e}_{it} = \widetilde{X}_{it} - \widetilde{\lambda}_i^\top \widetilde{F}_t$ is a

For the staggered adoption pattern, the observation matrix W is generated from the following scheme:

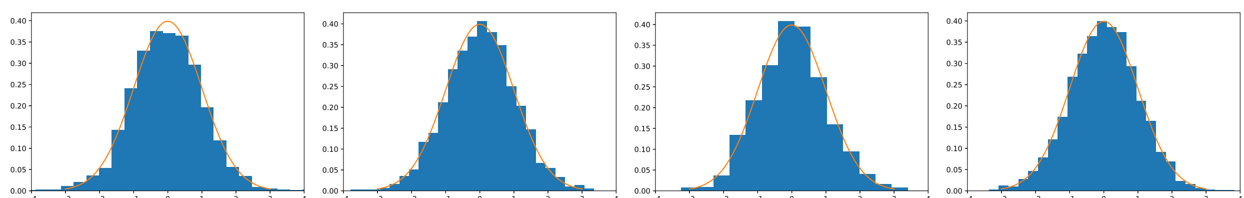
1. All the observations before T_0 are control observations. That is, for $t \leq T_0$, it holds $W_{it} = 1$.

2. After T_0, if unit i is not treated at time $t-1$, the probability for this unit to stay in the control group at time $t > T_0$ is $P(W_{it} = 1|\lambda_i, W_{i,t-1} = 1) = \frac{1}{1+4\exp(0.1\lambda_i)}$,[29] where λ_i is a scalar. If unit i is treated at time $t-1$, then it stays treated at time $t > T_0$.

Figures 16-18 in the Appendix show histograms of the standardized estimated factors, loadings and common components for randomly selected observed entries and missing entries. The histograms match the standard normal density function very well and support the validity of our asymptotic results in finite samples.



(a) (250, 500, 0.5, 0) (b) (250, 500, 0.9, 0) (c) (500, 250, 0.5, 0) (d) (500, 250, 0.9, 0)



(e) (250, 500, 0.5, 1) (f) (250, 500, 0.9, 1) (g) (500, 250, 0.5, 1) (h) (500, 250, 0.9, 1)

Figure 2: Randomly Missing: Histograms of estimated standardized common components. The normal density function is superimposed on the histograms. $P(W_{it} = 1|\lambda_i) = p$ for any i and t, where $p = 0.5$ and 0.9 in the simulation. The caption in the sub-figures denotes a tuple of (N, T, p, W_{it}).

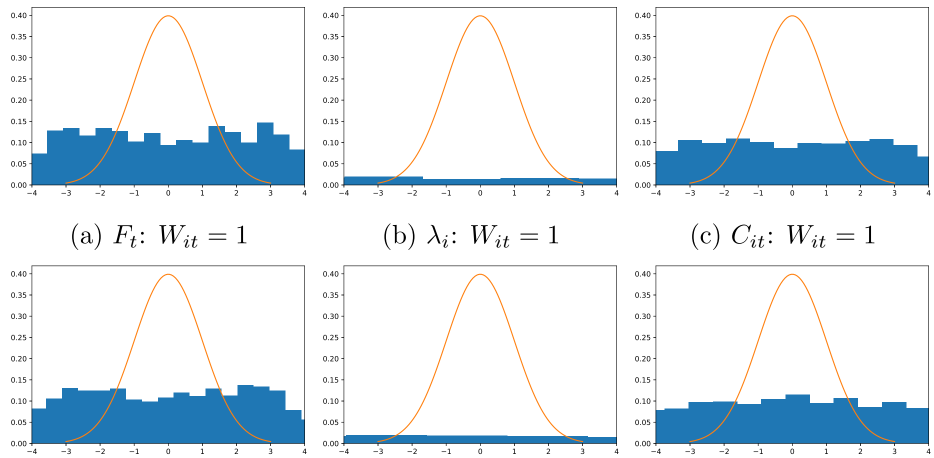
6.2 Estimation Without Reweighting or Variance Correction

Our simulations confirm that without proper reweighting the estimates are severely biased. Here, we plot the histograms of the standardized estimated factors, loadings and common components similar as in the last subsection for randomly missing data, but instead of using the reweighting

consistent estimator for e_{it}.

[29] Here we assume $S = \Lambda$.

scheme in (1) to estimate the sample covariance matrix and the weighted linear regression in (3) to estimate the factors, we use $\frac{1}{T}\widetilde{X}\widetilde{X}^\top$ as the sample covariance matrix and conventional PCA estimators. Figure 3 shows that the asymptotic distributions of the estimated factors, loadings and common components from the conventional PCA estimators in the presence of missing data. We can see that our method is critical to get the correct asymptotic distributions for the latent factor model.

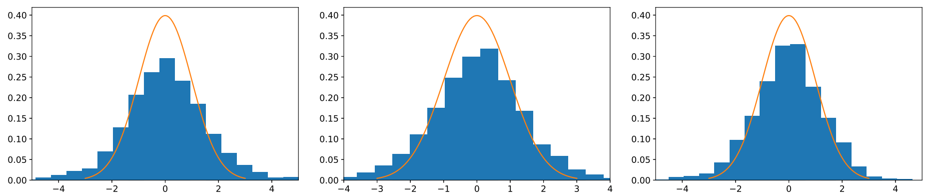


(a) F_t: $W_{it} = 1$ (b) λ_i: $W_{it} = 1$ (c) C_{it}: $W_{it} = 1$

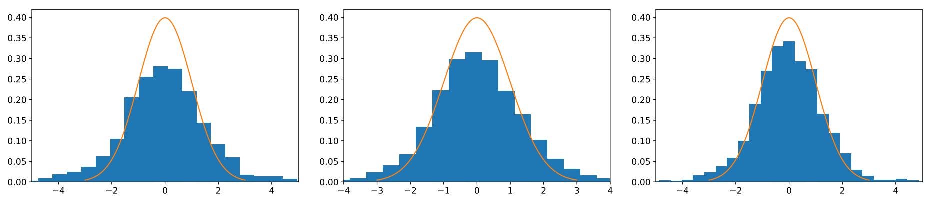
(d) F_t: $W_{it} = 0$ (e) λ_i: $W_{it} = 0$ (f) C_{it}: $W_{it} = 0$

Figure 3: Randomly Missing: Histograms of estimated standardized factors, loadings and common components, where the factor model is estimated with conventional PCA. The normal density function is superimposed on the histograms. The observation probability is 0.5, $N = 500$ and $T = 500$.

The simulations demonstrate that without the variance correction term the asymptotic standard errors are too small. We plot the histograms of the standardized estimated factors, loadings and common components for the randomly missing observation pattern, but without correcting the variances with the additional terms in Theorems 2-4. Without the variance correction term, the asymptotic distribution has a too low variance compared to the Monte Carlo simulations, as illustrated in Figure 4. Thus, the additional variance term cannot be ignored in the asymptotic distributions of the estimated factors, loadings, and common components.



(a) F_t: $W_{it} = 1$ (b) λ_i: $W_{it} = 1$ (c) C_{it}: $W_{it} = 1$



(d) F_t: $W_{it} = 0$ (e) λ_i: $W_{it} = 0$ (f) C_{it}: $W_{it} = 0$

Figure 4: Randomly Missing: Histograms of estimated standardized factors, loadings and common components, where the variances are not corrected by the additional variance terms in Theorems 2-4. The normal density function is superimposed on the histograms. The observation probability is 0.5, $N = 500$ and $T = 500$.

7 Empirical Study: Publication Effect on Anomaly Returns

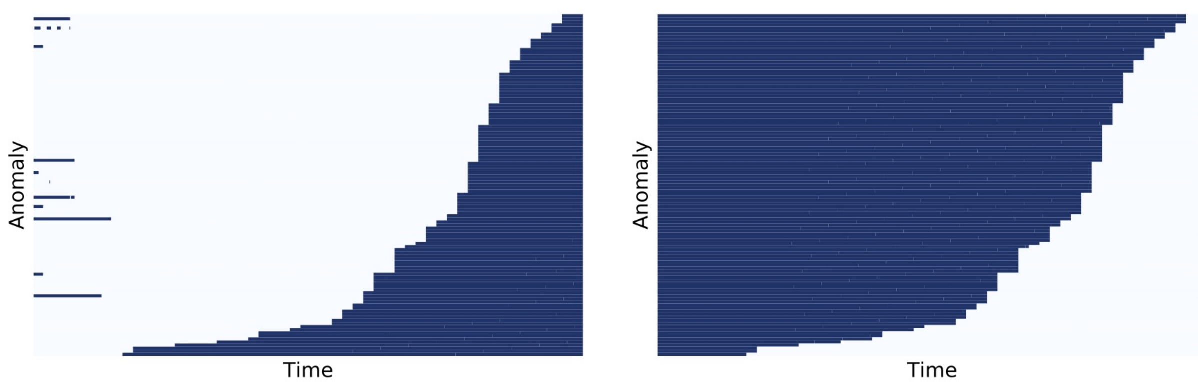
There is an ongoing debate in asset pricing whether academic publications result in the disappearance, reversion or attenuation of anomalies in equity returns. An anomaly describes a pattern in average returns that cannot be explained by a benchmark asset pricing model as for example the Capital Asset Pricing Model (CAPM). Schwert (2003) finds that anomalies including size effect, value effect, weekend effect and dividend yield effect seem to have disappeared or lost the predictive power after they were published. McLean and Pontiff (2016) suggest a 32% lower return from publication-informed trading. Harvey et al. (2016) suggest that the publication returns are biased upwards because of journals' preference for large t-statistics. Chen and Zimmermann (2018) measure that 12% of the anomalies' returns are due to the publication bias while 35% are due to mis-pricing that can be traded away after the anomalies are discovered.

In this paper, we study if the publication in an academic journal has a significant negative effect on the risk premium of anomaly portfolios. We use the data of Chen and Zimmermann (2018)[30] which contains monthly returns for characteristic-sorted quintile portfolios from July 1963

[30] We thank the authors for sharing the data. We refer to their paper for the details of the data collection. The data is available on the website `https://sites.google.com/site/chenandrewy/home?authuser=0`

to December 2015. Each anomaly is based on a firm-specific variable, e.g. the size or book-to-market ratio. All U.S. stocks are sorted into five quintile portfolios based on the cross-sectional rank order of the firm-specific characteristic values and the composition is regularly updated. Long-short portfolios that buy the highest quintile and sell the lowest quintile portfolio are a standard procedure to construct "risk factors" that exploit the risk premium in these strategies. Most of these strategies have a large average return, i.e. these are zero cost portfolios, that provide a positive average payoff with a high probability. Our data set consists of a panel with 111 long-short portfolios and 630 time-series observations. Appendix 9.1 contains a detailed description of all the anomaly portfolios sorted by the publication effect as described next.

We define the "treatment" as the publication of an anomaly strategy in an academic journal. The returns of portfolios before publication are the control observations, while the returns after publication serve as the treated observations. Figure 5 shows the observation patterns for the control and treated panels. The control panel in Figure 5a uses only returns before publications and treats the post-publication returns as missing values as indicated by the shaded entries.[31] This pattern is reversed for the treated panel in Figure 5b that only uses entries after the publication.



(a) Control panel

(b) Treated panel

Figure 5: Observation patterns for the control (before publication) and treated (after publication) panels. The shaded entries represent the missing entries.

Under the assumption of no-arbitrage and complete markets, there exists a unique stochastic discount factor (SDF) that can price all assets. In an approximate factor model this SDF is spanned by the latent factors which can explain well the cross-section of expected returns.[32] We assume that

[31]A small number of anomalies have also a few missing values at the beginning of the sample.

[32]See for example(Kozak et al., 2019; Kelly et al., 2018; Lettau and Pelger, 2018).

the stochastic discount factor does not change by the publication of anomalies, i.e. the same latent factors describe the returns before and after publication. However, the exposure of the assets to the SDF can be affected by the publication, i.e. the loadings with respect to the latent factors and hence also their risk premium can change. Since most observations are available before publication we estimate the latent factors from the control panel and use the regression method in Section 4.3.2 to estimate loadings and common components after publication. We compare the estimated returns without and with publication on the time periods after an anomaly has been published to study the following two questions:

1. Does publication decrease the average returns of an anomaly portfolio?

2. Does publication decrease the pricing error (alpha) against the popular CAPM model?

The first question is related to the observation[33] that the mean return is lowered after publication, i.e. the risk premium of the anomaly decreases. This can be due to mis-pricing of the portfolios and after investors become aware of this arbitrage opportunity it is traded away. The second question is related to the fishing for alphas argument, i.e. journals are only willing to publish an anomaly if it is significant relative to the most relevant benchmark model, the CAPM model, although it may be just noise. Hence, it is possible that an insignificant anomaly gets published due to the multiple hypothesis testing problem (Harvey et al., 2016). Instead of directly comparing the mean returns before and after publication, which could be different because of time effects, we estimate the counterfactuals, the returns if the anomaly had not been published, and compare the returns with and without publication on the same time periods to control for the time effect. Indeed, we show that the naive comparison of mean returns and pricing errors on the time periods before and after publication is much more likely to find an effect as the sample mean returns are in general lower on the latter part of the data set.

In our analysis we assume that the publication of anomalies does not depend on anomaly specific characteristics S, i.e. $P(W_{it}|S) = P(W_{it})$. One reason is that most portfolio specific characteristics are time-varying and hence cannot be used as a time-invariant covariate S. For example, the size portfolio includes by construction only stocks with similar size characteristics. However, other firm characteristics, e.g. their book-to-market ratio, are in general time-varying for this portfolio. It

[33]See (Schwert, 2003; McLean and Pontiff, 2016; Harvey et al., 2016; Chen and Zimmermann, 2018)

is possible that the publication depends on the time-series pattern of certain strategies, but our estimator in its current form only allows for cross-sectional, time-invariant variables to control for differences in the treatment probability.[34]

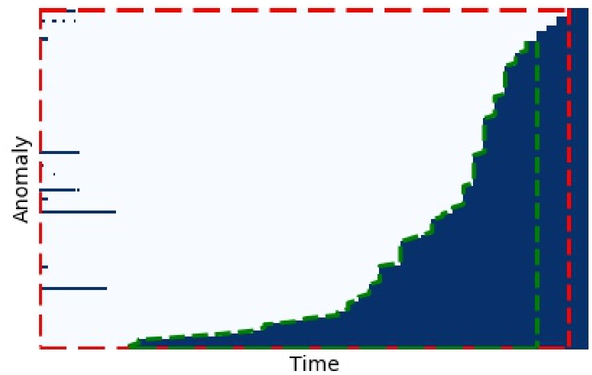


Figure 6: The red and green boxes indicate the time periods used to estimate the factor model before and after publication respectively.

We first estimate the latent factor model from the control returns before publication. We use the data until the end of 2013 which is the last year for which we observe unpublished anomalies, as indicated by the red box in Figure 6. This provides us with the common components of the control before and after publication and the latent factor time-series from 1963 to 2013. Note, that the latent factors are a weighted average of control returns. For the latter years, there are only very few control return time-series which results in noisy time-series for the latent factors.[35] In order to strike a balance between a precise estimation and using as much data as possible, our benchmark analysis will test for treatment effects until the year 2010. We confirm that our results are robust to changing the time horizon. Given the latent factor time series, we run regressions on the treated return time-series to obtain the loadings and common components for the treated returns after publication. Then, we calculate the mean returns or pricing errors on the time period after publication until 2010 for the control and treated data as indicated by the green box in 6.

We apply Theorem 6 to obtain the test statistics for the mean return and pricing error effect. For $Z = \vec{1}$ we test whether the mean return is significantly affected by the publication. For $Z = [\vec{1}, F_m]$, where F_m is the observed excess return of the market factor, we test whether the pricing error, the

[34]We are currently working on extending our theoretical and empirical framework to include time-varying cross-sectional features S_{it}. However this is beyond the scope of this paper.

[35]In fact, the latent factors from 2011 to 2013 are the weighted average of fewer than 11 control returns.

coefficient corresponding to the intercept $\vec{1}$, is significantly affected by the publication.

Figure 7 illustrates the counterfactual outcomes for three anomalies that experience a significant publication effect in their mean returns. Figure 13 collects the corresponding results for the 13 anomalies that exhibit the statistically strongest treatment effect on their mean returns and the five most prominent anomalies in the literature, namely size, value, investment, profitability, and momentum. The blue line plots the common component of returns based on the latent factors estimated on the control data. This means the blue line after the publication date are imputed values that serve as the counterfactual outcome. The orange line are the cumulative returns of the common component after publication based on the loadings estimated on the treated panel. The green line are simply the cumulative returns of the observed price process after publication. Note, that in an approximate factor model the risk premium should be fully captured by the common component and indeed the orange line is very close to the green line but less wiggly as it averages out some idiosyncratic noise. The large difference between the blue and orange lines confirms the statistical finding that the publication significantly affects their mean returns.

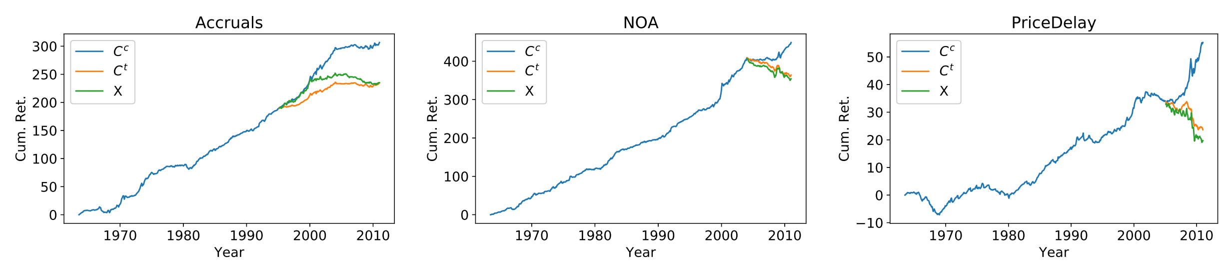
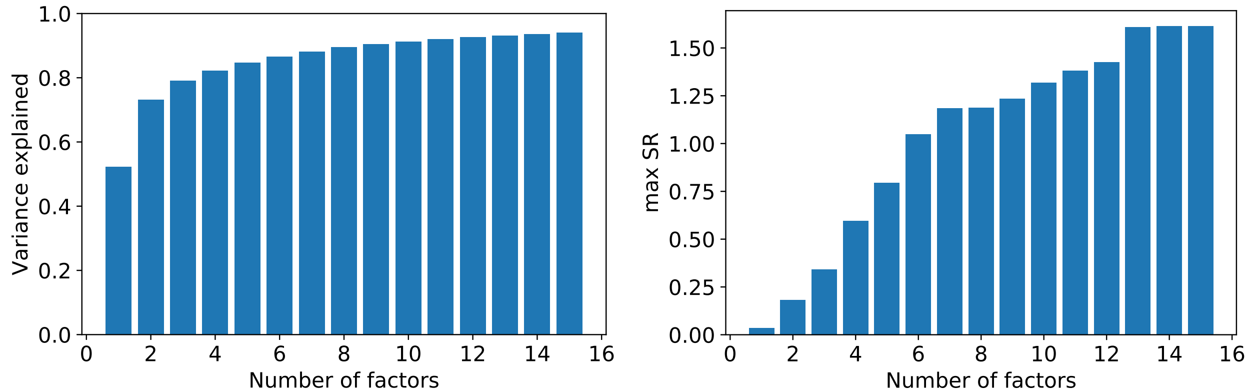


Figure 7: Publication effect: Cumulative returns of (in blue) control common component $\widetilde{C}_{it}^{ctrl}$ before publication and counterfactual after publication, (in orange) treatment common component $\widetilde{C}_{it}^{treat}$ after publication and (in green) observed X_{it} after publication. 10 latent factors.

Figure 8 shows the variation explained and the maximum Sharpe ratio of the SDF for different number of factors on the control data.[36] The marginal increase of the variation explained and the maximum Sharpe ratio with more than 10 factors is very small. A factor model with 6 to 10 factors seems to capture most of the information in this data set. If not stated otherwise we use 10 latent factors.

[36]The Sharpe ratio of the SDF is the maximum Sharpe ratio obtained by mean-variance optimization of all the estimated latent factors.



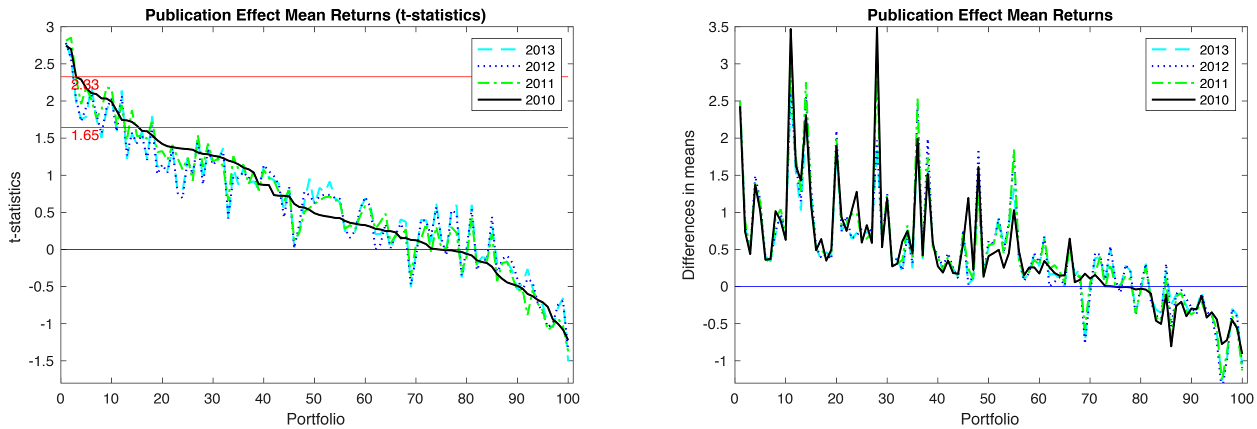
(a) Variation explained

(b) Max Sharpe-ratio

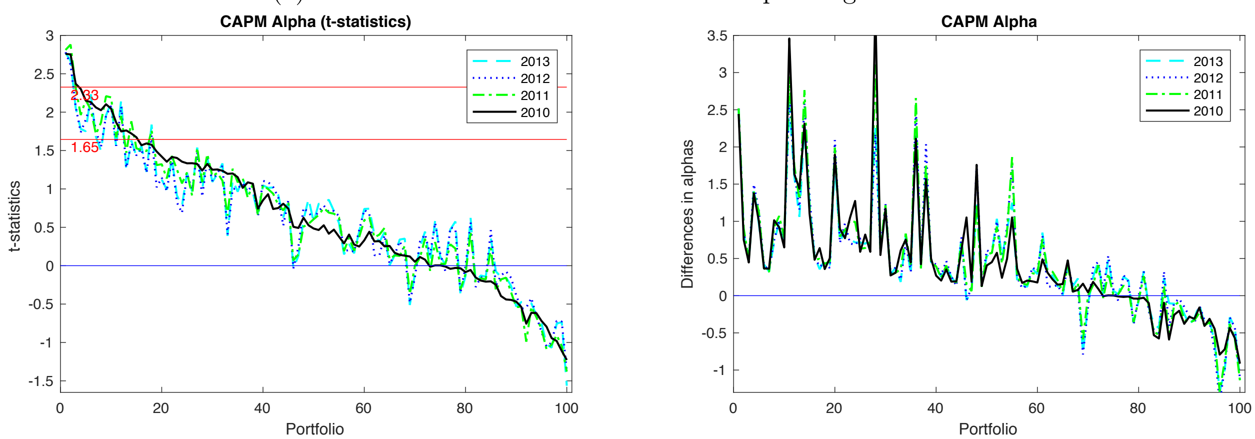
Figure 8: Variation explained and the Sharpe ratio of the SDF for different number of factors on the control data.

First, the majority of the anomaly portfolios have lower average returns and pricing errors after publication. Figure 9 collects the publication effect on mean returns and CAPM pricing errors for all anomalies for different horizons of the treatment effect. The left plots show the t-statistics, while the right side has the non-normalized differences. The portfolios are sorted by their t-statistics for the reference year 2010. Over 80% of the differences are positive, i.e. after publication, the anomaly risk premium is more likely to decrease. The results are very robust to the choice of the final year of the treatment effect and justify why we can focus our analysis on the year 2010. Importantly, for the 20% negative differences, the values are economically much smaller than the positive values and all statistically insignificant for a 95% significance level. On the other hand, the positive publication effect is economically large with a monthly return ranging from 0.5 to 3%.

Second, only around 14% of the publication effects are statistically significant. Using a one-sided 95% test shows that only 15 anomalies are statistically significantly affected by the publication, which holds for the mean returns and pricing errors. Due to the multiple testing problem this number has to be viewed as an upper bound. When correcting the critical value for multiple testing, there are even fewer portfolios with a significant publication effect. This result is not surprising as we are correctly taking into account the uncertainty in the estimation. The time periods after publication are relatively short, while the mean estimation itself is known to have high standard errors. In addition, there are relatively few control anomalies that have not been published in the later part of the data set. Hence, it cannot be avoided that the counterfactual outcome is relatively noisy which further increases the variance of our test statistic. In summary, because of the nature of the data, we show that the bar for classifying a publication effect as significant has to be put quite high.



(a) Differences in mean returns and corresponding t-statistics



(b) Differences in pricing errors (CAPM alpha) and the corresponding t-statistics

Figure 9: Publication effect: Differences and corresponding t-statistics with the last year to test the treatment effect ranging from 2010 to 2013. Portfolios are sorted by their t-statistics for differences in mean returns in 2010 in descending order. Left panels show the t-statstics for differences in means and pricing errors. Right panels show absolute differences in means and pricing errors.

Third, the pricing error and risk premium effects on the mean are very much aligned. The top and bottom plots of Figure 9 are very close to each other. This is not surprising as the long-short factors are constructed to be "market neutral", and hence most of their mean returns should not be explained by a market portfolio. Hence, most of our findings about mean returns directly carry over to CAPM alphas.

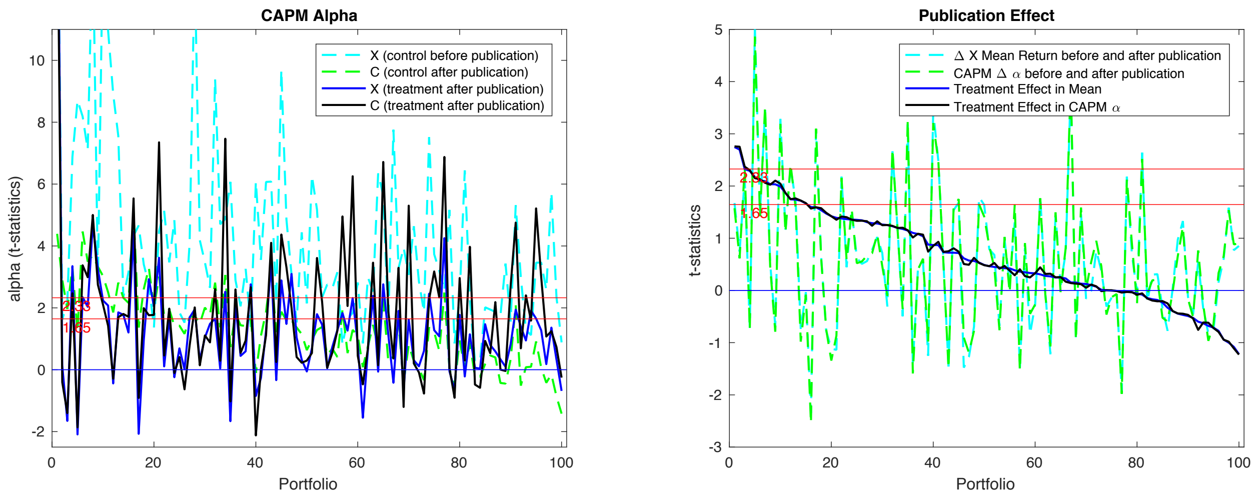
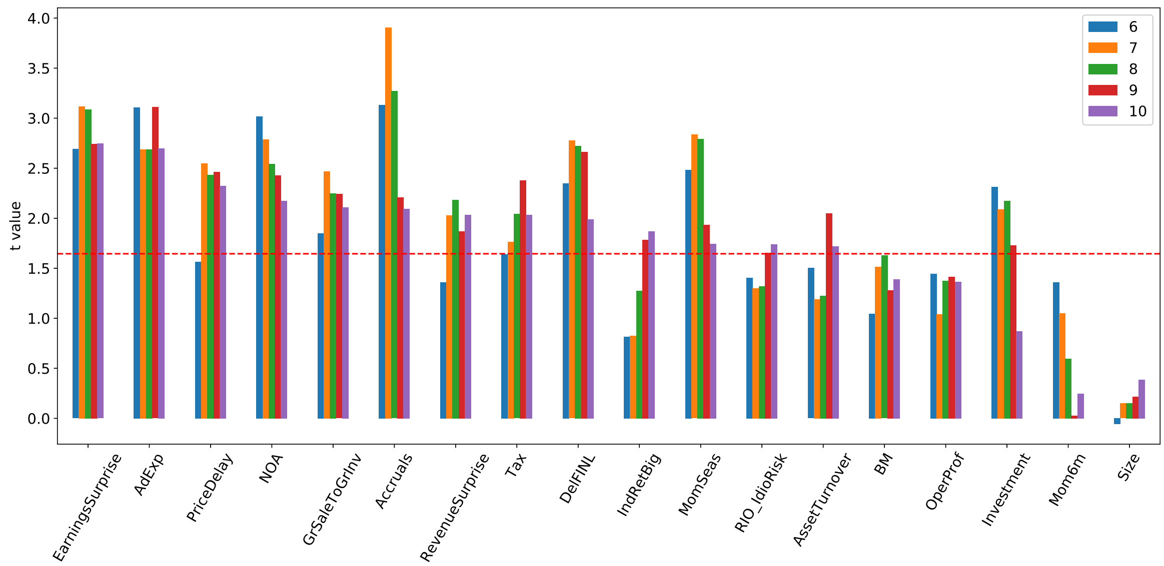
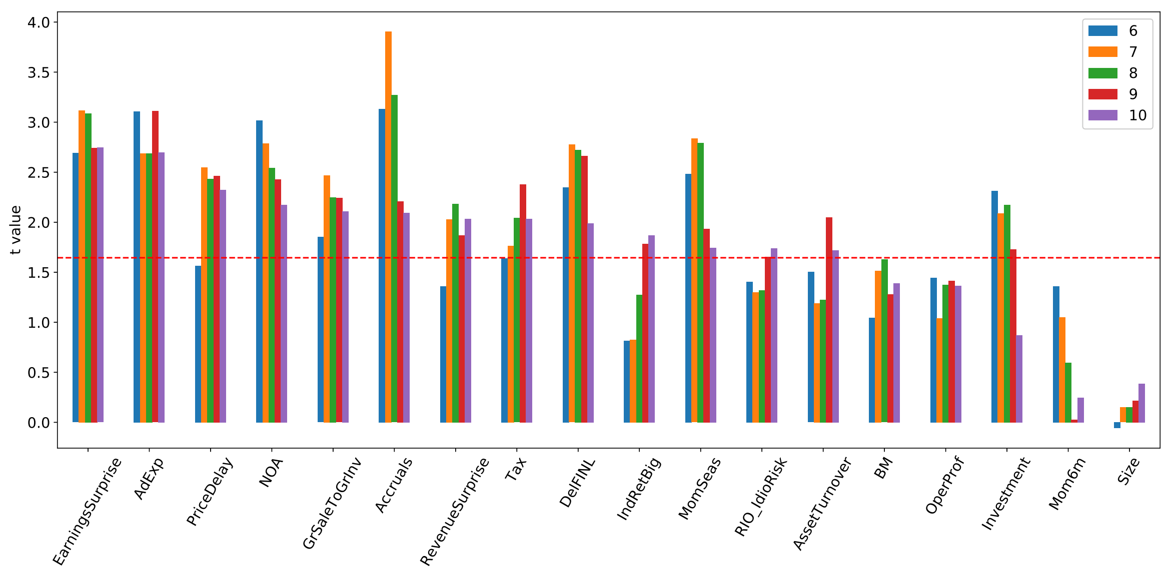


Figure 10: Publication effect. Left plot: Pricing error (CAPM alpha) t-statistics estimated on the control (without publication) observations before publication (light blue), treated (with publication) observations after publication (dark blue), estimated control observations after publication (green) and estimated treated observations after publication (black). Pricing errors have higher t-statistics when estimated before publication. Right plot: Naive treatment effect based on difference in mean returns or pricing errors before and after publication (light blue and green line) and synthetic control treatment effect using only observations after the publication (dark blue and black line). Portfolios are sorted by their t-statistics for differences in mean returns in 2010 in descending order. Red lines denote 1% and 5% critical values.

Fourth and importantly, a naive estimation of the publication effect leads to different and misleading conclusions. Figure 10 contrasts our treatment effect with a simple comparison of mean and pricing errors before and after publication without constructing a counterfactual. Note, that comparing the sample mean of returns before and after publication can suggest a larger publication effect if sample means are lower in the second part of the data, even for those anomalies that are not published. This is exactly what happens in our case. The left plot in Figure 10 shows the t-statistics for pricing errors based on the return date before publication. Not surprisingly, almost all anomalies are significant as otherwise they would not have been published. The dark blue line shows the result of the same regression but on the time periods after publication. Substantially fewer anomalies are significant. However, the before publication returns are a bad control as they neglect any time effects. In contrast, the green line depicting the results of the counterfactual, i.e. pricing errors if the anomaly had not been published but calculated on the time periods after its publication, is much lower. Note, that using the common component or observed returns for the regression of the treated data give similar results. The right plots confirms that a naive estimation, that simply compares means and pricing errors before and after publication, would suggest that more anomalies have a significant publication effect. The green and light blue line describe the naive approach that has more spikes above the critical values than our approach. These spikes also occur for very different anomalies compared to our approach. On a 1% confidence level the naive approach would suggest a significant publication effect for three times more anomaly portfolios.

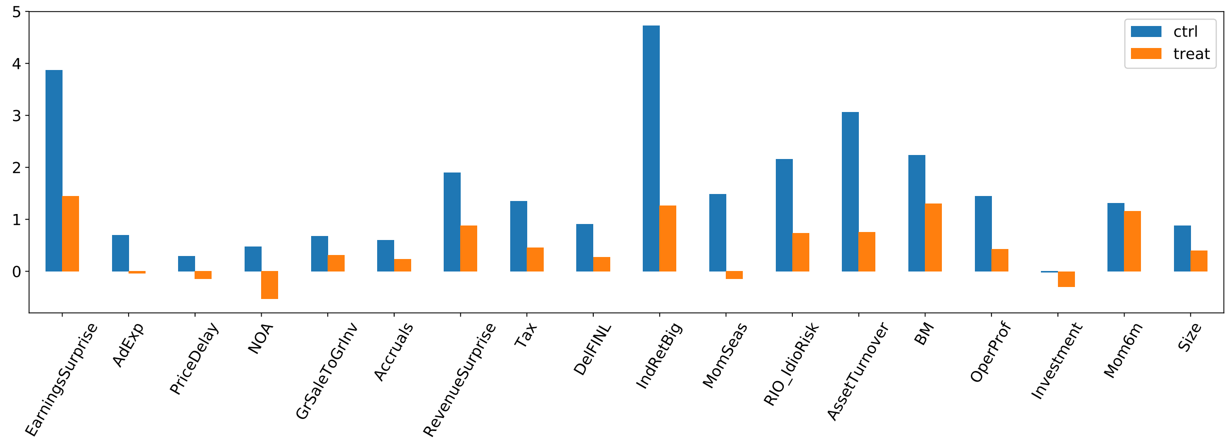


(a) t values for differences in mean returns $\bar{C}_i^{ctrl} - \bar{C}_i^{treat}$

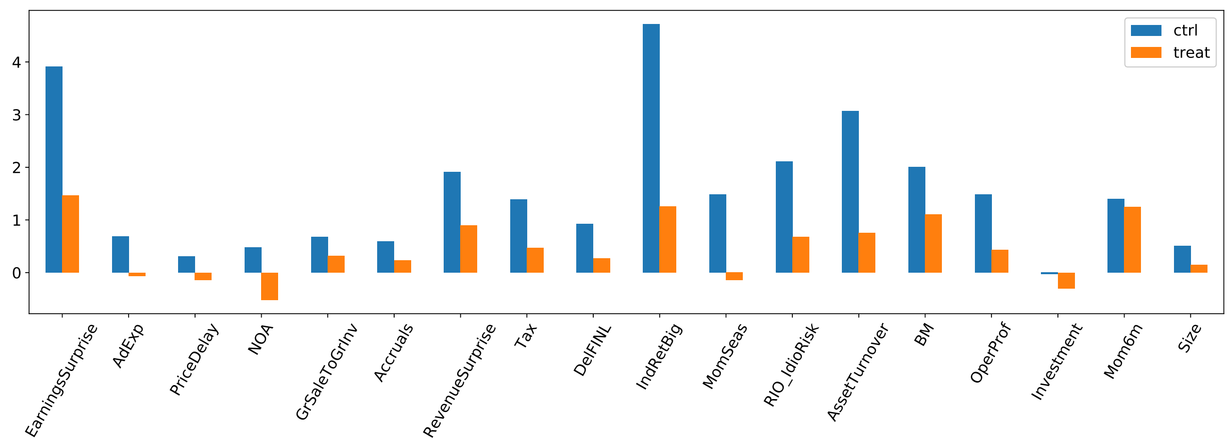


(b) t values for differences in CAPM pricing errors $\beta_{i,\vec{1}}^{ctrl} - \beta_{i,\vec{1}}^{treat}$

Figure 11: Publication effect: t-statistics for 6 to 10 latent factors. The red dashed-line is the critical value for a one-sided test of a negative publication effect. Results for the 13 anomalies with the statistically largest effects and the 5 most prominent anomalies.



(a) Mean returns $\bar{C}_i^{ctrl}$ and $\bar{C}_i^{treat}$



(b) CAPM α^{ctrl} and α^{treat}

Figure 12: Publication effect: Mean returns $\bar{C}_i^{ctrl}$ and $\bar{C}_i^{treat}$ and CAPM pricing errors $\beta_{i,\vec{1}}^{ctrl}$ and $\beta_{i,\vec{1}}^{treat}$ under a 10-factor model. Results for the 13 anomalies with the statistically largest effects and the 5 most prominent anomalies.

Fifth, we zoom in and present the detailed results for the 13 anomalies with the statistically largest effect and the 5 most prominent anomalies. Figure 11 shows the t-statistics for 6 to 10 latent factors spanning the SDF. It is apparent that both, the effect on mean returns and the pricing errors, are robust to the choice of the number of latent factors. Importantly, for sufficiently many latent factors neither of the "classical" anomalies size, value, profitability, investment, and momentum is significantly affected by their publication. This suggests, that these anomalies represent the systematic risk that requires a risk premium and which is part of the SDF. Some of the anomalies whose risk premium disappear after publication are less "standard", e.g. advertisement expenditure.

This is suggestive that these were either arbitrage opportunities that were traded away by informed investors or their detection was simply spurious.

Lastly, we highlight the distinction between statistical and economic significance. Figure 12 shows the mean returns and pricing errors for the same subset of anomalies as Figure 11. The statistically significant portfolios also have economically significant differences ranging from 0.5 to 3% monthly returns. However, an economically large difference does not necessarily turn into statistical significance when correctly accounting for the uncertainty. For example, the size portfolio experiences an effect of around 0.5% which is insignificant because of its volatile time-series as shown in Figure 13.

8 Conclusion

In this paper, we propose a method to estimate a latent factor model from partially observed panel data. The estimation is based on an adjusted covariance matrix estimated from the partially observed panel data. We derive the inference theory for the estimated factors, loadings, and common components. The asymptotic variance of the estimators is larger than that from the fully observed panel. In particular, there is an additional variance correction term in the asymptotic variance compared with the fully observed panel. Based on the inferential theory, we construct a test for the treatment effect for each unit at any time. In our empirical analysis of anomaly long-short portfolios we find that around 14% of the portfolios returns are significantly reduced by academic publication.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Aït-Sahalia, Y. and Xiu, D. (2018). Principal component analysis of high-frequency data. *Journal of the American Statistical Association*, pages 1–17.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2018). Matrix completion methods for causal panel data models. Technical report, National Bureau of Economic Research.

Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

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

Bai, J. and Ng, S. (2019). Matrix completion, counterfactuals, and factor analysis of missing data. *arXiv preprint arXiv:1910.06677*.

Bańbura, M. and Modugno, M. (2014). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data. *Journal of Applied Econometrics*, 29(1):133–160.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717.

Carvalho, C., Masini, R., and Medeiros, M. C. (2018). Arco: an artificial counterfactual approach for high-dimensional panel time-series data. *Journal of econometrics*, 207(2):352–380.

Chen, A. Y. and Zimmermann, T. (2018). Publication bias and the cross-section of stock returns.

Chen, Y., Fan, J., Ma, C., and Yan, Y. (2019). Inference and uncertainty quantification for noisy matrix completion. *arXiv preprint arXiv:1906.04159*.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

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

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on kalman filtering. *Journal of Econometrics*, 164(1):188–205.

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

Gagliardini, P., Ossola, E., and Scaillet, O. (2019). A diagnostic criterion for approximate factor structure. *Journal of Econometrics*.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, pages 315–331.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68.

Hirano, K., Imbens, G. W., and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71(4):1161–1189.

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

Jungbacker, B., Koopman, S. J., and Van der Wel, M. (2011). Maximum likelihood estimation for dynamic factor models with missing data. *Journal of Economic Dynamics and Control*, 35(8):1358–1368.

Kelly, B., Pruitt, S., and Su, Y. (2018). Characteristics are covariances: A unified model of risk and return. Technical report, National Bureau of Economic Research.

Kozak, S., Nagel, S., and Santosh, S. (2019). Shrinking the cross-section. *Journal of Financial Economics*.

Lettau, M. and Pelger, M. (2018). Factors that fit the time series and cross-section of stock returns. Technical report, National Bureau of Economic Research.

Li, K. T. (2017). Statistical inference for average treatment effects estimated by synthetic control methods.

Li, K. T. and Bell, D. R. (2017). Estimation of average treatment effects with panel data: Asymptotic theory and implementation. *Journal of Econometrics*, 197(1):65–75.

Masini, R. and Medeiros, M. C. (2018). Counterfactual analysis with artificial controls: Inference, high dimensions and nonstationarity. *High Dimensions and Nonstationarity (September 1, 2018)*.

Mazumder, R., Hastie, T., and Tibshirani, R. (2010). Spectral regularization algorithms for learning large incomplete matrices. *Journal of machine learning research*, 11(Aug):2287–2322.

McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.

Meng, X.-L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. *Biometrika*, 80(2):267–278.

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

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

Pelger, M. (2019). Large-dimensional factor modeling based on high-frequency observations. *Journal of Econometrics*, 208(1):23–42.

Pelger, M. and Xiong, R. (2018). State-varying factor models of large dimensions.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Schwert, G. W. (2003). Anomalies and market efficiency. *Handbook of the Economics of Finance*, 1:939–974.

Stock, J. H. and Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, volume 2, pages 415–525. Elsevier.

Su, L., Miao, K., and Jin, S. (2019). On factor models with random missing: Em estimation, inference, and cross validation.

Zhou, Y., Wilkinson, D., Schreiber, R., and Pan, R. (2008). Large-scale parallel collaborative filtering for the netflix prize. In *International conference on algorithmic applications in management*, pages 337–348. Springer.

9 Appendix

9.1 Empirical Results

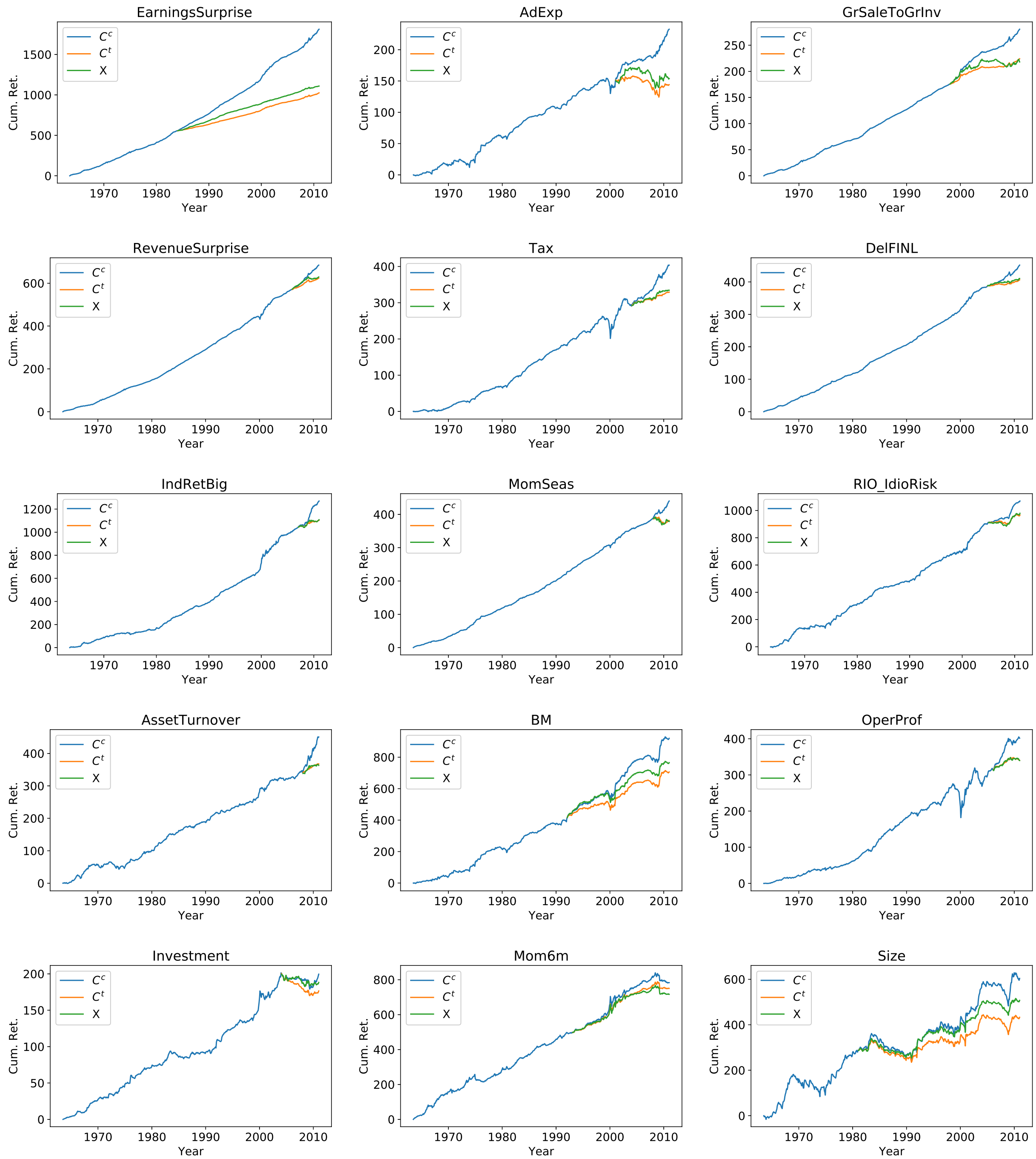


Figure 13: Publication effect: Cumulative returns of 1. (blue) control common component $\widetilde{C}_{it}^{ctrl}$ before publication and counterfactual after publication 2. (orange) treatment common component $\widetilde{C}_{it}^{treat}$ after publication, 3. (green) observed X_{it} after publication. 10 latent factors.

Acronym	Description	Authors	Start Date	Publication	$C^{ctrl} - C^{treat}$ t-stats	$C^{ctrl} - C^{treat}$ mean	$\alpha^{ctrl} - \alpha^{treat}$ t-stats	$\alpha^{ctrl} - \alpha^{treat}$ mean
EarnSurp	Earnings Surprise	Foster et al	1963/07	1984	2.75	2.43	2.76	2.44
AdExp	Advertising Expense	Chan et al	1963/07	2001	2.70	0.74	2.75	0.75
PriceDelay	Price delay	Hou and Moskowitz	1963/07	2005	2.32	0.44	2.37	0.45
KZ	Kaplan Zingales index	Lamont et al	1963/07	2001	2.29	1.37	2.30	1.38
NOA	Net Operating Assets	Hirshleifer et al	1963/07	2004	2.17	1.01	2.15	1.00
GrSaleToGrInv	Sales growth over inventory growth	Abarbanell and Bushee	1963/07	1998	2.11	0.36	2.13	0.37
Accruals	Accruals	Sloan	1964/06	1996	2.09	0.37	2.06	0.36
RevenueSurprise	Revenue Surprise	Jegadeesh and Livnat	1963/07	2006	2.04	1.02	2.03	1.01
Tax	Taxable income to income	Lev and Nissim	1963/07	2004	2.03	0.89	2.10	0.92
ChFinLiab	Change in financial liabilities	Richardson et al	1963/07	2005	1.99	0.63	2.05	0.65
IndRetBig	Industry return of big firms	Hou	1963/07	2007	1.87	3.47	1.87	3.46
Seasonality	Return Seasonality	Heston and Sadka	1963/07	2008	1.75	1.63	1.75	1.63
RIO_IdioRisk	Inst Own and Idio Vol	Nagel	1963/07	2005	1.74	1.43	1.76	1.44
AssetTurnover	Asset Turnover	Soliman	1963/07	2008	1.72	2.31	1.72	2.31
RD	R&D over market cap	Chan et al	1963/07	2001	1.67	1.10	1.67	1.11
DivOmit	Dividend Omission	Michaely Thaler Womack	1963/07	1995	1.60	0.49	1.57	0.48
ChNWC	Change in Net Working Capital	Soliman	1963/07	2008	1.59	0.64	1.59	0.64
GrGMToGrSales	Gross Margin growth over sales growth	Abarbanell and Bushee	1963/07	1998	1.54	0.35	1.58	0.36
Spinoff	Spinoffs	Cusatis et al	1963/07	1993	1.47	0.50	1.49	0.50
NetDebtPrice	Net debt to price	Penman Richardson Tuna	1963/07	2007	1.42	1.90	1.42	1.90
BM	Book to market	Fama and French	1963/07	1992	1.39	0.94	1.35	0.90
Mscore	Mohanram G-score	Mohanram	1964/01	2005	1.38	0.75	1.42	0.77
OperProf	operating profits / book equity	Fama and French	1963/07	2006	1.37	1.02	1.40	1.05
EarnSupBig	Earnings surprise of big firms	Hou	1963/07	2007	1.36	1.28	1.36	1.27
SurpriseRD	Unexpected R&D increase	Eberhart et al	1963/07	2004	1.35	0.59	1.33	0.59
Herf	Industry concentration (Herfindahl)	Hou and Robinson	1963/07	2006	1.34	0.83	1.34	0.82
CFPinc	Cash flow to market	Lakonishok et al	1963/07	1994	1.31	0.59	1.33	0.59
RoA	earnings / assets	Balakrishnan, Bartov, Faurel	1963/07	2010	1.29	3.51	1.24	3.86
GrSaleToGrOverhead	Sales growth over overhead growth	Abarbanell and Bushee	1963/07	1998	1.28	0.53	1.32	0.55
RIO_Turnover	Inst Own and Turnover	Nagel	1963/07	2005	1.26	1.19	1.25	1.17
ChDeprToPPE	Change in depreciation to gross PPE	Holthausen and Larcker	1963/07	1992	1.25	0.27	1.25	0.27
ChInventory	Inventory Growth	Thomas and Zhang	1963/07	2002	1.23	0.31	1.24	0.31
BMent	Enterprise component of BM	Penman Richardson Tuna	1963/07	2007	1.20	0.61	1.20	0.61
SalesToPrice	Sales-to-price	Barbee et al	1963/07	1996	1.18	0.75	1.20	0.75
ChATurn	Change in Asset Turnover	Soliman	1963/07	2008	1.15	0.45	1.15	0.45
OperLeverage	Operating Leverage	Novy-Marx	1963/07	2010	1.10	1.99	1.01	2.11
BPEBM	Leverage component of BM	Penman Richardson Tuna	1963/07	2007	1.08	0.42	1.09	0.42

Table 2: Summary statistics of the anomaly portfolios (publication effect until 2010).

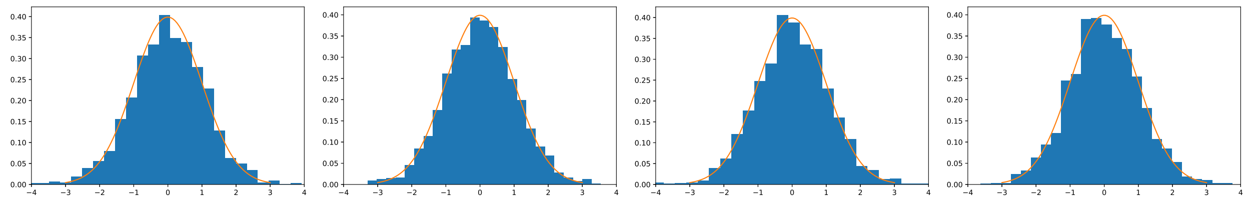
Label	Name	Authors	Start Date	Publication	$C^{ctrl} - C^{treat}$ t-stats	$C^{ctrl} - C^{treat}$ mean	$\alpha^{ctrl} - \alpha^{treat}$ t-stats	$\alpha^{ctrl} - \alpha^{treat}$ mean
IdioRisk	Idiosyncratic risk	Ang et al	1963/07	2006	1.04	1.52	1.07	1.57
Mom1813	Momentum-Reversal	De Bondt and Thaler	1963/07	1985	0.88	0.60	0.74	0.50
InvToRev	Investment	Titman et al	1963/07	2004	0.87	0.27	0.85	0.27
SalesGr	Revenue Growth Rank	Lakonishok et al	1963/07	1994	0.87	0.19	0.93	0.20
CompDebtI	Composite debt issuance	Lyandres Sun Zhang	1963/07	2008	0.73	0.35	0.74	0.35
ChInvestInd	Change in capital inv (ind adj)	Abarbanell and Bushee	1963/07	1998	0.72	0.18	0.75	0.19
DivInit	Dividend Initiation	Michaely Thaler Womack	1963/07	1995	0.72	0.17	0.81	0.19
Mom1m	Short term reversal	Jegedeesh	1963/07	1989	0.72	0.59	0.74	0.62
EffFrontier	Efficient frontier index	Nguyen and Swanson	1963/07	2009	0.61	1.19	0.51	1.05
ExchSwitch	Exchange Switch	Dharan and Ikenberry	1963/07	1995	0.57	0.22	0.49	0.19
Tangibility	Tangibility	Hahn and Lee	1963/07	2009	0.56	1.61	0.62	1.76
ChLTI	Change in long-term investment	Richardson et al	1963/07	2005	0.53	0.13	0.53	0.13
RIO_BM	Inst Own and BM	Nagel	1963/08	2005	0.49	0.41	0.48	0.41
ShareIs5	Share issuance (1 year)	Pontiff and Woodgate	1963/07	2008	0.46	0.45	0.46	0.45
Mom12m	Momentum (12 month)	Jegadeesh and Titman	1963/07	1993	0.45	0.49	0.53	0.58
IntanEP	Intangible return	Daniel and Titman	1963/07	2006	0.44	0.25	0.42	0.24
High52	52 week high	George and Hwang	1963/07	2004	0.43	0.45	0.47	0.49
MaxRet	Maximum return over month	Bali et al	1963/07	2010	0.42	1.03	0.38	1.05
Size	Size	Banz	1963/07	1981	0.39	0.47	0.29	0.36
RoE	net income / book equity	Haugen and Baker	1963/07	1996	0.36	0.15	0.41	0.17
EP	Earnings-to-Price Ratio	Basu	1963/07	1977	0.35	0.26	0.27	0.20
Mom36m	Long-run reversal	De Bondt and Thaler	1963/07	1985	0.33	0.26	0.25	0.19
ChPM	Change in Profit Margin	Soliman	1963/07	2008	0.33	0.18	0.33	0.18
OrderBacklog	Order backlog	Rajgopal et al	1970/12	2003	0.31	0.34	0.44	0.49
EarnCons	Earnings Consistency	Alwathainani	1963/07	2009	0.28	0.26	0.32	0.31
CFPcash	Operating Cash flows to price	Desai, Rajgopal, and Venkatachalam	1964/06	2004	0.27	0.18	0.32	0.22
ChNCOA	Change in Noncurrent Operating Assets	Soliman	1963/07	2008	0.25	0.15	0.25	0.15
Mom6m	Momentum (6 month)	Jegadeesh and Titman	1963/07	1993	0.25	0.15	0.26	0.16
Price	Price	Blume and Husic	1963/07	1972	0.23	0.65	0.17	0.47
ChCOA	Change in current operating assets	Richardson et al	1963/07	2005	0.15	0.06	0.14	0.05
IntanCFP	Intangible return	Daniel and Titman	1963/07	2006	0.15	0.09	0.13	0.08
MomRev	Momentum and LT Reversal	Chan and Kot	1963/07	2006	0.13	0.17	0.12	0.16
Leverage	Market leverage	Bhandari	1963/07	1988	0.12	0.11	0.05	0.04
ChBE	Sustainable Growth	Lockwood and Prombutr	1963/07	2010	0.10	0.16	0.11	0.19
NetPayoutYield	Net Payout Yield	Boudoukh et al	1963/07	2007	0.07	0.09	0.08	0.09
DivYield	Dividend Yield	Naranjo et al	1963/07	1998	0.01	0.01	-0.01	-0.01

Table 3: Summary statistics of the anomaly portfolios (publication effect until 2010).

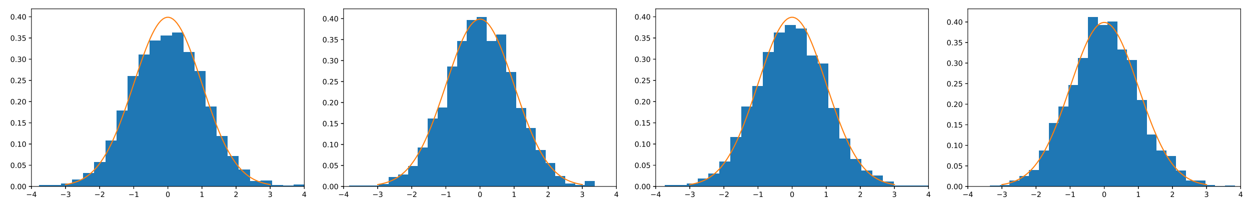
Label	Name	Authors	Start Date	Publication	$C^{ctrl} - C^{treat}$ t-stats	$C^{ctrl} - C^{treat}$ mean	$\alpha^{ctrl} - \alpha^{treat}$ t-stats	$\alpha^{ctrl} - \alpha^{treat}$ mean
IntanBM	Intangible return	Daniel and Titman	1967/06	2006	-0.01	-0.01	-0.01	-0.01
VolumeTrend	Volume Trend	Haugen and Baker	1963/07	1996	-0.01	0.00	-0.04	-0.02
ChCol	Change in current operating liabilities	Richardson et al	1963/07	2005	-0.03	-0.01	-0.03	-0.01
ProfitMargin	Profit Margin	Soliman	1963/07	2008	-0.03	-0.04	-0.03	-0.04
DolVol	Past trading volume	Brennan Chordia Subrahmanyam	1963/07	1998	-0.06	-0.03	-0.08	-0.04
GrCAPX	Change in capex (two years)	Anderson and Garcia-Feijoo	1963/07	2006	-0.08	-0.04	-0.06	-0.03
VarCF	Cash-flow variance	Haugen and Baker	1963/07	1996	-0.15	-0.09	-0.15	-0.10
Beta	CAPM beta	Fama and MacBeth	1963/07	1973	-0.17	-0.46	-0.19	-0.53
BetaSquared	CAPM beta squred	Fama and MacBeth	1963/07	1973	-0.18	-0.50	-0.21	-0.58
VolSD	Volume Variance	Chordia Roll Subrahmanyam	1963/07	2001	-0.23	-0.11	-0.20	-0.09
SinStock	Sin Stock (selection criteria)	Hong and Kacperczyk	1963/07	2009	-0.33	-0.80	-0.25	-0.59
IntanSP	Intangible return	Daniel and Titman	1963/07	2006	-0.40	-0.26	-0.40	-0.26
Illiquidity	AmihudÃ¿s illiquidity	Amihud	1963/07	2002	-0.44	-0.21	-0.44	-0.20
ZeroTrade	Days with zero trades	Liu	1963/07	2006	-0.47	-0.40	-0.45	-0.38
VolMkt	Volume to market equity	Haugen and Baker	1963/07	1996	-0.49	-0.30	-0.47	-0.28
IndMom	Industry Momentum	Grinblatt and Moskowitz	1963/07	1999	-0.53	-0.31	-0.54	-0.31
GrLTNOA	Growth in Long term net operating assets	Fairfield et al	1963/07	2003	-0.60	-0.13	-0.76	-0.16
StdTurnover	Turnover volatility	Chordia Roll Subrahmanyam	1963/07	2001	-0.63	-0.42	-0.61	-0.41
SEO	Public Seasoned Equity Offerings	Loughran and Ritter	1970/01	1995	-0.68	-0.35	-0.62	-0.31
ZScore	Altman Z-Score	Dichev	1963/07	1998	-0.71	-0.44	-0.72	-0.45
AccrualsBM	Book-to-market and accruals	Bartov and Kim	1967/05	2004	-0.77	-0.77	-0.79	-0.79
MomVol	Momentum and Volume	Lee and Swaminathan	1963/07	2000	-0.94	-0.72	-0.94	-0.72
ShareIs1	Share issuance (5 year)	Daniel and Titman	1963/07	2006	-1.00	-0.45	-0.98	-0.43
ChBEtoA		Richardson et al	1963/07	2005	-1.08	-0.56	-1.11	-0.57
PayYield	Payout Yield	Boudoukh et al	1963/07	2007	-1.22	-0.91	-1.23	-0.91
AssetGrowth	Asset Growth	Cooper et al	1963/07	2008	0.00	0.00	0.00	0.00
BetaTailRisk	Tail risk beta	Kelly and Jiang	1963/07	2014	0.00	0.00	0.00	0.00
ChTax	Change in Taxes	Thomas and Zhang	1963/07	2011	0.00	0.00	0.00	0.00
DivInd	Dividends	Hartzmark and Salomon	1963/07	2013	0.00	0.00	0.00	0.00
EntMult	Enterprise Multiple	Loughran and Wellman	1963/07	2011	0.00	0.00	0.00	0.00
GrAdExp	Growth in advertising expenses	Lou	1967/01	2014	0.00	0.00	0.00	0.00
GrEmp	Employment growth	Bazdresch, Belo and Lin	1963/07	2014	0.00	0.00	0.00	0.00
GrossProf	gross profits / total assets	Novy-Marx	1963/07	2013	0.00	0.00	0.00	0.00
InterMom	Intermediate Momentum	Novy-Marx	1963/07	2012	0.00	0.00	0.00	0.00
NumEarnIncrease	Number of consecutive earnings increases	Loh and Warachka	1963/07	2012	0.00	0.00	0.00	0.00
OrgCap	Organizational Capital	Eisfeldt and Papanikolaou	1964/12	2013	0.00	0.00	0.00	0.00
PctAcc	Percent Operating Accruals	Hafzalla et al	1964/06	2011	0.00	0.00	0.00	0.00

Table 4: Summary statistics of the anomaly portfolios (publication effect until 2010).

9.2 Simulation Results

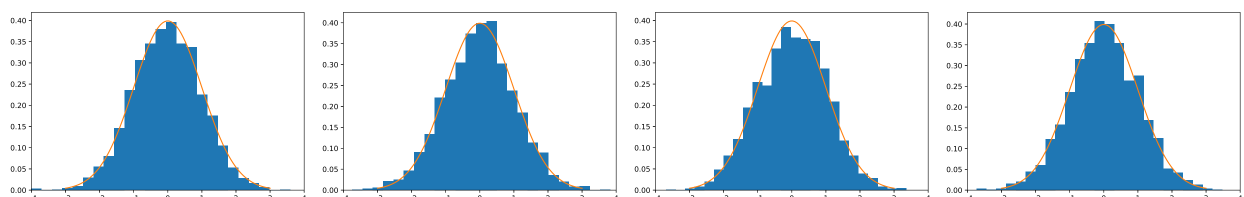


(a) (250, 500, 0.5, 0) (b) (250, 500, 0.9, 0) (c) (500, 250, 0.5, 0) (d) (500, 250, 0.9, 0)

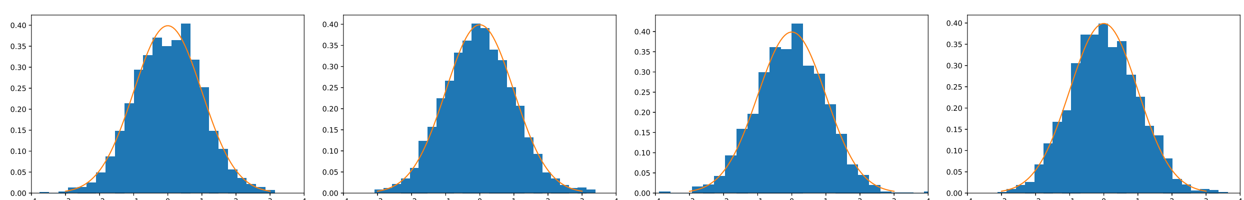


(e) (250, 500, 0.5, 1) (f) (250, 500, 0.9, 1) (g) (500, 250, 0.5, 1) (h) (500, 250, 0.9, 1)

Figure 14: Randomly Missing: Histograms of estimated standardized factors. The normal density function is superimposed on the histograms. $P(W_{it} = 1|\lambda_i) = p$ for any i and t, where $p = 0.5$ and 0.9 in the simulation. The caption in the sub-figures denotes a tuple of (N, T, p, W_{it}).

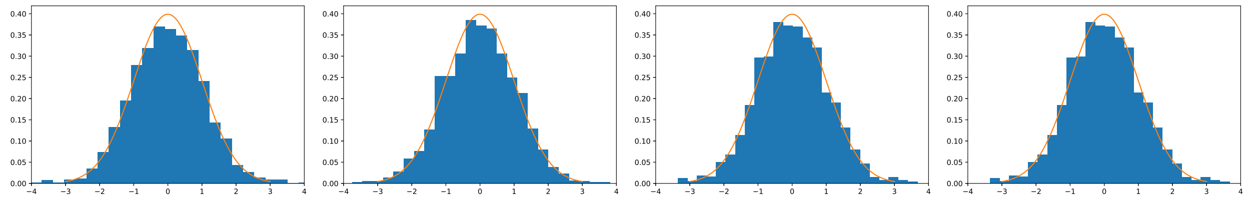


(a) (250, 500, 0.5, 0) (b) (250, 500, 0.9, 0) (c) (500, 250, 0.5, 0) (d) (500, 250, 0.9, 0)



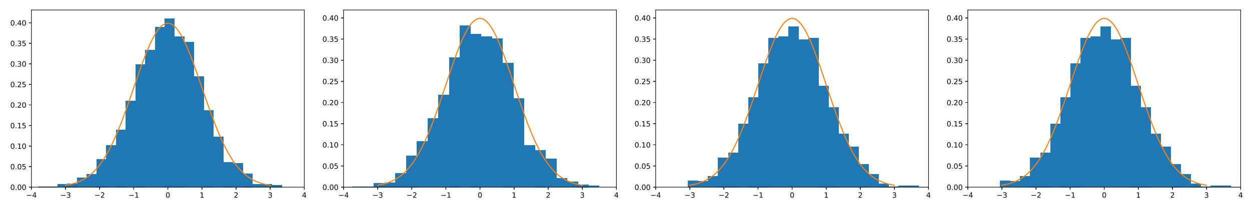
(e) (250, 500, 0.5, 1) (f) (250, 500, 0.9, 1) (g) (500, 250, 0.5, 1) (h) (500, 250, 0.9, 1)

Figure 15: Randomly Missing: Histograms of estimated standardized loadings. The normal density function is superimposed on the histograms. $P(W_{it} = 1|\lambda_i) = p$ for any i and t, where $p = 0.5$ and 0.9 in the simulation. The caption in the sub-figures denotes a tuple of (N, T, p, W_{it}).



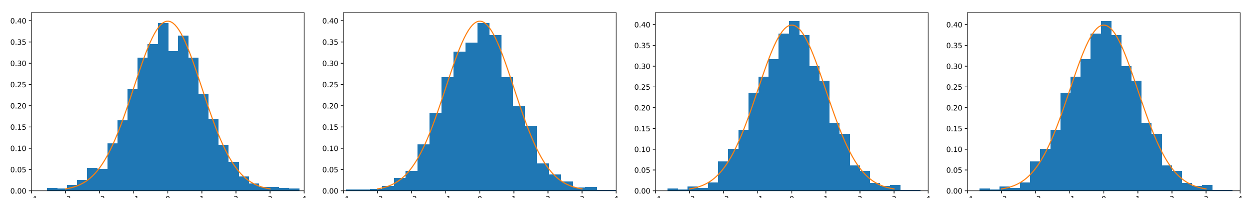
(a) (250, 500, 0) (b) (250, 500, 1) (c) (500, 250, 0) (d) (500, 250, 1)

Figure 16: Staggered Adoption: Histograms of estimated standardized factors. The normal density function is superimposed on the histograms. The caption in the sub-figures denotes a tuple of (N, T, W_{it}).



(a) (250, 500, 0) (b) (250, 500, 1) (c) (500, 250, 0) (d) (500, 250, 1)

Figure 17: Staggered Adoption: Histograms of estimated standardized loadings. The normal density function is superimposed on the histograms. The caption in the sub-figures denotes a tuple of (N, T, W_{it}).



(a) (250, 500, 0) (b) (250, 500, 1) (c) (500, 250, 0) (d) (500, 250, 1)

Figure 18: Staggered Adoption: Histograms of estimated standardized common components. The normal density function is superimposed on the histograms. The caption in the sub-figures denotes a tuple of (N, T, W_{it}).

9.3 Proofs

Denote $W_t \in \mathbb{R}^{N\times 1}$ as the t-th column in W and $\overline{W}_i \in \mathbb{R}^{T\times 1}$ as the i-th row in W; Similarly $e_t \in \mathbb{R}^{N\times 1}$ as the t-th column in e and $\overline{e}_i \in \mathbb{R}^{T\times 1}$ as the i-th row in e.

Plug $\widetilde{X} = (\Lambda F^{treat}) \odot W + e \odot W$ into

$$\left(\frac{1}{NT}(\widetilde{X}\widetilde{X}^\top) \odot Q^{(-1)}\right)\widetilde{\Lambda} = \widetilde{\Lambda}\widetilde{V},$$

and right multiply $\widetilde{V}^{-1}$ on both side, we have

$$\frac{1}{NT}\left((W \odot (\Lambda F^\top) + W \odot e)\left((F\Lambda^\top) \odot W^\top + e^\top \odot W^\top\right)\right)\widetilde{\Lambda}\widetilde{V}^{-1} = \widetilde{\Lambda}.$$

Note that (i,j)-th entry in $(W \odot (\Lambda F^\top))((F\Lambda^\top) \odot W^\top)$, $(W \odot (\Lambda F^\top))(e^\top \odot W^\top)$, $(W \odot e)((F\Lambda^\top) \odot W^\top)$ and $(W \odot e)(e^\top \odot W^\top)$ have

$$\begin{aligned}
\left((W \odot (\Lambda F^\top))((F\Lambda^\top) \odot W^\top)\right)_{ij} &= \lambda_i^\top F^\top diag(W_i \odot W_j) F\lambda_j \\
\left((W \odot (\Lambda F^\top))(e^\top \odot W^\top)\right)_{ij} &= e_i^\top diag(W_i \odot W_j) F\lambda_j \\
\left((W \odot e)((F\Lambda^\top) \odot W^\top)\right)_{ij} &= \lambda_i^\top F^\top diag(W_i \odot W_j) e_j \\
\left((W \odot e)(e^\top \odot W^\top)\right)_{ij} &= e_i^\top diag(W_i \odot W_j) e_j
\end{aligned}$$

Then, we have

$$\begin{aligned}
\widetilde{\lambda}_j = \frac{1}{NT}\widetilde{V}^{-1} &\left[\sum_{i=1}^N \widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i \odot W_j)F\lambda_j/q_{ij} + \sum_{i=1}^N \widetilde{\lambda}_i e_i^\top diag(W_i \odot W_j)F\lambda_j/q_{ij}\right. \\
&\left. + \sum_{i=1}^N \widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i \odot W_j)e_j/q_{ij} + \sum_{i=1}^N \lambda_i e_i^\top diag(W_i \odot W_j)Fe_j/q_{ij}\right] \qquad (24)
\end{aligned}$$

Denote $H_j = \frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^N \widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i \odot W_j)F/q_{ij}$.

From Equation (24), we have

$$\widetilde{\lambda}_j - H_j\lambda_j = \widetilde{V}^{-1}\left(\frac{1}{N}\sum_{i=1}^N \widetilde{\lambda}_i\gamma(i,j) + \frac{1}{N}\sum_{i=1}^N \widetilde{\lambda}_i\zeta_{ij} + \frac{1}{N}\sum_{i=1}^N \widetilde{\lambda}_i\eta_{ij} + \frac{1}{N}\sum_{i=1}^N \widetilde{\lambda}_i\xi_{ij}\right),$$

where $\gamma(i,j) = \mathbb{E}\left[\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} e_{it}e_{jt}\right]$ and

$$\begin{aligned}
\widetilde{\gamma}(i,j) &= \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} \mathbb{E}[e_{it}e_{jt}] = \gamma(i,j) \\
\zeta_{ij} &= \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} e_{it}e_{jt} - \widetilde{\gamma}(i,j) \\
\eta_{ij} &= \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} \lambda_i^\top F_t e_{jt} \\
\xi_{ij} &= \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} \lambda_j^\top F_t e_{it}.
\end{aligned}$$

from $Tq_{ij} = T \cdot |\mathcal{Q}_{ij}|/T = |\mathcal{Q}_{ij}|$. From Lemma 7, we have $\left\|\widetilde{V}^{-1}\right\| = O_p(1)$; From $\frac{1}{N}\widetilde{\Lambda}^\top\widetilde{\Lambda} = I_r$, we have $\frac{1}{N}\sum_{i=1}^N \left\|\widetilde{\lambda}_i\right\|^2 = O_p(1)$; From Assumption 2.1 that for all i and j, $\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_tF_t^\top \xrightarrow{P} \Sigma_F$. We have for any j

$$\|H_j\|^2 \leq \left\|\widetilde{V}^{-1}\right\|^2 \left(\frac{1}{N}\sum_{i=1}^N \left\|\widetilde{\lambda}_i\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^N \|\lambda_i\|^2\right)\left(\frac{1}{N}\sum_{i=1}^T \left\|\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_tF_t^\top\right\|^2\right) = O_p(1).$$

Furthermore, denote $H = \frac{1}{NT}\widetilde{V}^{-1}\widetilde{\Lambda}^\top\Lambda F^\top F$ and $\delta_{NT} = \min(\sqrt{N}, \sqrt{T})$ for the rest of the appendix.

Lemma 3. *Under Assumptions 1-4, we have for some $M_1 < \infty$, and for all N and T,*

1. $\frac{1}{N}\sum_{i=1}^N\sum_{i=1}^N \gamma(i,j)^2 \leq M_1$*, where* $\gamma(i,j) = \mathbb{E}\left[\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} e_{it}e_{jt}\right]$
2. $\mathbb{E}\left[\left(\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\lambda_i^\top \sum_{t\in\mathcal{Q}_{ij}} F_te_{jt}\right)^2\right] \leq M_1$

Proof. 1. Let $\rho(i,j) = \gamma(i,j)/\left[\mathbb{E}\left[\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} e_{it}^2\right]\mathbb{E}\left[\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} e_{jt}^2\right]\right]^{1/2}$
$= \gamma(i,j)/\left[\overline{\gamma}_{|\mathcal{Q}_{ij}|}(i,i)\overline{\gamma}_{|\mathcal{Q}_{ij}|}(j,j)\right]^{1/2}$, where $\overline{\gamma}_{|\mathcal{Q}_{ij}|}(i,i) = \mathbb{E}\left[\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} e_{it}^2\right]$. Then $|\rho(i,j)| \leq 1$ and $\rho(i,j)^2 \leq |\rho(i,j)|$. From Assumption 2.3.2, we have $|\overline{\gamma}_{|\mathcal{Q}_{ij}|}(i,i)| \leq M$ and $|\overline{\gamma}_{|\mathcal{Q}_{ij}|}(j,j)| \leq M$. We then have for all i and j,
$\gamma(i,j)^2 = \overline{\gamma}_{|\mathcal{Q}_{ij}|}(i,i)\overline{\gamma}_{|\mathcal{Q}_{ij}|}(j,j)\rho(i,j)^2 \leq M|\overline{\gamma}_{|\mathcal{Q}_{ij}|}(i,i)\overline{\gamma}_{|\mathcal{Q}_{ij}|}(j,j)|^{1/2}|\rho(i,j)| = M|\gamma(i,j)|$ and

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \gamma(i,j)^2 &= \frac{M}{N}\sum_{i=1}^N\sum_{j=1}^N |\overline{\gamma}_{|\mathcal{Q}_{ij}|}(i,i)\overline{\gamma}_{|\mathcal{Q}_{ij}|}(j,j)|^{1/2}|\rho(i,j)| \\ &\leq \frac{M}{N}\sum_{i=1}^N\sum_{j=1}^N |\gamma(i,j)| \leq M^2,\end{aligned}$$

where the last inequality follows from Assumption 2.3.2.

2.

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\lambda_i^\top \sum_{t\in\mathcal{Q}_{ij}} F_te_{jt}\right)^2\right] \leq \mathbb{E}[\|\lambda_i\|^2]\cdot\mathbb{E}\left[\left(\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}} F_te_{jt}\right)^2\right] \leq \overline{\lambda}^2 M$$

by Assumption 2 and the independence of Λ with F and e.

□

Lemma 4. *Under Assumptions 1-4, let $\delta_{NT}^2 = \min(N,T)$, we have*

$$\delta_{NT}^2\left(\frac{1}{N}\sum_{j=1}^N \left\|\widetilde{\lambda}_j - H_j\lambda_j\right\|^2\right) = O_p(1), \tag{25}$$

where $H_j = \frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^N \widetilde{\lambda}_i\lambda_i^\top F^\top diag(W_i \odot W_j)F/q_{ij}$.

Proof of Lemma 4. From Cauchy-Schwartz inequality, we have $\left\|\widetilde{\lambda}_j - H_j\lambda_j\right\|^2 \leq 4\left\|\widetilde{V}^{-1}\right\|^2 (a_j+b_j+c_j+d_j)$, where

$$\begin{aligned} a_j &= \frac{1}{N^2}\left\|\sum_{i=1}^{N}\widetilde{\lambda}_i\gamma(i,j)\right\|^2 \\ b_j &= \frac{1}{N^2}\left\|\sum_{i=1}^{N}\widetilde{\lambda}_i\zeta_{ij}\right\|^2 \\ c_j &= \frac{1}{N^2}\left\|\sum_{i=1}^{N}\widetilde{\lambda}_i\eta_{ij}\right\|^2 \\ d_j &= \frac{1}{N^2}\left\|\sum_{i=1}^{N}\widetilde{\lambda}_i\xi_{ij}\right\|^2 \end{aligned}$$

From $\frac{1}{N^2}\left\|\sum_{i=1}^{N}\widetilde{\lambda}_i\gamma(i,j)\right\|^2 \leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\gamma(i,j)^2\right)$, we have

$$\frac{1}{N}\sum_{j=1}^{N}a_j \leq \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma(i,j)^2\right) = O_p\left(\frac{1}{N}\right),$$

by Lemma 3.1.

Similar as the proof of Theorem 1 in Bai and Ng (2002),

$$\frac{1}{N}\sum_{j=1}^{N}b_j \leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i\right\|^2\right)\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}\left(\sum_{j=1}^{N}\zeta_{ij}\zeta_{lj}\right)^2\right)^{1/2},$$

$\mathbb{E}\left[\sum_{j=1}^{N}\zeta_{ij}\zeta_{lj}\right]^2 \leq N^2\max_{i,j}\mathbb{E}|\zeta_{ij}|^4$ and

$$\mathbb{E}|\zeta_{ij}|^4 = \frac{1}{|\mathcal{Q}_{ij}|^2}\mathbb{E}\left|\frac{1}{|\mathcal{Q}_{ij}|^{1/2}}\sum_{t\in\mathcal{Q}_{ij}}(e_{it}e_{jt} - \mathbb{E}[e_{it}e_{jt}])\right|^4 \leq \frac{M}{|\mathcal{Q}_{ij}|^2} = O_p\left(\frac{1}{T^2}\right)$$

by Assumption 2.3.5. Thus, $\frac{1}{N}\sum_{j=1}^{N}b_j = O_p\left(\frac{1}{T}\right)$.

Note that

$$\begin{aligned} c_j &= \frac{1}{N^2}\left\|\sum_{i=1}^{N}\widetilde{\lambda}_i\eta_{ij}\right\|^2 = \frac{1}{N^2}\left\|\frac{1}{T}\sum_{i=1}^{N}\left(\widetilde{\lambda}_i\lambda_i^\top\sum_{t\in\mathcal{Q}_{ij}}F_te_{jt}/\widetilde{q}_{ij}\right)\right\|^2 \\ &\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i\right\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{|\mathcal{Q}_{ij}|}\left(\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\lambda_i^\top\sum_{i\in\mathcal{Q}_{ij}}F_te_{jt}\right)^2\right) = O_p\left(\frac{1}{T}\right) \end{aligned}$$

by Lemma 3.2. Then $\frac{1}{N}\sum_{j=1}^{N}c_j = O_p\left(\frac{1}{T}\right)$. Similarly, we can show $\frac{1}{N}\sum_{j=1}^{N}d_j = O_p\left(\frac{1}{T}\right)$. Then

$$\frac{1}{j}\sum_{j=1}^{N}\left\|\widetilde{\lambda}_j - H_j\lambda_j\right\|^2 \leq 4\left\|\widetilde{V}^{-1}\right\|^2\frac{1}{N}\sum_{j=1}^{N}(a_j+b_j+c_j+d_j) = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right).$$

□

Lemma 5. *Under Assumptions 1-4, $H_j - H = O_p(1/\delta_{NT})$*

Proof. Note that

$$H_j - H = \frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^{N}\widetilde{\lambda}_i\lambda_i^\top\left(\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_tF_t^\top - \frac{1}{T}\sum_{t=1}^{T}F_tF_t^\top\right)$$

From Assumption 2.1, $\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_tF_t^\top - \frac{1}{T}\sum_{t=1}^{T}F_tF_t^\top = \Sigma_F + O_p\left(\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\right) - \left(\Sigma_F + O_p\left(\frac{1}{\sqrt{T}}\right)\right) = O_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\frac{1}{\delta_{NT}}\right)$ from Assumption 1, Assumption 2.1 and $\lim_{T\to\infty}|\mathcal{Q}_{ij}|/T > 0$. $\widetilde{V} = O_p(1)$ follows from Lemma 7 and $\widetilde{V}^{-1} = O_p(1)$ follows from Assumptions 1 and 2. $\widetilde{\lambda}_i = O_p(1)$ by construction and $\lambda_i = O_p(1)$ from Assumption 2.2. Thus,

$$H_j - H = O_p(1/\delta_{NT})$$

□

Proof of Theorem 1.

$$\frac{1}{N}\sum_{j=1}^{N}\left\|\widetilde{\lambda}_j - H\lambda_j\right\|^2 \le \frac{1}{N}\sum_{j=1}^{N}\left\|\widetilde{\lambda}_j - H_j\lambda_j\right\|^2 + \frac{1}{N}\sum_{j=1}^{N}\|(H_j - H)\lambda_j\|^2$$

The first term $\frac{1}{N}\sum_{j=1}^{N}\left\|\widetilde{\lambda}_j - H_j\lambda_j\right\|^2 = O_p\left(1/\delta_{NT}^2\right)$ from Lemma 4. The second term $\frac{1}{N}\sum_{j=1}^{N}\|(H_j - H)\lambda_j\|^2 = O_p\left(1/\delta_{NT}^2\right)$ following $H_j - H = O_p(1/\delta_{NT})$ from Lemma 5 and Assumption 2.2. Thus,

$$\frac{1}{N}\sum_{j=1}^{N}\left\|\widetilde{\lambda}_j - H\lambda_j\right\|^2 = O_p\left(\frac{1}{\delta_{NT}^2}\right).$$

□

Lemma 6. *Assume Assumptions 1-4 hold, we have*

1. $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\gamma(i,j) = O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right)$, *where* $\gamma(i,j) = \mathbb{E}\left[\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}e_{it}e_{jt}\right]$
2. $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\zeta_{ij} = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right)$
3. $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\eta_{ij} = O_p\left(\frac{1}{\sqrt{T}}\right)$
4. $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\xi_{ij} = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right)$

Proof. 1.

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\gamma(i,j) = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H\lambda_i)\gamma(i,j) + \frac{1}{N}\sum_{i=1}^{N}H\lambda_i\gamma(i,j)$$

Since

$$\mathbb{E}\left\|\sum_{i=1}^{N}\lambda_i\gamma(i,j)\right\| \le \mathbb{E}\sum_{i=1}^{N}\|\lambda_i\gamma(i,j)\| = \sum_{i=1}^{N}\mathbb{E}[\|\lambda_i\|]\gamma(i,j)] = O(1)$$

by Assumptions 2 and 3.2, we have $\frac{1}{N}\sum_{i=1}^{N} H\lambda_i\gamma(i,j) = O_p(1)$. Furthermore,

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H\lambda_i)\gamma(i,j)\right\| &\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i - H\lambda_i\right\|^2\right)^{1/2}\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}\gamma(i,j)^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\delta_{NT}}\right)\frac{1}{\sqrt{N}}O(1) = O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right)\end{aligned}$$

followed from $\sum_{s=1}^{T}\gamma_N(s,t)^2 \leq M\sum_{s=1}^{T}|\gamma_N(s,t)|$ by the argument in the proof of Lemma 3, Assumptions 1 and 3.2.

2.

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\zeta_{ij} = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H\lambda_i)\zeta_{ij} + \frac{1}{N}\sum_{i=1}^{N}H\lambda_i\zeta_{ij}$$

Note that

$$\frac{1}{N}\sum_{i=1}^{N}\zeta_{ij}^2 \quad \leq \quad \frac{1}{N}\sum_{i=1}^{N}\frac{1}{|\mathcal{Q}_{ij}|}\left[\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}(e_{it}e_{jt} - \mathbb{E}[e_{it}e_{jt}])\right]^2 = O_p\left(\frac{1}{T}\right)$$

from Assumption 2.3.5. Thus,

$$\left\|\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H\lambda_i)\zeta_{ij}\right\| \leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i - H\lambda_i\right\|^2\right)^{1/2}\frac{1}{\sqrt{N}}\left(\frac{1}{N}\sum_{i=1}^{N}\zeta_{ij}^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right).$$

Furthermore,

$$\frac{1}{N}\sum_{i=1}^{N}H\lambda_i\zeta_{ij} = \frac{1}{\sqrt{N}}H\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}\lambda_i\left(e_{it}e_{jt} - \mathbb{E}[e_{it}e_{jt}]\right) = O_p\left(\frac{1}{\sqrt{NT}}\right)$$

following $H = O_p(1)$, Assumption 2.2 ($\|\lambda_i\| = O_p(1)$), and Assumption 3.1.

3.

$$\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\eta_{ij} = \frac{1}{N}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H_i\lambda_i)\eta_{ij} + \frac{1}{N}\sum_{i=1}^{N}H\lambda_i\eta_{ij}$$

Note that

$$\frac{1}{N}\sum_{i=1}^{N}H\lambda_i\eta_{ij} = \frac{1}{N}\sum_{i=1}^{N}H\lambda_i\lambda_i^\top\frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}}F_te_{jt} = O_p\left(\frac{1}{\sqrt{T}}\right)$$

following Assumptions 2.2 and 2.4. Furthermore,

$$\left\|\frac{1}{N}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H_i\lambda_i)\eta_{ij}\right\| \leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i - H_i\lambda_i\right\|^2\right)^{1/2}\frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}\eta_{ij}^2\right)^{1/2} = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right)$$

followed from Lemma 3.2.

4.

$$
\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\xi_{ij} &= \frac{1}{NT}\sum_{i=1}^{N}\widetilde{\lambda}_i\lambda_j^\top F^\top diag(W_i\odot W_j)e_i/q_{ij} = \frac{1}{NT}\sum_{i=1}^{N}\widetilde{\lambda}_i e_i^\top diag(W_i\odot W_j)F\lambda_j/q_{ij}\\
&= \frac{1}{NT}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H\lambda_i)e_i^\top diag(W_i\odot W_j)F\lambda_j/q_{ij} + \frac{1}{NT}\sum_{i=1}^{N}H\lambda_i e_i^\top diag(W_i\odot W_j)F\lambda_j/q_{ij}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\left\|\frac{1}{NT}\sum_{i=1}^{N}(\widetilde{\lambda}_i - H_i\lambda_i)e_i^\top diag(W_i\odot W_j)F\lambda_j/q_{ij}\right\|\\
\leq &\left(\max\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left\|(\widetilde{\lambda}_i - H\lambda_i)\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}F_t e_{it}\right\|\right)^{1/2}\|\lambda_j\|\\
= &\, O_p\left(\frac{1}{\sqrt{T}}\right)O_p\left(\frac{1}{\delta_{NT}}\right)O_p(1) = O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right)
\end{aligned}
$$

followed from Theorem 1, Assumption 2.2, and Assumption 2.4. Furthermore,

$$
\begin{aligned}
&\left\|\frac{1}{NT}\sum_{i=1}^{N}H\lambda_i e_i^\top diag(W_i\odot W_j)F\lambda_j/q_{ij}\right\|\\
\leq &\, \|H\|\left(\max\frac{1}{\sqrt{N|\mathcal{Q}_{ij}|}}\right)\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}\lambda_i F_t^\top e_{it}\right\|\|\lambda_j\| = O_p\left(\frac{1}{\sqrt{NT}}\right),
\end{aligned}
$$

following $H = O_p(1)$, Assumption 2.2 and Assumption 3.2.

□

Lemma 7. *Assume Assumptions 1 and 2 hold. As $T, N\to\infty$,*

1. $\frac{1}{T}\widetilde{\Lambda}^\top\left(\frac{1}{NT}(\widetilde{X}\widetilde{X}^\top)\odot Q^{(-1)}\right)\widetilde{\Lambda} = \widetilde{V}\xrightarrow{P} V$,
2. $\frac{1}{NT^2}\widetilde{\Lambda}^\top\left(\left((\Lambda F^\top)\odot W\right)\left((F\Lambda^\top)\odot W^\top\right)\odot Q^{(-1)}\right)\widetilde{F} = \widetilde{V}\xrightarrow{P} V$
3. $\frac{1}{NT^2}\widetilde{\Lambda}^\top\left(\Lambda F^\top F\Lambda^\top\right)\widetilde{\Lambda} = \widetilde{V}\xrightarrow{P} V$

where $V = diag(v_1, v_2, \cdots, v_r)$ are the eigenvalues of $\Sigma_\Lambda\Sigma_F$.

Proof of Lemma 7. The proof is similar to the proof of (R12) on page 1175 in Stock and Watson (2002a). Let γ denote $N\times 1$ vector and let $\Gamma = \{\gamma|\gamma^\top\gamma/N = 1\}$, $R(\gamma) = \frac{1}{NT^2}\gamma^\top\left((\widetilde{X}\widetilde{X}^\top)\odot Q^{(-1)}\right)\gamma$, $\widetilde{R}(\gamma) = \frac{1}{NT^2}\gamma^\top\left(\left((\Lambda F^\top)\odot W\right)\left((F\Lambda^\top)\odot W^\top\right)\odot Q^{(-1)}\right)\gamma$ and $R^*(\gamma) = \frac{1}{NT^2}\gamma^\top\Lambda F^\top F\Lambda^\top\gamma$. We follow similar steps as Stock and Watson (2002a) and can sequentially show

(R2) $\sup_{\gamma\in\Gamma}\frac{1}{NT^2}\gamma^\top\left(\left((W\odot e)(e^\top\odot W^\top)\right)\odot Q^{(-1)}\right)\gamma\xrightarrow{P} 0$

(R5) $\sup_{\gamma\in\Gamma}\frac{1}{NT^2}|\gamma^\top\left(\left(((W\odot e)(F\Lambda^\top)\odot W^\top)\right)\odot Q^{(-1)}\right)\gamma|\xrightarrow{P} 0$

(R6) $\sup_{\gamma\in\Gamma}|R(\gamma) - \widetilde{R}(\gamma)|\xrightarrow{P} 0$ and $\sup_{\gamma\in\Gamma}|R(\gamma) - R^*(\gamma)|\xrightarrow{P} 0$

Proof. We have the decomposition

$$R(\gamma) - R^*(\gamma) = R(\gamma) - \widetilde{R}(\gamma) + \widetilde{R}(\gamma) - R^*(\gamma)$$

For $R(\gamma) - \widetilde{R}(\gamma)$, we have

$$R(\gamma) - \widetilde{R}(\gamma) = \tfrac{1}{NT^2}\gamma^\top \left(((W \odot e)(e^\top \odot W^\top)) \odot 1/\widetilde{\Pi} \right) \gamma + \tfrac{2}{NT^2}\gamma^\top \left((((W \odot e)(F\Lambda^\top) \odot W^\top)) \odot Q^{(-1)} \right) \gamma$$

and

$$\begin{aligned} \sup_{\gamma\in\Gamma} |R(\gamma) - \widetilde{R}(\gamma)| \quad &\leq \quad \sup_{\gamma\in\Gamma} \frac{1}{NT^2} |\gamma^\top \left(\left((W \odot e)(e^\top \odot W^\top) \right) \odot Q^{(-1)} \right) \gamma| \\ &\quad + \sup_{\gamma\in\Gamma} \frac{2}{NT^2} |\gamma^\top \left(\left((W \odot e)((F\Lambda^\top) \odot W^\top) \right) \odot Q^{(-1)} \right) \gamma| \to 0. \end{aligned}$$

For $\widetilde{R}(\gamma) - R^*(\gamma)$, we have for any $\gamma \in \Gamma$

$$\begin{aligned} \widetilde{R}(\gamma) - R^*(\gamma) \quad &= \quad \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \gamma_i \gamma_j \lambda_i^\top \left(\frac{1}{|\mathcal{Q}_{ij}|} \sum_{t\in\mathcal{Q}_{ij}} F_t F_t^\top - \frac{1}{T} \sum_{t=1}^T F_t F_t^\top \right) \lambda_j \\ &\leq \quad \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \gamma_i^2 \gamma_j^2 \right)^{1/2} \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\lambda_i^\top \Xi_{ij} \lambda_j \right)^2 \right)^{1/2} \\ &= \quad \left(\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\lambda_i^\top \Xi_{ij} \lambda_j \right)^2 \right)^{1/2} \end{aligned}$$

where $\Xi_{ij} = \frac{1}{|\mathcal{Q}_{ij}|} \sum_{t\in\mathcal{Q}_{ij}} F_t F_t^\top - \frac{1}{T} \sum_{t=1}^T F_t F_t^\top$ and $\gamma^\top \gamma / N = 1$ for any $\gamma \in \Gamma$. Since $\Xi_{ij} = o_p(1)$ for all (i, j) and $\|\lambda_i\| = O_p(1)$ for all i, $\lambda_i^\top \Xi_{ij} \lambda_j = o_p(1)$. Then $\left(\lambda_i^\top \Xi_{ij} \lambda_j \right)^2 = o_p(1)$ and $\frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\lambda_i^\top \Xi_{ij} \lambda_j \right)^2 = o_p(1)$. Thus,

$$\sup_{\gamma\in\Gamma} |\widetilde{R}(\gamma) - R^*(\gamma)| \xrightarrow{P} 0$$

and

$$\sup_{\gamma\in\Gamma} |R(\gamma) - R^*(\gamma)| \xrightarrow{P} 0.$$

□

(R7) $|\sup_{\gamma\in\Gamma} R(\gamma) - \sup_{\gamma\in\Gamma} \widetilde{R}(\gamma)| \xrightarrow{P} 0$ and $|\sup_{\gamma\in\Gamma} R(\gamma) - \sup_{\gamma\in\Gamma} R^*(\gamma)| \xrightarrow{P} 0$

(R8) $\sup_{\gamma\in\Gamma} R^*(\gamma) \xrightarrow{P} v_1$, where v_1 is the largest eigenvalue of $\Sigma_F \Sigma_\Lambda$

(R9) $\sup_{\gamma\in\Gamma} R(\gamma) \xrightarrow{P} v_1$

(R10) Let $\widetilde{\Lambda}_1 = \arg\sup_{\gamma\in\Gamma} R(\gamma)$; then $\widetilde{R}(\widetilde{\Lambda}_1) \xrightarrow{P} v_1$ and $R^*(\widetilde{\Lambda}_1) \xrightarrow{P} v_1$

(R11) Let $\underline{\widetilde{\Lambda}}_1$ denote the first column of $\widetilde{\Lambda}$ and let $S_1 = sign(\underline{\widetilde{\Lambda}}_1, \underline{\Lambda}_1)$, meaning $S_1 = 1$ if $\underline{\widetilde{\Lambda}}_1^\top \underline{\Lambda}_1 \geq 0$ and $S_1 = -1$ if $\underline{\widetilde{\Lambda}}_1^\top \underline{\Lambda}_1 < 0$. Then $S_1 \underline{\widetilde{\Lambda}}_1^\top \underline{\Lambda}_1 (\Lambda^\top \Lambda / N)^{-1/2} \xrightarrow{P} l_1^\top$, where $l_1 = (1, 0, \cdots, 0)^\top$.

(R12) Suppose that the $N \times r$ matrix $\widetilde{\Lambda}$ is formed as the r ordered eigenvectors of $(X \odot W)(\widetilde{X}^\top \odot W^\top)$ normalized as $\widetilde{\Lambda}^\top \widetilde{\Lambda}/N = I_r$. Let S denote $S = diag(sign(\widetilde{\Lambda}^\top \Lambda))$. Then $S\widetilde{\Lambda}^\top \Lambda(\Lambda^\top \Lambda/N)^{-1/2} \xrightarrow{P} I_r$.

(R13) For $j = 1, 2, \cdots, r$, $R(\widetilde{\Lambda}_j) \xrightarrow{P} v_j$, $\widetilde{R}(\widetilde{\Lambda}_j) \xrightarrow{P} v_j$ and $R^*(\widetilde{\Lambda}_j) \xrightarrow{P} v_j$.

Proof. The result for $R(\widetilde{\Lambda}_1) \xrightarrow{P} v_1$, $\widetilde{R}(\widetilde{\Lambda}_1) \xrightarrow{P} v_1$ and $R^*(\widetilde{\Lambda}_1) \xrightarrow{P} v_1$ is given in (R9) and (R10). The results for the other columns mimic the steps in (R8)-(R10), for the other principal components, that is, by maximizing $R(\cdot)$ and $R^*(\cdot)$ sequentially using orthonormal subspaces of Γ. □

Note that Lemma 7.1 has

$$\frac{1}{NT^2}\widetilde{\Lambda}^\top \left((\widetilde{X}\widetilde{X}^\top) \odot Q^{(-1)}\right)\widetilde{\Lambda} = diag(R(\widetilde{\Lambda}_1), \cdots, R(\widetilde{\Lambda}_r)) \to diag(v_1, \cdots v_r)$$

from (R13); Lemma 7.2 has

$$\frac{1}{NT^2}\widetilde{\Lambda}^\top \left(\left(\left(W \odot (\Lambda F^\top)\right)\left((F\Lambda^\top) \odot W^\top\right)\right) \odot Q^{(-1)}\right)\widetilde{\Lambda} = diag(\widetilde{R}(\widetilde{\Lambda}_1), \cdots, \widetilde{R}(\widetilde{\Lambda}_r)) \to diag(v_1, \cdots v_r)$$

from (R13); Lemma 7.3 has

$$\frac{1}{NT^2}\widetilde{\Lambda}^\top \left(\Lambda F^\top F \Lambda^\top\right)\widetilde{\Lambda} = diag(R^*(\widetilde{\Lambda}_1), \cdots, R^*(\widetilde{\Lambda}_r)) \xrightarrow{P} diag(v_1, \cdots v_r)$$

from (R13). □

Lemma 8. *Under Assumptions 1 and 2,*

1. $\frac{1}{N}\widetilde{\Lambda}^\top \Lambda \xrightarrow{P} Q$, *where Q is invertible, $Q = V^{1/2}\Upsilon\Sigma_F^{-1/2}$, diagonal entries of $V = diag(v_1, v_2, \cdots, v_r)$ are the eigenvalues of $\Sigma_F^{1/2}\Sigma_\Lambda\Sigma_F^{1/2}$, and Υ is the corresponding eigenvector matrix such that $\Upsilon^\top \Upsilon = I$.*

2. $H^{-1} \xrightarrow{P} Q^\top$, *where $H = \frac{1}{NT}\widetilde{V}^{-1}\widetilde{\Lambda}^\top \Lambda F^\top F$.*

Proof. 1. The proof is similar to the proof of Proposition 1 in Bai (2003). Multiple $\widetilde{\Sigma}\widetilde{\Lambda} = \widetilde{\Lambda}\widetilde{V}$ by $\frac{1}{N}\left(\frac{F^\top F}{T}\right)^{1/2}\Lambda$, then we have

$$\frac{1}{N}\left(\frac{F^\top F}{T}\right)^{1/2}\Lambda^\top \widetilde{\Sigma}\widetilde{\Lambda} = \left(\frac{F^\top F}{T}\right)^{1/2}\frac{\Lambda^\top \widetilde{\Lambda}}{N}\widetilde{V}$$

and then

$$\left(\frac{F^\top F}{T}\right)^{1/2}\frac{\Lambda^\top \Lambda}{N}\left(\frac{F^\top F}{T}\right)\frac{\Lambda^\top \widetilde{\Lambda}}{N} + d_{NT} = \left(\frac{F^\top F}{T}\right)^{1/2}\frac{\Lambda^\top \widetilde{\Lambda}}{N}\widetilde{V},$$

where $d_{NT} = \frac{1}{N}\left(\frac{F^\top F}{T}\right)^{1/2}\Lambda^\top \tilde{d}_{NT}\widetilde{\Lambda}$ and $\tilde{d}_{NT}$ has

$$\begin{aligned}\tilde{d}_{NT,ij} &= \lambda_i^\top \left(\frac{1}{|\mathcal{Q}_{ij}|}F^\top diag(W_i \odot W_j)F - \frac{1}{T}F^\top F\right)\lambda_j + \frac{1}{|\mathcal{Q}_{ij}|}e_i^\top diag(W_i \odot W_j)F\lambda_j \\ &+ \frac{1}{|\mathcal{Q}_{ij}|}\lambda_i^\top F^\top diag(W_i \odot W_j)e_j + \frac{1}{|\mathcal{Q}_{ij}|}e_i^\top diag(W_i \odot W_j)e_j\end{aligned}$$

From Assumption 2.1, $\frac{1}{|\mathcal{Q}_{ij}|}F^\top diag(W_i \odot W_j)F - \frac{1}{T}F^\top F = O_p\left(\frac{1}{\delta_{NT}}\right)$ and then

$$\frac{1}{N}\Lambda^\top \tilde{d}_{NT} = O_p\left(\frac{1}{\delta_{NT}}\right)$$

following Lemma 6. The remaining steps to show $\frac{1}{N}\widetilde{\Lambda}^\top \Lambda \xrightarrow{P} Q$ are exactly the same as that in Proposition 1 in Bai (2003).

2. Note that

$$H = \frac{1}{NT}\widetilde{V}^{-1}\widetilde{\Lambda}^\top \Lambda F^\top F \xrightarrow{P} V^{-1}Q\Sigma_F = V^{-1}V^{1/2}\Upsilon\Sigma_F^{-1/2}\Sigma_F = V^{-1/2}\Upsilon\Sigma_F^{1/2} = (Q^\top)^{-1}$$

□

Proof of Theorem 2. By Lemma 6, we have

$$\widetilde{\lambda}_j - H_j\lambda_j = O_p\left(\frac{1}{\sqrt{N}\delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{T}\delta_{NT}}\right).$$

When $\sqrt{T}/N \to 0$, the limiting distribution is determined by the third term. Thus,

$$\begin{aligned}\sqrt{T}(\widetilde{\lambda}_j - H_j\lambda_j) &= \widetilde{V}^{-1}\frac{1}{N}\sum_{i=1}^N \sqrt{\frac{T}{|\mathcal{Q}_{ij}|}}H_i\lambda_i\lambda_i^\top \frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} + o_p(1)\\ &= \widetilde{V}^{-1}\frac{1}{N}H\sum_{i=1}^N \sqrt{\frac{T}{|\mathcal{Q}_{ij}|}}\lambda_i\lambda_i^\top \frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} + o_p(1)\end{aligned}$$

following $H_i - H = O_p\left(\frac{1}{\delta_{NT}}\right)$. From Assumption 3.3,

$$\frac{1}{N}\sum_{i=1}^N \sqrt{\frac{T}{|\mathcal{Q}_{ij}|}}\lambda_i\lambda_i^\top \frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} \xrightarrow{d} N(0, \Phi_j).$$

From Lemma 8, $H \to (Q^{-1})^\top$ and from Lemma 7, $\widetilde{V}^{-1} \xrightarrow{P} V$. From Slutsky's theorem,

$$\widetilde{V}^{-1}\frac{1}{N}H\sum_{i=1}^N \sqrt{\frac{T}{|\mathcal{Q}_{ij}|}}\lambda_i\lambda_i^\top \frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} \xrightarrow{d} N(0, V^{-1}(Q^{-1})^\top\Phi_j Q^{-1}V^{-1}).$$

A consistent estimate for the asymptotic variance $V^{-1}(Q^{-1})^\top\Phi_j Q^{-1}V^{-1}$ is shown in Lemma 9.

Furthermore,

$$\sqrt{T}(\widetilde{\lambda}_j - H\lambda_j) = \sqrt{T}(\widetilde{\lambda}_j - H_j\lambda_j) + \sqrt{T}(H_j - H)\lambda_j + o_p(1).$$

From Lemma 5, $H_j - H = O_p\left(\frac{1}{\sqrt{T}}\right)$. Then $\sqrt{T}(H_j - H)\lambda_j = O_p(1)$ from Assumption 2.2 and $\sqrt{T}(H_j - H)\lambda_j$ contributes to the asymptotic distribution of $\widetilde{\lambda}_j$.

From the definition of $H = \frac{1}{NT}\widetilde{V}^{-1}\widetilde{\Lambda}^\top\Lambda F^\top F$ and $H_j = \frac{1}{N}\widetilde{V}^{-1}\sum_{i=1}^N \widetilde{\lambda}_i\lambda_i^\top \frac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_tF_t^\top$, we can write $H_j - H$ as $H_j - H = \frac{1}{NT}\widetilde{V}^{-1}\sum_{i=1}^N\sum_{t=1}^T y_{it,j}\widetilde{\lambda}_i\lambda_i^\top F_tF_t^\top$, where $y_{it,j} = \frac{T-|\mathcal{Q}_{ij}|}{|\mathcal{Q}_{ij}|}$ for $t \in \mathcal{Q}_{ij}$, $y_{it,j} = -1$ for $t \notin \mathcal{Q}_{ij}$ and $\sum_{i=1}^N\sum_{t=1}^T y_{it,j} = 0$.

Note that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{T=1}^{T} y_{it,j}\widetilde{\lambda}_i\lambda_i^\top F_tF_t^\top = \underbrace{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it,j}H\lambda_i\lambda_i^\top F_tF_t^\top}_{\mathrm{I}} + \underbrace{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it,j}(\widetilde{\lambda}_i - H\lambda_i)\lambda_i^\top F_tF_t^\top}_{\mathrm{II}}$$

For the second term II, we have

$$\begin{aligned}
\mathrm{II} &\leq \left(\frac{1}{N}\sum_{i=1}^{N}\left\|\widetilde{\lambda}_i - H\lambda_i\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\left(\frac{1}{T}\sum_{t=1}^{T} y_{it,j}F_tF_t^\top\right)\lambda_j\right\|^2\right)^{1/2} \\
&= O_p(1)O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\sqrt{T}}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right)
\end{aligned}$$

following Theorem 1 and $\frac{1}{T}\sum_{t=1}^{T} y_{it,j}F_tF_t^\top = O_p\left(\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\right) = O_p\left(\frac{1}{\sqrt{T}}\right)$. Thus, the second term is smaller than the first term.

From Assumption 3.5,

$$\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it,j}\lambda_i\lambda_i^\top F_tF_t^\top \xrightarrow{d} N(0, \Xi_{F,j}).$$

From Slutsky's theorem,

$$\frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it,j}\lambda_i\lambda_i^\top F_tF_t^\top\lambda_j \xrightarrow{d} N\left(0, (\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j \otimes I)\right).$$

and

$$\frac{1}{N\sqrt{T}}\widetilde{V}^{-1}H\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it,j}\lambda_i\lambda_i^\top F_tF_t^\top\lambda_j \xrightarrow{d} N\left(0, V^{-1}(Q^{-1})^\top(\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j \otimes I)Q^{-1}V^{-1}\right).$$

Lemma 10 shows a consistent estimator for the asymptotic variance $V^{-1}(\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j \otimes I)V^{-1}$.

Furthermore, $\sqrt{T}(\widetilde{\lambda}_j - H_j\lambda_j)$ and $\sqrt{T}(H_j - H)\lambda_j$ are asymptotic independent because the randomness of $\sqrt{T}(\widetilde{\lambda}_j - H_j\lambda_j)$ comes from F_te_{jt} while the randomness of $\sqrt{T}(H_j - H)\lambda_j$ comes from $y_{it,j}\widetilde{\lambda}_i\lambda_i^\top F_tF_t^\top$. Then we have

$$\sqrt{T}(\widetilde{\lambda}_j - H\lambda_j) \xrightarrow{d} N(0, \underbrace{V^{-1}(Q^{-1})^\top\Phi_jQ^{-1}V^{-1}}_{\Gamma_{\lambda_j,1}} + \underbrace{V^{-1}(Q^{-1})^\top(\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j \otimes I)Q^{-1}V^{-1}}_{\Gamma_{\lambda_j,2}})$$

The plug-in estimators of $\Gamma_{\lambda_j,1}$ and $\Gamma_{\lambda_j,2}$, denoted as $\widetilde{\Gamma}_{\lambda_j,1}$ and $\widetilde{\Gamma}_{\lambda_j,2}$, are provided in Lemmas 9 and 10 respectively. Let $\widetilde{\Gamma}_{\lambda_j} = \widetilde{\Gamma}_{\lambda_j,1} + \widetilde{\Gamma}_{\lambda_j,2}$ From Slusky's Theorem,

$$\sqrt{T}\widetilde{\Gamma}_{\lambda_j}^{-1/2}(\widetilde{\lambda}_j - H\lambda_j) \xrightarrow{d} N(0, I_r)$$

□

Lemma 9. *Assume there are finitely many nonzeros in each row of* $\Sigma_{e_j} = \mathbb{E}[e_j e_j^T]$ *and we know the set* Ω_{e_j} *of nonzero indices in* Σ_{e_j}*. Under the assumptions of Theorem 2 we have*

$$\widetilde{\Gamma}_{\lambda_j,1} = AVar(\sqrt{T}(\widetilde{\lambda}_j - H_j\lambda_j)) + o_p(1),$$

where

$$\widetilde{\Gamma}_{\lambda_j,1} = \frac{T}{N^2}\widetilde{V}^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\left(\frac{1}{|\mathcal{Q}_{ij}||\mathcal{Q}_{lj}|}\sum_{s\in\mathcal{Q}_{ij},t\in\mathcal{Q}_{lj},(s,t)\in\Omega_{e_j}}\widetilde{F}_s\widetilde{F}_t^\top\widetilde{e}_{js}\widetilde{e}_{jt}\right)\widetilde{\lambda}_l\widetilde{\lambda}_l^\top\widetilde{V}^{-1}$$

and $\widetilde{e}_{it} = X_{it} - \widetilde{\lambda}_i^\top\widetilde{F}_t$ *for the observed* X_{it}.

Proof. $\widetilde{e}_{it} = X_{it} - \widetilde{\lambda}_i^\top\widetilde{F}_t$ is a consistent estimator for e_{it} for $(i,t) \in \{(i,t) : W_{it} = 1\}$ because $\widetilde{F}_t$ and $\widetilde{\lambda}_i$ are consistent estimators for $(H^\top)^{-1}F_t$ and $H\lambda_i$ following Theorems 2 and 3. Recall

$$\sqrt{T}(\widetilde{\lambda}_j - H_j\lambda_j) = \widetilde{V}^{-1}\frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{T}{|\mathcal{Q}_{ij}|}}H_i\lambda_i\lambda_i^\top\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{Q}_{ij}}F_te_{jt} + o_p(1).$$

Note that X_{it} is observed for $t \in \mathcal{Q}_{ij}$ so $\widetilde{e}_{it}$ is a consistent estimator for e_{it} for $t \in \mathcal{Q}_{ij}$. Then for each i and l, a consistent estimator for the asymptotic covariance between $\frac{1}{\sqrt{|\mathcal{Q}_{ij}|}}\sum_{t\in\mathcal{O}_{ij}}(H^\top)^{-1}F_te_{jt}$ and $\frac{1}{\sqrt{|\mathcal{Q}_{lj}|}}\sum_{t\in\mathcal{O}_{lj}}(H^\top)^{-1}F_te_{jt}$ is

$$\frac{1}{|\mathcal{Q}_{ij}||\mathcal{Q}_{lj}|}\sum_{s\in\mathcal{Q}_{ij},t\in\mathcal{Q}_{lj},(s,t)\in\Omega_{e_j}}\widetilde{F}_s\widetilde{F}_t^\top\widetilde{e}_{js}\widetilde{e}_{jt}.$$

Together with $\widetilde{\lambda}_i$ to be the consistent estimator for $H_i\lambda_i$ and $H\lambda_i$, a consistent estimator for the asymptotic variance of $\sqrt{N}(\widetilde{\lambda}_i - H_i\lambda_i)$ is

$$\widetilde{\Gamma}_{\lambda_j,1} = \frac{T}{N^2}\widetilde{V}^{-1}\sum_{i=1}^{N}\sum_{l=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\left(\frac{1}{|\mathcal{Q}_{ij}||\mathcal{Q}_{lj}|}\sum_{s\in\mathcal{Q}_{ij},t\in\mathcal{Q}_{lj},(s,t)\in\Omega_{e_j}}\widetilde{F}_s\widetilde{F}_t^\top\widetilde{e}_{js}\widetilde{e}_{jt}\right)\widetilde{\lambda}_l\widetilde{\lambda}_l^\top\widetilde{V}^{-1}$$

□

Lemma 10. *Under the Assumptions in Theorem 2 and* $F_tF_t^\top$ *and* $\lambda_i\lambda_i^\top$ *are ergodic in mean, we have*

$$\widetilde{\Gamma}_{\lambda_j,2} = AVar\left(\frac{1}{\sqrt{T}N}\widetilde{V}^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it,j}H\lambda_i\lambda_i^\top F_tF_t^\top\lambda_j\right) + o_p(1),$$

where $\widetilde{\Gamma}_{\lambda_j,2} = \frac{1}{NT^2}\widetilde{V}^{-1}\left[\widetilde{A}_1 + \widetilde{A}_2 + \widetilde{A}_3 + \widetilde{A}_4\right]\widetilde{V}^{-1}$ *with*
$\widetilde{A}_1 = \left(\sum_{i=1}^N\sum_{t=1}^T y_{it,j}^2\right)\left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\widetilde{F}_t\widetilde{F}_t^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_t\widetilde{F}_t^\top\lambda_i\widetilde{\lambda}_i^\top\right)$,

$$\widetilde{A}_2 = \begin{cases}\left(\sum_{t=1}^T\sum_{i=1}^N\sum_{l\neq i}y_{it,j}y_{lt,j}\right)\left(\frac{1}{T}\sum_{t=1}^T\widetilde{F}_t\widetilde{F}_t^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_t\widetilde{F}_t^\top\right), & if\,\lambda_i\text{ is independent}\\ \sum_{t=1}^T\sum_{i=1}^N\sum_{l\neq i}y_{it,j}y_{lt,j}\left(\frac{1}{T(N-|\rho|)}\sum_{s=1}^T\sum_{m=\max(1,1-\rho),\text{ where }\rho=l-i}^{\min(N,N-\rho)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_s\widetilde{F}_s^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_s\widetilde{F}_s^\top\widetilde{\lambda}_{m+\rho}\widetilde{\lambda}_{m+\rho}^\top\right), & otherwise\end{cases}$$

$$\widetilde{A}_3 = \begin{cases}\left(\sum_{t=1}^T\sum_{s\neq t}\sum_{i=1}^N y_{it,j}y_{is,j}\right)\left(\frac{1}{N}\sum_{m=1}^N\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\left(\frac{1}{T}\sum_{u=1}^T\widetilde{F}_u\widetilde{F}_u^\top\right)\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\left(\frac{1}{T}\sum_{u=1}^T\widetilde{F}_u\widetilde{F}_u^\top\right)\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\right), & if\,F_t\text{ is independent}\\ \sum_{t=1}^T\sum_{s\neq t}\sum_{i=1}^N y_{it,j}y_{is,j}\left(\frac{1}{N(T-|\tau|)}\sum_{m=1}^N\sum_{u=\max(1,1-\tau),\text{ where }\tau=s-t}^{\min(T,T-\tau)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_u\widetilde{F}_u^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_{u+\tau}\widetilde{F}_{u+\tau}^\top\lambda_m\widetilde{\lambda}_m^\top\right), & otherwise\end{cases}$$

$$\widetilde{A}_4 = \begin{cases} \left(\sum_{t=1}^T \sum_{s\neq t} \sum_{i=1}^N \sum_{l\neq i} y_{it,j} y_{ls,j}\right) \left(\frac{1}{T}\sum_{t=1}^T \widetilde{F}_t \widetilde{F}_t^\top\right) \widetilde{\lambda}_j \widetilde{\lambda}_j^\top \left(\frac{1}{T}\sum_{t=1}^T \widetilde{F}_t \widetilde{F}_t^\top\right) & if F_t, \lambda_i \text{ are independent} \\ \sum_{t=1}^T \sum_{s\neq t} \sum_{i=1}^N \sum_{l\neq i} y_{it,j} y_{ls,j} \left(\frac{1}{T-|\tau|} \sum_{u=\max(1,1-\tau),\ where\ \tau=s-t}^{\min(T,T-\tau)} \widetilde{F}_u \widetilde{F}_u^\top \widetilde{\lambda}_j \widetilde{\lambda}_j^\top \widetilde{F}_{u+\tau} \widetilde{F}_{u+\tau}^\top\right) & if \lambda_i \text{ is independent} \\ \sum_{t=1}^T \sum_{s\neq t} \sum_{i=1}^N \sum_{l\neq i} y_{it,j} y_{ls,j} \left(\frac{1}{N-|\rho|} \sum_{m=\max(1,1-\rho),\ where\ \rho=j-i}^{\min(N,N-\rho)} \widetilde{\lambda}_m \widetilde{\lambda}_m^\top \left(\frac{1}{T}\sum_{u=1}^T \widetilde{F}_u \widetilde{F}_u^\top\right) \widetilde{\lambda}_j \widetilde{\lambda}_j^\top \left(\frac{1}{T}\sum_{u=1}^T \widetilde{F}_u \widetilde{F}_u^\top\right) \widetilde{\lambda}_{m+\rho} \lambda_{m+\rho}^\top\right) & if F_t \text{ is independent} \\ \sum_{t=1}^T \sum_{s\neq t} \sum_{i=1}^N \sum_{l\neq i} y_{it,j} y_{ls,j} \left(\frac{1}{(N-|\rho|)(T-|\tau|)} \sum_{m=\max(1,1-\rho),\ where\ \rho=j-i}^{\min(N,N-\rho)} \sum_{u=\max(1,1-\tau),\ where\ \tau=s-t}^{\min(T,T-\tau)} \widetilde{\lambda}_m \widetilde{\lambda}_m^\top \widetilde{F}_u \widetilde{F}_u^\top \widetilde{\lambda}_j \widetilde{\lambda}_j^\top \widetilde{F}_{u+\tau} \widetilde{F}_{u+\tau}^\top \lambda_{m+\rho} \lambda_{m+\rho}^\top\right) & otherwise \end{cases}$$

Proof. We have

$$\begin{aligned} & Cov\left(\frac{1}{\sqrt{T}N}\widetilde{V}^{-1}\sum_{i=1}^N\sum_{t=1}^T y_{it,j} H\lambda_i\lambda_i^\top F_t F_t^\top \lambda_j, \frac{1}{\sqrt{T}N}\widetilde{V}^{-1}\sum_{i=1}^N\sum_{t=1}^T y_{it,j} H\lambda_i\lambda_i^\top F_t F_t^\top \lambda_j\right) \\ = \quad & \frac{1}{TN^2}\widetilde{V}^{-1}\left[A_1 + A_2 + A_3 + A_4 - A_5\right]\widetilde{V}^{-1}, \end{aligned}$$

where

$$\begin{aligned} A_1 &= \sum_{i=1}^N\sum_{t=1}^T y_{it,j}^2 \mathbb{E}\left[H\lambda_i\lambda_i^\top \mathbb{E}\left[F_t F_t^\top \lambda_j\lambda_j^\top F_t F_t^\top\right]\lambda_i\lambda_i^\top H^\top\right] \\ A_2 &= \sum_{i=1}^N\sum_{l\neq i}\sum_{t=1}^T y_{it,j} y_{lt,j} \mathbb{E}\left[H\lambda_i\lambda_i^\top \mathbb{E}\left[F_t F_t^\top \lambda_j\lambda_j^\top F_t F_t^\top\right]\lambda_l\lambda_l^\top H^\top\right] \\ A_3 &= \sum_{i=1}^N\sum_{t=1}^T\sum_{s\neq t} y_{it,j} y_{is,j} \mathbb{E}\left[H\lambda_i\lambda_i^\top \mathbb{E}\left[F_t F_t^\top \lambda_j\lambda_j^\top F_s F_s^\top\right]\lambda_i\lambda_i^\top H^\top\right] \\ A_4 &= \sum_{i=1}^N\sum_{l\neq i}\sum_{t=1}^T\sum_{s\neq t} y_{it,j} y_{ls,j} \mathbb{E}\left[H\lambda_i\lambda_i^\top \mathbb{E}\left[F_t F_t^\top \lambda_j\lambda_j^\top F_s F_s^\top\right]\lambda_l\lambda_l^\top H^\top\right] \\ A_5 &= \left(\sum_{i=1}^N\sum_{t=1}^T y_{it,j}\mathbb{E}\left[H\lambda_i\lambda_i^\top\right]\mathbb{E}\left[F_t F_t^\top\right]\lambda_j\right)\left(\sum_{i=1}^N\sum_{t=1}^T y_{it,j}\mathbb{E}\left[H\lambda_i\lambda_i^\top\right]\mathbb{E}\left[F_t F_t^\top\right]\lambda_j\right)^\top \end{aligned}$$

First note that $\widetilde{F}_t$ is a consistent estimator for $(H^\top)^{-1}F_t$ and $\widetilde{\lambda}_i$ is a consistent estimator for $H\lambda_i$. If we plug in $\widetilde{F}_s$ for F_s and $\widetilde{\lambda}_i$ for λ_i, all the rotation matrices canceled out in A_1 to A_5.

Since $F_s F_s^\top$ is ergodic in mean and $\lambda_i\lambda_i^\top$ is ergodic in mean, we have $\frac{1}{T}\sum_{s=1}^T HF_sF_s^\top - \mathbb{E}\left[HF_sF_s^\top\right] = o_p(1)$, $\frac{1}{N}\sum_{i=1}^N \lambda_i\lambda_i^\top - \mathbb{E}\left[\lambda_i\lambda_i^\top\right] = o_p(1)$, $\frac{1}{T}\sum_{t=1}^T F_tF_t^\top\lambda_j\lambda_j^\top F_tF_t^\top - \mathbb{E}[F_tF_t^\top\lambda_j\lambda_j^\top F_tF_t^\top] = o_p(1)$, $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T H\lambda_i\lambda_i^\top F_tF_t^\top\lambda_j\lambda_j^\top F_tF_t^\top\lambda_i\lambda_i^\top H^\top - \mathbb{E}\left[\lambda_i\lambda_i^\top F_tF_t^\top\lambda_j\lambda_j^\top F_tF_t^\top\lambda_i\lambda_i^\top H^\top\right] = o_p(1)$.

When F_t is independent, $\mathbb{E}[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top|\lambda_j] = \mathbb{E}[F_tF_t^\top]\lambda_j\lambda_j^\top\mathbb{E}[F_sF_s^\top]$. When λ_i is independent, $\mathbb{E}[H\lambda_i\lambda_i^\top\mathbb{E}[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top]\lambda_l\lambda_l^\top H^\top] = \mathbb{E}[H\lambda_i\lambda_i^\top]\mathbb{E}[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top]\mathbb{E}[\lambda_l\lambda_l^\top H^\top]$. When F_t is independent and λ_i is independent,
$\mathbb{E}[H\lambda_i\lambda_i^\top\mathbb{E}[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top|\lambda_j]\lambda_l\lambda_l^\top H^\top] = \mathbb{E}[H\lambda_i\lambda_i^\top]\mathbb{E}[F_tF_t^\top]\lambda_j\lambda_j^\top\mathbb{E}[F_tF_t^\top]\mathbb{E}[\lambda_l\lambda_l^\top H^\top]$.

When F_t is independent,
$\frac{1}{N}\sum_{i=1}^N H\lambda_i\lambda_i^\top\left(\frac{1}{T}\sum_{t=1}^T F_tF_t^\top\right)\lambda_j\lambda_j^\top\left(\frac{1}{T}\sum_{t=1}^T F_tF_t^\top\right)\lambda_i\lambda_i^\top H^\top - \mathbb{E}[H\lambda_i\lambda_i^\top\mathbb{E}[F_tF_t^\top]\lambda_j\lambda_j^\top\mathbb{E}[F_tF_t^\top]\lambda_i\lambda_i^\top H^\top] = o_p(1)$.

When λ_i is dependent,
$\frac{1}{T(N-|\rho|)}\sum_{s=1}^T\sum_{m=\max(1,1-\rho),\text{ where }\rho=l-i}^{\min(N,N-\rho)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_s\widetilde{F}_s^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_s\widetilde{F}_s^\top\widetilde{\lambda}_{m+\rho}\widetilde{\lambda}_{m+\rho}^\top$
$-\mathbb{E}\left[H\lambda_i\lambda_i^\top\mathbb{E}\left[F_tF_t^\top\lambda_j\lambda_j^\top F_tF_t^\top\right]\lambda_l\lambda_l^\top H^\top\right] = o_p(1)$.

When F_t is dependent, $\frac{1}{N(T-|\tau|)}\sum_{m=1}^{N}\sum_{u=\max(1,1-\tau),\text{ where }\tau=s-t}^{\min(T,T-\tau)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_u\widetilde{F}_u^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_{u+\tau}\widetilde{F}_{u+\tau}^\top\lambda_m\widetilde{\lambda}_m^\top -$ $\mathbb{E}\left[H\lambda_i\lambda_i^\top\mathbb{E}\left[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top\right]\lambda_i\lambda_i^\top H^\top\right] = o_p(1)$.

When λ_i is independent and F_t is dependent, $\frac{1}{T-|\tau|}\sum_{u=\max(1,1-\tau),\text{ where }\tau=s-t}^{\min(T,T-\tau)}\widetilde{F}_u\widetilde{F}_u^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_{u+\tau}\widetilde{F}_{u+\tau}^\top$ $-\mathbb{E}\left[H\lambda_i\lambda_i^\top\mathbb{E}\left[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top\right]\lambda_l\lambda_l^\top H^\top\right] = o_p(1)$.

When λ_i is dependent and F_t is independent, $\frac{1}{N-|\rho|}\sum_{m=\max(1,1-\rho),\text{ where }\rho=j-i}^{\min(N,N-\rho)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\left(\frac{1}{T}\sum_{u=1}^{T}\widetilde{F}_u\widetilde{F}_u^\top\right)\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\left(\frac{1}{T}\sum_{u=1}^{T}\widetilde{F}_u\widetilde{F}_u^\top\right)\widetilde{\lambda}_{m+\rho}\lambda_{m+\rho}^\top$ $-\mathbb{E}\left[H\lambda_i\lambda_i^\top\mathbb{E}\left[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top\right]\lambda_l\lambda_l^\top H^\top\right] = o_p(1)$.

When λ_i is dependent and F_t is dependent, $\frac{1}{(N-|\rho|)(T-|\tau|)}\sum_{m=\max(1,1-\rho),\text{ where }\rho=j-i}^{\min(N,N-\rho)}\sum_{u=\max(1,1-\tau),\text{ where }\tau=s-t}^{\min(T,T-\tau)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_u\widetilde{F}_u^\top\widetilde{\lambda}_j\widetilde{\lambda}_j^\top\widetilde{F}_{u+\tau}\widetilde{F}_{u+\tau}^\top\lambda_{m+\rho}\lambda_{m+\rho}^\top -$ $\mathbb{E}\left[H\lambda_i\lambda_i^\top\mathbb{E}\left[F_tF_t^\top\lambda_j\lambda_j^\top F_sF_s^\top\right]\lambda_l\lambda_l^\top H^\top\right] = o_p(1)$

This completes the proof for Lemma 10 for all scenarios.

$\square$

Lemma 11. *Under Assumptions 1-3, we have*

1. $\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H_i\lambda_i\right)e_{it} = O_p(1/\delta_{NT}^2)$
2. $\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H\lambda_i\right)e_{it} = O_p(1/\delta_{NT}^2)$
3. $\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\lambda}_i - H\lambda_i\right)\lambda_i^\top = o_p(1/\delta_{NT})$
4. $\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\lambda}_i - H\lambda_i\right)\widetilde{\lambda}_i^\top = o_p(1/\delta_{NT})$

Proof. 1. The proof is very similar as Lemma B.1 in Bai (2003).

$$
\begin{aligned}
&\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H_i\lambda_i\right)e_{it}\\
= \quad & \widetilde{V}^{-1}\left[\frac{1}{N^2}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\sum_{l=1}^{N}\widetilde{\lambda}_l\gamma(l,i)e_{it} + \frac{1}{N^2}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\sum_{l=1}^{N}\widetilde{\lambda}_l\zeta_{li}e_{it}\right.\\
= \quad & \left.\frac{1}{N^2}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\sum_{l=1}^{N}\widetilde{\lambda}_l\eta_{li}e_{it} + \frac{1}{N^2}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\sum_{l=1}^{N}\widetilde{\lambda}_l\xi_{li}e_{it}\right]\\
= \quad & \widetilde{V}^{-1}\left[\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}\right]
\end{aligned}
$$

Since $\underline{p} \le P(W_{it}=1|S)$ by Assumption 1.3 (then $\frac{1}{P(W_{it}=1|S)} \le \frac{1}{\underline{p}}$), $\mathrm{I} = O_p(1/\delta_{NT}^2)$, $\mathrm{II} = O_p(1/\delta_{NT}^2)$, $\mathrm{III} = O_p(1/\delta_{NT}^2)$ and $\mathrm{IV} = O_p(1/\delta_{NT}^2)$ can be shown similar as Bai (2003) given Lemma 3 and under Assumptions 1-3.

2.

$$
\begin{aligned}
\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H\lambda_i\right)e_{it} \quad = \quad & \frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H_i\lambda_i\right)e_{it}\\
& + \frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}(H_i - H)\lambda_i e_{it}
\end{aligned}
$$

The first term is $o_p(1/\delta_{NT})$ by Lemma 11.1. By Assumption 8, the second term is $O_p(1/\delta_{NT}^2)$. Thus, $\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H\lambda_i\right)e_{it} = O_p(1/\delta_{NT}^2)$.

3.
$$\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\lambda}_i - H\lambda_i\right)\lambda_i^\top = \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\lambda}_i - H_i\lambda_i\right)\lambda_i^\top + \frac{1}{N}\sum_{i=1}^{N}\left(H_i - H\right)\lambda_i\lambda_i^\top$$

$\frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{\lambda}_i - H_i\lambda_i\right)\lambda_i^\top = O_p(1/\delta_{NT}^2)$ can be shown similar as Bai (2003) Lemma B.2 and under Assumptions 1-3.

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\left(H_i - H\right)\lambda_i\lambda_i^\top &= \widetilde{V}^{-1}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}\widetilde{\lambda}_l\lambda_l^\top\left(\frac{1}{|\mathcal{Q}_{li}|}\sum_{s\in\mathcal{Q}_{li}}F_sF_s^\top - \frac{1}{T}\sum_{s=1}^{T}F_sF_s^\top\right)\lambda_i\lambda_i \\
&= \widetilde{V}^{-1}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}(\widetilde{\lambda}_l - H\lambda_l)\lambda_l^\top\left(\frac{1}{|\mathcal{Q}_{li}|}\sum_{s\in\mathcal{Q}_{li}}F_sF_s^\top - \frac{1}{T}\sum_{s=1}^{T}F_sF_s^\top\right)\lambda_i\lambda_i \\
&\quad + \widetilde{V}^{-1}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{l=1}^{N}H\lambda_l\lambda_l^\top\left(\frac{1}{|\mathcal{Q}_{li}|}\sum_{s\in\mathcal{Q}_{li}}F_sF_s^\top - \frac{1}{T}\sum_{s=1}^{T}F_sF_s^\top\right)\lambda_i\lambda_i,
\end{aligned}$$

where the second term is $o_p(1/\sqrt{T})$ by Assumption 7. The first term is $O_p(1/\delta_{NT}^2)$ by 1 and Assumption 3.7.

4. From Lemma 11.3 and Theorem 1

□

Proof of Theorem 3. Decomposing

$$\widetilde{F}_t = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{P(W_{it}=1|S)}X_{it}W_{it}\widetilde{\lambda}_i = \frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}X_{it}\widetilde{\lambda}_i,$$

we have

$$\begin{aligned}
\widetilde{F}_t &= \tfrac{1}{N}\textstyle\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}(\lambda_i^\top F_t + e_{it})\widetilde{\lambda}_i \\
&= \left(\tfrac{1}{N}\textstyle\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\widetilde{\lambda}_i\lambda_i^\top\right)F_t + \tfrac{1}{N}\textstyle\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\widetilde{\lambda}_i e_{it} \\
&= \left(\tfrac{1}{N}\textstyle\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\widetilde{\lambda}_i\lambda_i^\top\right)F_t + \tfrac{1}{N}\textstyle\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}H\lambda_i e_{it} + \tfrac{1}{N}\textstyle\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\left(\widetilde{\lambda}_i - H\lambda_i\right)e_{it}
\end{aligned}$$

where the last term is $o_p(1/\delta_{NT})$ by Lemma 11.1. From Assumption 3.4, $\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\lambda_i e_{it} \xrightarrow{d} N(0,\Gamma_t)$. From Slutsky's theorem and Lemma 8,

$$\frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}H\lambda_i e_{it} \xrightarrow{d} N(0,(Q^{-1})^\top\Gamma_t Q^{-1}).$$

A consistent estimate for the asymptotic variance $(Q^{-1})^\top\Gamma_t Q^{-1}$ is shown in Lemma 12.

Denote $G_t = \frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\widetilde{\lambda}_i\lambda_i^\top$. We have

$$\widetilde{F}_t - (H^{-1})^\top F_t = \left(\widetilde{F}_t - G_tF_t\right) + \left(G_t - \frac{1}{N}\widetilde{\Lambda}^\top\Lambda\right)F_t + \left(\frac{1}{N}\widetilde{\Lambda}^\top\Lambda - (H^{-1})^\top\right)F_t$$

Note that $\widetilde{\Lambda}^\top\widetilde{\Lambda}/N = I_r$, we have

$$\frac{1}{N}\widetilde{\Lambda}^\top\Lambda - (H^{-1})^\top = \frac{1}{N}\widetilde{\Lambda}^\top\left(\Lambda H^\top - \widetilde{\Lambda}\right)(H^{-1})^\top = o_p\left(\frac{1}{\delta_{NT}}\right)$$

from Lemma 11.3.

Note that

$$G_t - \frac{1}{N}\widetilde{\Lambda}^\top\Lambda = \frac{1}{N}\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\widetilde{\lambda}_i\lambda_i^\top - \frac{1}{N}\sum_{i=1}^N\widetilde{\lambda}_i\lambda_i^\top = \frac{1}{N}\sum_{i=1}^N v_{i,t}\widetilde{\lambda}_i\lambda_i^\top,$$

where $v_{i,t} = \frac{1}{P(W_{it}=1|S)} - 1$ for $i\in\mathcal{O}_t$ and $v_{i,t} = -1$ for $i\notin\mathcal{O}_t$. Note that

$$\frac{1}{N}\sum_{i=1}^N v_{i,t}\widetilde{\lambda}_i\lambda_i^\top = \underbrace{\frac{1}{N}\sum_{i=1}^N v_{i,t}H\lambda_i\lambda_i^\top}_{\mathrm{I}} + \underbrace{\frac{1}{N}\sum_{i=1}^N v_{i,t}\left(\widetilde{\lambda}_i - H\lambda_i\right)\lambda_i^\top}_{\mathrm{II}}$$

For the second term II, we have

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^N v_{i,t}\left(\widetilde{\lambda}_i - H\lambda_i\right)\lambda_i^\top\right\| &\leq \left(\frac{1}{N}\sum_{i=1}^N v_{i,t}\left\|\widetilde{\lambda}_i - H\lambda_i\right\|^2\right)^{1/2}\left(\frac{1}{N}\sum_{i=1}^N v_{i,t}\left\|\lambda_i\right\|^2\right)^{1/2} \\ &= O_p\left(\frac{1}{\delta_{NT}}\right)O_p\left(\frac{1}{\sqrt{N}}\right) = O_p\left(\frac{1}{\delta_{NT}^2}\right)\end{aligned}$$

from Cauchy-Schwarz inequality, Theorem 2, Assumption 1.4 and Assumption 3.2.

Thus, the first term I is the leading term in $\frac{1}{N}\sum_{i=1}^N v_{i,t}\widetilde{\lambda}_i\lambda_i^\top$. From Assumption 3.6, $\frac{1}{\sqrt{N}}\sum_{i=1}^N v_{i,t}\lambda_i\lambda_i^\top \xrightarrow{d} N(0,\Theta_{\Lambda,t})$. From Slutsky's theorem,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N v_{i,t}\lambda_i\lambda_i^\top F_t \xrightarrow{d} N(0,(F_t^\top\otimes I)\Theta_{\Lambda,t}(F_t\otimes I))$$

and

$$\frac{1}{\sqrt{N}}H\sum_{i=1}^N v_{i,t}\lambda_i\lambda_i^\top F_t \xrightarrow{d} N(0,(Q^{-1})^\top(F_t^\top\otimes I)\Theta_{\Lambda,t}(F_t\otimes I)Q^{-1})$$

A consistent estimator for the asymptotic variance of $(Q^{-1})^\top\Theta_{\Lambda,t}Q^{-1}$ is shown in Lemma 13.

$\sqrt{N}(\widetilde{F}_t - G_tF_t)$ and $\sqrt{N}(G_t - (H^\top)^{-1})F_t$ are asymptotic independent because the randomness of $\sqrt{N}(\widetilde{F}_t - G_tF_t)$ comes from the time series average of $H_i\lambda_i e_{it}$ while the randomness of $\sqrt{T}(G_t - (H^\top)^{-1})F_t$ comes from $v_{i,t}\widetilde{\lambda}_i\lambda_i^\top$. Then, we have

$$\sqrt{N}(\widetilde{F}_t - (H^{-1})^\top F_t) \xrightarrow{d} N(0,\underbrace{(Q^{-1})^\top\Gamma_tQ^{-1}}_{\Theta_{F_t,1}} + \underbrace{(Q^{-1})^\top(F_t^\top\otimes I)\Theta_{\Lambda,t}(F_t\otimes I)Q^{-1}}_{\Theta_{F_t,2}}).$$

The plug-in estimators of $\Theta_{F_t,1}$ and $\Theta_{F_t,2}$, denoted as $\widetilde{\Theta}_{F_t,1}$ and $\widetilde{\Theta}_{F_t,2}$, are provided in Lemmas 12 and 13 respectively. Let $\widetilde{\Theta}_{F_t} = \widetilde{\Theta}_{F_t,1} + \widetilde{\Theta}_{F_t,2}$.

From Slusky's Theorem,

$$\sqrt{N}\widetilde{\Theta}_{F_t}^{-1/2}(\widetilde{F}_t - (H^\top)^{-1}F_t) \xrightarrow{d} N(0,I_r)$$

□

Lemma 12. *Assume there are finitely many nonzeros in each row of $\Sigma_{e_t} = \mathbb{E}[e_t e_t^T]$ and we know the set Ω_{e_t} of nonzero indices in Σ_{e_t}. Under the assumptions in Theorem 3 we have*

$$\widetilde{\Theta}_{F_t,1} = AVar(\sqrt{N}(\widetilde{F}_t - G_t F_t)) + o_p(1),$$

where

$$\widetilde{\Theta}_{F_t,1} = \frac{1}{N} \sum_{i \in \mathcal{O}_t, l \in \mathcal{O}_t, (i,l) \in \Omega_{e_t}} \frac{1}{\widetilde{P}(W_{it} = 1|\Lambda)\widetilde{P}_t(W_{lt} = 1|\Lambda, \mathcal{F}_{t-1})} \widetilde{\lambda}_i \widetilde{\lambda}_l^\top \widetilde{e}_{it} \widetilde{e}_{lt},$$

$\widetilde{e}_{it} = \widetilde{X}_{it} - \widetilde{\lambda}_i^\top \widetilde{F}_t$ *for observed* X_{it} *and* $\widetilde{P}(W_{it} = 1|S)$ *is a consistent estimate for* $P(W_{it} = 1|S)$.

Proof. If X_{it} is observed, $\widetilde{e}_{it}$ is a consistent estimator for e_{it} following the same reasoning as Lemma 9. Recall

$$\sqrt{N}(\widetilde{F}_t - G_t F_t) = \frac{1}{\sqrt{N}} \sum_{i \in \mathcal{O}_t} \frac{1}{P(W_{it} = 1|S)} H_i \lambda_i e_{it} + o_p(1).$$

$\widetilde{\lambda}_i$ is a consistent estimator for $H_i \lambda_i$, then a consistent estimator for the asymptotic variance of $\frac{1}{\sqrt{N}} \sum_{i \in \mathcal{O}_t} \frac{1}{P(W_{it}=1|S)} H_i \lambda_i e_{it}$ is

$$\frac{1}{N} \sum_{i \in \mathcal{O}_t, l \in \mathcal{O}_t, (i,l) \in \Omega_{e_t}} \frac{1}{\widetilde{P}(W_{it} = 1|\Lambda, \mathcal{F}_{t-1})\widetilde{P}(W_{lt} = 1|\Lambda, \mathcal{F}_{t-1})} \widetilde{\lambda}_i \widetilde{\lambda}_l^\top \widetilde{e}_{it} \widetilde{e}_{lt},$$

where $\widetilde{P}(W_{it} = 1|S)$ is a consistent estimate for $P(W_{it} = 1|S)$. $\square$

Lemma 13. *Under the same assumptions in Theorem 3, if $F_t F_t^\top$ and $\lambda_i \lambda_i^\top$ are ergodic in mean, we have*

$$\widetilde{\Theta}_{F_t,2} = AVar\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_{i,t} H \lambda_i \lambda_i^\top F_t \right) + o_p(1),$$

where $\widetilde{\Theta}_{F_t,2} = \frac{1}{N}\left(\widetilde{B}_1 + \widetilde{B}_2\right)$ *with* $\widetilde{B}_1 = \left(\sum_{i=1}^N v_{i,t}^2\right)\left(\frac{1}{N}\sum_{i=1}^N \widetilde{\lambda}_i \widetilde{\lambda}_i^\top \widetilde{F}_t \widetilde{F}_t^\top \widetilde{\lambda}_i \widetilde{\lambda}_i^\top\right)$ *and*

$$\widetilde{B}_2 = \begin{cases} \left(\sum_{i=1}^N \sum_{l \neq i} v_{i,t} v_{l,t}\right) \widetilde{F}_t \widetilde{F}_t^\top, & \textit{if}\, \lambda_i \textit{ is independent} \\ \left(\sum_{i=1}^N \sum_{l \neq i} v_{i,t} v_{l,t}\right) \left(\frac{1}{N-|\rho|} \sum_{m=\max(1,1-\rho),\ \textit{where}\ \rho = l-i}^{\min(N,N-\rho)} \widetilde{\lambda}_m \widetilde{\lambda}_m^\top \widetilde{F}_t \widetilde{F}_t^\top \widetilde{\lambda}_{m+\rho} \widetilde{\lambda}_{m+\rho}^\top\right), & \textit{otherwise} \end{cases}$$

Proof. We have

$$Cov\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N v_{i,t} H \lambda_i \lambda_i^\top F_t, \frac{1}{\sqrt{N}} \sum_{i=1}^N v_{i,t} H \lambda_i \lambda_i^\top F_t | F_t \right) = \frac{1}{N}\left(B_1 + B_2\right),$$

where

$$\begin{aligned} B_1 &= \sum_{i=1}^N v_{i,t}^2 \mathbb{E}[H \lambda_i \lambda_i^\top F_t F_t^\top \lambda_i \lambda_i^\top H^\top | F_t] \\ B_2 &= \sum_{i=1}^N \sum_{l \neq i} v_{i,t} v_{l,t} \mathbb{E}[H \lambda_i \lambda_i^\top F_t F_t^\top \lambda_l \lambda_l^\top H^\top | F_t] \end{aligned}$$

For term B_1, we can consistently estimate $\mathbb{E}[H \lambda_i \lambda_i^\top F_t F_t^\top \lambda_i \lambda_i^\top H^\top | F_t]$ by $\frac{1}{N}\sum_{i=1}^N \widetilde{\lambda}_i \widetilde{\lambda}_i^\top \widetilde{F}_t \widetilde{F}_t^\top \widetilde{\lambda}_i \widetilde{\lambda}_i^\top$.

For term B_2, we separate the case that λ_i is independent from the case that λ_i is dependent.

When λ_i is independent, $\mathbb{E}[H\lambda_i\lambda_i^\top F_t F_t^\top \lambda_l \lambda_l^\top H^\top | F_t] = \mathbb{E}[H\lambda_i\lambda_i^\top] F_t F_t^\top \mathbb{E}[\lambda_i\lambda_i^\top H^\top]$. Since $\lambda_i\lambda_i^\top$ is ergodic in mean, we have $\widetilde{F}_t\widetilde{F}_t^\top - \mathbb{E}[H\lambda_i\lambda_i^\top F_t F_t^\top \lambda_l\lambda_l^\top H^\top | F_t] = o_p(1)$.

If λ_i is dependent, we have $\frac{1}{N-|\rho|}\sum_{m=\max(1,1-\rho),\text{ where }\rho=l-i}^{\min(N,N-\rho)} \widetilde{\lambda}_m\widetilde{\lambda}_m^\top \widetilde{F}_t\widetilde{F}_t^\top \widetilde{\lambda}_{m+\rho}\widetilde{\lambda}_{m+\rho}^\top$ $-\mathbb{E}[H\lambda_i\lambda_i^\top F_t F_t^\top \lambda_l\lambda_l^\top H^\top | F_t] = o_p(1)$.

□

Proof of Theorem 4. From $\widetilde{C}_{it} = \widetilde{\lambda}_i^\top \widetilde{F}_t$ and $C_{it} = \lambda_i^\top F_t$, we have

$$\widetilde{C}_{it} - C_{it} = \lambda_i^\top H^\top(\widetilde{F}_t - (H^\top)^{-1}F_t) + (\widetilde{\lambda}_i - H\lambda_i)^\top \widetilde{F}_t + o_p(1/\delta_{NT})$$

The second term can be written as

$$\begin{aligned}(\widetilde{\lambda}_i - H\lambda_i)^\top \widetilde{F}_t &= (\widetilde{\lambda}_i - H\lambda_i)^\top (H^\top)^{-1}F_t + (\widetilde{\lambda}_i - H\lambda_i)^\top(\widetilde{F}_t - (H^\top)^{-1}F_t) \\ &= (\widetilde{\lambda}_i - H\lambda_i)^\top (H^\top)^{-1}F_t + o_p(1/\delta_{NT})\end{aligned}$$

Thus,

$$\widetilde{C}_{it} - C_{it} = \lambda_i^\top H^\top(\widetilde{F}_t - (H^\top)^{-1}F_t) + (\widetilde{\lambda}_i - H\lambda_i)^\top (H^\top)^{-1}F_t + o_p(1/\delta_{NT}).$$

Following Theorem 3 in Bai (2003), we can show that $H^\top H = \left(\frac{\Lambda^\top\Lambda}{N}\right)^{-1} + O_p\left(\frac{1}{\delta_{NT}^2}\right)$. Then,

$$\begin{aligned}\delta_{NT}\lambda_i^\top H^\top(\widetilde{F}_t - (H^\top)^{-1}F_t) &= \frac{\delta_{NT}}{N}\lambda_i^\top H^\top H\left(\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\lambda_i e_{it} + \sum_{i=1}^N v_{i,t}\lambda_i\lambda_i^\top F_t\right) + o_p(1) \\ &= \frac{\delta_{NT}}{N}\lambda_i^\top \left(\frac{\Lambda^\top\Lambda}{N}\right)^{-1}\left(\sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\lambda_i e_{it} + \sum_{i=1}^N v_{i,t}\lambda_i\lambda_i^\top F_t\right) + o_p(1)\end{aligned}$$

and

$$\begin{aligned}&F_t^\top H^{-1}(\widetilde{\lambda}_i - H\lambda_i) \\ =\ & \tfrac{\delta_{NT}}{T}F_t^\top\left(\tfrac{F^\top F}{T}\right)^{-1}\left(\tfrac{\Lambda^\top\widetilde{\Lambda}}{N}\right)^{-1}\widetilde{V}\widetilde{V}^{-1}H\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \lambda_i\lambda_i^\top \tfrac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} + \tfrac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T y_{it,j}\lambda_i\lambda_i^\top F_t F_t^\top\right) + o_p(1) \\ =\ & \tfrac{\delta_{NT}}{T}F_t^\top\left(\tfrac{F^\top F}{T}\right)^{-1}\left(\tfrac{\Lambda^\top\Lambda}{N}\right)^{-1}\left(\tfrac{1}{N}\textstyle\sum_{i=1}^N \lambda_i\lambda_i^\top \tfrac{1}{|\mathcal{Q}_{ij}|}\sum_{t\in\mathcal{Q}_{ij}} F_t e_{jt} + \tfrac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T y_{it,j}\lambda_i\lambda_i^\top F_t F_t^\top\right) + o_p(1)\end{aligned}$$

following $\left(\frac{\Lambda^\top\widetilde{\Lambda}}{N}\right)^{-1} = H^\top + o_p(1)$ in Lemma 8. $\widetilde{F}_t - (H^\top)^{-1}F_t$ and $\widetilde{\lambda}_i - H\lambda_i$ are asymptotic independent because the former is the average of cross-section random variables and $\frac{1}{N}\sum_{i=1}^N v_{i,t}\widetilde{\lambda}_i\lambda_i^\top$ with $\sum_{i=1}^N v_{i,t} = 0$ and the latter the average of time-series random variables and $\frac{1}{NT}\sum_{l=1}^N\sum_{T=1}^T y_{lt,i}\widetilde{\lambda}_l\lambda_l^\top F_t F_t^\top$ with $\sum_{l=1}^N\sum_{t=1}^T y_{lt,i} = 0$. Then we have

$$\begin{aligned}\delta_{NT}(\widetilde{C}_{it} - C_{it}) \xrightarrow{d} N\Bigg(0, &\frac{\delta_{NT}^2}{N}\lambda_i^\top\Sigma_\Lambda^{-1}(\Gamma_t + (F_t^\top\otimes I)\Theta_{\Lambda,t}(F_t\otimes I))\Sigma_\Lambda^{-1}\lambda_i \\ &+\frac{\delta_{NT}^2}{T}F_t^\top\Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_j + (\lambda_j^\top\otimes I)\Xi_{F,j}(\lambda_j\otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_t\Bigg)\end{aligned}$$

From Theorem 2, the plug-in consistent estimator for the asymptotic variance of $\delta_{NT}(\widetilde{\lambda}_i - H\lambda_i)$ is $\frac{\delta_{NT}^2}{T}\widetilde{\Gamma}_{\lambda_i}$. From Theorem 3, the plug-in consistent estimator for the asymptotic variance of $\delta_{NT}(\widetilde{F}_t - (H^\top)^{-1}F_t)$ is $\frac{\delta_{NT}^2}{N}\widetilde{\Theta}_{F_t}$. Together with $\widetilde{F}_t$ to be the consistent estimator for $(H^\top)^{-1}F_t$ and $\widetilde{\lambda}_i$ to be the consistent estimator for $H\lambda_i$, the consistent estimator for th asymptotic variance of $\delta_{NT}(\widetilde{C}_{it} - C_{it})$ is $\frac{\delta_{NT}^2}{T}\widetilde{\lambda}_i^\top \widetilde{\Theta}_{F_t}\widetilde{\lambda}_i + \frac{\delta_{NT}^2}{N}\widetilde{F}_t^\top \widetilde{\Gamma}_{\lambda_i}\widetilde{F}_t$. Then we have

$$\left(\frac{1}{T}\widetilde{\lambda}_i^\top \widetilde{\Theta}_{F_t}\widetilde{\lambda}_i + \frac{1}{N}\widetilde{F}_t^\top \widetilde{\Gamma}_{\lambda_i}\widetilde{F}_t\right)^{-1/2}(\widetilde{C}_{it} - C_{it}) \xrightarrow{d} N(0,1).$$

$\square$

Proof of Lemma 1. The average of common components $\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T \left(\widetilde{C}_{it} - C_{it}\right)$ has

$$\begin{aligned}\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T \left(\widetilde{C}_{it} - C_{it}\right) &= \lambda_i^\top H^\top \frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T (\widetilde{F}_t - (H^\top)^{-1}F_t) \\ &\quad + (\widetilde{\lambda}_i - H\lambda_i)^\top (H^\top)^{-1}\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T F_t + o_p(1/\delta_{NT})\end{aligned}$$

From Theorem 3,

$$\sqrt{N}(\widetilde{F}_t - (H^\top)^{-1}F_t) = \sqrt{N}(\widetilde{F}_t - G_tF_t) + \sqrt{N}(G_t - (H^\top)^{-1})F_t + o_p(1)$$

From Assumption 5.1, we have $\frac{1}{\sqrt{N(T-T_{0,i})}}\sum_{t=T_{0,i}+1}^T \sum_{i\in\mathcal{O}_t}\frac{1}{P(W_{it}=1|S)}\lambda_i e_{it} = o_p(1)$ and then

$$\frac{\sqrt{N}}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T (\widetilde{F}_t - G_tF_t) = o_p(1)$$

From Assumption 5.2, we have $\frac{\sqrt{N}}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T \left(\frac{1}{N}\sum_{i=1}^N \frac{W_{it}\lambda_i\lambda_i^\top}{P(W_{it}=1|S)} - \frac{1}{N}\sum_{i=1}^N \lambda_i\lambda_i^\top\right)F_t \xrightarrow{d} N(0,\Theta_{\Lambda,i})$ and then

$$\begin{aligned}\frac{\sqrt{N}}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T (G_t - (H^\top)^{-1})F_t &= \frac{1}{\sqrt{N}(T-T_{0,i})}H\sum_{t=T_{0,i}+1}^T\sum_{i=1}^N v_{i,t}\lambda_i\lambda_i^\top F_t + o_p(1) \\ &\xrightarrow{d} N(0,(Q^{-1})^\top \Theta_{\Lambda,i}Q^{-1})\end{aligned}$$

From the proof of Theorems 2 and 4,

$$\frac{1}{\sqrt{T}}\left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T F_t\right)^\top H^{-1}(\widetilde{\lambda}_i - H\lambda_i) \xrightarrow{d} N\left(0, \mu_F^\top \Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_j + (\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j\otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}\mu_F\right)$$

Following the same argument as Theorems 2-4, the asymptotic variance of $\lambda_i^\top H^\top \sum_{t=t_0+1}^{t_1}(\widetilde{F}_t - (H^\top)^{-1}F_t)$ is asymptotic independent of the asymptotic variance of $(\widetilde{\lambda}_i - H\lambda_i)^\top (H^\top)^{-1}\sum_{t=t_0+1}^{t_1} F_t$. Then we have

$$\begin{aligned}\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T \left(\widetilde{C}_{it} - C_{it}\right) &\xrightarrow{d} N\Big(0, \lambda_i^\top \Sigma_\Lambda^{-1}\Theta_{\Lambda,i}\Sigma_\Lambda^{-1}\lambda_i \\ &\qquad + \mu_F^\top \Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_j + (\lambda_j^\top \otimes I)\Xi_{F,j}(\lambda_j\otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}\mu_F\Big)\end{aligned}$$

$\square$

Lemma 14. *Under the same assumptions in Theorem 3 we have*

$$\widetilde{\Theta}_{\Lambda,i} = AVar\left(\frac{1}{\sqrt{N}(T-T_{0,i})}\sum_{t=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}H\lambda_i\lambda_i^\top F_t\right) + o_p(1),$$

where $\widetilde{\Theta}_{\Lambda,i} = \frac{1}{N}\left(\widetilde{B}_{i,1} + \widetilde{B}_{i,2}\right)$ *with*
$\widetilde{B}_{i,1} = \frac{1}{(T-T_{0,i})^2}\sum_{t=T_{0,i}+1}^{T}\sum_{s=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}v_{i,s}\left(\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\widetilde{F}_t\widetilde{F}_s^\top\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\right)$*, and when* λ_i *is independent,*

$$\widetilde{B}_{i,2} = \frac{1}{(T-T_{0,i})^2}\sum_{t=T_{0,i}+1}^{T}\sum_{s=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}v_{i,s}\widetilde{F}_t\widetilde{F}_s^\top;$$

otherwise,

$$\widetilde{B}_{i,2} = \frac{1}{(T-T_{0,i})^2}\sum_{t=T_{0,i}+1}^{T}\sum_{s=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}v_{i,s}\left(\frac{1}{N-|\rho|}\sum_{m=\max(1,1-\rho),\ \textit{where}\ \rho=l-i}^{\min(N,N-\rho)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_t\widetilde{F}_s^\top\widetilde{\lambda}_{m+\rho}\widetilde{\lambda}_{m+\rho}^\top\right)$$

Proof. We have

$Cov\left(\frac{1}{\sqrt{N}(T-T_{0,i})}\sum_{t=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}H\lambda_i\lambda_i^\top F_t, \frac{1}{\sqrt{N}(T-T_{0,i})}\sum_{t=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}H\lambda_i\lambda_i^\top F_t|F\right) = \frac{1}{N}\left(B_1 + B_2\right),$

where

$$\begin{aligned} B_1 &= \frac{1}{(T-T_{0,i})^2}\sum_{t=T_{0,i}+1}^{T}\sum_{s=T_{0,i}+1}^{T}\sum_{i=1}^{N} v_{i,t}v_{i,s}\mathbb{E}[H\lambda_i\lambda_i^\top F_tF_s^\top\lambda_i\lambda_i^\top H^\top|F] \\ B_2 &= \frac{1}{(T-T_{0,i})^2}\sum_{t=T_{0,i}+1}^{T}\sum_{s=T_{0,i}+1}^{T}\sum_{i=1}^{N}\sum_{l\neq i} v_{i,t}v_{l,s}\mathbb{E}[H\lambda_i\lambda_i^\top F_tF_s^\top\lambda_l\lambda_l^\top H^\top|F] \end{aligned}$$

For term B_1, we can consistently estimate $\mathbb{E}[H\lambda_i\lambda_i^\top F_tF_s^\top\lambda_i\lambda_i^\top H^\top|F]$ by $\frac{1}{N}\sum_{i=1}^{N}\widetilde{\lambda}_i\widetilde{\lambda}_i^\top\widetilde{F}_t\widetilde{F}_s^\top\widetilde{\lambda}_i\widetilde{\lambda}_i^\top$.

For term B_2, we separate the case that λ_i is independent from the case that λ_i is dependent.

When λ_i is independent, $\mathbb{E}[H\lambda_i\lambda_i^\top F_tF_s^\top\lambda_l\lambda_l^\top H^\top|F] = \mathbb{E}[H\lambda_i\lambda_i^\top]F_tF_s^\top\mathbb{E}[\lambda_i\lambda_i^\top H^\top]$. Since $\lambda_i\lambda_i^\top$ is ergodic in mean, we have $\widetilde{F}_t\widetilde{F}_s^\top - \mathbb{E}[H\lambda_i\lambda_i^\top F_tF_s^\top\lambda_l\lambda_l^\top H^\top|F] = o_p(1)$.

If λ_i is dependent, we have $\frac{1}{N-|\rho|}\sum_{m=\max(1,1-\rho),\text{ where }\rho=l-i}^{\min(N,N-\rho)}\widetilde{\lambda}_m\widetilde{\lambda}_m^\top\widetilde{F}_t\widetilde{F}_s^\top\widetilde{\lambda}_{m+\rho}\widetilde{\lambda}_{m+\rho}^\top$
$-\mathbb{E}[H\lambda_i\lambda_i^\top F_tF_s^\top\lambda_l\lambda_l^\top H^\top|F] = o_p(1)$. □

Proof of Lemma 2. We can decompose the estimated loadings $\widetilde{\lambda}_i^{treat}$ by

$$\begin{aligned} \widetilde{\lambda}_i^{treat} &= \left(\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t\widetilde{F}_t^\top\right)^{-1}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_tX_{it}^{treat} \\ &= \left(\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t\widetilde{F}_t^\top\right)^{-1}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_tF_t^\top\lambda_i^{treat} + \left(\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_t\widetilde{F}_t^\top\right)^{-1}\sum_{t=T_{0,i}+1}^{T}\widetilde{F}_te_{it}^{treat} \end{aligned}$$

Denote the population and estimated factors from $T_{0,i}+1$ to T as $F_{(T_{0,i}+1):T}, \widetilde{F}_{(T_{0,i}+1):T} \in \mathbb{R}^{(T-T_{0,i})\times r}$ From Theorem 3, for any t, $\widetilde{F}_t - (H^\top)^{-1}F_t = O_p\left(\frac{1}{\delta_{NT}}\right)$. Then we have

$$\begin{aligned}
\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T} &= \tfrac{1}{T-T_{0,i}}\left(F_{(T_{0,i}+1):T}H^{-1} + O_p\left(\tfrac{1}{\delta_{NT}}\right)\right)^\top\left(F_{(T_{0,i}+1):T}H^{-1} + O_p\left(\tfrac{1}{\delta_{NT}}\right)\right) \\
&= \tfrac{1}{T-T_{0,i}}(H^\top)^{-1}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}H^{-1} + O_p\left(\tfrac{1}{\delta_{NT}}\right)
\end{aligned}$$

From Assumption 2.1, $\frac{1}{T-T_{0,i}}FF^\top_{(T_{0,i}+1):T}FF_{(T_{0,i}+1):T}$ is invertible, thus,

$$\left(\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1} = H\left(\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\right)^{-1}H^\top + O_p\left(\tfrac{1}{\delta_{NT}}\right)$$

and for any t and s

$$\begin{aligned}
&\widetilde{F}^\top_t\left(\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\widetilde{F}_s \\
=\ &\left(F_tH^{-1} + O_p\left(\tfrac{1}{\delta_{NT}}\right)\right)\left(H\left(\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\right)^{-1}H^\top + O_p\left(\tfrac{1}{\delta_{NT}}\right)\right)\left((H^\top)^{-1}F_s + O_p\left(\tfrac{1}{\delta_{NT}}\right)\right) \\
=\ &F^\top_t\left(\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\right)^{-1}F_s + O_p\left(\tfrac{1}{\delta_{NT}}\right) \qquad (26)
\end{aligned}$$

We regress $X_{i,T_{0,i}+1:T}$ on $\widetilde{F}_{(T_{0,i}+1):T}$ to get $\widetilde{\lambda}^{treat}_i$, that is

$$\begin{aligned}
\widetilde{\lambda}^{treat}_i &= \left(\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\widetilde{F}^\top_{(T_{0,i}+1):T}X_{i,T_{0,i}+1:T} \\
&= \left(\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\widetilde{F}^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\lambda^{treat}_i \\
&\quad + \left(\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\widetilde{F}^\top_{(T_{0,i}+1):T}e_{i,T_{0,i}+1:T},
\end{aligned}$$

where the second term is the estimation error. Then for $\widetilde{C}^{treat}_{it} = \widetilde{F}^\top_t\widetilde{\lambda}^{treat}_i$, we have

$$\begin{aligned}
\widetilde{C}^{treat}_{it} &= \widetilde{F}^\top_t\left(\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\widetilde{F}^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\lambda^{treat}_i \\
&\quad + \widetilde{F}^\top_t\left(\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\widetilde{F}^\top_{(T_{0,i}+1):T}e_{i,T_{0,i}+1:T}
\end{aligned}$$

From Equation (26), for the first term in $\widetilde{C}^{treat}_{it}$, we have

$$\begin{aligned}
&\widetilde{F}^\top_t\left(\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\lambda^{treat}_i \\
=\ &F^\top_t\left(\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\right)^{-1}\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\lambda^{treat}_i + O_p\left(\tfrac{1}{\delta_{NT}}\right) \\
=\ &F^\top_t\lambda^{treat}_i + O_p\left(\tfrac{1}{\delta_{NT}}\right) = C^{treat}_{it} + O_p\left(\tfrac{1}{\delta_{NT}}\right)
\end{aligned}$$

and the leading term in the error is $(\widetilde{F}_t - (H^\top)^{-1}F_t)^\top H\lambda^{treat}_i$. For the second term in $\widetilde{C}^{treat}_{it}$, we have

$$\begin{aligned}
&\widetilde{F}_{(T_{0,i}+1):T}\left(\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}\widetilde{F}_{(T_{0,i}+1):T}\right)^{-1}\tfrac{1}{T-T_{0,i}}\widetilde{F}^\top_{(T_{0,i}+1):T}e_{T_{0,i}+1:T} \\
=\ &F_{(T_{0,i}+1):T}\left(\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}F_{(T_{0,i}+1):T}\right)^{-1}\tfrac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}e_{T_{0,i}+1:T} + o_p\left(\tfrac{1}{\delta_{NT}}\right)
\end{aligned}$$

following the estimation error for $\widetilde{F}_t$ is $O_p\left(\frac{1}{\delta_{NT}}\right)$, $\frac{1}{T-T_{0,i}}F^\top_{(T_{0,i}+1):T}e_{T_{0,i}+1:T} = O_p\left(\frac{1}{\sqrt{T-T_{0,i}}}\right)$ and $\left(\widetilde{F}_{(T_{0,i}+1):T} - F_{(T_{0,i}+1):T}H^{-1}\right)^\top e_{T_{0,i}+1:T} = o_p\left(\frac{1}{\delta_{NT}}\right)$ following the same argument as Lemma 6. Thus, we have

$$
\begin{aligned}
\widetilde{C}_{it}^{treat} - C_{it}^{treat} &= (\widetilde{F}_t - (H^\top)^{-1}F_t)^\top H\lambda_i^{treat} \\
&\quad + F_t^\top \left(\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T F_tF_t^\top\right)^{-1}\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T F_te_{it} + o_p\left(\frac{1}{\delta_{NT}}\right)
\end{aligned}
$$

Since the estimation of $(\widetilde{F}_t-(H^\top)^{-1}F_t)^\top$ comes from the control observations and $\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T F_te_{it}$ is determined by treated observations, they are asymptotically independent. Together with $\frac{1}{T-T_{0,i}}\sum_{t=T_{0,i}+1}^T F_tF_t^\top \xrightarrow{P} \Sigma_F$ from Assumption 2.1 and $\frac{1}{\sqrt{T-T_{0,i}}}\sum_{t=T_{0,i}+1}^T F_te_{it} \xrightarrow{d} N(0,\Psi_i)$ from Assumption 4, we have

$$
\begin{aligned}
\sqrt{T-T_{0,i}}(\widetilde{C}_{it}^{treat} - C_{it}^{treat}) &\xrightarrow{d} N\Big(0, \tfrac{T-T_{0,i}}{T}(\lambda_i^{treat})^\top\Sigma_\Lambda^{-1}(\Gamma_t + (F_t^\top\otimes I)\Theta_{\Lambda,t}(F_t\otimes I))\Sigma_\Lambda^{-1}\lambda_i^{treat} \\
&\qquad + F_t^\top\Sigma_F^{-1}\Psi_i\Sigma_F^{-1}F_t\Big)
\end{aligned}
$$

□

Lemma 15. *Suppose Assumptions 1-5 hold, $T-T_{0,i}\to\infty$, t_1-t_0 is finite, and let $\delta^2_{N,T-T_{0,i}} = \min(N, T-T_{0,i})$,*

$$
\delta_{NT}((Z^\top Z)^{-1}Z^\top M^{ctrl}Z(Z^\top Z)^{-1} + M_Z^{ctrl})^{-1/2}\left(\widetilde{\beta}_i^{ctrl} - \beta_i^{ctrl}\right) \xrightarrow{d} N(0, I) \quad (27)
$$

$$
\delta_{N,T-T_{0,i}}((Z^\top Z)^{-1}Z^\top M^{treat}Z(Z^\top Z)^{-1} + M_Z^{treat})^{-1/2}\left(\widetilde{\beta}_i^{treat} - \beta_i^{treat}\right) \xrightarrow{d} N(0, I) \quad (28)
$$

where $(Z^\top M^{ctrl}Z)^{-1/2}$ ($(Z^\top M^{treat}Z)^{-1/2}$) is the square root of $Z^\top M^{ctrl}Z$ ($Z^\top M^{treat}Z$) and M^{ctrl} (M^{treat}) is a $(T-T_{0,i})\times(T-T_{0,i})$ matrix with

$$
\begin{aligned}
M^{ctrl}_{t-T_{0,i},t-T_{0,i}} &= AVar(\delta_{N,T-T_{0,i}}(\widetilde{C}^{ctrl}_{i,t} - C^{ctrl}_{i,t})) \\
&= \frac{\delta^2_{NT}}{T}F_t^\top\Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_i + ((\lambda_i^{ctrl})^\top\otimes I)\Xi_{F,i}(\lambda_i^{ctrl}\otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_t \\
M^{ctrl}_{t-T_{0,i},s-T_{0,i}} &= Cov(\delta_{N,T-T_{0,i}}(\widetilde{C}^{ctrl}_{i,t} - C^{ctrl}_{i,t}), \delta_{N,T-T_{0,i}}(\widetilde{C}^{ctrl}_{i,s} - C^{ctrl}_{i,s})) \\
&= \frac{\delta^2_{NT}}{T}F_t^\top\Sigma_F^{-1}\Sigma_\Lambda^{-1}(\Phi_i + ((\lambda_i^{ctrl})^\top\otimes I)\Xi_{F,i}(\lambda_i^{ctrl}\otimes I))\Sigma_\Lambda^{-1}\Sigma_F^{-1}F_s \\
M^{ctrl}_{Z,lm} &= \frac{\delta^2_{N,T-T_{0,i}}}{N}(\lambda_i^{ctrl})^\top\Sigma_\Lambda^{-1}\Theta_{\Lambda,i,Z,lm}\Sigma_\Lambda^{-1}\lambda_i^{ctrl}
\end{aligned}
$$

and

$$
\begin{aligned}
M^{treat}_{t-T_{0,i},t-T_{0,i}} &= AVar(\delta_{N,T-T_{0,i}}(\widetilde{C}^{treat}_{i,t} - C^{treat}_{i,t})) = \frac{\delta^2_{N,T-T_{0,i}}}{T-T_{0,i}}F_t^\top\Sigma_F^{-1}\Psi_i\Sigma_F^{-1}F_t \\
M^{treat}_{t-T_{0,i},s-T_{0,i}} &= Cov(\delta_{N,T-T_{0,i}}(\widetilde{C}^{treat}_{i,t} - C^{treat}_{i,t}), \delta_{N,T-T_{0,i}}(\widetilde{C}^{treat}_{i,s} - C^{treat}_{i,s})) = \frac{\delta^2_{N,T-T_{0,i}}}{T-T_{0,i}}F_t^\top\Sigma_F^{-1}\Psi_i\Sigma_F^{-1}F_s \\
M^{treat}_{Z,lm} &= \frac{\delta^2_{N,T-T_{0,i}}}{N}(\lambda_i^{treat})^\top\Sigma_\Lambda^{-1}\Theta_{\Lambda,i,Z,lm}\Sigma_\Lambda^{-1}\lambda_i^{treat}
\end{aligned}
$$

Proof of Lemma 15. Since

$$\widetilde{\beta}_i^{ctrl} - \beta_i^{ctrl} = (Z^\top Z)^{-1} Z^\top (\widetilde{C}_{i,(T_{0,i}+1):T}^{ctrl} - C_{i,(T_{0,i}+1):T}^{ctrl})$$

and

$$\widetilde{\beta}_i^{treat} - \beta_i^{treat} = (Z^\top Z)^{-1} Z^\top (\widetilde{C}_{i,(T_{0,i}+1):T}^{treat} - C_{i,(T_{0,i}+1):T}^{treat}),$$

the proof is a direct extension of Theorem 4 and Lemma 2. The term $(Z^\top Z)^{-1} Z^\top M^{ctrl} Z (Z^\top Z)^{-1}$ comes from $(\widetilde{\lambda}_i^{ctrl} - H\lambda_i^{ctrl})^\top (H^\top)^{-1} F_t$. The term M_Z^{ctrl} (and M_Z^{treat}) comes from $(\widetilde{F}_t - (H^\top)^{-1} F_t)^\top H \lambda_i^{ctrl}$ (and $(\widetilde{F}_t - (H^\top)^{-1} F_t)^\top H \lambda_i^{treat}$), which follows Assumptions 5 that the variance correction term dominates. The term $(Z^\top Z)^{-1} Z^\top M^{treat} Z (Z^\top Z)^{-1}$ comes from $F_t^\top \left(\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^T F_t F_t^\top \right)^{-1} \frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^T F_t e_{it}$.

□

Proof of Theorem 6. Note that

$$(\widetilde{\beta}_i^{ctrl} - \widetilde{\beta}_i^{treat}) - (\beta_i^{ctrl} - \beta_i^{treat}) = (Z^\top Z)^{-1} Z^\top ((\widetilde{C}_{i,(T_{0,i}+1):T}^{ctrl} - C_{i,(T_{0,i}+1):T}^{ctrl}) - (\widetilde{C}_{i,(T_{0,i}+1):T}^{treat} - C_{i,(T_{0,i}+1):T}^{treat}))$$

We have

$$\begin{aligned} & ((\widetilde{C}_{i,t}^{ctrl} - C_{i,t}^{ctrl}) - (\widetilde{C}_{i,t}^{treat} - C_{i,t}^{treat})) \\ = \; & (\widetilde{F}_t - (H^\top)^{-1} F_t)^\top H (\lambda_i^{ctrl} - \lambda_i^{treat}) + (\widetilde{\lambda}_i^{ctrl} - H\lambda_i^{ctrl})^\top (H^\top)^{-1} F_t \\ & - F_t^\top \left(\tfrac{1}{T-T_{0,i}} \textstyle\sum_{t=T_{0,i}+1}^T F_t F_t^\top \right)^{-1} \tfrac{1}{T-T_{0,i}} \textstyle\sum_{t=T_{0,i}+1}^T F_t e_{it} + o_p\left(\tfrac{1}{\delta_{NT}}\right) \end{aligned}$$

$(\widetilde{\lambda}_i^{ctrl} - H\lambda_i^{ctrl})^\top (H^\top)^{-1} F_t$ and $F_t^\top \left(\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^T F_t F_t^\top \right)^{-1} \frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^T F_t e_{it}$ are asymptotically independent because the first term is determined by the errors on the control panel while the second term is determined by the errors on the treated panel.

Thus, from Lemma 2 and Theorem 4,

$$\begin{aligned} & M_{t-T_{0,i},t-T_{0,i}} \\ = \; & AVar(\delta_{N,T-T_{0,i}}((\widetilde{\lambda}_i^{ctrl} - H\lambda_i^{ctrl})^\top (H^\top)^{-1} F_t - F_t^\top (\frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^T F_t F_t^\top)^{-1} \frac{1}{T-T_{0,i}} \sum_{t=T_{0,i}+1}^T F_t e_{it})) \\ = \; & \frac{\delta_{N,T-T_{0,i}}^2}{T} F_t^\top \Sigma_F^{-1} \Sigma_\Lambda^{-1} (\Phi_i + ((\lambda_i^{ctrl})^\top \otimes I) \Xi_{F,i} (\lambda_i^{ctrl} \otimes I)) \Sigma_\Lambda^{-1} \Sigma_F^{-1} F_t + \frac{\delta_{N,T-T_{0,i}}^2}{T-T_{0,i}} F_t^\top \Sigma_F^{-1} \Psi_i \Sigma_F^{-1} F_t \end{aligned}$$

Moreover,

$$\begin{aligned} & M_{t-T_{0,i},s-T_{0,i}} \\ = \; & Cov(\delta_{N,T-T_{0,i}}((\widetilde{\lambda}_i^{ctrl} - H\lambda_i^{ctrl})^\top (H^\top)^{-1} F_t - F_t^\top (\frac{1}{T-T_{0,i}} \sum_{u=T_{0,i}+1}^T F_u F_u^\top)^{-1} \frac{1}{T-T_{0,i}} \sum_{u=T_{0,i}+1}^T F_u e_{iu}), \\ & \delta_{N,T-T_{0,i}}((\widetilde{\lambda}_i^{ctrl} - H\lambda_i^{ctrl})^\top (H^\top)^{-1} F_s - F_s^\top (\frac{1}{T-T_{0,i}} \sum_{u=T_{0,i}+1}^T F_u F_u^\top)^{-1} \frac{1}{T-T_{0,i}} \sum_{u=T_{0,i}+1}^T F_u e_{iu}) \\ = \; & \frac{\delta_{N,T-T_{0,i}}^2}{T} F_t^\top \Sigma_F^{-1} \Sigma_\Lambda^{-1} (\Phi_i + ((\lambda_i^{ctrl})^\top \otimes I) \Xi_{F,i} (\lambda_i^{ctrl} \otimes I)) \Sigma_\Lambda^{-1} \Sigma_F^{-1} F_s + \frac{\delta_{N,T-T_{0,i}}^2}{T-T_{0,i}} F_t^\top \Sigma_F^{-1} \Psi_i \Sigma_F^{-1} F_s \end{aligned}$$

We also need to consider the asymptotic distribution for $(\widetilde{F}_t - (H^\top)^{-1}F_t)^\top H(\lambda_i^{ctrl} - \lambda_i^{treat})$, which is driven by the variance correction term (the term in Assumption 5.3) from Assumption 5.

$$
\begin{aligned}
M_{Z,lm} &= Cov(\delta_{N,T-T_{0,i}}(Z^\top Z)^{-1}Z_l(\widetilde{F}_{l+T_{0,i}} - (H^\top)^{-1}F_{l+T_{0,i}})^\top H(\lambda_i^{ctrl} - \lambda_i^{treat}), \\
&\quad \delta_{N,T-T_{0,i}}(Z^\top Z)^{-1}Z_m(\widetilde{F}_{m+T_{0,i}} - (H^\top)^{-1}F_{m+T_{0,i}})^\top H(\lambda_i^{ctrl} - \lambda_i^{treat})) \\
&= \frac{\delta^2_{N,T-T_{0,i}}}{N}(\lambda_i^{ctrl} - \lambda_i^{treat})^\top \Sigma_\Lambda^{-1}\Theta_{\Lambda,i,Z,lm}\Sigma_\Lambda^{-1}(\lambda_i^{ctrl} - \lambda_i^{treat})
\end{aligned}
$$

Thus,

$$
\delta_{N,T-T_{0,i}}\left((Z^\top Z)^{-1}(Z^\top MZ)(Z^\top Z)^{-1} + M_Z\right)^{-1/2}\left((\widetilde{\beta}_i^{ctrl} - \widetilde{\beta}_i^{treat}) - (\beta_i^{ctrl} - \beta_i^{treat})\right) \xrightarrow{d} N(0, I)
$$

$\square$