Spillovers under Common Shocks in Panel Data: Estimation and Inference*

Qiyang Yu†

Draft Version: November 13, 2026

Click here for latest version

Abstract

Economic outcomes reflect both private effects and spillovers across firms, industries, or regions. Distinguishing these effects is difficult when interaction networks are unknown and unobserved common shocks affect both outcomes and covariates. A high-dimensional spillover network must then be distinguished from co-movement generated by common shocks. This paper develops estimation and inference for bilateral spillover and common private effects with unknown networks and unobserved common shocks. Unobserved common shocks affect outcomes through interactive fixed effects. Identification exploits a factor structure in covariates and exogenous idiosyncratic variation under sparsity conditions. Combining factor projection with debiased LASSO and post-selection OLS, we establish asymptotic normality and simultaneous spillover inference while accounting for factor-estimation and regularization errors. An application to U.S. industry-level R&D reveals an asymmetric spillover network with both positive and negative effects across industries. In addition, the results highlight the importance of accounting for common shocks: the number of significant spillover links falls by 42%, and the estimated private R&D return declines and becomes less statistically significant.

Keywords: Spillover effects; interactive fixed effects; high-dimensional parameters; panel data; R&D Spillovers.

JEL classification: C23, C38, C55, C12, O33.

*I am deeply grateful to my advisor, Bin Chen, for her continued guidance and support. I also thank Yuefeng Han, Nese Yildiz, Yukun Ma, Xun Lu, Liangjun Su and participants at the econometrics seminar at Rochester, the 2026 Asian Meeting of the Econometric Society, Hong Kong–China, the 2026 Midwest Econometrics Group Meeting, the 2026 Canadian Econometric Study Group Meeting, and the 2026 New York State Economics Association Conference for helpful comments and suggestions. All remaining errors are solely mine.

†Department of Economics, University of Rochester. Email: qyu13@ur.rochester.edu.

1 Introduction

Economic outcomes depend not only on a unit's own characteristics but also on those of others. A student's academic achievement depends on classmates as well as individual behaviors (Sacerdote, 2001; Duflo et al., 2011; Barwick et al., 2026). A firm's productivity and market performance respond to innovations by its peers and competitors (Jaffe, 1986; Bloom et al., 2013; Arora et al., 2021). Financial shocks originating elsewhere can affect regional economic activity (Ilzetzki and Jin, 2021; Camara, 2025). Understanding the magnitude and direction of these spillovers is important for assessing how economic changes propagate across individuals, firms, and regions. This paper develops estimation and inference methods for private and bilateral spillover effects in panel data with unknown interaction networks and unobserved common shocks.

Recovering spillover effects from panel data poses two related econometric challenges. First, the relationships through which spillovers operate are often only partially observed. In studies of R&D spillovers, the literature identifies several channels, including knowledge diffusion, product-market competition, and supplier-customer relationships (Bloom et al., 2013; Lucking et al., 2019; Arque-Castells and Spulber, 2022). These channels need not exhaust the ways firms affect one another. Researchers may also lack the information needed to measure even the recognized channels accurately. Therefore, using pre-specified networks to estimate spillover effects faces two distinct problems. A network that misrepresents the intended channel can yield misleading estimates of spillovers through that channel. Even if the specified network accurately represents its intended channel, omitting other channels that generate correlated spillover exposures can cause its coefficient to absorb their effects. The coefficient may then reflect a combination of channels rather than the effect of the specified channel alone, as emphasized by Ritzwoller (2025). This paper treats the spillover network as unknown and estimates bilateral effects directly. In this case, each unit has up to $N-1$ spillover coefficients, which could be too many to estimate reliably without further restrictions.

Second, common shocks can affect both outcomes and covariates. Their influence may therefore be mistaken for private or spillover effects (Manski, 1993). These shocks may be unobserved by researchers, making their influence even harder to distinguish from private and spillover effects. When the network is also unknown, this confounding creates an identification difficulty: the same observed comovement may be explained by different combinations of networks and common shocks. Hence, identification requires restrictions that distinguish spillover transmission from common-shock variation while allowing the interaction network to remain unknown.

Allowing for an unknown spillover network and unobserved common shocks also creates distinct challenges for inference. For bilateral effects, regularization used to estimate the high-dimensional network introduces bias, while accounting for unobserved common shocks introduces additional estimation uncertainty. Both must be controlled to construct valid confidence intervals. For private effects, errors from the unit-specific spillover regressions accumulate when information is pooled across units. Consistency of each estimated spillover vector therefore does not ensure that its contribution to the pooled estimator is negligible at the $\sqrt{NT}$ scale. Simultaneous inference poses a further difficulty. Testing whether a unit receives any spillovers involves a growing number of potential sending units and requires approximating the distribution of the maximum estimation error. Asymptotic normality for individual coefficients alone does not provide this approximation.

To distinguish bilateral spillovers from common-shock variation, the model combines a sparse spillover network with a factor structure in the covariates and interactive fixed effects in outcomes. Each bilateral coefficient captures the combined effect of the underlying interaction channels, allowing spillovers to differ across pairs without prescribing network weights. The factor structure separates the common components of the covariates from their idiosyncratic variation. Under exogeneity conditions, this idiosyncratic variation identifies private and bilateral spillover effects after removing the common factor space. Sparsity restricts the number of relevant spillover sources for each unit while leaving their identities unknown. Together, these restrictions permit the spillover network to be distinguished from common correlated effects.

Building on this identification strategy, the estimation procedure combines factor projection with debiasing for bilateral spillovers and post-selection for common private effects. Principal component analysis estimates the factor space separately from each covariate panel. The resulting factor estimates are combined and projected out of outcomes and regressors before estimating bilateral spillovers by LASSO. Following the debiased-LASSO approach of van de Geer et al. (2014), a correction removes the leading regularization bias and yields an estimator with trackable distribution. For private effects, to control the accumulated errors from unit-specific spillover regressions, the procedure uses adaptive LASSO to select relevant spillover regressors and refits their coefficients by least squares before pooling. This post-selection approach draws on the literature on inference after high-dimensional variable selection (Belloni and Chernozhukov, 2013; Belloni et al., 2014; Hansen and Liao, 2019; Okui et al., 2025).

The theoretical analysis establishes inference while accounting jointly for estimation of the common factor space and the high-dimensional spillover network. A technical difficulty arises because factor estimates obtained from different covariates recover overlapping factor spaces. Their com-

bined Gram matrix that enters the projection matrix can therefore have a rank-deficient limit, at which the pseudo-inverse is discontinuous. This prevents direct continuous-mapping-theorem-based arguments for bounding terms involving factor-projection errors. To address this issue, I adopt the leave-one-out analysis from high-dimensional statistics literature (Abbe et al., 2022; Chen et al., 2021, among others) to bound the factor-estimation terms entering the spillover and private-effect estimators. The analysis also establishes uniform model-selection consistency of adaptive LASSO across units, providing the basis for controlling accumulated spillover-estimation errors in the pooled post-selection estimator. These results yield root-T asymptotic normality for bilateral spillovers and root-NT asymptotic normality for private effects, together with consistent variance estimators. A Gaussian approximation for the maximum estimation error across incoming links supports a simulation-based test of whether a unit receives any spillovers.

Inter-industry R&D spillovers provide a natural setting for the proposed method: multiple interaction channels complicate the specification of network weights, while common technological and macroeconomic shocks can jointly affect R&D investment and output. Using a panel of U.S. industries constructed from Compustat data, the analysis estimates an augmented production function in which output depends on an industry's own R&D stock, capital, labor, and R&D stocks in other industries. Accounting for common factors substantially changes the results. The estimated private R&D elasticity falls from 0.034 under industry fixed effects to 0.014, and is no longer statistically significant. This result is consistent with Eberhardt et al. (2013). The number of bilateral links detected at a 5% false discovery rate also falls from 203 to 118, a reduction of approximately 42%. These results suggest that common shocks confound estimates of private R&D returns and make the spillover network appear denser when left unaccounted for.

The remaining network reveals substantial differences in the sources, recipients, and signs of estimated spillovers. Of the 118 detected links, 77 are negative and 41 are positive. Industrial machinery and computer equipment, together with electronic and electrical equipment, are the largest sources. Their detected effects extend to manufacturing and nonmanufacturing industries. The largest recipients are stone, clay, glass, and concrete products, building-material retail, and textiles, highlighting exposure in relatively low-technology manufacturing and materials distribution. Receiving industries differ in their exposure: some receive predominantly negative effects, while others receive sizable effects of both signs that partly offset in the aggregate. Estimating bilateral relationships reveals how spillover effects vary across industry pairs, including which industries are the main sources and recipients and whether incoming effects reinforce or offset one another.

This paper is related to the literature that recovers spillover networks from panel data without prescribing proximity weights. Manresa (2013) proposes a pooled LASSO estimator for an unknown, time-invariant interaction structure and studies convergence rates in a simplified sparse spillover model. Okui et al. (2025) allows the linkage structure to change at an unknown breakpoint and develops inference for private effects using double machine learning. This paper addresses an additional identification problem that arises when an unknown spillover network coexists with unobserved common shocks. A factor structure in the covariates, together with exogeneity and rank conditions on their idiosyncratic components, provides restrictions that separate these sources of variation. Building on this identification argument, the paper establishes estimation and inference for bilateral spillovers and common private effects, as well as simultaneous tests of spillover reception, while accounting for errors from estimating both the factor space and the network.

The proposed methodology also relates to panel data methods for unobserved common factors. One strand accounts for common factors through common correlated effects (CCE) methods using cross-sectional averages or through principal component analysis (PCA) to estimate the factor space (Pesaran, 2006; Westerlund and Urbain, 2015; Cui et al., 2022). Another estimates interactive effects jointly with slope coefficients through least-squares or nuclear-norm methods (Bai, 2009; Moon and Weidner, 2026). These approaches provide methods for handling common correlated effects, while the present setting also requires recovering an unknown, high-dimensional spillover network. Closely related work by Higgins and Martellosio (2023) combines interactive effects with penalized estimation of spillovers represented by pre-specified network matrices. The distinction concerns what is learned about the interaction structure: selecting among given matrices retains their restrictions on bilateral relationships, whereas the proposed approach directly estimates the individual links under sparsity. A factor structure in the covariates separates common variation from the idiosyncratic variation used to identify these links, allowing the analysis to recover bilateral spillover effects alongside common private effects.

The remainder of the paper proceeds as follows. Section 2 introduces the baseline latent-network model with exogeneous spillover effects and common correlated effects. Section 3 presents the estimation procedure for this latent-network setup, and Section 4 establishes its large-sample properties and inference results. Section 5 reports simulation results to evaluate finite-sample performance. Section 6 presents empirical applications to study R&D spillover effects among industries. Section 7 concludes. Additional notations used throughout the paper are summarized in Subsection 1.1. The appendices contain proofs, technical assumptions, model extensions, and supplementary simulation and empirical results.

1.1 Notations

Throughout the paper, I use the following notations. For a vector $\mathbf{a} = (a_1, \ldots, a_n)$, $\|\mathbf{a}\|_p$ denote the ℓ_p-norm of a for $p \geq 0$. In particular, $\|\mathbf{a}\|_0 = \sum_{i=1}^n \mathbb{1}\{a_i \neq 0\}$ denotes the sparsity of $\mathbf{a}$ and $\|\mathbf{a}\|_\infty = \max_{1\leq i\leq n} |a_i|$ denotes the max-norm of $\mathbf{a}$. For a matrix $\mathbf{A} = (a_{ij})_{i=1,j=1}^{m,n} \in \mathbb{R}^{m\times n}$, $\mathbf{A} \succ 0$ indicates $\mathbf{A}$ is positive definite. $\|\mathbf{A}\|_2$ denotes the operator norm of $\mathbf{A}$, $\|\mathbf{A}\|_\infty$ denotes the max row norm: $\|\mathbf{A}\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^n |a_{ij}|$, and $\|\mathbf{A}\|_{\max} = \max_{1\leq i\leq m, 1\leq j\leq n} |a_{ij}|$ denotes the max-norm of $\mathbf{A}$. $\|\mathbf{A}\|_*$ denotes the nuclear norm of $\mathbf{A}$, i.e. the sum of the singular values of $\mathbf{A}$. I use $\mathbf{A}_{i:}$ to denote the i-th row of $\mathbf{A}$ and $\mathbf{A}_{:j}$ to denote the j-th column of $\mathbf{A}$. $\mathbf{I}_T$ denotes the $T \times T$ identity matrix. $\text{diag}\left(\{a_i\}_{i=1}^N\right)$ denotes the $N \times N$ diagonal matrix with the i-th diagonal element being a_i. Given a set S, $|S|$ denotes its cardinality. For a scalar random variable Z, $\|Z\|_{\psi_2} = \inf\{c > 0 : \mathbb{E}\left[\exp(Z^2/c^2)\right] \leq 2\}$ denotes its sub-Gaussian norm. For a random vector $\mathbf{Z} \in \mathbb{R}^d$, I define $\|\mathbf{Z}\|_{\psi_2} = \sup_{\|\mathbf{a}\|_2=1} \|\mathbf{a}'\mathbf{Z}\|_{\psi_2}$. For deterministic sequences a_{NT} and $b_{NT} > 0$, $a_{NT} = \mathcal{O}\left(b_{NT}\right)$ means that $|a_{NT}|/b_{NT}$ remains bounded as $N, T \to \infty$, while $a_{NT} = \mathbf{o}\left(b_{NT}\right)$ means that this ratio converges to zero. For a random sequence Z_{NT}, $Z_{NT} = \mathcal{O}_p\left(b_{NT}\right)$ means that Z_{NT}/b_{NT} is bounded in probability, while $Z_{NT} = \mathbf{o}_p\left(b_{NT}\right)$ means that this ratio converges to zero in probability.

2 Model

This section presents a panel data model with an unknown spillover network and unobserved common factors, explains its identification, and illustrates its economic interpretation using R&D spillovers. The private effect and coefficients on control variables are common across units, while bilateral spillover effects may differ across pairs of units. Extensions allowing for heterogeneous private effects and observed interaction channels are presented in Appendices D and A, respectively.

2.1 A spillover model with unobserved interaction channels

Consider a panel of N units observed over T periods. Unit i's outcome depends on its own covariates and on a covariate of other units that generates spillovers:

$$y_{it} = \rho x_{it} + \sum_{j\neq i}^{N} \gamma_{ij} x_{jt} + \sum_{k=1}^{K} \beta_k w_{it,k} + e_{it}, \tag{1}$$

Here, x_{it} is the spillover-generating covariate, while $w_{it,k}$ denotes a control variable whose effects

are confined to unit i. The coefficient ρ measures the private effect of x_{it}, and γ_{ij} measures the effect of unit j's covariate on unit i's outcome, holding other covariates and common shocks fixed. The parameters of primary interest are the common private effect ρ and the bilateral spillover effects $\{\gamma_{ij} : i \neq j\}$. The private effect and control coefficients are common across units; bilateral effects may differ in magnitude and sign and need not be symmetric.

To account for unobserved shocks shared across units, let

$$e_{it} = \lambda_i' f_t + \varepsilon_{it},$$

where $f_t \in \mathbb{R}^{r\times 1}$ collects latent common factors and λ_i contains unit-specific loadings. The interactive component allows the same shock to affect units differently.

The common factors may also influence the covariates. A technological development, for instance, can jointly affect firms' output and investment. To accommodate this dependence, suppose that:

$$\begin{aligned} x_{it} &= \eta_{x,1,i}' f_t + b_{x,i}' f_{x,t} + u_{x,it} \\ w_{it,k} &= \eta_{w,k,i}' f_t + b_{w,k,i}' f_{w,k,t} + u_{w,k,it}, \quad k = 1, \ldots, K. \end{aligned} \tag{2}$$

The factors f_t enter both the outcome and covariate equations. The additional factors $f_{x,t}$ and $f_{w,k,t}$ generate common variation specific to each covariate and do not enter the outcome equation directly. They may affect outcomes through the included covariates. The terms $u_{x,it}$ and $u_{w,k,it}$ are the idiosyncratic components.

The complete model is therefore given by:

$$\begin{aligned} y_{it} &= \rho x_{it} + \sum_{j\neq i}^{N} \gamma_{ij} x_{jt} + \sum_{k=1}^{K} \beta_k w_{it,k} + \lambda_i' f_t + \varepsilon_{it}, \\ x_{it} &= \eta_{x,1,i}' f_t + b_{x,i}' f_{x,t} + u_{x,it}, \\ w_{it,k} &= \eta_{w,k,i}' f_t + b_{w,k,i}' f_{w,k,t} + u_{w,k,it}, \quad k = 1, \ldots, K. \end{aligned} \tag{3}$$

To address the curse of dimensionality caused by the large number of potential spillover effects, we impose a unit-wise sparsity structure on the spillover matrix: each unit receives spillovers mainly from a limited number of other units, although their identities are unknown. This restriction captures the idea that, after accounting for common shocks and observed controls, a unit's outcome depends primarily on a limited set of other units. For example, benefiting from another firm's R&D

may require both relevant knowledge and the capacity to absorb it, limiting the set of spillover sources. The model leaves the identities of these sources unrestricted. Sparsity makes estimation feasible when the $N-1$ potential spillover regressors are numerous relative to the time dimension. The asymptotic analysis specifies the restrictions on sparsity required for estimation and inference.

The baseline specification focuses on factors shared by outcomes and covariates because these factors generate the confounding considered here. Appendix C extends the analysis to additional factors that enter only the outcome equation.

Remark 2.1 (Reduced-form interpretation)**.** *Bilateral coefficients can also summarize effects arising from simultaneous interactions. For example, Barwick et al. (2026) model students' app usage as depending on their own characteristics, roommates' characteristics, and roommates' contemporaneous usage. Allowing additionally for unobserved common shocks, consider*

$$y_t = Ay_t + Bx_t + \Lambda f_t + \varepsilon_t.$$

When A is time-invariant and $I - A$ is invertible, the reduced form is

$$y_t = \Gamma x_t + \widetilde{\Lambda} f_t + \widetilde{\varepsilon}_t, \qquad \Gamma = (I-A)^{-1}B, \quad \widetilde{\Lambda} = (I-A)^{-1}\Lambda, \quad \widetilde{\varepsilon}_t = (I-A)^{-1}\varepsilon_t.$$

The transformation preserves the interactive-effects structure: the same common factors enter with transformed loadings. The off-diagonal entries of Γ combine the effects of other units' characteristics with their propagation through endogenous interactions. When the reduced-form coefficients and errors satisfy the maintained restrictions, the proposed model identifies these combined effects without separately recovering the underlying interaction channels.[1]

Remark 2.2 (Flexibility of the interactive effects)**.** *The interactive component $\lambda_i' f_t$ accommodates several familiar forms of unobserved heterogeneity. Additive individual and time effects arise by setting $f_t = (1, d_t)'$ and $\lambda_i = (a_i, 1)'$, which gives $\lambda_i' f_t = a_i + d_t$. Time-varying group effects, as in Bonhomme and Manresa (2015), arise when f_t collects group-specific time effects and λ_i indicates unit i's group membership. Beyond these special cases, unrestricted loadings allow units to respond differently to several common shocks without requiring units to share the same exposure or belong to discrete groups. The specification therefore captures a broad range of cross-sectional dependence through a low-dimensional common component.*

[1]The estimation and inference theory assumes exact sparsity and cross-sectional independence of the idiosyncratic errors. Additional simulations in Appendix E suggest that the proposed methods also perform well under approximate sparsity and weak cross-sectional dependence in the outcome errors.

Remark 2.3 (Outcome-specific factors)**.** *The main analysis focuses on latent factors that enter both the outcome and the regressors because these shared factors generate the endogeneity problem studied in this paper. Factors that enter only the outcome equation do not create the same endogeneity channel. Appendix C shows that, under the stated conditions, allowing for such outcome-specific factors does not change the identification argument or invalidate the proposed estimation and inference procedures, although it increases their asymptotic variance. The appendix also develops a double-refinement procedure that explicitly projects out the outcome-specific factor space to improve efficiency, together with its inferential theories and simulation results.*

2.1.1 Motivating example: R&D spillover

A firm's R&D can affect other firms through several channels. Bloom et al. (2013) and Lucking et al. (2019) distinguish technology spillovers from product-market rivalry, while Arque-Castells and Spulber (2022) emphasizes that firms also obtain external knowledge through voluntary technology transfers. These channels motivate allowing bilateral effects to differ in sign and magnitude. In the proposed model, γ_{ij} captures the net effect of firm j's R&D on firm i's outcome across the underlying channels, without distinguishing uncompensated knowledge spillovers from effects transmitted through technology-market transactions.

Empirically, identifying R&D spillovers is challenging for two reasons. First, the channels through which knowledge and competitive effects propagate are only imperfectly observed and may differ across firms. Measures of technological proximity or product market overlap are potential proxies for the true interaction network. Second, firms are simultaneously exposed to common shocks—such as technological waves, aggregate demand fluctuations, or financial conditions—that affect both their R&D decisions and their outcomes, and their effects vary across firms. As a result, standard panel regressions, including two-way fixed effects and simple pre-specified networks, may confound true spillovers with correlated responses to unobserved common factors.

The framework developed in this paper provides a unified approach to address these issues. In the latent-network specification, y_{it} in Model (3) corresponds to firm i's measure of performance (e.g., productivity or market value), x_{it} corresponds to its R&D stock, and the spillover coefficients γ_{ij} capture the aggregate effect of firm j's R&D on firm i's outcome. The factor f_t corresponds to the collection of unobserved common shocks that affect both the firm performance and firms' R&D stocks. $f_{x,t}$ corresponds to shocks that affect the performance of firms only through the R&D investment decision, such as R&D tax policy and subsidies, university outputs in research, or visa availablity to foreign researchers. $w_{it,k}$ includes additional firm controls such as employment and

capital intensity and $f_{w,k,t}$ represnts the control-specific shocks, such as labor market and financing conditions. Sparsity of the spillover matrix reflects the economic idea that each firm is directly affected by only a limited set of technologically, competitively or geographically related peers.

2.2 Identification of spillover parameters

Identifying spillover effects requires separating the effects of other units' covariates from responses to unobserved common shocks. Without additional restrictions, changes in the spillover coefficients can be offset by changes in the factor loadings, leaving observed outcomes unchanged. We first illustrate the resulting observational equivalence and then explain how the maintained covariate structure supplies identifying variation.

Suppressing controls, write the outcome equation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Gamma}' + \mathbf{F}\boldsymbol{\Lambda}' + \boldsymbol{\varepsilon}, \qquad \gamma_{ii} = \rho.$$

Suppose that unit j's covariate lies in the common factor space: $\mathbf{X}_{:j} = \mathbf{F}a$ for some $a \in \mathbb{R}^r$. For any $c \in \mathbb{R}^N$ with $c_j = 0$, define

$$\widetilde{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma} - ce_j', \qquad \widetilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda} + ca',$$

where e_j is the j-th standard basis vector. These parameters yield the same outcome equation:

$$\mathbf{X}\widetilde{\boldsymbol{\Gamma}}' + \mathbf{F}\widetilde{\boldsymbol{\Lambda}}' = \mathbf{X}\boldsymbol{\Gamma}' + \mathbf{F}\boldsymbol{\Lambda}'.$$

The restriction $c_j = 0$ preserves the common diagonal coefficient ρ. Moreover, the transformation changes at most one spillover coefficient per row, so both representations can satisfy the maintained sparsity restriction. The ambiguity therefore concerns the spillover coefficients themselves and can arise even when private effects are homogeneous and the network is sparse.

More generally, the same ambiguity can arise when a linear combination of fewer than N columns of $\mathbf{X}$ lies in the common factor space. Such combinations exist with probability one when $N > (T-r)$ and $\mathbf{X}$ has row rank greater than $T-r$: there exist at least one linear combination of the common-factor time series that can be represented as some linear combination of at most $(T-r) < N$ columns of $\mathbf{X}$. Variation attributed to spillovers can then be reassigned to the interactive effects without changing observed outcomes. This example establishes the need for restrictions that

distinguish these two sources of variation.

The factor structure imposed on the covariates provides a way to achieve this separation. Let

$$g_t = (f_t', f_{x,t}', f_{w,1,t}', \ldots, f_{w,K,t}')'$$

collect the factors in the covariate equations, which contains the factors that affect the outcome, directly or indirectly. Once the space is identified from the covariate panels, projecting onto its orthogonal complement removes the common components of both the outcome and the regressors.

Let Π_g denote population linear projection onto the span of g_t, and define

$$\widetilde{y}_{it} = y_{it} - \Pi_g y_{it}.$$

When the idiosyncratic covariate components and the outcome errors are orthogonal to g_t, the projected equation is

$$\widetilde{y}_{it} = \rho u_{x,it} + \sum_{j \neq i} \gamma_{ij} u_{x,jt} + \sum_{k=1}^{K} \beta_k u_{w,k,it} + \varepsilon_{it}.$$

Thus, coefficients are identified from covariate variation that remains after removing the common factors.

Proposition 2.1 (Sufficient conditions for identification)**.** *Suppose Model* (3) *holds and:*

1. *The population projection Π_g is identified from the covariate process.*
2. *The idiosyncratic covariate components and ε_{it} are orthogonal to g_t.*
3. *For every i, the vector*

$$u_{it} = (u_{x,1t}, \ldots, u_{x,Nt}, u_{w,1,it}, \ldots, u_{w,K,it})'$$

satisfies

$$\mathbb{E}\left[u_{it}\varepsilon_{it}\right] = 0, \qquad \mathbb{E}\left[u_{it}u_{it}'\right] \succ 0.$$

Then $\mathbf{\Gamma}$ and $\beta = (\beta_1, \ldots, \beta_K)'$, including the common private effect ρ, are uniquely identified.

The result follows directly from the population moment conditions of the projected equation: exogeneity and nonsingularity of $\mathbb{E}\left[u_{it}u_{it}'\right]$ uniquely determine the coefficients for each unit. The proof is therefore omitted.

Proposition 2.1 shows that private and bilateral spillover effects are identified from the idiosyncratic variation remaining after removal of the common components. Orthogonality to the factors preserves this variation under projection, while exogeneity and a nonsingular second-moment matrix distinguish the effects of different regressors.

Remark 2.4 (Time fixed effects and network identification)**.** *Similar observational equivalences arise even with simple additive time fixed effects. In a specification allowing heterogeneous private effects, as in Manresa (2013), write*

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\Gamma}' + \boldsymbol{\delta}\mathbf{1}' + \boldsymbol{\varepsilon}.$$

For any $a \in \mathbb{R}^N$, the transformation

$$\widetilde{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma} + \mathbf{1}a', \qquad \widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta} - \mathbf{X}a$$

leaves fitted outcomes unchanged. Choosing a with sparse entries allows this ambiguity to arise within a class of sparse networks, provided both coefficient matrices are admissible. Heterogeneous private effects thus allow observational equivalence without requiring combination of less than N columns of $\mathbf{X}$ lying in the common factor space. This observation complements Manresa's discussion of time effects absorbing aggregate spillovers: adding time fixed effects does not by itself separate bilateral spillovers from aggregate effects.

3 Estimation

This section presents a three-step estimation procedure for the model in Section 2:

1. Estimate the factor space from the covariates and project it out of outcomes and regressors.
2. Estimate bilateral spillover effects from the projected equations using LASSO, with a debiasing step for inference.
3. Estimate the common private effect and coefficients on control variables after accounting for spillovers using a post-selection procedure.

Implementing these steps requires accounting for both the estimation of the factor space and the regularization of spillover coefficients. Section 3.1 develops the procedure, explaining how factor projection, debiasing, and post-selection address these challenges. Section 3.2 then considers alter-

native estimation methods for this model and explains the advantages of the proposed procedure.

3.1 Estimation procedure

The factor structure of the covariates provides a starting point for estimating the model: it allows common-shock variation to be isolated before estimating spillover and private effects. I begin by removing the common factor components from outcomes and regressors. If the factor space were observed, this would amount to projecting each variable onto its orthogonal complement. In practice, I estimate this space by applying principal component analysis (PCA) separately to each covariate matrix. Let $\mathbf{X}$ and $\mathbf{W}_k$ denote the $T \times N$ matrices collecting observations of x_{it} and $w_{it,k}$, respectively. The following algorithm describes the factor-estimation step.

Algorithm 1 (PCA on each covariate)**.**

1. *Initialize* $\widehat{\mathbf{G}} = \emptyset$.
2. *For each* $\mathbf{Q} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K\}$*:*
   (a) *Let* $\widehat{\mathbf{G}}_{\mathbf{Q}}$ *be* $\sqrt{T}$ *times the matrix of eigenvectors corresponding to the* $r_{\mathbf{Q}}$ *largest eigenvalues of* $\mathbf{Q}\mathbf{Q}'/(NT)$.
   (b) *Update* $\widehat{\mathbf{G}} \leftarrow [\,\widehat{\mathbf{G}}, \widehat{\mathbf{G}}_{\mathbf{Q}}\,]$.

Using the estimated factors, define the residual maker

$$\mathbf{M}_{\widehat{\mathbf{G}}} = \mathbf{I}_T - \widehat{\mathbf{G}}(\widehat{\mathbf{G}}'\widehat{\mathbf{G}})^{-}\widehat{\mathbf{G}}',$$

where the superscript $-$ denotes the Moore–Penrose inverse. Premultiplying outcomes and regressors by $\mathbf{M}_{\widehat{\mathbf{G}}}$ removes their components in the estimated factor space.

Remark 3.1 (Extracting factors from multiple covariates)**.** *Algorithm 1 estimates the factor space separately from each covariate matrix. Alternative PCA-based approaches include applying PCA to the sum of the covariate matrices or to the concatenated matrix* $[\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K]$*. Although all three approaches exploit the factor structure in the covariates, the proposed procedure permits valid inference under weaker restrictions on the relative growth of* N *and* T *under the assumptions in Section 4.1. This advantage comes with a technical complication: factor estimates extracted from different covariates share common factors* $\mathbf{F}$ *asymptotically, so their concatenation becomes asymptotically rank deficient. Therefore, naively decomposing the projection matrix of the concatenated*

factor matrix can lead to a diverging rate due to the pseudo-inverse on rank-degenerate matrices. Appendix I explains how my analysis addresses this issue.

Separate factor extraction also facilitates empirical interpretation. Comparing the estimated factor spaces across covariates provides evidence on the common variation they share. Section 6 illustrates this comparison in the R&D application.

The second step estimates bilateral spillover effects from these projected variables. For each unit, there are $N-1$ potential spillover regressors, which may exceed the number of time observations. I use LASSO to exploit the sparsity of the spillover network. However, its kinked ℓ_1 penalty introduces regularization bias and complicates the analysis of the asymptotic distribution. I therefore apply a debiasing correction to obtain asymptotically normal spillover estimators under the conditions established in Section 4, following the debiased LASSO approach of van de Geer et al. (2014) and Zhang and Zhang (2014).

For unit i, let $\boldsymbol{\gamma}_{i:} = (\gamma_{i1}, \ldots, \gamma_{i,i-1}, \gamma_{i,i+1}, \ldots, \gamma_{iN})'$ collect its $N-1$ bilateral spillover coefficients. Let $\mathbb{Z}_i = [\mathbf{X}_i, \mathbf{W}_{1,i}, \ldots, \mathbf{W}_{K,i}]$ be the $T \times (1+K)$ matrix of own covariates, and let $\mathbf{X}_{-i}$ be the $T \times (N-1)$ matrix obtained by removing column i from $\mathbf{X}$. Before estimating spillovers, I partial out the projected own covariates from the projected outcome and spillover regressors. The following algorithm combines this step with LASSO estimation and debiasing.

Algorithm 2 (Interaction network estimations: Factor-projection + Debiased LASSO)**.**

1. *(Factor Projection) Estimate* $\mathbf{G}$ *by Algorithm 1 and compute the projection matrix* $\mathbf{M}_{\widehat{\mathbf{G}}}$.
2. *(LASSO) For each* i*, given tuning parameter* υ_i*, the spillover effects* $\boldsymbol{\gamma}_{i:}$ *can be estimated by solving the following LASSO problem:*

$$\widehat{\boldsymbol{\gamma}}_{i:} = \underset{\boldsymbol{\gamma}_{i:}}{\operatorname{argmin}} \frac{1}{2T} \left\| \mathbf{M}_{\widehat{\mathbb{Z}},i} \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{Y}_i - \mathbf{M}_{\widehat{\mathbb{Z}},i} \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{X}_{-i} \boldsymbol{\gamma}_{i:} \right\|_2^2 + \upsilon_i \|\boldsymbol{\gamma}_{i:}\|_1, \tag{4}$$

   where $\mathbb{Z}_i = [\mathbf{X}_i, \mathbf{W}_{1,i}, \ldots, \mathbf{W}_{K,i}]$ *and* $\mathbf{M}_{\widehat{\mathbb{Z}},i} = \mathbf{I}_T - \mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{Z}_i (\mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{Z}_i)^{-1} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}}$ *is the residual maker of* $\mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{Z}_i$.

3. *(Debiased LASSO) Let* $\widehat{\boldsymbol{\Theta}} = \operatorname{diag}\left(\left\{ \left(\mathbf{X}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{X}_i / T \right)^{-1} \right\}_{i=1}^{N} \right)$. *For each* i*, debias the LASSO estimate* $\widehat{\boldsymbol{\gamma}}_{i:}$ *by*

$$\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} = \widehat{\boldsymbol{\gamma}}_{i:} + \frac{1}{T} \widehat{\boldsymbol{\Theta}}_{-i} \mathbf{X}_{-i}' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{M}_{\widehat{\mathbb{Z}},i} \mathbf{M}_{\widehat{\mathbf{G}}} \left(\mathbf{Y}_i - \mathbf{X}_{-i} \widehat{\boldsymbol{\gamma}}_{i:} \right), \tag{5}$$

where $\widehat{\boldsymbol{\Theta}}_{-i}$ is the submatrix of $\widehat{\boldsymbol{\Theta}}$ that drops the i-th row and column.

$\widehat{\boldsymbol{\Theta}}$ is an estimator of the precision matrix of $\mathbf{M}_{\mathbf{G}}\mathbf{X}$, i.e. $\boldsymbol{\Theta} = \left(\mathbb{E}\left[\mathbf{X}'\mathbf{M}_{\mathbf{G}}\mathbf{X}/T\right]\right)^{-1}$, and the debiasing step corrects for the bias induced by the ℓ_1-penalty in the LASSO estimator. In Section 4, I show that, under suitable conditions, the debiased LASSO estimator $\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}$ is asymptotically normal and can be used for inference on spillover effects.

In practice, the tuning parameter υ_i can be selected by cross-validation, BIC, or other data-driven methods in the literature.

The estimation and inference of private effects requires additional care. While the spillover coefficients $\boldsymbol{\gamma}_{i:}$ remain high-dimensional and heterogeneous across i by construction, the private effects are common across units so that the estimation errors from LASSO estimators can severely contaminate their estimation. For the purpose of valid inference, I propose the following post-selection-based estimator that adopts ideas from post-selection inference (e.g., Belloni and Chernozhukov, 2013; Belloni et al., 2014; Hansen and Liao, 2019; Okui et al., 2025).

Algorithm 3 (Private effects: post-selection estimator)**.**

1. *Implement Step 1 and 2 of Algorithm 2 to obtain the residual maker $\mathbf{M}_{\widehat{\mathbf{G}}}$ and the LASSO estimator $\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}$ for each i.*
2. *For each i, implement the following steps to obtain $\left\{\widehat{\gamma}_{ij}^{(ps)}, j \neq i\right\}$:*
   (a) *Given $\upsilon_i^{(al)}$, run the following adaptive LASSO regression to select the non-zero γ_{ij}:*

$$\left\{\widehat{\gamma}_{ij}^{(al)}\right\}_{j:\widehat{\gamma}_{ij}\neq 0} = \underset{\{\gamma_{ij}\}_{j:\widehat{\gamma}_{ij}\neq 0}}{\operatorname{argmin}} \frac{1}{2T}\left\|\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{Y}_i - \sum_{j:\widehat{\gamma}_{ij}\neq 0}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_j\cdot\gamma_{ij}\right\|_2^2 + \upsilon_i^{(al)}\sum_{j:\widehat{\gamma}_{ij}\neq 0}\frac{|\gamma_{ij}|}{|\widehat{\gamma}_{ij}^{(db)}|}. \tag{6}$$

   Let $\widehat{S}_i = \left\{j : \widehat{\gamma}_{ij}^{(al)} \neq 0\right\}$ be the selected support set.
   (b) *Run the following post-selection OLS regression to obtain the final estimator of γ_{ij}:*

$$\left\{\widehat{\gamma}_{ij}^{(ps)}\right\}_{j\in\widehat{S}_i} = \underset{\{\gamma_{ij}\}_{j\in\widehat{S}_i}}{\operatorname{argmin}} \frac{1}{T}\left\|\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{Y}_i - \sum_{j\in\widehat{S}_i}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_j\cdot\gamma_{ij}\right\|_2^2,$$

   and $\widehat{\gamma}_{ij}^{(ps)} = 0$ for $j \notin \widehat{S}_i$.

3. Let $\theta = [\rho, \beta_1, \ldots, \beta_K]'$. The estimator of θ is given by

$$\widehat{\theta} = \left(\sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \left(\mathbf{Y}_i - \sum_{j \neq i} \mathbf{X}_j \cdot \widehat{\gamma}_{ij}^{(ps)} \right). \tag{7}$$

The post-selection procedure reduces the regularization bias that spillover estimation introduces into the estimator of θ. Without this step, inference on θ requires the rate condition $N/T \to 0$. Section 4 establishes valid inference under milder rate conditions using post-selection. This result concerns the common private effects; inference on individual spillover coefficients instead relies on the debiased LASSO estimator.

Remark 3.2. *As discussed in Remark 2.2, the interactive fixed-effects structure nests the standard two-way fixed-effects model. When two-way fixed effects are present, estimation can be made more efficient by removing them prior to implementing the factor-projection estimator (Bai, 2009). In particular, when spillover channels are observed, individual and time fixed effects can be eliminated by residualizing the data with respect to unit and time dummies before applying the factor-projection step.*

In contrast, when the spillover network is unobserved, demeaning across individuals is no longer innocuous, as it does not commute with the high-dimensional spillover components. In this setting, only time demeaning remains valid, while time fixed effects can alternatively be absorbed by the interactive factor structure and removed through the projection step. See Appendix B for further discussion.

For clarity, the analysis in this paper focuses on the factor-projection estimator without prior one- or two-way demeaning. Extending the results to incorporate demeaning is conceptually straightforward but technically involved, and is therefore left for future work.

3.2 Comparison with alternative methods

In this section, I briefly discuss the differences and connections between the proposed estimation methods and alternative methods in the literature.

Instrumental variable (IV) approach: IV regression is widely used to address endogeneity in spillover regressions and, in principle, can deal with unobserved common shocks if valid instruments are available. In practice, however, finding instruments that are both relevant and exogenous is often difficult, and the exogeneity condition is typically untestable.

Moreover, even with valid instruments, IV estimators may be inefficient in panel settings with strong cross-sectional dependence. When both the endogenous regressor and the instrument load on common factors, the effective variation in the instrument could be driven by time-series rather than cross-sectional fluctuations, so that the IV estimator may converge at the slower root-T rate instead of the optimal root-NT rate in homogeneous panels. By explicitly modeling and removing the common factors, the proposed projection-based estimators exploit both the time-series and cross-sectional information and can attain root-NT efficiency under suitable conditions. See Appendix B for detailed derivations.

Common correlated effects (CCE) estimators: The CCE estimator proposed by Pesaran (2006) is a widely used approach for addressing common correlated effects in panel data models due to its simplicity and ease of implementation. It has been extensively studied in the econometrics literature (e.g., Chudik et al., 2011; Chudik and Pesaran, 2015; Westerlund and Urbain, 2015; Westerlund, 2018). Conceptually, the CCE approach follows the same principle as the proposed estimators in this paper: it seeks to control for latent common factors to eliminate cross-sectional dependence. However, instead of estimating the factor space via principal component analysis, the CCE method uses cross-sectional averages of the dependent and independent variables as proxies for the latent factors.

Theoretical comparisons between PCA-based approaches and CCE estimators have been developed in Westerlund and Urbain (2015), which shows that both methods require similar rate conditions for consistent estimation and valid inference in pooled OLS settings. Nevertheless, several differences are relevant in the present framework. First, inference results for the CCE estimator are typically established under the condition that the number of regressors exceeds the number of factors, which may be restrictive in models with limited covariates. Second, the CCE approach relies on the informativeness of cross-sectional averages for recovering the factor space. This requires the average loading matrix of the dependent and independent variables to be non-singular. Such a condition may fail when factor loadings have similar cross-sectional means across variables, in which case the cross-sectional averages provide little variation for identifying the common factors. Moreover, Westerlund and Urbain (2015) shows that the CCE estimator can be sensitive to heterogeneity in factor loadings: its bias increases with the variance of loadings.

Given these considerations, I adopt PCA-based approaches to estimate the factor space, which do not rely on cross-sectional averages and are more robust to heterogeneity in factor loadings.

Nuclear norm regularization and iterative approach: Bai (2009) studies panel models with

interactive fixed effects and proposes an iterative procedure to jointly estimate slope coefficients and latent factors, while Moon and Weidner (2026) introduces nuclear-norm regularization as a convex relaxation to obtain a reliable initialization and improve numerical stability. These approaches are attractive for their robustness and flexibility on dependency between factors and regressors, but they are computationally intensive, as they require repeated singular-value decompositions in large panels. In the proposed model, each iteration would additionally require re-estimating a high-dimensional spillover regression for every unit, further increasing the computational burden.

More importantly for the present context, as discussed in Section 2.2, leaving the dependency between factors and regressors unrestricted creates an identification issue for spillover effects γ_{ij}. By explicitly modeling the factor structure in the covariate that generate spillovers, the proposed approach separates these two sources of dependence and delivers identifiability and computationally tractable estimators, when the interaction network is latent.

4 Theory

This section establishes the large-sample properties of the estimators as N and T tend to infinity. Section 4.1 states the assumptions for the theoretical analysis. Under these conditions, Section 4.2 develops inference for bilateral spillover effects using the debiased LASSO estimator, including simultaneous tests of whether a unit receives any spillovers. Section 4.3 then establishes asymptotic normality and a consistent covariance estimator for the common private-effect and control coefficients, accounting for estimation of the spillover network through post-selection.

4.1 Assumptions

The following assumptions are maintained throughout Section 4. For each t, let $\boldsymbol{\varepsilon}_t$, $\mathbf{U}_{x,t}$ and $\mathbf{U}_{w,k,t}$ denote the $N \times 1$ vectors stacking ε_{it}, $u_{x,it}$ and $u_{w,k,it}$, respectively. Let $\mathbb{U}_t \in \mathbb{R}^{N\times(K+1)}$ stacks $\mathbf{U}_{x,t}$ and $\mathbf{U}_{w,k,t}$ for $k = 1, \ldots, K$, and $\mathbb{B}$ stacks loadings on $\mathbf{X}_t$ and $\mathbf{W}_{k,t}$ for $k = 1, \ldots, K$. Finally, define the support $S_i = \{j : \gamma_{ij} \neq 0\}$ and let $s_i = |S_i|$ denote the number of nonzero spillover coefficients for unit i.

Assumption 1. *(Common factors) The factor process $\mathbf{G}_t$ satisfies:*

- *(Sub-Gaussian) There exists a constant $C > 0$ such that with $\|\mathbf{G}_t - E(\mathbf{G}_t)\|_{\psi_2} < C$.*
- *(Second moment) There exists a positive definite matrix Σ_G such that $\mathbb{E}[\mathbf{G}_t\mathbf{G}_t'] = \Sigma_{\mathbf{G}}$ for*

some constant $c > 0$.

Assumption 2. *Factor loadings* $\mathbf{\Lambda}$, $\mathbf{B}$ *and* $\mathbb{B}$ *satisfy:*

- *(Sub-Gaussian) There exists a constant* $C > 0$ *such that with*

$$\left\|\widetilde{\mathbf{\Lambda}}_{:j}\right\|_{\psi_2} \quad \text{and} \quad \left\|\widetilde{\mathbb{B}}^{(x)}_{:j}\right\|_{\psi_2} < C$$

  for all j, *where* $\widetilde{\mathbf{\Lambda}}_{:j} = \mathbf{\Lambda}_{:j} - \mathbb{E}\left[\mathbf{\Lambda}_{:j}\right]$ *and similarly for* $\widetilde{\mathbb{B}}_{:j}$.

- *(Second moment) There exists a positive definite matrix* Σ_λ *such that* $\mathbb{E}\left[\mathbf{\Lambda}_{:i}\mathbf{\Lambda}'_{:i}\right] = \Sigma_\lambda$ *with bounded eigenvalues. Similar conditions hold for* $\mathbf{B}_{:i}$.

Assumption 3. *(Idiosyncratic errors) The idiosyncratic errors* $\mathbb{U}$ *and* $\boldsymbol{\varepsilon}$ *satisfy:*

- *(Sub-Gaussian) For all* t, *there exists a constant* $C > 0$ *such that with*

$$\max\{\|\boldsymbol{\varepsilon}_t\|_{\psi_2}, \|\mathbf{U}_{x,t}\|_{\psi_2}, \|\mathbf{U}_{w,1,t}\|_{\psi_2}, \ldots, \|\mathbf{U}_{w,K,t}\|_{\psi_2}\} < C.$$

- *(Cross-sectional Independence)* $\{\mathbb{U}_i, \boldsymbol{\varepsilon}_i\}$ *are i.i.d. across* i.
- *(Exogeneity)* $\boldsymbol{\varepsilon}$ *is independent of* $\{\mathbf{G}, \mathbf{\Lambda}, \mathbb{B}, \mathbb{U}\}$ *and* $\mathbb{U}$ *is independent of* $\{\mathbf{G}, \mathbf{\Lambda}, \mathbb{B}\}$.
- *(No Collinearity) There exists a positive definite matrix* $\Sigma_\mathbb{U}$ *such that* $\mathbb{E}\left[\mathbb{U}_{it}\mathbb{U}'_{it}\right] = \Sigma_\mathbb{U}$ *with bounded eigenvalues.*
- *(M.D.S.)* $\mathbb{E}\left[\boldsymbol{\varepsilon}_t | \boldsymbol{\varepsilon}_{t-1}\right] = 0$.

Assumption 4. *(Weak serial dependence)* $(\mathbf{G}_t, \boldsymbol{\varepsilon}_t, \mathbf{U}_{x,t}, \mathbf{U}_{w,1,t}, \ldots, \mathbf{U}_{w,K,t})$ *is stationary and* α*-mixing such that the mixing coefficient satisfies*

$$\alpha(m) \leq \exp\left(-c_0 m^\nu\right),$$

for some constant $c_0 > 0$ *and* $\nu > 0$.

Assumption 5. *(Rate conditions)*

- *As* $N, T \to \infty$, $\frac{\sqrt{N}\log^\delta N}{T} + \frac{\sqrt{T}\log^\delta N}{N} \to 0$ *for some* $\delta > \frac{3}{2}$.
- *The number of factors* r_G *is fixed and* $r_G < \min\{N, T\}$. *Similar conditions hold for* K.
- $s_i = \mathcal{O}(1)$ *for each* i *and* $\min_i \min_{j \in S_i} |\gamma_{ij}| \geq c_0$ *for some constant* $c_0 > 0$.

These assumptions are standard in the literature of factor models, high-dimensional regressions and interactive fixed effects. Sub-Gaussian and α-mixing assumptions in Assumption 1 to Assumption 4 provide convenient tail and weak serial dependence conditions; related conditions appear in, among others, Fan et al. (2011, 2013); Huang et al. (2022); Gao and Tsay (2024). Assumption 2 implies that factor loadings are random variables, which is a common setup in interactive fixed effects literature (Pesaran, 2006; Rücker et al., 2025; Lu and Su, 2025).[2] Assumption 2 allows the loadings on the same set of factors $\mathbf{F}$, $\left(\boldsymbol{\Lambda}, \boldsymbol{\eta}_x, \boldsymbol{\eta}_{w,k}\right)$, are correlated, and allows loadings to be cross-sectionally correlated, which relaxes the assumption of cross-sectional independence in Pesaran (2006). In addition, Assumption 2 imposes strong factor structures on factor models, which can be relaxed to allow for weak factors, see Huang et al. (2022); Gao and Tsay (2024); Bai and Ng (2023), at the cost of stronger dimensionality restrictions that depend on the factor strength.

Assumption 3 imposes cross-sectional independence on the idiosyncratic errors in the factor and outcome equations, a common condition in the CCE literature (Pesaran, 2006; Rücker et al., 2025; Lu and Su, 2025). Together with the remaining assumptions, this condition makes the contributions of factor-estimation and spillover-estimation errors asymptotically negligible in the pooled estimator. No analytical bias correction is therefore required. Cross-sectional independence is also essential to the leave-one-out argument developed in Appendix I. As mentioned in Remark 3.1, Algorithm 1 estimates shared factors separately across covariates, making the combined factor matrix $\widehat{\mathbf{G}}$ asymptotically rank deficient. The pseudoinverse of $\widehat{\mathbf{G}}'\widehat{\mathbf{G}}/T$ is not continuous at the limit when its rank changes, so direct expansions may yield divergent error bounds (Karabiyik et al., 2017). The leave-one-out argument avoids this problem by exploiting independence between a unit's idiosyncratic errors and factor estimates computed without that unit. Appendix I develops the leave-one-out argument, while Remark 4.1 outlines possible modifications to the estimation procedures to accommodate cross-sectional dependence.

Assumption 5 restricts the relative growth of the panel dimensions, the number of factors, and network sparsity. Ignoring logarithmic factors, it requires $\sqrt{T}/N \to 0$, rather than the stronger condition $T/N \to 0$ commonly imposed for inference in CCE and PCA-based estimation of panel models with interactive fixed effects (Pesaran, 2006; Rücker et al., 2025; Lu and Su, 2025, among others). Projecting the interactive fixed effects onto the complement of the PCA-estimated factor space introduces two bias terms in the normalized pooled estimator. Each is of order $\sqrt{T/N}$, so

[2]The random sub-Gaussian loading assumption is used only in the leave-one-out analysis in Appendix I. The remaining arguments extend readily to nonrandom loadings under suitable factor-strength conditions. Additional simulations yield similar finite-sample results with nonrandom loadings.

the stronger condition makes both asymptotically negligible. However, under Algorithm 1 and the cross-sectional i.i.d. condition in Assumption 3, the means of these terms cancel, leaving a mean-zero leading term and an asymptotically negligible remainder. The weaker rate condition is therefore sufficient. This cancellation does not generally occur when factors are extracted from the sum or concatenation of the covariate matrices. See Equation (33) and (34) in the proof of Theorem 4.2 in Appendix G for details. The bounded-sparsity condition $s_i = \mathcal{O}(1)$ is imposed to simplify the analysis. It could be relaxed to allow s_i to grow slowly with T, subject to additional rate restrictions.

4.2 Spillover effect estimators

In this section, I show that the debiased LASSO estimator for the spillover effects $\boldsymbol{\gamma}_{i:}$ defined in (5) is asymptotically normal and can be used for inference.

To begin with, the debiased LASSO estimator $\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}$ can be decomposed as

$$\sqrt{T}\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right) = \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\Theta}}_{-i}\mathbf{X}_{-i}'\widetilde{\mathbf{M}}\boldsymbol{\varepsilon}_i + \frac{1}{\sqrt{T}}\widehat{\boldsymbol{\Theta}}_{-i}\mathbf{X}_{-i}'\widetilde{\mathbf{M}}\mathbf{F}\boldsymbol{\lambda}_i' - \sqrt{T}\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}} - \mathbf{I}_{N-1}\right)\left(\widehat{\boldsymbol{\gamma}}_{-i} - \boldsymbol{\gamma}_{-i}\right), \tag{8}$$

where $\widetilde{\mathbf{M}} = \mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}}_i}\mathbf{M}_{\widehat{\mathbf{G}}}$ and $\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}} = \frac{1}{T}\mathbf{X}_{-i}'\widetilde{\mathbf{M}}\mathbf{X}_{-i}$.

The decomposition separates the leading stochastic term from errors due to factor estimation and regularization. Controlling these errors requires convergence rates for both the precision matrix estimator and the initial LASSO estimator. I first establish the rate for the precision matrix estimator, which affects the debiasing correction.

Define $\boldsymbol{\Theta} = \left(\mathbb{E}\left[\mathbf{X}'\mathbf{M}_{\mathbf{G}}\mathbf{X}/T\right]\right)^{-1} \in \mathbb{R}^{N\times N}$ the precision matrix of the spillover covariate $\mathbf{X}_t$ after projecting out the common factors $\mathbf{G}_t$.

Proposition 4.1. *Under Assumption 1 to 5, the precision matrix estimator $\widehat{\boldsymbol{\Theta}}$ defined in Step 3 of Algorithm 2 satisfies*

$$\left\|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N}\right).$$

The remaining regularization error also depends on the accuracy of the initial LASSO estimator. The next proposition establishes the consistency and its convergence rate.

Proposition 4.2. *Under Assumption 1 to 5, suppose that $\upsilon_i \geq C\left(\sqrt{\frac{\log(N)}{T}} + \frac{1}{N}\right)$ for some $C > 0$,*

the initial LASSO estimator $\widehat{\boldsymbol{\gamma}}_{i:}$ defined in Step 2 of Algorithm 2 satisfies

$$\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\|_2 = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N}\right), \qquad \textit{for all } i = 1, \ldots, N.$$

Together with the factor-estimation bounds in Appendix H, these convergence rates make the remainder terms negligible at the scales required for individual and simultaneous inference. The debiased estimator therefore admits an asymptotic representation driven by the idiosyncratic regressors and outcome errors. The next proposition establishes this representation, together with asymptotic normality for linear combinations and Gaussian approximation for the maximum norm.

Proposition 4.3. *Under Assumption 1 to 5, suppose that $\upsilon_i \geq C\left(\sqrt{\frac{\log(N)}{T}} + \frac{1}{N}\right)$ for some $C > 0$, we have*

$$\sqrt{T}\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right) = \frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}'\boldsymbol{\varepsilon}_i + r_i, \qquad \textit{for all } i = 1, \ldots, N,$$

where r_i is asymptotically negligible. And we have

$$\sqrt{T}d'\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right) \xrightarrow{d} N(0, d'\boldsymbol{\Theta}_{-i}d\sigma_{\varepsilon,i}^2), \tag{9}$$

$$\sup_{a\geq 0}\left|\mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right) - \mathbb{P}\left(\|\boldsymbol{\zeta}_i\|_\infty \leq a\right)\right| \to 0, \tag{10}$$

where $\|d\|_1 = \mathcal{O}(1)$ and $\boldsymbol{\zeta}_i \sim N(0, \sigma_{\varepsilon,i}^2\boldsymbol{\Theta}_{-i})$.

Inference depends on the unknown covariance matrix $\sigma_{\varepsilon,i}^2\Theta_{-i}$. Proposition 4.1 establishes the convergence rate of its precision matrix component. It remains to estimate the disturbance variance $\sigma_{\varepsilon,i}^2$. Define $\widehat{\sigma}_{\varepsilon,i}^2 = \frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i$ where $\widehat{\boldsymbol{\varepsilon}}_i = \mathbf{M}_{\widehat{\mathbf{G}}}\left(\mathbf{Y}_i - \mathbf{X}_{-i}\widehat{\boldsymbol{\gamma}}_{i:} - \mathbb{Z}_i\widehat{\theta}\right)$. The following proposition establishes the convergence rate of this residual-based estimator.

Proposition 4.4. *Under the same assumptions in Proposition 4.3, for $i = 1, \ldots, N$, we have*

$$\widehat{\sigma}_{\varepsilon,i}^2 - \sigma_{\varepsilon,i}^2 = \mathcal{O}_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right) = \mathbf{o}_p(1).$$

The proposition 4.1 and 4.4 justify replacing the unknown covariance components with their estimates in both distributional approximations. Theorem 4.1 therefore delivers feasible inference for linear combinations of spillover effects and for their maximum norm. Let $\widehat{\mathcal{I}}_i = \left\|\widehat{\sigma}_{\varepsilon,i}\widehat{\boldsymbol{\Theta}}_{-i}^{1/2}\boldsymbol{\xi}\right\|_\infty$ where $\boldsymbol{\xi} \sim N(0, \mathbf{I}_{N-1})$.

Theorem 4.1. *Under the same assumptions as in Proposition 4.3, and assuming that the initial LASSO estimates are consistent, then for any $i = 1, \ldots, N$ and any vector $d \in \mathbb{R}^N$ such that $\|d\|_1 = \mathcal{O}(1)$, $\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}$ obtained from Algorithm 2 satisfies*

$$\sqrt{T} d' \left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:} \right) / \sqrt{d' \widehat{\Theta}_{-i} d \widehat{\sigma}_{\varepsilon,i}^2} \xrightarrow{d} N(0,1), \tag{11}$$

Let $\widehat{\mathcal{I}}_i = \left\| \widehat{\sigma}_{\varepsilon,i} \widehat{\Theta}_{-i}^{1/2} \boldsymbol{\xi} \right\|_\infty$ where $\boldsymbol{\xi} \sim N(0, \mathbf{I}_{N-1})$, we have

$$\sup_{a \geq 0} \left| \mathbb{P}\left(\sqrt{T} \left\| \widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:} \right\|_\infty \leq a \right) - \mathbb{P}\left(\widehat{\mathcal{I}}_i \leq a \right) \right| \to 0. \tag{12}$$

Theorem 4.1 enables us to conduct inference on individual spillover effects γ_{ij} by setting d as a vector with 1 at the j^{th} entry and 0 elsewhere. The Gaussian approximation result enables simultaneous inference on $\boldsymbol{\gamma}_{i:}$. For instance, one could test whether unit i receives any spillovers by

$$H_0 : \|\boldsymbol{\gamma}_{i:}\|_\infty = 0 \quad \text{vs} \quad H_1 : \|\boldsymbol{\gamma}_{i:}\|_\infty > 0.$$

Critical values for $\|\widehat{\mathcal{I}}_i\|_\infty$ could be obtained by a Gaussian simulation procedure (Chernozhukov et al., 2017, 2023; Fan et al., 2024). Given test size α, simulation size B, the precision matrix estimator $\widehat{\Theta}_{-i}$ and the error variance estimator $\widehat{\sigma}_\varepsilon^2$, a simple simulation procedure for the max test is as follows:

1. For $b = 1, \ldots, B$, draw $\boldsymbol{\xi}^{(b)} = (\xi_1^{(b)}, \ldots, \xi_{N-1}^{(b)}) \sim N(0, \mathbf{I}_{N-1})$ and compute

$$\widehat{\mathcal{I}}_{b,i} = \|\widehat{\sigma}_\varepsilon \widehat{\Theta}_{-i}^{1/2} \boldsymbol{\xi}^{(b)}\|_\infty. \tag{13}$$

2. Estimate the $(1-\alpha)$ quantile of $\|\mathcal{Z}_i\|_\infty$ by the $(1-\alpha)$ empirical quantile of $\{\widehat{\mathcal{I}}_{b,i}\}_{b=1}^B$, denoted by $\widehat{c}_{1-\alpha}$. That is,

$$\widehat{c}_{1-\alpha,i} = \inf \left\{ c : \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\left\{ \widehat{\mathcal{I}}_{b,i} \leq c \right\} \geq 1 - \alpha \right\}.$$

3. Reject $H_0 : \|\boldsymbol{\gamma}_{i:}\|_\infty = 0$ if $\sqrt{T} \|\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}\|_\infty > \widehat{c}_{1-\alpha,i}$.

Together, these results provide feasible inference on individual bilateral effects and joint tests of incoming spillovers, accounting for both regularization and estimation of the common factor space.

4.3 Private effect estimators

Estimating the private effects $\boldsymbol{\theta}$ requires controlling spillover estimation errors after pooling across units. Although the initial LASSO estimators are consistent, their regularization errors can accumulate and affect inference at the $\sqrt{NT}$ scale. The post-selection procedure addresses this problem by selecting relevant spilloer regressors through adaptive LASSO and refitting their coefficients by least squares.

Because the private effect estimator pools information across units, the anaylysis requires correct selection uniformly across the unit-specific regressions. Recall that $\widehat{S}_i$ is the support set selected by the adaptive LASSO in Algorithm 3. The following proposition establishes that all these support sets are recovered correctly with probability approaching one.

Proposition 4.5. *Under the same conditions in Theorem 4.1, and further assuming that the tuning parameter for the adaptive LASSO, $\upsilon_i^{(al)}$, satisfies*

$$\upsilon_i^{(al)} \to 0 \text{ and } T\upsilon_i^{(al)} \geq T^c, \quad c > 0,$$

for all i, then we have

$$\mathbb{P}\left(\widehat{S}_i = S_i \text{ for all } i = 1, \ldots, N\right) \to 1.$$

This proposition allows the limiting distribution of $\widehat{\boldsymbol{\theta}}$ to be derived on the event of correct selection. On this event, the post-selection estimator coincides with the estimator constructed using the true support sets. Although estimation of the retained spillover coefficients and the factor spaces still contributes error, the next theorem establishes that their aggregate contribution is negligible at the $\sqrt{NT}$ scale, and therefore $\widehat{\boldsymbol{\theta}}$ is asymptotically normal.

Let

$$\mathbb{U}_{it} = (u_{x,it}, u_{w,1,it}, \ldots, u_{w,K,it})'$$

collect the idiosyncratic components of the own regressor and controls after projecting out the common factors.

Theorem 4.2. *Under the same assumptions in Proposition 4.5, the estimator $\widehat{\theta}$ from Algorithm 3 satisfies*

$$\sqrt{NT}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} N\left(0, \Sigma_{\mathbb{U}\varepsilon}\right),$$

where

$$\Sigma_{\mathbb{U}\varepsilon} = \left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\mathbb{U}_{it}\mathbb{U}_{it}'\right]\right)^{-1}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_{\varepsilon,i}^{2}\mathbb{E}\left[\mathbb{U}_{it}\mathbb{U}_{it}'\right]\right)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\mathbb{U}_{it}\mathbb{U}_{it}'\right]\right)^{-1}.$$

To make inference feasible, the limiting covariance matrix can be estimated using the projected regressors and estimated residuals:

$$\widehat{\mathbf{\Sigma}}_{\mathbb{U}\varepsilon} = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\left(\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i/T\right)\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}. \tag{14}$$

Corollary 4.1. *Under the same conditions in Theorem 4.2,* $\widehat{\mathbf{\Sigma}}_{\mathbb{U}\varepsilon} \xrightarrow{p} \mathbf{\Sigma}_{\mathbb{U}\varepsilon}$

Theorem 4.2 and Corollary 4.1 establish feasible inference at the $\sqrt{NT}$ rate despite estimation of the spillover network and common factors. Post-selection relaxes the rate restrictions needed to control accumulated spillover-estimation errors. In particular, using the initial LASSO spillover estimates directly requires the stronger condition $N/T \to 0$, which is not required by the proposed procedure. The simulations in Section 5 illustrate the finite-sample benefit: post-selection brings confidence-interval coverage for the common private effect closer to the nominal level across all designs considered.

Inference results for both spillover and private effects rely on controlling errors from projecting onto the estimated factor space. As discussed in Remark 3.1, factor estimates extracted from different covariates recover overlapping spaces, so their combined matrix can have a rank-deficient limit. The resulting discontinuity of the pseudoinverse prevents a direct CMT-based argument for bounding these errors. To overcome this issue, I adopt the leave-one-out analysis of Abbe et al. (2022) and Chen et al. (2021). The argument uses auxiliary leave-one-out factor estimates constructed after omitting the relevant units to exploit cross-sectional independence, and controls the difference between these projections and the full-sample projection. These bounds establish the required control of factor-estimation errors in the spillover and private-effect estimators; Appendix I provides the details.

Remark 4.1 (Cross-sectional dependence in the idiosyncratic regressors)**.** *Assumption 3 in Section 4.1 simplifies inference for both spillover and private effects by ruling out cross-sectional dependence in the idiosyncratic components of the regressors. Allowing weak dependence in these components would require modifications to both estimation procedures.*

For spillover effects, the precision matrix of the regressors after partialling out the common factors and own covariates would generally be non-diagonal and specific to the receiving unit. It could be estimated using nodewise regressions as in van de Geer et al. (2014), subject to suitable sparsity and convergence conditions.

For private effects, cross-sectional dependence would remove the orthogonality between the own regressors and the spillover regressors used in the current argument. Consequently, post-selection alone would no longer ensure that spillover-estimation errors are negligible after aggregation across units. A possible solution is to apply sample splitting and cross-fitting over the time dimension, under an additional assumption of independence over time as in Okui et al. (2025). Factor-estimation bias would require further treatment, potentially through an analytical bias correction in the spirit of Bai (2009). Developing these extensions and establishing their asymptotic properties are left for future research.

5 Simulation

This section evaluates inference on bilateral spillovers and common slope coefficients when the spillover network is unknown. Section 5.1 describes the data-generating process and implementation. Section 5.2 reports coverage and rejection rates, examines network recovery under multiple testing, and assesses the role of post-selection in inference on the private effect. Appendix E examines robustness to weak cross-sectional correlation in the outcome errors, partial factor overlap across covariates, and approximate sparsity of the spillover network.

5.1 Data-generating process

Set $K = 1$ and $(r, r_x, r_w) = (3, 1, 1)$ and generate outcomes and regressors as follows:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X} \cdot \rho + \mathbf{X}\boldsymbol{\Omega}' \cdot \rho + \mathbf{W} \cdot \beta + \mathbf{1}_T \mathbf{a}' + \mathbf{d}\mathbf{1}_N' + \mathbf{f}\boldsymbol{\lambda}' + \boldsymbol{\varepsilon} \\
\mathbf{X} &= \mathbf{1}_T \mathbf{a}_x' + \mathbf{d}\boldsymbol{\eta}_{d,x}' + \mathbf{f}\boldsymbol{\eta}_{f,x}' + \mathbf{f}_x \mathbf{b}_x' + \mathbf{U}_x, \\
\mathbf{W} &= \mathbf{1}_T \mathbf{a}_w' + \mathbf{d}\boldsymbol{\eta}_{d,w}' + \mathbf{f}\boldsymbol{\eta}_{f,w}' + \mathbf{f}_w \mathbf{b}_w' + \mathbf{U}_w.
\end{aligned}$$

The coefficients are set to $\rho = \rho = \beta = 1$. The vectors $\mathbf{a} \in \mathbb{R}^N$ and $\boldsymbol{\lambda} \in \mathbb{R}^N$ represent individual fixed effects and factor loadings in the outcome equation, respectively. Both are generated as $\boldsymbol{\Sigma}_N^{1/2}\widetilde{\mathbf{a}}$, where $\widetilde{\mathbf{a}} \overset{i.i.d.}{\sim} \text{Unif}[-1, 1]$ and $\boldsymbol{\Sigma}_N$ is a Toeplitz covariance matrix with (i, j)-th element $0.5^{|i-j|}$.

Define the loading matrices $\mathbf{\Lambda} = [\mathbf{a}, \mathbf{1}_N, \boldsymbol{\lambda}]$, $\mathbf{\Lambda}_x = [\mathbf{a}_x, \boldsymbol{\eta}_{d,x}, \boldsymbol{\eta}_{f,x}]$, and $\mathbf{\Lambda}_w = [\mathbf{a}_w, \boldsymbol{\eta}_{d,w}, \boldsymbol{\eta}_{f,w}]$ for $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{W}$, respectively. To induce endogeneity through common factors, I generate $\mathbf{\Lambda}_x = \mathbf{\Lambda} + \widetilde{\mathbf{\Lambda}}_x$ and $\mathbf{\Lambda}_w = \mathbf{\Lambda} + \widetilde{\mathbf{\Lambda}}_w$, where the perturbations $\widetilde{\mathbf{\Lambda}}_x$ and $\widetilde{\mathbf{\Lambda}}_w$ have i.i.d. entries from Unif$[-1, 1]$.

The full set of common factors is given by $\mathbf{G} = [\mathbf{d}, \mathbf{f}, \mathbf{f}_x, \mathbf{f}_w] \in \mathbb{R}^{T \times 4}$, generated as $\widetilde{\mathbf{G}}\mathbf{\Sigma}_f^{1/2}$, where $\widetilde{\mathbf{G}}$ has i.i.d. $N(1, 1)$ entries and $\mathbf{\Sigma}_f$ is a 4×4 Toeplitz covariance matrix with (i, j)-th element $0.5^{|i-j|}$. This construction ensures that the regressors and the outcome share correlated factor structures. The idiosyncratic errors ε_{it} are drawn from $N(0, \sigma_i^2)$ with heterogeneous variances $\sigma_i^2 \overset{i.i.d.}{\sim}$ Unif$[0.5, 2]$. The idiosyncratic components in the regressor factor models are generated as $\mathbb{U}_{it} = (u_{x,it}, u_{w,it})' = \sigma_i \widetilde{\mathbb{U}}_{it}$, where $\widetilde{\mathbb{U}}_{it} \sim N(0, \mathbf{\Sigma}_{\mathbb{U}})$ and $\mathbf{\Sigma}_{\mathbb{U}}$ is a 2×2 Toeplitz covariance matrix with off-diagonal elements equal to 0.5. Although the DGP includes two-way fixed effects, the simulated data are not demeaned prior to estimation, allowing the proposed procedures to remove these components through factor projection.

The spillover matrix $\mathbf{\Omega}$ is sparse: $\omega_{i,i-1} = \omega_{i,i+1} = 0.5$ whenever the column index lies in $\{1, \ldots, N\}$, and all other entries are zero. I then set the first row $\mathbf{\Omega}_{1:} = 0$, so that unit 1 receives no spillovers. Since $\rho = 1$, the matrix of bilateral coefficients is $\mathbf{\Gamma} = \mathbf{\Omega}$. The estimator treats this network as unknown.

I use 2,000 repetitions for each combination $(N, T) \in \{50, 200\} \times \{100, 200, 300\}$. The tuning parameters υ_i and $\upsilon_i^{(al)}$ are selected by BIC. For each design, I construct 95% confidence intervals for the spillover coefficients $\gamma_{12} = 0$ and $\gamma_{21} = 0.5$, the common private effect $\rho = 1$, and the control coefficient $\beta = 1$. I also test whether units 1 and 2 receive any spillovers using the row-wise tests in Theorem 4.1. The null $\|\boldsymbol{\gamma}_{1:}\|_\infty = 0$ is true, whereas $\|\boldsymbol{\gamma}_{2:}\|_\infty = 0.5$, so the two rejection rates measure size and power, respectively.

5.2 Simulation results

Table 1 reports coverage for individual spillover coefficients and rejection rates for the row-wise tests. Coverage for the zero coefficient γ_{12} ranges from 94% to 95% across designs. For the nonzero coefficient γ_{21}, coverage is 91% at $(N, T) = (50, 100)$, rises to 93% at the two larger values of T when $N = 50$, and reaches 95% in all three designs with $N = 200$. Thus, the intervals are close to their nominal level in most designs.

The row-wise test rejects the true null for unit 1 in 6–7% of repetitions, indicating modest overrejection relative to the nominal 5% level. For unit 2, power is 93% at $(N, T) = (50, 100)$ and 77% at $(200, 100)$; it reaches 100% in all designs with $T \geq 200$. The results show that a longer time

dimension therefore improves the ability to detect incoming spillovers.

Figures 1 and 2 extend the analysis to recovery of the full network for $N = 50$ and $T = 100$ and 300, respectively. Each figure compares the true support with entry-wise rejection frequencies over 2,000 repetitions. I compare pointwise tests at the nominal 5% level with the Benjamini–Hochberg (BH; Benjamini and Hochberg, 1995) and Benjamini–Yekutieli (BY; Benjamini and Yekutieli, 2001) procedures, both targeting a 5% false discovery rate (FDR). Here, FDR is the average fraction of rejected nulls that correspond to zero spillover coefficients, with this fraction set to zero when there are no rejections. Power is the average fraction of nonzero coefficients detected.

At $T = 100$, the pointwise tests recover the banded pattern but also reject zero coefficients throughout the network. Both multiple-testing adjustments reduce these false detections. BH achieves 80.1% power, but its empirical FDR is 10.6%, above the target level. BY reduces the FDR to 2.3%, at the cost of lower power of 67.2%. At $T = 300$, both procedures detect nearly all nonzero coefficients: power reaches 99.8% for BH and 99.2% for BY. Their empirical FDRs are 7.4% and 1.1%, respectively. In these two designs, BY keeps the FDR below the target level, and its power cost becomes small as T increases. The distinction between pointwise inference and inference across the full network therefore matters even when individual confidence intervals have coverage close to 95%.

Table 2 shows the results of inference on the common private effect ρ and control coefficient β. The non-post-selection estimator (Non-PS) uses the initial LASSO spillover estimates from Step 2 of Algorithm 2 without the refinement in Algorithm 3, while the proposed post-selection estimator (PS) implements that refinement.

Post-selection improves coverage for ρ in every design. When $N = 50$, Non-PS coverage remains at 83–84%, whereas PS coverage increases from 90% at $T = 100$ to 95% at $T = 300$. When $N = 200$, the corresponding ranges are 87–91% and 90–94%. The gains for β are smaller: coverage ranges from 92% to 94% under Non-PS and from 92% to 95% under PS, with no uniform improvement across designs. These results show that post-selection is particularly useful for inference on the private effect.

6 Empirical application

I apply the proposed framework to estimate the private and bilateral spillover effects of R&D across industries. An industry's output may respond both to R&D undertaken elsewhere and

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Coverage rates for γ_{12} and γ_{21}						
γ_{12}	0.94	0.94	0.95	0.94	0.94	0.95
γ_{21}	0.91	0.93	0.93	0.95	0.95	0.95
Rejection rates of the row-wise tests						
$H_0 : \|\boldsymbol{\gamma}_{1:}\|_\infty = 0$	0.07	0.06	0.07	0.07	0.06	0.06
$H_0 : \|\boldsymbol{\gamma}_{2:}\|_\infty = 0$	0.93	1.00	1.00	0.77	1.00	1.00

Table 1: Coverage and rejection rates for spillover inference. Columns report (N, T). The upper panel reports empirical coverage of nominal 95% confidence intervals for $\gamma_{12} = 0$ and $\gamma_{21} = 0.5$. The lower panel reports rejection rates of nominal 5% row-wise tests. Unit 1 receives no spillovers, so its rejection rate measures size; unit 2 receives nonzero spillovers, so its rejection rate measures power. Estimation uses Algorithm 2 with PCA-Each.

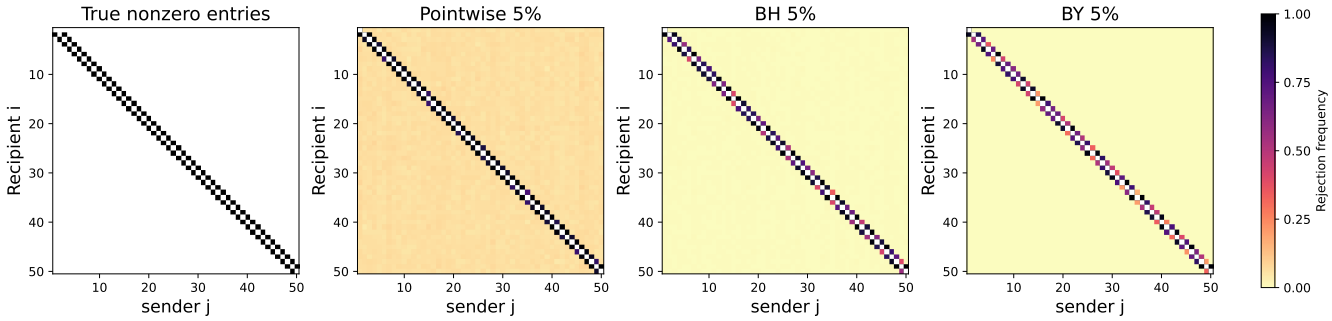


Figure 1: Network recovery for $(N, T) = (50, 100)$ over 2,000 repetitions. Rows index recipients and columns index senders. The first panel shows the true nonzero spillover coefficients. The remaining panels show rejection frequencies for pointwise tests at the nominal 5% level and the Benjamini–Hochberg (BH) and Benjamini–Yekutieli (BY) procedures, both targeting a 5% FDR. Darker cells indicate more frequent rejections.

to common shocks that also affect input choices. Distinguishing these sources of comovement is therefore important for estimating the productivity effects of R&D. The model allows me to estimate the spillover network without specifying proximity weights, while controlling for common shocks through the factor structure of the inputs.

This application builds on the concern raised by Eberhardt et al. (2013) that cross-sectional dependence can distort estimates of the private return to R&D. Their analysis accounts for this dependence through latent factors. Here, I estimate bilateral spillover effects separately from common correlated effects, allowing me to examine both the private return to R&D and the pattern

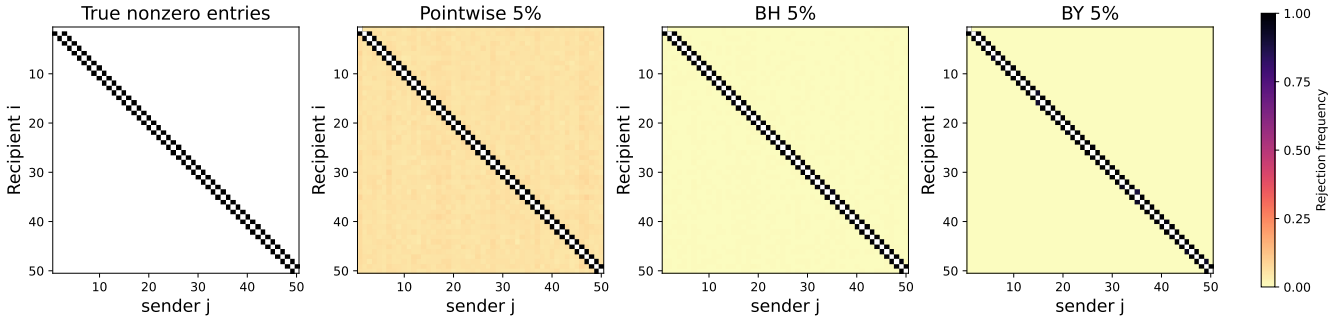


Figure 2: Network recovery for $(N, T) = (50, 300)$ over 2,000 repetitions. Rows index recipients and columns index senders. The first panel shows the true nonzero spillover coefficients. The remaining panels show rejection frequencies for pointwise tests at the nominal 5% level and the Benjamini–Hochberg (BH) and Benjamini–Yekutieli (BY) procedures, both targeting a 5% FDR. Darker cells indicate more frequent rejections.

	(N,T)	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
ρ	Non-PS	0.83	0.84	0.84	0.87	0.90	0.91
	PS	0.90	0.93	0.95	0.90	0.93	0.94
β	Non-PS	0.92	0.94	0.93	0.94	0.94	0.94
	PS	0.92	0.94	0.95	0.93	0.94	0.94

Table 2: Empirical coverage of nominal 95% confidence intervals for the common private effect $\rho = 1$ and control coefficient $\beta = 1$. Columns report (N, T). Non-PS uses the initial LASSO spillover estimates without post-selection; PS implements the post-selection refinement in Algorithm 3.

of transmission across industries.

I use a balanced panel of 41 two-digit SIC industries constructed from Compustat Quarterly data, covering 108 quarters over 1991–2017. Table 3 lists the industries in the sample, and Figure 3 reports summary statistics for the main variables. I first specify the production function and examine the evidence for common factors in the inputs, then report estimates of the private effects and the bilateral spillover network.

I assume industries produce according to a classic Cobb-Douglas production function augmented with R&D capital as in Manresa (2013) and Okui et al. (2025):

$$Y_{it} = A_{it} C_{it}^{\beta_1} L_{it}^{\beta_2} K_{it}^{\rho} \prod_{j \neq i} K_{jt}^{\gamma_{ij}} e^{\varepsilon_{it}}, \tag{15}$$

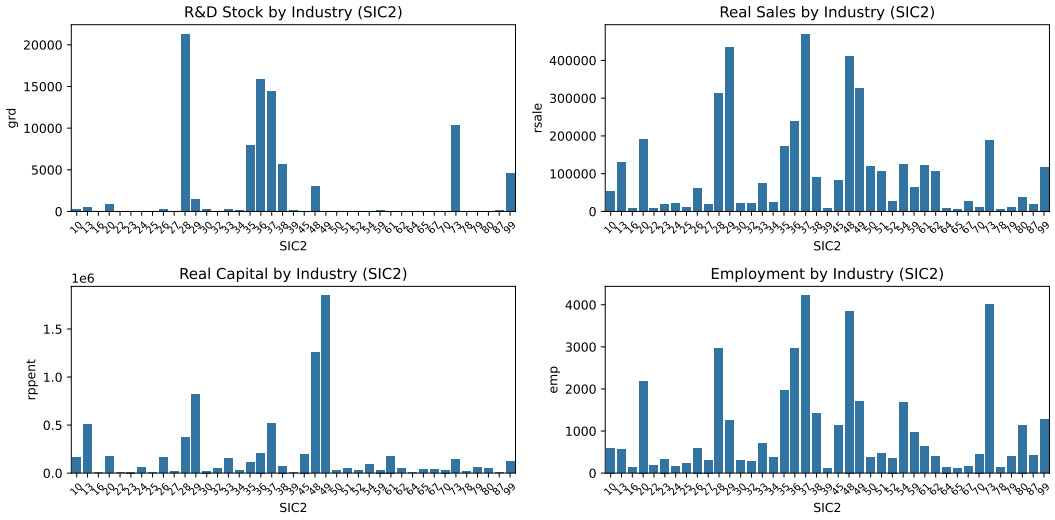


Figure 3: Means of industry-level R&D stock, real sales, real capital and employment

where Y_{it} is the real output of industry i at time t measured by real sales, C_{it} is captial, L_{it} is labor, and K_{it} is the R&D capital stock. The term $\prod_{j\neq i} K_{jt}^{\gamma_{ij}}$ captures the R&D spillover effects from other industries, where γ_{ij} measures the spillover effect from industry j to industry i. The technological progress component A_{it} is an industry-specfic shock to productivity that take the following interactive fixed effects structure:

$$A_{it} = \exp\left(\alpha_i + f_t'\lambda_i\right).$$

Substituting A_{it} into (15) and taking logs yield the following estimating equation:

$$y_{it} = c_{it}\cdot\beta_1 + l_{it}\cdot\beta_2 + k_{it}\cdot\rho + \sum_{j\neq i} k_{jt}\cdot\gamma_{ij} + \alpha_i + f_t'\lambda_i + \varepsilon_{it}. \tag{16}$$

The cross-sectional dependence in this model arises from two sources: the R&D spillover structure $\sum_{j\neq i} k_{jt}\cdot\gamma_{ij}$ and the unobserved common correlated effects $f_t'\lambda_i$. I further assume that the input variables are driven by the common shocks f_t and additional shocks specific to each input:

$$z_{it} = \alpha_{z,i} + f_t'\eta_{z,i} + f_{z,t}'b_{z,i} + u_{z,it}, \quad z\in\{c,l,k\}. \tag{17}$$

Equation (17) allows each input to depend on factors shared with productivity and additional

input-specific factors, with $u_{z,it}$ capturing industry-specific variation. Similar input-factor structures appear in the framework of Eberhardt et al. (2013) discussed above and in Eberhardt and Teal (2013, 2020) and Ertur and Musolesi (2017). These studies use the structure to account for technological and macroeconomic shocks that affect both productivity and input choices, with heterogeneous loadings allowing units to respond differently.

I impose sparsity on the industry-specific spillover coefficients, γ_{ij}, conditional on observed controls and common factors. An industry's ability to benefit from external R&D depends on the relevance of that knowledge to its production and its capacity to absorb it (Cohen and Levinthal, 1989). Consistent with selective transmission, Acemoglu et al. (2016b) document concentrated innovation flows across technological fields, and Bernstein and Nadiri (1988) find substantial differences in industries' importance as sources of interindustry R&D spillovers. These findings motivate a specification in which a limited set of source industries drives each recipient's spillover exposure.

I estimate the production function in (16) using the proposed algorithm, where the factor space is estimated using Algorithm 1. The estimator exploits the factor structure in (17) to control for unobserved common shocks that jointly affect industry productivity and input choices. The data are first demeaned and detrended over time to remove industry fixed effects and trend effects, after which the factor projection is applied to purge the remaining common components. The number of latent factors in the proposed specification is selected using the eigenvalue-ratio criterion of Ahn and Horenstein (2013). Figure 4 reports the scree and eigenvalue-ratio plots for each input, which indicate the numbers of factors are 3, 3, 2 for R&D stock, capital, and employment, respectively.

Table 4 quantifies the explanatory power of the selected factors. After removing industry means and linear trends, three factors explain 83.1% of the pooled within-industry variation in log R&D stock, three explain 72.4% in log real capital, and two explain 68.6% in log employment. Thus, a small number of common components captures most of the variation in each input panel, while leaving nontrivial residual input variation. The combined covariate factor space also explains 72.6% of the variation in log real sales, with industry-specific loadings. Although the factors are estimated from the inputs, their common components are therefore closely associated with output. These results support the empirical relevance of a parsimonious factor representation in this application.

Figures 5 and 6 complement Table 4 by illustrating heterogeneous industry exposure and shared variation across inputs. Figure 5 shows substantial differences in factor loadings across industries. Since the factors are estimated separately for each input and are identified up to rotations, their numbered components need not correspond one-to-one. Figure 6 therefore compares their

estimated spaces using canonical correlations. The first canonical correlations range from 0.96 to 0.97, and the second from 0.85 to 0.92, indicating strong alignment along two directions in each pair of factor spaces. These patterns are consistent with shared components across R&D stock, capital, and employment, with heterogeneous industry loadings.

To assess the importance of accounting for common shocks, I also report estimates of the common production elasticities (ρ, β_1, β_2) obtained from a conventional industry-fixed-effects OLS regression that ignores the interactive factor structure. Because this benchmark estimator does not orthogonalize the regressors with respect to the common shocks, inference is conducted using the time-domain cross-fitting procedure as in Okui et al. (2025), which mitigates bias arising from cross-sectional dependence in R&D stock, as discussed in Remark 4.1.

Table 5 reports estimates of the common production elasticities (ρ, β_1, β_2) obtained from the proposed IFE estimator and from a conventional fixed-effects OLS benchmark. The estimated R&D elasticity ρ under IFE is close to zero and statistically insignificant. This contrasts with the positive and significant OLS estimate, and is consistent with the findings of Eberhardt et al. (2013). It indicates that ignoring common correlated effects tends to inflate estimated private returns, reinforcing the importance of jointly modeling spillovers and latent factor structures when assessing the productivity impact of R&D.

To characterize the estimated spillover network, I focus on the directed effects $\widehat{\gamma}_{ij}$ that are statistically significant according to Theorem 4.1. With 41 industries, the network contains $41 \times (41-1) =$ 1,640 possible directed links. Testing all links separately would create many false discoveries, so I control the false discovery rate (FDR) at the 5% level using the procedure of Benjamini and Yekutieli (2001). The FDR screen leaves 118 significant directed links.

Table 6 summarizes participation in the estimated network. An industry is a sender if its R&D stock has at least one significant positive or negative effect on another industry. It is a receiver if at least one other industry's R&D stock significantly affects its output. By this definition, 33 industries are senders and 30 are receivers. 25 industries perform both roles, eight are senders only, five are receivers only, and three have no significant links.

Table 7 turns to their magnitudes and sources across the 30 receiving industries. It shows substantial heterogeneity in the magnitude and sign of received R&D spillovers. Among the 30 industries, 23 have negative net coefficient sums. Stone, clay, glass, and concrete products (SIC 32) have the largest sum of absolute incoming coefficients: negative contributions total 7.19 in magnitude and positive contributions total 1.24, yielding a net sum of -5.95. Large gross contributions can also

coexist with a modest net sum. For hotels and lodging (SIC 70), positive and negative contributions have magnitudes of 4.16 and 3.55, respectively, leaving a net sum of 0.62. The net measure alone therefore understates the strength of this industry's estimated bilateral relationships. Electronics (SIC 36) is the largest negative source for stone, clay, glass, and concrete products, while machinery (SIC 35) is the largest negative source for textiles (SIC 22); the importance of these source industries also appears in Manresa (2013). Separately, the joint receiver test rejects the absence of incoming spillovers for 20 of the 30 industries at the 5% level.

Figure 7 places these received spillovers in the wider network by plotting the significant links between industries. Each node is a SIC2 industry. Link color gives the sign of $\widehat{\gamma}_{ij}$, and link width increases with its magnitude. Node color indicates whether the industry is a net sender or a net receiver. Let $\widehat{\gamma}_{ij}^{\mathrm{FDR}}$ equal $\widehat{\gamma}_{ij}$ for links that pass the FDR screen and zero otherwise. To measure the difference between outgoing and incoming exposure, define

$$SO_i = \sum_{j \neq i} |\widehat{\gamma}_{ji}^{\mathrm{FDR}}| - \sum_{j \neq i} |\widehat{\gamma}_{ij}^{\mathrm{FDR}}|.$$

A positive value of SO_i means that industry i transmits more spillovers, in absolute value, than it receives; a negative value means that it receives more than it transmits. Node size increases with $|SO_i|$. Thus, a large red node need not generate positive spillovers: its outgoing effects can be predominantly negative.

The estimated network is sparse and contains spillovers of both signs. There are 118 significant links, accounting for 7.2% of all possible directed pairs. Among these links, 77 are negative and 41 are positive. Controlling only for industry fixed effects yields 203 significant directed links, compared with 118 after accounting for common factors—a reduction of approximately 42%. This difference indicates that the estimated spillover network is sensitive to common-shock adjustment and is consistent with common shocks being partly attributed to bilateral spillovers when only industry fixed effects are included.

The most prominent senders are Industrial Machinery and Computer Equipment (SIC 35) and Electronic Equipment (SIC 36). These industries rank first and second by the sum of absolute outgoing coefficients, with 12 and 6 significant links, respectively, all negative. Both transmit negative spillovers to textiles (SIC 22), stone, clay, glass, and concrete products (SIC 32), fabricated metals (SIC 34), and building-material retail (SIC 52). Their prominence as spillover sources is consistent with Manresa (2013). One possible explanation for the negative signs is that innovations

in machinery and electronics change the demand for existing products and inputs. For example, new equipment may enable material-saving production or substitution toward different inputs, reducing sales in some supplying industries. Innovation may also strengthen competitors that adopt these technologies. Such displacement effects can coexist with positive knowledge spillovers, as emphasized by Bloom et al. (2013).

On the receiving side, as shown in Table 7, SIC 32 has the largest total absolute incoming effect, followed by building-material retail (SIC 52), textiles (SIC 22), lodging (SIC 70), and durable-goods wholesale (SIC 50). The predominantly negative links to SIC 32, 52, and 22 are compatible with the input-substitution mechanism described above: changes in material use can affect both producers and distributors. Durable-goods wholesale receives its largest negative coefficient from Instruments and Related Products (SIC 38); improvements in inventory management or direct distribution could reduce demand for some intermediary services. Lodging receives effects of both signs and has no significant outgoing links. Its positive links from air transportation and recreation have a natural interpretation through complementary demand for travel and accommodation. Its negative links, including the largest one from SIC 38, may instead reflect competitive pressure or shifts in expenditure toward alternative services. More generally, the transmission of supply and demand changes across industries is consistent with the evidence on input-output linkages in Acemoglu et al. (2016a).

The estimated network reveals a complex pattern of inter-industry R&D spillovers, with substantial differences in their sources, recipients, and signs. These bilateral effects summarize the net influence of the economic relationships connecting industries. The resulting patterns provide empirical benchmarks for future structural models that endogenize R&D investment and explore how spillover networks form across industries.

Table 3: SIC2 code descriptions of industries in the sample

SIC2 Codes	Descriptions
10	Metal Mining
13	Oil And Gas Extraction
16	Heavy Cnstrctn, Except Building Construction - Contractors
20	Food And Kindred Products
22	Textile Mill Products
23	Apparel, Finished Prdcts From Fabrics & Similar Materials
24	Lumber And Wood Products, Except Furniture
25	Furniture And Fixtures
26	Paper And Allied Products
27	Printing, Publishing And Allied Industries
28	Chemicals And Allied Products
29	Petroleum Refining And Related Industries
30	Rubber And Miscellaneous Plastic Products
32	Stone, Clay, Glass, And Concrete Products
33	Primary Metal Industries
34	Fabricated Metal Prdcts, Except Machinery & Transport Eqpmnt
35	Industrial And Commercial Machinery And Computer Equipment
36	Electronic, Elctrcl Eqpmnt & Cmpnts, Excpt Computer Eqpmnt
37	Transportation Equipment
38	Mesr/Anlyz/Cntrl Instrmnts; Photo/Med/Opt Gds; Watchs/Clocks
39	Miscellaneous Manufacturing Industries
45	Transportation By Air
48	Communications
49	Electric, Gas And Sanitary Services
50	Wholesale Trade - Durable Goods
51	Wholesale Trade - Nondurable Goods
52	Building Matrials, Hrdwr, Garden Supply & Mobile Home Dealrs
54	Food Stores
59	Miscellaneous Retail
61	Nondepository Credit Institutions
62	Security & Commodity Brokers, Dealers, Exchanges & Services
64	Insurance Agents, Brokers And Service
65	Real Estate
67	Holding And Other Investment Offices
70	Hotels, Rooming Houses, Camps, And Other Lodging Places
73	Business Services
78	Motion Pictures
79	Amusement And Recreation Services
80	Health Services
87	Engineering, Accounting, Research, Management & Related Svcs
99	Nonclassifiable Establishments

Source: NAICS Association.

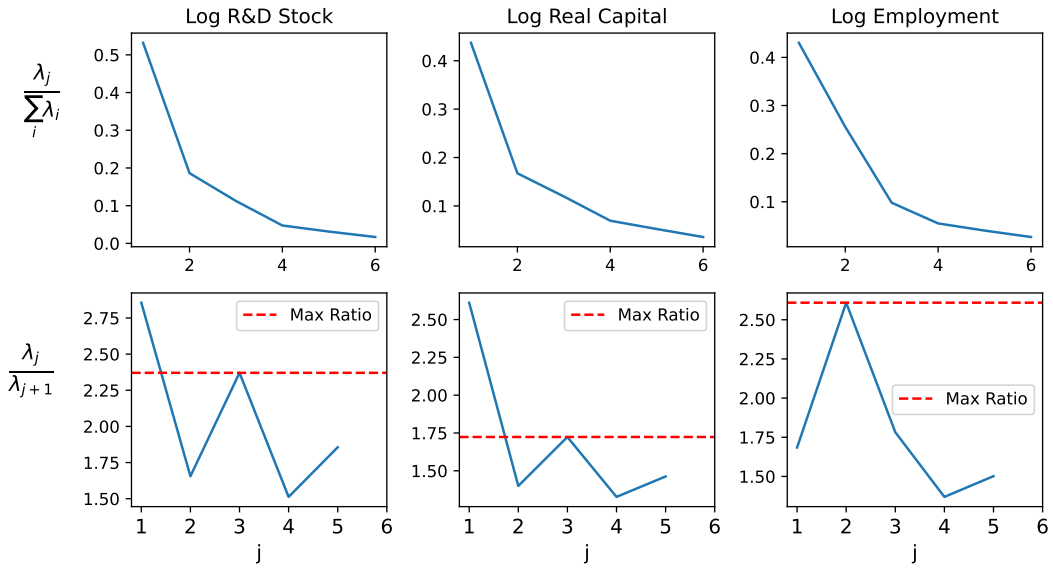


Figure 4: Scree plot and eigenvalue ratio plot for selecting number of factors. The first row of panels plots the scree plot of $\lambda_j / \sum_i \lambda_i$ against j, where λ_j is the j-th largest eigenvalue of the sample covariance matrix of the covariate indicated by the column header. The second row of panels plots the eigenvalue ratio $\lambda_j / \lambda_{j+1}$ against j. The red horizontal dash line indicates the maximum eigenvalue ratio except for the first one. The bottom plot suggests $(3, 3, 2)$ factors for log R&D stock, log real capital, and log employment, respectively.

Table 4: Variance explained by covariate factors

	Number of factors	Explained Variance	Residual Variance
Panel A: Covariates			
Log R&D stock	3	83.08%	16.92%
Log real capital	3	72.41%	27.59%
Log employment	2	68.58%	31.42%
Panel B: Outcome			
Log real sales	8	72.60%	27.40%

Notes: Variance shares pool within-industry variation over industries and quarters after removing industry means and linear trends. Explained variance is $(100(1 - \text{SSR}/\text{SST}))\%$; residual variance is $100\,\text{SSR}/\text{SST}$. Log real sales is projected jointly on all eight covariate factors, with industry-specific loadings.

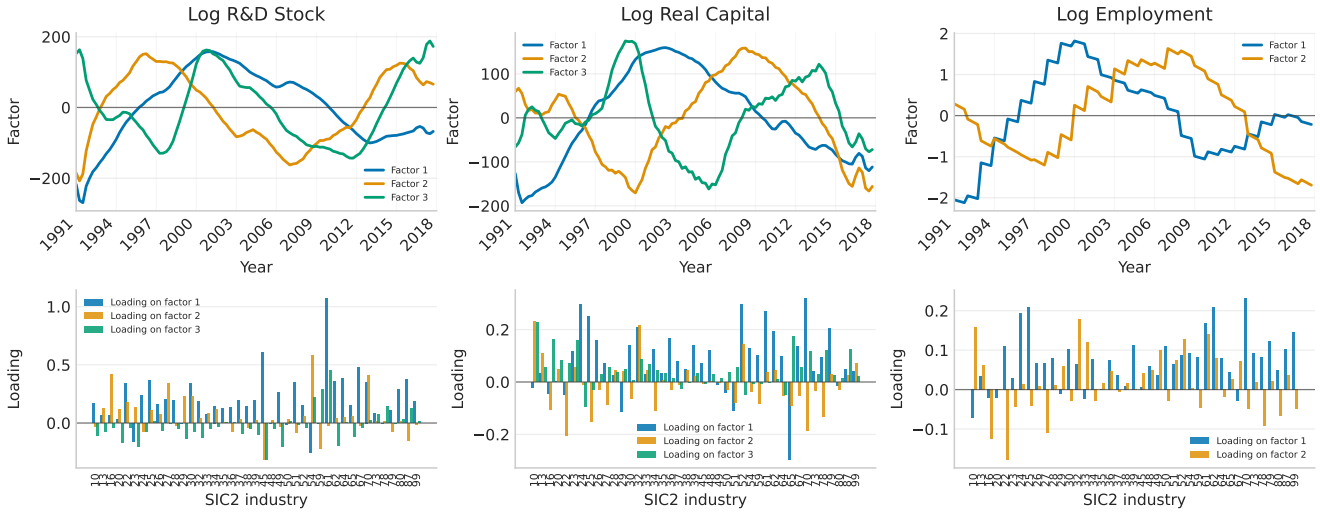


Figure 5: Estimated covariate-specific factors and industry loadings. The top panels plot the estimated factors for log R&D stock, log real capital, and log employment. The bottom panels plot the corresponding SIC2 industry loadings. The factors are estimated from the quarterly industry panel after removing industry fixed effects.

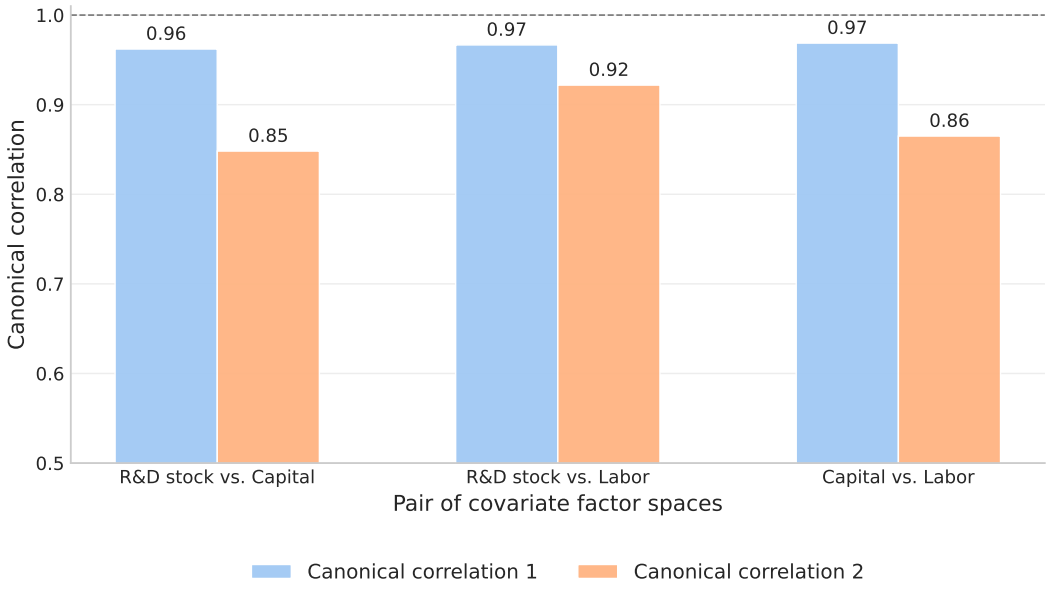


Figure 6: Canonical correlations between estimated covariate factor spaces. Each bar reports a canonical correlation for a pair of factor spaces estimated from log R&D stock, log real capital, and log employment. Values close to one indicate that the covariate-specific factors span similar spaces, consistent with the common-factor structure in (17).

Table 5: Coefficient estimates of common parameters in production function

	IFE	Industry FE	Industry FE
R&D Stock	0.014	0.034	0.034
	(0.013)	(0.006)	(0.007)
Capital	0.387	0.449	0.477
	(0.021)	(0.017)	(0.018)
Labor	0.28	0.5	0.549
	(0.029)	(0.024)	(0.027)
Time Cross-fitting	-	No	Yes

This table reports the estimates of common parameters on R&D stock, capital, and labor in the production function (16) using different estimators. The IFE columns report the estimates from the proposed interactive fixed effects estimator that accounts for the common correlated effects. The Industry FE column reports the estimates from the standard OLS estimator with individual fixed effects that ignores the common correlated effects, with and without cross-fitting along the time dimension discussed in Remark 4.1. Cross-fitting along time dimension is applied to address potential correlations between industry R&D stocks. Standard errors are reported in parentheses.

Table 6: Number of significant spillover relationships after controlling FDR at 5% level

		Sender		
		Yes	No	Total
Receiver	Yes	25	5	30
	No	8	3	11
	Total	33	8	41

Notes: This table reports the number of significant spillover relationships among industries after controlling the false discovery rate (FDR) at 5% level using the Benjamini and Yekutieli (2001) procedure. A significant spillover relationship from industry j to industry i is defined as $\widehat{\gamma}_{ij}$ being statistically significant based on Theorem 4.1 after controlling FDR. An industry is "Receiver" if it receives spillovers from any other industry, either postive or negative, while it is an "Sender" if it transmits spillovers to any other industry, either positive or negative.

Table 7: Received R&D spillovers across industries

SIC2	Receiving industry	Receiver p-value	Positive	Negative magnitude	Net	Largest positive source	Largest negative source
10	Metal Mining	0.136	0.0000	0.4393	-0.4393	—	Lumber And Wood Products, Except Furniture (SIC 24) (-0.4393)
13	Oil And Gas Extraction	0.038	0.0000	0.2005	-0.2005	—	Heavy Cnstrctn, Except Building Construction - Contractors (SIC 16) (-0.2005)
22	Textile Mill Products	< 0.01	0.3664	3.4795	-3.1131	Food And Kindred Products (SIC 20) (+0.2341)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.8668)
23	Apparel, Finished Prdcts From Fabrics & Similar Materials	0.164	1.1744	0.0000	1.1744	Paper And Allied Products (SIC 26) (+0.6485)	—
24	Lumber And Wood Products, Except Furniture	0.026	0.1278	0.6802	-0.5524	Insurance Agents, Brokers And Service (SIC 64) (+0.1278)	Electronic, Elctrcl Eqpmnt & Cmpnts, Excpt Computer Eqpmnt (SIC 36) (-0.6802)
26	Paper And Allied Products	< 0.01	0.0000	0.4982	-0.4982	—	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.4982)
27	Printing, Publishing And Allied Industries	0.058	0.4395	0.0000	0.4395	Paper And Allied Products (SIC 26) (+0.4395)	—
29	Petroleum Refining And Related Industries	0.626	0.2107	0.3879	-0.1772	Miscellaneous Retail (SIC 59) (+0.2107)	Lumber And Wood Products, Except Furniture (SIC 24) (-0.3879)
30	Rubber And Miscellaneous Plastic Products	< 0.01	0.1626	1.0613	-0.8987	Food And Kindred Products (SIC 20) (+0.1626)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.6037)
32	Stone, Clay, Glass, And Concrete Products	< 0.01	1.2416	7.1888	-5.9472	Paper And Allied Products (SIC 26) (+0.6580)	Electronic, Elctrcl Eqpmnt & Cmpnts, Excpt Computer Eqpmnt (SIC 36) (-1.9162)
33	Primary Metal Industries	< 0.01	0.2989	1.9274	-1.6285	Nondepository Credit Institutions (SIC 61) (+0.2989)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-1.3179)
34	Fabricated Metal Prdcts, Except Machinery & Transport Eqpmnt	< 0.01	0.2316	1.8638	-1.6322	Food And Kindred Products (SIC 20) (+0.1394)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.3846)
35	Industrial And Commercial Machinery And Computer Equipment	0.061	0.0000	0.1620	-0.1620	—	Holding And Other Investment Offices (SIC 67) (-0.1620)
38	Mesr/Anlyz/Cntrl Instrmnts; Photo/Med/Opt Gds; Watchs/Clocks	0.296	0.0000	0.0740	-0.0740	—	Heavy Cnstrctn, Except Building Construction - Contractors (SIC 16) (-0.0740)
39	Miscellaneous Manufacturing Industries	0.023	0.0000	0.4324	-0.4324	—	Wholesale Trade - Nondurable Goods (SIC 51) (-0.4324)
45	Transportation By Air	< 0.01	0.3707	1.1895	-0.8187	Food And Kindred Products (SIC 20) (+0.2366)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.5693)

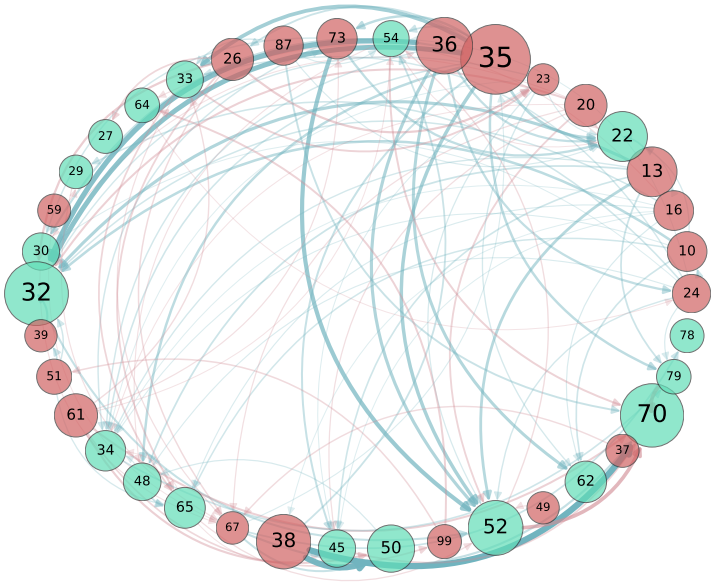
Continued on next page

Received R&D spillovers under the baseline IFE specification (continued)

SIC2	Receiving industry	Receiver p-value	Positive	Negative magnitude	Net	Largest positive source	Largest negative source
48	Communications	< 0.01	0.1840	1.6443	-1.4603	Food Stores (SIC 54) (+0.1840)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.7754)
50	Wholesale Trade - Durable Goods	< 0.01	0.6179	2.8009	-2.1830	Nondepository Credit Institutions (SIC 61) (+0.6179)	Mesr/Anlyz/Cntrl Instrmnts; Photo/Med/Opt Gds; Watchs/Clocks (SIC 38) (-2.2388)
51	Wholesale Trade - Nondurable Goods	0.107	0.3667	0.0000	0.3667	Nonclassifiable Establishments (SIC 99) (+0.3667)	—
52	Building Matrials, Hrdwr, Garden Supply & Mobile Home Dealrs	< 0.01	0.9272	5.7783	-4.8510	Food And Kindred Products (SIC 20) (+0.4958)	Business Services (SIC 73) (-1.4976)
54	Food Stores	< 0.01	0.4263	1.2759	-0.8495	Building Matrials, Hrdwr, Garden Supply & Mobile Home Dealrs (SIC 52) (+0.2891)	Mesr/Anlyz/Cntrl Instrmnts; Photo/Med/Opt Gds; Watchs/Clocks (SIC 38) (-0.4896)
62	Security & Commodity Brokers, Dealers, Exchanges & Services	0.030	0.0000	1.8953	-1.8953	—	Electronic, Elctrcl Eqpmnt & Cmpnts, Excpt Computer Eqpmnt (SIC 36) (-1.0018)
64	Insurance Agents, Brokers And Service	0.070	0.7258	0.0000	0.7258	Nonclassifiable Establishments (SIC 99) (+0.5445)	—
65	Real Estate	0.074	0.9405	1.3074	-0.3669	Paper And Allied Products (SIC 26) (+0.4813)	Communications (SIC 48) (-0.5118)
67	Holding And Other Investment Offices	< 0.01	1.5914	0.0000	1.5914	Paper And Allied Products (SIC 26) (+0.3414)	—
70	Hotels, Rooming Houses, Camps, And Other Lodging Places	< 0.01	4.1610	3.5457	0.6153	Amusement And Recreation Services (SIC 79) (+1.2994)	Mesr/Anlyz/Cntrl Instrmnts; Photo/Med/Opt Gds; Watchs/Clocks (SIC 38) (-2.3906)
73	Business Services	< 0.01	0.2205	0.9799	-0.7595	Nondepository Credit Institutions (SIC 61) (+0.2205)	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.7524)
78	Motion Pictures	0.314	0.0000	0.4557	-0.4557	—	Holding And Other Investment Offices (SIC 67) (-0.4557)
79	Amusement And Recreation Services	< 0.01	0.0000	1.8550	-1.8550	—	Industrial And Commercial Machinery And Computer Equipment (SIC 35) (-0.7698)
99	Nonclassifiable Establishments	0.024	0.4222	0.0000	0.4222	Nondepository Credit Institutions (SIC 61) (+0.4222)	—

Notes: Industries with zero positive and zero negative received spillovers after the BY screen are omitted; 30 receiving industries remain. Magnitudes and source rankings use debiased coefficients passing the 5% BY screen across all 1,640 off-diagonal links within each specification. Positive and negative columns sum retained positive coefficients and absolute negative coefficients; net is their difference. Source cells show industry names, SIC codes, and coefficients (not standard errors). A dash means no BY-detected source of that sign. The receiver p-value tests $H_{0i} : \|\gamma_{i:}\|_\infty = 0$ using the Gaussian maximum approximation for the full debiased row. Values below 0.01 display as < 0.01.

Figure 7: Estimated R&D spillover structure among industries



Notes: Nodes are SIC2 industries, and arrows run from the R&D sender to the receiver. Only links passing the Benjamini and Yekutieli (2001) FDR screen at 5% are shown; industries with no significant links are omitted. Red nodes have greater total absolute outgoing than incoming effects, and cyan nodes have greater incoming than outgoing effects. Node size increases with the absolute difference between these totals, $|SO_i|$, rather than their sum. Red edges indicate positive coefficients and blue edges indicate negative coefficients; edge width increases with coefficient magnitude. Node color describes the balance of outgoing and incoming exposure, not the sign of the industry's effects.

7 Conclusion

This paper develops estimation and inference procedures for panel data models with an unknown spillover network and common correlated effects. Exploiting a factor structure in the covariates, I account for latent common shocks and estimate a row-wise sparse matrix of bilateral spillover coefficients without requiring pre-specified network matrices. These coefficients capture the aggregate effects of underlying spillover channels within each bilateral relationship. The proposed debiasing and post-selection procedures deliver asymptotically valid inference for bilateral spillover and common private effects, accounting for both factor-estimation and regularization errors.

In the application, I estimate an industry-level R&D spillover network in a production-function framework. Factor spaces estimated separately from R&D stock, capital, and employment exhibit substantial overlap, providing evidence of common shocks shared across these inputs. After

accounting for latent common factors, the estimated private return to R&D becomes small and statistically insignificant, in line with Eberhardt et al. (2013). The recovered network reveals heterogeneous cross-industry spillovers, with predominantly negative estimated effects, which may reflect competition or resource reallocation in the market.

Overall, the results highlight the importance of jointly modeling spillovers and common correlated effects, and show that failing to account for latent common shocks can lead to distorted inference on both the magnitude and the structure of economic externalities.

References

Abbe, E., Fan, c., and Wang, K. (2022). An ℓ_p theory of pca and spectral clustering. *The Annals of Statistics*, 50(4):2359–2385.

Acemoglu, D., Akcigit, U., and Kerr, W. (2016a). Networks and the macroeconomy: An empirical exploration. *NBER Macroeconomics Annual*, 30(1):273–335.

Acemoglu, D., Akcigit, U., and Kerr, W. R. (2016b). Innovation network. *Proceedings of the National Academy of Sciences*, 113(41):11483–11488.

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

Arora, A., Belenzon, S., and Sheer, L. (2021). Knowledge spillovers and corporate investment in scientific research. *American Economic Review*, 111(3):871–898.

Arque-Castells, P. and Spulber, D. F. (2022). Measuring the private and social returns to r&d: Unintended spillovers versus technology markets. *Journal of Political Economy*, 130(7):1860–1918.

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. (2023). Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2):1893–1916.

Barwick, P. J., Chen, S., Fu, C., and Li, T. (2026). Digital distractions with peer influence: The

impact of mobile app usage on academic and labor market outcomes. *The Quarterly journal of economics*, 141(1):1–49.

Belloni, A. and Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19(2):521–547.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2):608–650.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, pages 1165–1188.

Bernstein, J. I. and Nadiri, M. I. (1988). Interindustry R&D spillovers, rates of return, and production in high-tech industries. *American Economic Review*, 78(2):429–434.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.

Camara, S. (2025). Spillovers of us interest rates: Monetary policy & information effects. *Journal of International Economics*, 154:104059.

Chang, J., Chen, X., and Wu, M. (2024). Central limit theorems for high dimensional dependent data. *Bernoulli*, 30(1):712–742.

Chen, Y., Chi, Y., Fan, J., and Ma, C. (2021). Spectral methods for data science: A statistical perspective. *Foundations and Trends in Machine Learning*, 14(5):566–806.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309 – 2352.

Chernozhukov, V., Chetverikov, D., and Koike, Y. (2023). Nearly optimal central limit theorem and

bootstrap approximations in high dimensions. *The Annals of Applied Probability*, 33(3):2374–2425.

Chudik, A. and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of econometrics*, 188(2):393–420.

Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 14(1).

Cohen, W. M. and Levinthal, D. A. (1989). Innovation and learning: The two faces of R&D. *The Economic Journal*, 99(397):569–596.

Cui, G., Norkutė, M., Sarafidis, V., and Yamagata, T. (2022). Two-stage instrumental variable estimation of linear panel data models with interactive effects. *The Econometrics Journal*, 25(2):340–361.

Duflo, E., Dupas, P., and Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American economic review*, 101(5):1739–1774.

Eberhardt, M., Helmers, C., and Strauss, H. (2013). Do spillovers matter when estimating private returns to r&d? *Review of Economics and Statistics*, 95(2):436–448.

Eberhardt, M. and Teal, F. (2013). Structural change and cross-country growth empirics. *The World Bank Economic Review*, 27(2):229–271.

Eberhardt, M. and Teal, F. (2020). The magnitude of the task ahead: Macro implications of heterogeneous technology. *Review of Income and Wealth*, 66(2):334–360.

Ertur, C. and Musolesi, A. (2017). Weak and strong cross-sectional dependence: A panel data analysis of international technology diffusion. *Journal of Applied Econometrics*, 32(3):477–503.

Fan, J., Liao, Y., and Mincheva, M. (2011). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of statistics*, 39(6):3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4):603–680.

Fan, J., Lou, Z., and Yu, M. (2024). Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, 119(546):1076–1088.

Gao, Z. and Tsay, R. S. (2024). Supervised dynamic pca: Linear dynamic forecasting with many predictors. *Journal of the American Statistical Association*, pages 1–15.

Griliches, Z. (1981). Market value, r&d, and patents. *Economics letters*, 7(2):183–187.

Hansen, C. and Liao, Y. (2019). The factor-lasso and k-step bootstrap approach for inference in high-dimensional economic applications. *Econometric Theory*, 35(3):465–509.

Higgins, A. and Martellosio, F. (2023). Shrinkage estimation of network spillovers with factor structured errors. *Journal of Econometrics*, 233(1):66–87.

Huang, D., Jiang, F., Li, K., Tong, G., and Zhou, G. (2022). Scaled pca: A new approach to dimension reduction. *Management Science*, 68(3):1678–1695.

Huang, J., Ma, S., and Zhang, C.-H. (2008). Adaptive lasso for sparse high-dimensional regression models. *statistica sinica*, pages 1603–1618.

Ilzetzki, E. and Jin, K. (2021). The puzzling change in the international transmission of us macroeconomic policy shocks. *Journal of international economics*, 130:103444.

Jaffe, A. B. (1986). Technological opportunity and spillovers of r&d: Evidence from firms' patents, profits, and market value. *The American Economic Review*, 76(5):984–1001.

Karabiyik, H., Reese, S., and Westerlund, J. (2017). On the role of the rank condition in cce estimation of factor-augmented panel regressions. *Journal of Econometrics*, 197(1):60–64.

Lu, X. and Su, L. (2025). Two-way mean group estimators for heterogeneous panel models with fixed t. *arXiv preprint arXiv:2508.10302*.

Lucking, B., Bloom, N., and Van Reenen, J. (2019). Have r&d spillovers declined in the 21st century? *Fiscal Studies*, 40(4):561–590.

Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Unpublished Manuscript. CEMFI, Madrid*, 1.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Merlevède, F., Peligrad, M., and Rio, E. (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields*, 151(3-4):435–474.

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

Okui, R., Sun, Y., and Wang, W. (2025). Recovering latent linkage structures and spillover effects with structural breaks in panel data models. *arXiv preprint arXiv:2501.09517*.

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

Ritzwoller, D. (2025). Regression adjustments for disentangling spillover effects. Stanford Mimeo.

Rücker, M., Vogt, M., Linton, O., and Walsh, C. (2025). Estimation and inference in high-dimensional panel data models with interactive fixed effects. *Quantitative Economics*, 16(4):1457–1509.

Rudelson, M. and Vershynin, R. (2013). Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly journal of economics*, 116(2):681–704.

van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166 – 1202.

Vershynin, R. (2024). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press.

Westerlund, J. (2018). Cce in panels with general unknown factors. *The Econometrics Journal*, 21(3):264–276.

Westerlund, J. and Urbain, J.-P. (2015). Cross-sectional averages versus principal components. *Journal of Econometrics*, 185(2):372–377.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1):217–242.

Zhao, P. and Yu, B. (2006). On model selection consistency of lasso. *The Journal of Machine Learning Research*, 7:2541–2563.

Appendix A: Observed interaction channels: model and inference

Observed measures of economic proximity allow researchers to specify the channels through which spillovers may operate. For example, technological similarity and product-market overlap provide distinct measures of firms' exposure to other firms' R&D. Even when these measures are available, common shocks can affect both outcomes and spillover-generating covariates, confounding the estimated effects. This appendix develops the factor-projection approach for this structured setting.

The observed-channel specification replaces the bilateral coefficients of the latent-network model with a finite set of coefficients on proximity-weighted covariates. Under the maintained specification and exogeneity conditions, factor projection followed by pooled least squares permits estimation and inference on these coefficients and on private effects. Their interpretation as distinct economic mechanisms additionally depends on the validity of the chosen channel specification. I present the model, estimation procedure, and asymptotic theory, then evaluate finite-sample performance and revisit the firm-level R&D application of Bloom et al. (2013) and Lucking et al. (2019).

A.1 Model

When spillover channels are observed, the bilateral spillover coefficient is parameterized as

$$\gamma_{ij} = \sum_{l=1}^{L} \gamma_l \omega_{l,ij},$$

where $\omega_{l,ij}$ is an observed proximity measure for channel l. Let $\mathbf{\Omega}_l = (\omega_{l,ij})_{i,j=1}^{N}$ have zero diagonal. The outcome equation is

$$y_{it} = \rho x_{it} + \sum_{l=1}^{L} \gamma_l \sum_{j \neq i}^{N} \omega_{l,ij} x_{jt} + \sum_{k=1}^{K} \beta_k w_{it,k} + e_{it}. \tag{A.1}$$

As in the main model, $e_{it} = f_t' \lambda_i + \varepsilon_{it}$, where f_t collects latent common shocks and λ_i allows their effects to differ across units. The parameters of interest are the private effects ρ, channel-specific spillover effects γ_l and the direct effects of covariates β_k.

A.2 Estimation

Stacking over time, the model is

$$\mathbf{Y}_i = \sum_{l=1}^{L} \mathbf{X}\boldsymbol{\Omega}_{l,i:} \cdot \gamma_l + \mathbf{X}_i \cdot \rho + \sum_{k=1}^{K} \mathbf{W}_{i,k} \cdot \beta_k + \mathbf{F}\lambda_i' + \boldsymbol{\varepsilon}_i, \quad i = 1, \ldots, N, \tag{A.2}$$

$$\mathbf{X} = \mathbf{F}\boldsymbol{\eta}_x' + \mathbf{F}_x\mathbf{B}_x' + \mathbf{U}_x, \tag{A.3}$$

$$\mathbf{W}_k = \mathbf{F}\boldsymbol{\eta}_{w,k}' + \mathbf{F}_{w,k}\mathbf{B}_{w,k}' + \mathbf{U}_{w,k}, \quad k = 1, \ldots, K, \tag{A.4}$$

where $\boldsymbol{\Omega}_{l,i:} = (\omega_{ij,l})_{j=1}^{N} \in \mathbb{R}^{N\times 1}$.

For this specification, factor projection is implemented by PCA on summed covariates. Define $\mathbf{Z} := \mathbf{X} + \sum_{k=1}^{K} \mathbf{W}_k$. By the factor structure in (A.3)–(A.4),

$$\mathbf{Z} = \mathbf{F}\left(\boldsymbol{\eta}_x + \sum_{k=1}^{K} \boldsymbol{\eta}_{w,k}\right)' + \mathbf{F}_x\mathbf{B}_x' + \sum_{k=1}^{K} \mathbf{F}_{w,k}\mathbf{B}_{w,k}' + \mathbf{U}_x + \sum_{k=1}^{K} \mathbf{U}_{w,k}. \tag{A.5}$$

Algorithm 4 (PCA on summed covariates)**.**

1. *Compute* $\mathbf{Z} = \mathbf{X} + \sum_{k=1}^{K} \mathbf{W}_k$.
2. *Let* $\widehat{\mathbf{G}}$ *be* $\sqrt{T}$ *times the eigenvectors corresponding to the* r_G *largest eigenvalues of* $\mathbf{Z}\mathbf{Z}'/(NT)$.

With $\widehat{\mathbf{G}}$, define $\mathbf{M}_{\widehat{\mathbf{G}}} = \mathbf{I}_T - \widehat{\mathbf{G}}(\widehat{\mathbf{G}}'\widehat{\mathbf{G}})^{-}\widehat{\mathbf{G}}'$, where the superscript $-$ denotes the Moore–Penrose inverse. After projection by $\mathbf{M}_{\widehat{\mathbf{G}}}$, pooled OLS yields:

Algorithm 5 (Estimation with observed spillover channels)**.**

1. *(Factor projection) Estimate* $\mathbf{G}$ *by Algorithm 4 and compute* $\mathbf{M}_{\widehat{\mathbf{G}}}$.
2. *(Pooled OLS) Collect the regressors and coefficients as*

$$\check{\mathbb{Z}}_i = [\mathbf{X}\boldsymbol{\Omega}_{1,i:}, \ldots, \mathbf{X}\boldsymbol{\Omega}_{L,i:}, \mathbf{X}_i, \mathbf{W}_{1,i}, \ldots, \mathbf{W}_{K,i}],$$

$$\delta = (\gamma_1, \ldots, \gamma_L, \rho, \beta_1, \ldots, \beta_K)'.$$

Then

$$\widehat{\delta} = \left(\sum_{i=1}^{N} \check{\mathbb{Z}}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\check{\mathbb{Z}}_i\right)^{-1} \sum_{i=1}^{N} \check{\mathbb{Z}}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{Y}_i. \tag{A.6}$$

Remark A.1 (PCA choices across latent and structured cases)**.** *The latent-network estimators in Algorithm 2 use Algorithm 1 (PCA on each covariate), while the structured special case in Algorithm 5 uses Algorithm 4 (PCA on summed covariates). In finite samples, PCA-Sum is more parsimonious and tends to be more efficient in short panels, whereas PCA-Each is typically more robust when T is relatively large or when covariates have heterogeneous factor structures. The distinction is not only finite-sample. If Algorithm 4 were used in Algorithms 2 and 3, valid inference for the private-effect estimator would require the stronger rate condition $T/N \to 0$, which would be stronger than the rate condition required for LASSO estimators in Theorem 4.1. By contrast, with Algorithm 1, private-effect inference is valid under weaker conditions in Section 4.*

Remark A.2 (Nonlinear spillover measures)**.** *In applications such as Bloom et al. (2013), Arque-Castells and Spulber (2022), and Lucking et al. (2019), nonlinear spillover indices can be treated as additional regressors in the same projection-based procedure. Let*

$$z_{it} = \big(\log(Spill_{it}),\ \log G_{it},\ \log X'_{it}\big)'.$$

Then estimation proceeds as in Section A.2 after factor extraction from the factor-rich regressor.

A.3 Theory

Define

$$\check{\mathbb{U}}_i = (\mathbf{U}_x\mathbf{\Omega}_{1,i:}, \ldots, \mathbf{U}_x\mathbf{\Omega}_{L,i:}, \mathbf{U}_x, \mathbf{U}_{w,1,i}, \ldots, \mathbf{U}_{w,K,i}) \in \mathbb{R}^{T\times(L+K+1)}.$$

The following assumption imposes regularity on the observed interaction matrices and on the second moment of the projected regressors.

Assumption 6. *(Observed spillover matrices) The observed spillover matrices $\mathbf{\Omega}_l$ for $l = 1, \ldots, L$ satisfy:*

1. *For each $l = 1, \ldots, L$, $\|\mathbf{\Omega}_l\|_\infty < M$ for some constant $M < \infty$;*
2. *For each $i = 1, \ldots, N$, the matrix $\frac{1}{T}\check{\mathbb{U}}_i'\check{\mathbb{U}}_i$ is positive definite almost surely with its minimum eigenvalue bounded away from zero by some constant $c > 0$.*

The first condition rules out explosive network amplification and ensures that the variance of the projected regressors is well behaved. The second condition is a standard full-rank requirement guaranteeing identification after removing the factor space.

Theorem A.1. *Under Assumptions 1–5 in Section 4.1 and Assumption 6 above, and if* $T\log(N)/N \to 0$*, the estimator in* (A.6) *satisfies*

$$\sqrt{NT}\left(\widehat{\delta}-\delta\right) \xrightarrow{d} N\left(0, \mathbf{\Sigma}_{\check{\mathbb{U}}\boldsymbol{\varepsilon}}\right), \tag{A.7}$$

where

$$\Sigma_{\check{\mathbb{U}}\varepsilon} = \left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\check{\mathbb{U}}_{it}\check{\mathbb{U}}_{it}'\right]\right)^{-1}\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sigma_{\varepsilon,i}^2\mathbb{E}\left[\check{\mathbb{U}}_{it}\check{\mathbb{U}}_{it}'\right]\right)\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\check{\mathbb{U}}_{it}\check{\mathbb{U}}_{it}'\right]\right)^{-1}.$$

Corollary A.1. *Let* $\widehat{\boldsymbol{\varepsilon}}_i = \mathbf{M}_{\widehat{\mathbf{G}}}\left(\mathbf{Y}_i - \check{\mathbb{Z}}_i\widehat{\delta}\right)$*,* $\widehat{\sigma}_{\varepsilon,i} = \frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i$*, and* $\widehat{\mathbf{\Sigma}}_{\check{\mathbb{U}}} = \frac{1}{NT}\sum_i \check{\mathbb{Z}}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\check{\mathbb{Z}}_i$*. Define*

$$\widehat{\mathbf{\Sigma}}_{\check{\mathbb{U}}\boldsymbol{\varepsilon}} := \widehat{\mathbf{\Sigma}}_{\check{\mathbb{U}}}^{-1}\left(\frac{1}{NT}\sum_i \check{\mathbb{Z}}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\check{\mathbb{Z}}_i\widehat{\sigma}_{\varepsilon,i}\right)\widehat{\mathbf{\Sigma}}_{\check{\mathbb{U}}}^{-1}. \tag{A.8}$$

Under the same conditions as Theorem A.1,

$$\widehat{\mathbf{\Sigma}}_{\check{\mathbb{U}}\boldsymbol{\varepsilon}} \xrightarrow{p} \mathbf{\Sigma}_{\check{\mathbb{U}}\boldsymbol{\varepsilon}}.$$

A.4 Simulation

I examine the finite-sample performance of inference with an observed spillover matrix. Across these designs, I set $K = 1$ and $(r, r_x, r_w) = (3, 1, 1)$. The models are generated as follows:

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\cdot\rho + \mathbf{X}\mathbf{\Omega}'\cdot\gamma + \mathbf{W}\cdot\beta + \mathbf{1}_T\mathbf{a}' + \mathbf{d}\mathbf{1}_N' + \mathbf{f}\boldsymbol{\lambda}' + \boldsymbol{\varepsilon}\\
\mathbf{X} &= \mathbf{1}_T\mathbf{a}_x' + \mathbf{d}\boldsymbol{\eta}_{d,x}' + \mathbf{f}\boldsymbol{\eta}_{f,x}' + \mathbf{f}_x\mathbf{b}_x' + \mathbf{U}_x,\\
\mathbf{W} &= \mathbf{1}_T\mathbf{a}_w' + \mathbf{d}\boldsymbol{\eta}_{d,w}' + \mathbf{f}\boldsymbol{\eta}_{f,w}' + \mathbf{f}_w\mathbf{b}_w' + \mathbf{U}_w.
\end{aligned}$$

The coefficients are set to $\rho = \gamma = \beta = 1$. The vectors $\mathbf{a} \in \mathbb{R}^N$ and $\boldsymbol{\lambda} \in \mathbb{R}^N$ represent individual fixed effects and factor loadings in the outcome equation, respectively. Both are generated as $\mathbf{\Sigma}_N^{1/2}\widetilde{\mathbf{a}}$, where $\widetilde{\mathbf{a}} \overset{i.i.d.}{\sim} \text{Unif}[-1, 1]$ and $\mathbf{\Sigma}_N$ is a Toeplitz covariance matrix with (i, j)-th element $0.5^{|i-j|}$. Define the loading matrices $\mathbf{\Lambda} = [\mathbf{a}, \mathbf{1}_N, \boldsymbol{\lambda}]$, $\mathbf{\Lambda}_x = [\mathbf{a}_x, \boldsymbol{\eta}_{d,x}, \boldsymbol{\eta}_{f,x}]$, and $\mathbf{\Lambda}_w = [\mathbf{a}_w, \boldsymbol{\eta}_{d,w}, \boldsymbol{\eta}_{f,w}]$ for $\mathbf{Y}$, $\mathbf{X}$, and $\mathbf{W}$, respectively. To induce endogeneity through common factors, I generate $\mathbf{\Lambda}_x = \mathbf{\Lambda} + \widetilde{\mathbf{\Lambda}}_x$ and $\mathbf{\Lambda}_w = \mathbf{\Lambda} + \widetilde{\mathbf{\Lambda}}_w$, where the perturbations $\widetilde{\mathbf{\Lambda}}_x$ and $\widetilde{\mathbf{\Lambda}}_w$ have i.i.d. entries from $\text{Unif}[-1, 1]$.

The full set of common factors is given by $\mathbf{G} = [\mathbf{d}, \mathbf{f}, \mathbf{f}_x, \mathbf{f}_w] \in \mathbb{R}^{T\times 4}$, generated as $\mathbf{\Sigma}_f^{1/2}\widetilde{\mathbf{G}}$, where $\widetilde{\mathbf{G}}$ has i.i.d. $N(1,1)$ entries and $\mathbf{\Sigma}_f$ is a Toeplitz covariance matrix with (i,j)-th element $0.5^{|i-j|}$. This construction ensures that the regressors and the outcome share correlated factor structures. The idiosyncratic errors ε_{it} are drawn from $N(0,\sigma_i^2)$ with heterogeneous variances $\sigma_i^2 \overset{i.i.d.}{\sim} \text{Unif}[0.5, 2]$. The idiosyncratic components in the regressor factor models are generated as $\mathbb{U}_{it} = (u_{x,it}, u_{w,it})' = \sigma_i\widetilde{\mathbb{U}}_{it}$, where $\widetilde{\mathbb{U}}_{it} \sim N(0, \mathbf{\Sigma}_{\mathbb{U}})$ and $\mathbf{\Sigma}_{\mathbb{U}}$ is a 2×2 Toeplitz covariance matrix with off-diagonal elements equal to 0.5. Although the DGP includes two-way fixed effects, the simulated data are not demeaned prior to estimation, allowing the proposed procedures to remove these components through factor projection.

I consider two settings for $\mathbf{\Omega}$ are considered:

(1) (Sparse) The i^{th} row of $\mathbf{\Omega}$ is generated such that $\omega_{ii-1} = \omega_{ii+1} = 0.5$ and all other entries are zero.

(2) (Dense) $\mathbf{\Omega}$ is generated as the Toeplitz matrix with its (i,j)-th entry being $0.5^{|i-j|}$ for $i \neq j$ and 0 for $i = j$.

(N,T) are set in $\{50, 200, 1000\} \times \{30, 100, 200\}$. In each design, I construct the 95% confidence intervals for (γ, ρ, β) using the proposed estimator in Algorithm 5, with factor projection implemented by Algorithm 4. For comparison, I also estimate the model with two-way fixed effects estimators, which are obtained by pooled OLS of the doubly-demeaned data. The histogram of the proposed estimator (IFE) and the two-way fixed effects estimator (TWFE) at $(T,N) = (50, 100)$ under the dense $\mathbf{\Omega}$ setting is shown in Figure A.1. It can be observed that the proposed estimator is approximately normal with the proposed variance estimator, while the two-way fixed effects estimator is seriously biased.

Table A.1 reports the empirical coverage rates of the confidence intervals under the proposed estimator. Overall, coverage is close to the nominal 95% level across designs. In the sparse-network design, the coverage rates lie mostly between 0.90 and 0.95 for all three parameters. In the dense-network design, coverage remains accurate in most cases, although it deteriorates somewhat when N is small and T is relatively large due to large T/N. Overall, the finite-sample results are consistent with Theorem A.1 and support the validity of the proposed inference procedure in the structured observed-network case.

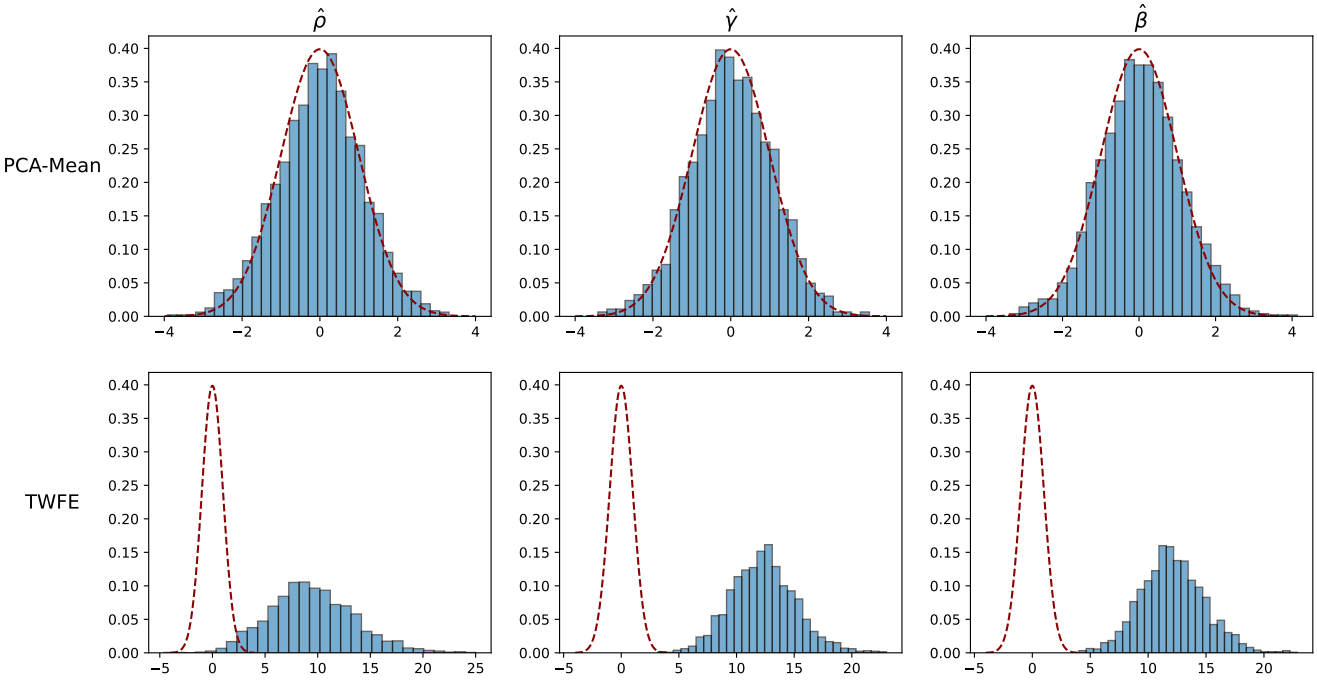


Figure A.1: Histogram of the proposed estimator (IFE) and the two-way fixed effects estimator (TWFE) for parameters in the homogeneous panel model with observed spillover channels at $(T, N) = (50, 100)$ under the dense $\mathbf{\Omega}$ setting. The red dashed line represents the standard normal density.

N	50	50	50	200	200	200	1000	1000	1000
T	30	100	200	30	100	200	30	100	200
Sparse $\mathbf{\Omega}$									
γ	0.93	0.92	0.91	0.93	0.94	0.94	0.93	0.95	0.93
ρ	0.93	0.93	0.9	0.92	0.95	0.94	0.93	0.94	0.95
β	0.92	0.92	0.9	0.93	0.94	0.95	0.92	0.95	0.95
Dense $\mathbf{\Omega}$									
γ	0.9	0.88	0.84	0.93	0.93	0.93	0.92	0.95	0.94
ρ	0.92	0.93	0.91	0.92	0.94	0.94	0.94	0.94	0.94
β	0.92	0.92	0.9	0.93	0.94	0.95	0.92	0.95	0.94

Table A.1: Coverage rates of the confidence intervals for parameters under the proposed estimator in the homogeneous panel model with observed spillover channels.

A.5 Revisiting technology spillovers and market competition

Understanding how R&D affects firm performance through both own innovation and external spillovers is a central question in the economics of innovation. Beyond the private returns to R&D, firms are influenced by the innovative activities of technologically related competitors and by intensified product-market competition. Using firm-level panel data, Bloom et al. (2013) show that R&D generates positive knowledge spillovers through technological proximity, but also negative business-stealing effects through product-market rivalry, with the former dominating so that the social returns to R&D exceed the private returns. These findings are confirmed and extended by Lucking et al. (2019) using a longer sample and richer measures of technological and market proximity.

A key empirical challenge in this literature is that both R&D investment and firm value are driven by unobserved aggregate and sectoral shocks, such as macroeconomic conditions or waves of technological change. If these common factors have heterogeneous effects across firms, standard fixed-effects specifications may confound common shocks with spillover effects, leading to biased estimates and potentially overstating the role of network interactions and the private effects. This concern is the correlated effects issue discussed by Manski (1993), whereby correlated outcomes may arise from shared environments rather than genuine cross-sectional spillovers.

A.5.1 Data and specification

To address this issue, I re-examine the technology-spillover and market-competition channels of R&D using the proposed exogenous-spillover model with common correlated effects. I employ the same firm-level dataset as in Lucking et al. (2019), constructed from Compustat Fundamentals, Compustat Segments, and the NBER Patent Data Project, which contains annual accounting information for publicly traded U.S. firms from 1980-2015 and patent data from 1976-2006. This dataset provides an ideal laboratory for assessing how explicitly controlling for latent common factors alters the estimated private returns, knowledge spillovers, and competitive effects of R&D. Table A.2 reports descriptive statistics for the main variables used in the analysis.

I evaluate the following linearized value function introduced by Griliches (1981) and modified by Bloom et al. (2013), augmented with common correlated effects:

$$
\begin{aligned}
\log\left(\frac{V}{A}\right)_{it} &= a_i + d_t + \gamma_1 \log SpillTech_{it} + \gamma_2 \log SpillMkt_{it} \\
&\quad + \rho \cdot \left(\frac{G}{A}\right)_{it} + x'_{it}\beta + f'_t\lambda_i + \varepsilon_{it}, \\
SpillTech_{it} &= \sum_{j\neq i}\left(\frac{G}{A}\right)_{jt} \omega_{ij}^{(tech)}, \\
SpillMkt_{it} &= \sum_{j\neq i}\left(\frac{G}{A}\right)_{jt} \omega_{ij}^{(mkt)}, \\
z_{it} &= a_{z,i} + d_{z,t} + f'_t\eta_{z,i} + f'_{z,t}b_{z,i} + u_{z,it}, \\
z &\in \left\{\log SpillTech_{it}, \log SpillMkt_{it}, \left(\frac{G}{A}\right), x_1, \ldots, x_K\right\},
\end{aligned}
\tag{A.9}
$$

where V/A is Tobin's Q, which equals the ratio of market value to the stock of non R&D assets, and G is the R&D stock. $\omega_{ij}^{(tech)}$ and $\omega_{ij}^{(mkt)}$ are the measures of technological proximity and product market proximity between firms i and j that will be specified later. x_{it} contains other control variables including employment and industry sales as well as the polynomial terms $\left(\frac{G}{A}\right)_{it}^{p}$ up to $p = 6$ to improve the accuracy of the approximation[3].

Following Bloom et al. (2013) and Lucking et al. (2019), I construct the technology proximity

[3]The private effect term in the linearized model by Griliches (1981) is $\log\left(1 + \rho \cdot \frac{G}{A}\right)$. By Taylor expansion, $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$. Higher order terms can be included for higher accuracy.

$\omega_{ij}^{(tech)}$ using the Jaffe (1986) measure and Mahalanobis generalization introduced in Bloom et al. (2013). Specfically, Let $T_i = (T_{i1}, \ldots, T_{i426})'$ to be the vector of the share of firm i's patents in each of the 426 technology classes. The Jaffe measure of technology proximity between firm i and j is defined as the correlation between their patenting activities:

$$\omega_{ij}^{(tech,J)} = \frac{T_i' T_j}{\|T_i\|_2 \|T_j\|_2}.$$

The Mahalanobis generalization further accounts for the covariance structure of the technology classes. Let $T = (T_i, \ldots, T_N)' \in \mathbb{R}^{N \times 426}$ be the matrix of patenting shares of all firms, and denote $\widetilde{T}$ to be the row-centered version of T and $\check{T}$ to be the column-centered version of T. The Mahalanobis technology proximity is defined as:

$$\mathbf{\Omega}^{(tech,M)} = \widetilde{T} \left(\check{T}' \check{T} \right) \widetilde{T}'.$$

The product market proximity $\omega_{ij}^{(mkt)}$ is constructed analogously using firms' sales shares across different product segments as in Bloom et al. (2013) and Lucking et al. (2019). However, as discussed in Lucking et al. (2019), A time-invariant measure based on full-sample sales may be endogenous, since future product-market positions are likely correlated with current performance, while a purely pre-sample measure may be misleading if firms reorient their product mix over time. To address these concerns, I follow Lucking et al. (2019) and construct time-varying market-proximity measures using firms's sales shares from the preceding five years. Let $S_{it} = (S_{i1t}, \ldots, S_{iMt})'$ be the vector of firm i's sales shares across M product segments 5 years before year t. The Jaffe measure of product market proximity is defined as:

$$\mathbf{\Omega}_t^{(mkt,J)} = \widetilde{S}_t \widetilde{S}_t',$$

and the Mahalanobis generalization is defined as:

$$\mathbf{\Omega}_t^{(mkt,M)} = \widetilde{S}_t \left(\check{S}_t' \check{S}_t \right) \widetilde{S}_t'.$$

A.5.2 Estimation and results

Before estimating the homogeneous panel model in (A.9), all variables are first residualized by firm and time dummies to remove firm and time fixed effects. The model is then estimated by Algorithm

5, treating the nonlinear spillover terms $\log SpillTech_{it}$ and $\log SpillMkt_{it}$ as observed covariates, as discussed in Remark A.2. Latent factors are estimated by Algorithm 4 for higher efficiency in the current short panel data. The number of latent factors is selected by the eigenvalue-ratio criterion of Ahn and Horenstein (2013).

For comparison, I also report estimates from the conventional two-way fixed effects (TWFE) estimator, which controls only for additive firm and time effects and therefore ignores cross-sectional dependence induced by unobserved common shocks.

Table A.3 reports estimates of the technology-spillover, market-competition, and private R&D effects under alternative proximity measures and model specifications. Columns labeled "TWFE" correspond to the conventional two-way fixed effects estimator, while columns labeled "IFE" report results from the proposed interactive fixed effects approach that controls for unobserved common shocks. The row "Incl. poly." indicates whether high-order polynomials in R&D stock (up to order six) are included as in Bloom et al. (2013) and Lucking et al. (2019).

Consistent with Bloom et al. (2013) and Lucking et al. (2019), the TWFE estimates indicate statistically significant positive technology spillovers and negative market-competition effects, with social returns to R&D exceeding private returns. Once common correlated effects are controlled for using the IFE estimator, the magnitudes of both spillover effects are substantially attenuated.

For knowledge spillovers, the IFE estimates remain statistically significant under the Mahalanobis proximity measure, but lose significance under the Jaffe measure. This contrast suggests that conventional proximity measures may differ in their ability to capture economically meaningful spillover channels once common shocks are accounted for.

In contrast, the market-competition effects become statistically insignificant under IFE across specifications, despite being strongly negative under TWFE. This pattern indicates that the apparent competition effects in standard estimates may largely reflect exposure to common shocks rather than genuine competitive spillovers.

In contrast, the estimated private return to R&D does not decline after accounting for common factors. In several specifications, the IFE estimates are comparable to or even larger than their TWFE counterparts, suggesting that the bias induced by common correlated effects may differ across private and spillover components.

Overall, these results highlight the importance of accounting for latent common shocks when estimating R&D spillovers, as failing to do so can lead to systematically inflated estimates of

cross-firm interactions.

Table A.2: Descriptive statistics of key variables

	Mean	Median	Std
Tobin's Q	3.41	1.83	4.26
R&D Stock	915.36	46.25	3640.29
Capital	2593.43	122.93	10212.88
R&D Stock / Capital	1.05	0.4	1.8
Employees	15300.71	1692	43373.05

Notes: This table reports the descriptive statistics of key variables used in Section A.5. It contains annual observations of 3,315 firms from 1980 to 2015.

A.5.3 Robustness: lagged spillovers

As a robustness check, I estimate regressions with lagged spillover variables to address the potential concern of reverse causality. I consider the following specification:

$$
\begin{aligned}
\log\left(\frac{V}{A}\right)_{it} &= a_i + d_t + \gamma_1 \log SpillTech_{i(t-1)} + \gamma_2 \log SpillMkt_{i(t-1)} \\
&\quad + \rho \cdot \left(\frac{G}{A}\right)_{i(t-1)} + \sum_{q=2}^{7} \rho_q \left(\frac{G}{A}\right)^q_{i(t-1)} + x'_{it}\beta_c + x'_{i(t-1)}\beta_l + f'_t\lambda_{c,i} + f'_{t-1}\lambda_{l,i} + \varepsilon_{it}, \\
SpillTech_{it} &= \sum_{j\neq i} \left(\frac{G}{A}\right)_{jt} \omega_{ij}^{(tech)}, \\
SpillMkt_{it} &= \sum_{j\neq i} \left(\frac{G}{A}\right)_{jt} \omega_{ij}^{(mkt)}, \\
z_{it} &= a_{z,i} + d_{z,t} + f'_t\eta_{z,i} + f'_{z,t}b_{z,i} + u_{z,it}, \\
z &\in \left\{\log SpillTech, \log SpillMkt, \left(\frac{G}{A}\right), x_1, \ldots, x_K\right\},
\end{aligned}
\tag{A.10}
$$

Compared with Model (A.9) in Section A.5, the main difference is that the spillover variables are lagged by one period. In addition, the control variables and common factors are also included in both contemporaneous and lagged forms. For the estimation, common factors are first estimated contemporaneously from z_{it}, and then estimated factors are included in the second step regression with both contemporaneous and lagged forms.

Table A.3: Coefficient estimates of market value regressions

	TWFE	IFE	TWFE	IFE	TWFE	IFE	TWFE	IFE
Log $SpillTech$	0.635	0.283	0.398	0.088	0.645	0.341	0.413	0.141
	(0.043)	(0.123)	(0.034)	(0.104)	(0.044)	(0.124)	(0.034)	(0.105)
Log $SpillMkt$	-0.128	-0.026	-0.085	-0.034	-0.112	-0.026	-0.077	-0.04
	(0.015)	(0.031)	(0.011)	(0.02)	(0.015)	(0.032)	(0.011)	(0.021)
G/A	0.668	0.734	0.614	0.741	0.108	0.15	0.108	0.149
	(0.108)	(0.037)	(0.108)	(0.037)	(0.006)	(0.007)	(0.006)	(0.007)
Prox. Measure	Mahal.	Mahal.	Jaffe	Jaffe	Mahal.	Mahal.	Jaffe	Jaffe
Inclu. Poly.	Yes	Yes	Yes	Yes	No	No	No	No

Notes: This table reports the coefficient estimates of model (A.9) using different estimators and different proximity measures. The dependent variable is the log Tobin's Q. The independent variables of interest are the log technology spillover ($\log SpillTech$), the log market competition effect ($\log SpillMkt$), and the R&D intensity (G/A). The TWFE columns report the estimates from the standard two-way fixed effects estimator that ignores the common correlated effects. The IFE columns report the estimates from the proposed interactive fixed effects estimator, after demeaning the data to remove unobserved heterogeneity. The proximity measures used are Jaffe and Mahalanobis (Mahal.) as discussed in the main text. The columns with "Incl. poly." equal to "Yes" include polynomial terms of R&D intensity up to order 6 as additional controls as in Bloom et al. (2013) and Lucking et al. (2019). Standard errors are reported in parentheses.

The results, reported in Table A.4, are broadly consistent with those obtained from the contemporaneous specification. In particular, controlling for common correlated effects continues to attenuate both types of spillover estimates. The knowledge spillover effects remain positive and statistically significant, although their magnitudes are reduced relative to TWFE estimates. In contrast, the market competition effects are substantially weakened, becoming close to zero (in some cases slightly positive), and are no longer statistically significant.

These findings suggest that the main conclusions in Section A.5 are possibly not biased by reverse causality. More importantly, they reinforce the role of common correlated effects as a primary source of bias in conventional estimates, highlighting the importance of accounting for latent common factors for credible inference on spillover effects.

Table A.4: Coefficient estimates of market value regressions with lagged spillover variables

	TWFE	IFE	TWFE	IFE	TWFE	IFE	TWFE	IFE
log $SpillTech$	0.493	0.146	0.327	0.15	0.509	0.158	0.345	0.167
	(0.042)	(0.015)	(0.033)	(0.012)	(0.042)	(0.015)	(0.033)	(0.012)
log $SpillSIC$	-0.11	-0.015	-0.081	0.023	-0.1	0.004	-0.076	0.032
	(0.015)	(0.037)	(0.01)	(0.024)	(0.016)	(0.038)	(0.01)	(0.024)
G/A	0.374	0.487	0.373	0.482	0.116	0.089	0.119	0.088
	(0.027)	(0.038)	(0.027)	(0.038)	(0.005)	(0.008)	(0.005)	(0.008)
Prox. Measure	Mahal.	Mahal.	Jaffe	Jaffe	Mahal.	Mahal.	Jaffe	Jaffe
Inclu. Poly.	Yes	Yes	Yes	Yes	No	No	No	No

Notes: This table reports the coefficient estimates of model (A.10) using different estimators and different proximity measures. The dependent variable is the log Tobin's Q. The independent variables of interest are the log technology spillover (log $SpillTech$), the log market competition effect (log $SpillMkt$), and the R&D intensity (G/A). The TWFE columns report the estimates from the standard two-way fixed effects estimator that ignores the common correlated effects. The IFE columns report the estimates from the proposed interactive fixed effects estimator, after demeaning the data to remove unobserved heterogeneity. The proximity measures used are Jaffe and Mahalanobis (Mahal.) as discussed in the main text. The columns with "Incl. poly." equal to "Yes" include polynomial terms of R&D intensity up to order 6 as additional controls as in Bloom et al. (2013) and Lucking et al. (2019). Standard errors are reported in parentheses.

Appendix B: Additional remarks on estimations

Two-way fixed effects estimators for heterogeneous panel models. As discussed in Remark 2.2, interactive fixed effects represented by $\mathbf{F}\Lambda'$ includes two-way fixed effects as a special case. For the homogeneous panel data models where the spillover channels are observed, one can incorporate two-way fixed effects in addition to the interactive fixed effects to improve efficiency, i.e.

$$e_{it} = a_i + d_t + f_t'\lambda_i + \varepsilon_{it}.$$

Two-way fixed effects can be also incorporated into the factor model for covariates $\mathbf{X}$ and $\mathbf{W}_k$ in a similar manner. One can demean the regressors $\mathbb{Z}$ and outcome $\mathbf{Y}$ variables to remove two-way fixed effects before applying the proposed estimation algorithm.

For the models where the spillover channels are unobserved, the same procedure can not be naively applied due to the heterogeneity of the pairwise spillover coefficients. Thus, regressing the outcome variable demeaned over individuals $\mathbf{Y}\mathbf{M}_{\mathbf{1}_N}$ on $N-1$ spillover characteristics demeaned over individuals $(\mathbf{X}\mathbf{M}_{\mathbf{1}_N})_{:,-i}$ does not yield the same model as demeaning over individuals in the original model. Specifically, starting from the simple model with fixed effects where spillover channel is not known, ignoring other covariates for ease of exposition, if we demean over individuals, we have

$$\begin{aligned}
\mathbf{Y}\mathbf{M}_{\mathbf{1}_N} &= \mathbf{X}\boldsymbol{\Gamma}'\mathbf{M}_{\mathbf{1}_N} + \mathbf{F}\boldsymbol{\Lambda}'\mathbf{M}_{\mathbf{1}_N} + \mathbf{d}_T\mathbf{1}_N'\mathbf{M}_{\mathbf{1}_N} + \boldsymbol{\varepsilon}\mathbf{M}_{\mathbf{1}_N} \\
&= \mathbf{X}\left(\boldsymbol{\Gamma}'\mathbf{M}_{\mathbf{1}_N}\right) + \mathbf{F}\boldsymbol{\Lambda}'\mathbf{M}_{\mathbf{1}_N} + \boldsymbol{\varepsilon}\mathbf{M}_{\mathbf{1}_N} \\
&\neq \left(\mathbf{X}\mathbf{M}_{\mathbf{1}_N}\right)\boldsymbol{\Gamma}' + \mathbf{F}\boldsymbol{\Lambda}'\mathbf{M}_{\mathbf{1}_N} + \boldsymbol{\varepsilon}\mathbf{M}_{\mathbf{1}_N},
\end{aligned}$$

while demeaning over time is still valid:

$$\begin{aligned}
\mathbf{M}_{\mathbf{1}_T}\mathbf{Y} &= \mathbf{M}_{\mathbf{1}_T}\mathbf{X}\boldsymbol{\Gamma}' + \mathbf{M}_{\mathbf{1}_T}\mathbf{F}\boldsymbol{\Lambda}' + \mathbf{M}_{\mathbf{1}_T}\mathbf{a}_N\mathbf{1}_T' + \mathbf{M}_{\mathbf{1}_T}\boldsymbol{\varepsilon} \\
&= \mathbf{M}_{\mathbf{1}_T}\mathbf{X}\boldsymbol{\Gamma}' + \mathbf{M}_{\mathbf{1}_T}\mathbf{F}\boldsymbol{\Lambda}' + \mathbf{M}_{\mathbf{1}_T}\boldsymbol{\varepsilon}.
\end{aligned}$$

Therefore, whereas two-way fixed effects can be incorporated into the model with unobserved spillover channels, one can only demean along the time dimension to remove the individual fixed effects before applying the proposed estimation algorithm. The time fixed effects can be absorbed by the common factors in the interactive fixed effects and be removed in the projection step.

Comparison with IV regressions. In the empirical literature, IV approach has been widely used to address the endogeneity issue in estimating spillover effects. IV approach is preferable

as a valid IV approach addresses general endogeneity issues, including endogeneity caused by unobserved common factors. However, valid instruments are often diffcult to find in practice as the exogeneity condition is generally untestable. In addition, IV approach could be less efficient than the proposed methods in homogeneous panel model settings if the IV itself displays strong cross-sectional dependence. Consider the following simple homogeneous panel model with interactive fixed effects:

$$\begin{aligned} y_{it} &= x_{it}\beta + f_t\lambda_i + \varepsilon_{it}, \\ x_{it} &= f_t\eta_i + g_t b_i + u_{it}, \\ z_{it} &= g_t c_i + v_{it}, \end{aligned}$$

where all variables are scalars for ease of analysis. Assume that $\mathbb{E}\left[\varepsilon_{it}|c_i, g_t, v_{it}\right] = \mathbb{E}\left[f_t|c_i, g_t, v_{it}\right] = 0$ and hence z_{it} is a valid instrument for the endogeneous variable x_{it}. The 2SLS estimator of β satisfies

$$\widehat{\beta}^{(2SLS)} - \beta = \frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}\left(f_t\lambda_i + \varepsilon_{it}\right)}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_{it}x_{it}},$$

while the denominator converges to a non-zero constant under regularity conditions, the first term of the numerator can be shown that

$$\begin{aligned} \frac{1}{NT}\sum_i\sum_t z_{it}f_t\lambda_i &= \frac{1}{NT}\sum_i\sum_t \left(g_t c_i + v_{it}\right) f_t\lambda_i = \\ &= \left(\frac{1}{N}\sum_i c_i\lambda_i\right)\left(\frac{1}{T}\sum_t g_t f_t\right) + \frac{1}{NT}\sum_i\sum_t v_{it}f_t\lambda_i, \\ &= \left(\frac{1}{N}\sum_i c_i\lambda_i\right) O_p\left(T^{-1/2}\right) + O_p\left((NT)^{-1/2}\right). \end{aligned}$$

If c_i and λ_i have non-zero means or are correlated, then $\frac{1}{N}\sum_i c_i\lambda_i = O_p(1)$ and hence the convergence rate of the 2SLS estimator is $O_p\left(T^{-1/2}\right)$, which is slower than the optimal rate $O_p\left((NT)^{-1/2}\right)$. By explicitly modeling and removing the common factors, the proposed projection-based estimators exploit both the time-series and cross-sectional information and can attain root-NT efficiency under suitable conditions.

Appendix C: Outcome-specific factors

This appendix explores the case of outcome-specific factors that enter the outcome equation but not the regressor equations.

Robustness of the original algorithm

The main text studies common factors shared by the outcome, the spillover variable, and the controls. This appendix allows additional factors that enter only the outcome equation. Under exogeneity with respect to the idiosyncratic regressor components, these factors enter the first-order scores of the original estimators without changing their convergence rates. The representation for the private-effect estimator requires the stronger rate condition $\sqrt{T/N} \to 0$.

The expanded model, written in $T \times N$ matrix form, is

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\boldsymbol{\Gamma}' + \sum_{k=1}^{K} \beta_k \mathbf{W}_k + \mathbf{F}\boldsymbol{\Lambda}' + \mathbf{F}_{\mathbf{Y}}\boldsymbol{\Lambda}_{\mathbf{Y}}' + \boldsymbol{\varepsilon} \\
\mathbf{X} &= \mathbf{F}\boldsymbol{\eta}_{\mathbf{X}}' + \mathbf{F}_{\mathbf{X}}\mathbf{B}_{\mathbf{X}}' + \mathbf{U}_{\mathbf{X}} := \mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}' + \mathbf{U}_{\mathbf{X}} \\
\mathbf{W}_k &= \mathbf{F}\boldsymbol{\eta}_{\mathbf{W}_k}' + \mathbf{F}_{\mathbf{W}_k}\mathbf{B}_{\mathbf{W}_k}' + \mathbf{U}_{\mathbf{W}_k} := \mathbf{G}_{\mathbf{W}_k}\mathbf{C}_{\mathbf{W}_k}' + \mathbf{U}_{\mathbf{W}_k}, \quad k = 1, \ldots, K,
\end{aligned}$$

where $(\boldsymbol{\Gamma})_{ij} = \gamma_{ij}$ for $i \neq j$ are spillover effects, and $(\boldsymbol{\Gamma})_{ii} = \rho$ is the homogeneous private effect of the spillover variable on the outcome. $\mathbf{F}_{\mathbf{Y}}$ is a $T \times r_{\mathbf{Y}}$ matrix of outcome-specific factors, and $\boldsymbol{\Lambda}_{\mathbf{Y}}$ is a $N \times r_{\mathbf{Y}}$ matrix of factor loadings.

Extend Assumptions 1 and 2 to $\mathbf{F}_{\mathbf{Y}}$ and $\boldsymbol{\Lambda}_{\mathbf{Y}}$. The next theorem characterizes how outcome-specific factors enter the first-order representations of the original estimators.

Theorem C1. Under the same extended assumptions as Theorem 4.2, and if $\sqrt{T/N} \to 0$, we have

$$\sqrt{T}\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right) = \frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}'\mathbf{M}_{\widehat{\mathbf{G}}}\widetilde{\boldsymbol{\varepsilon}}_i + \mathbf{o}_p(1), \tag{18}$$

where $\widetilde{\boldsymbol{\varepsilon}}_i = \mathbf{F}_{\mathbf{Y}}\boldsymbol{\Lambda}_{\mathbf{Y},i} + \boldsymbol{\varepsilon}_i$. For the private-effect estimator,

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \left(\frac{1}{NT}\sum_{i=1}^{N}\mathbb{U}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{U}_i\right)^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{U}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\left(\mathbf{L}_i + \boldsymbol{\varepsilon}_i\right) + \mathbf{o}_p(1), \tag{19}$$

where $\mathbf{L}_i$ is the ith column of

$$\mathbf{L} := \mathbf{F}_{\mathbf{Y}}\boldsymbol{\Lambda}_{\mathbf{Y}}'\mathbf{M}_{\mathbb{C}},$$

and $\mathbf{M}_{\mathbb{C}} \in \mathbb{R}^{N\times N}$ depends only on the regressor factor loadings and satisfies $\|\mathbf{M}_{\mathbb{C}}\|_2 = \mathcal{O}_p(1)$. Moreover,

$$\left\| \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbb{U}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{U}_i \right)^{-1} \right\|_2 = \mathcal{O}_p(1), \qquad \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathbb{U}_i' \mathbf{M}_{\widehat{\mathbf{G}}} (\mathbf{L}_i + \boldsymbol{\varepsilon}_i) \right\|_2 = \mathcal{O}_p(1).$$

Theorem C1 shows that outcome-specific factors do not change the convergence rates of the original estimators. For spillover estimation, they enter the composite regression error. For private-effect estimation, they contribute the additional component $\mathbf{L}_i$ to the first-order score. The representation uses the same PCA-Each projection as the estimator. The Monte Carlo exercise below evaluates the normal approximation and standard-error calibration for this projected score.

Double-refinement algorithm

If more efficient estimation is desired, I provide a double-refinement algorithm in the spirit of the single refinement studied by Cui et al. (2022).

Let $\mathbb{Z}_i = (\mathbf{X}_i, \mathbf{W}_{i,1}, \ldots, \mathbf{W}_{i,K}) \in \mathbb{R}^{T\times(K+1)}$. For a matrix $\mathbf{A}$, write $\mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{+}\mathbf{A}'$ and $\mathbf{M}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$, where $(\cdot)^{+}$ denotes the Moore–Penrose inverse. The algorithm proceeds as follows.

Step 1: Estimate the regressor factor space. Estimate the union of the factor spaces in X and W, denoted by $\widehat{\mathbf{G}}$, using Algorithm 1 in the main text. Form

$$\mathbf{M}_{\widehat{\mathbf{G}}} = \mathbf{I} - \widehat{\mathbf{G}}(\widehat{\mathbf{G}}'\widehat{\mathbf{G}})^{+}\widehat{\mathbf{G}}'.$$

Define $\dot{\mathbf{Y}} = \mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{Y}$ and analogously define $\dot{\mathbf{X}}$, $\dot{\mathbf{W}}$, and $\dot{\mathbb{Z}}_i$.

Step 2: Obtain preliminary coefficient estimates. For each i, estimate the off-diagonal spillover coefficients by plain LASSO after partialling out the private regressors:

$$\widehat{\boldsymbol{\gamma}}_{i:}^{(0)} = \underset{b\in\mathbb{R}^{N-1}}{\operatorname{argmin}} \left\{ \frac{1}{2T} \left\| \mathbf{M}_{\dot{\mathbb{Z}}_i} (\dot{\mathbf{Y}}_i - \dot{\mathbf{X}}_{-i} b) \right\|_2^2 + \lambda_i^{(0)} \|b\|_1 \right\}.$$

Given $\widehat{\boldsymbol{\Gamma}}^{(0)}$, estimate the private effects by pooled least squares:

$$\widehat{\boldsymbol{\theta}}^{(0)} = \left(\sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \left(\mathbf{Y}_i - \mathbf{X}_{-i} \widehat{\boldsymbol{\gamma}}_{i:}^{(0)} \right).$$

This preliminary stage uses neither debiasing nor post-selection OLS.

Step 3: Estimate the outcome-specific factor for the first time. Construct the ith column of the preliminary residual matrix as

$$\dot{\mathbf{r}}_i^{(0)} = \mathbf{M}_{\widehat{\mathbf{G}}} \left(\mathbf{Y}_i - \mathbf{X}_{-i} \widehat{\mathbf{\Gamma}}_{i,-i}^{(0)} - \mathbb{Z}_i \widehat{\boldsymbol{\theta}}^{(0)} \right).$$

Apply PCA to $\dot{\mathbf{r}}^{(0)} = (\dot{\mathbf{r}}_1^{(0)}, \ldots, \dot{\mathbf{r}}_N^{(0)})$ and retain $r_{\mathbf{Y}}$ components, producing $\widehat{\mathbf{F}}_{\mathbf{Y}}^{(1)}$. Set

$$\widehat{\mathbf{Q}}^{(1)} = \left[\widehat{\mathbf{G}}, \widehat{\mathbf{F}}_{\mathbf{Y}}^{(1)} \right], \qquad \mathbf{M}_{\widehat{\mathbf{Q}}^{(1)}} = \mathbf{I} - \widehat{\mathbf{Q}}^{(1)} \left(\widehat{\mathbf{Q}}^{(1)\prime} \widehat{\mathbf{Q}}^{(1)} \right)^{+} \widehat{\mathbf{Q}}^{(1)\prime}.$$

Step 4: Conduct first-refinement inference and select the spillover support. With $\ddot{\mathbf{A}}^{(1)} = \mathbf{M}_{\widehat{\mathbf{Q}}^{(1)}} \mathbf{A}$, re-estimate each row of $\mathbf{\Gamma}$ by plain LASSO and denote the result by $\widehat{\boldsymbol{\gamma}}_{i:}^{(1,L)}$. Run adaptive LASSO on the support of $\widehat{\boldsymbol{\gamma}}_{i:}^{(1,L)}$, using $1/|\widehat{\gamma}_{ij}^{(1,L)}|$ as the penalty weight:

$$\widehat{\boldsymbol{\gamma}}_{i:}^{(1,AL)} = \underset{b \in \mathbb{R}^{N-1}}{\operatorname{argmin}} \left\| \frac{1}{2NT} \mathbf{M}_{\ddot{\mathbb{Z}}_i^{(1)}} \left(\ddot{\mathbf{Y}}_i^{(1)} - \ddot{\mathbf{X}}_{-i}^{(1)} b \right) \right\|_F^2 + \upsilon_i^{(1,AL)} \sum_{j=1}^{N-1} \frac{|b_j|}{|\widehat{\gamma}_{ij}^{(1,L)}|}.$$

and define

$$\widehat{S}_i^{(1)} = \operatorname{supp} \left(\widehat{\boldsymbol{\gamma}}_{i:}^{(1,AL)} \right).$$

Refit the selected model by OLS:

$$\widehat{\boldsymbol{\gamma}}_{i:,\widehat{S}_i^{(1)}}^{(1,PS)} = \left(\ddot{\mathbf{X}}_{\widehat{S}_i^{(1)}}^{(1)\prime} \mathbf{M}_{\ddot{\mathbb{Z}}_i^{(1)}} \ddot{\mathbf{X}}_{\widehat{S}_i^{(1)}}^{(1)} \right)^{-1} \ddot{\mathbf{X}}_{\widehat{S}_i^{(1)}}^{(1)\prime} \mathbf{M}_{\ddot{\mathbb{Z}}_i^{(1)}} \ddot{\mathbf{Y}}_i^{(1)}, \qquad \widehat{\boldsymbol{\gamma}}_{i:,\widehat{S}_i^{(1)c}}^{(1,PS)} = 0.$$

Finally, use $\widehat{\mathbf{\Gamma}}^{(1,PS)}$ to obtain

$$\widehat{\boldsymbol{\theta}}^{(1)} = \left(\sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(1)}} \mathbb{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(1)}} \left(\mathbf{Y}_i - \mathbf{X}_{-i} \widehat{\boldsymbol{\gamma}}_{i:}^{(1,PS)} \right).$$

Step 5: Re-estimate the outcome-specific factor. Use the first-refinement estimates to construct

$$\dot{\mathbf{r}}_i^{(1)} = \mathbf{M}_{\widehat{\mathbf{G}}} \left(\mathbf{Y}_i - \mathbf{X}_{-i} \widehat{\boldsymbol{\gamma}}_{i:}^{(1,PS)} - \mathbb{Z}_i \widehat{\boldsymbol{\theta}}^{(1)} \right).$$

PCA applied to $\dot{\mathbf{r}}^{(1)}$ gives the updated factor estimate $\widehat{\mathbf{F}}_{\mathbf{Y}}^{(2)}$. Define

$$\widehat{\mathbf{Q}}^{(2)} = \left[\widehat{\mathbf{G}}, \widehat{\mathbf{F}}_{\mathbf{Y}}^{(2)}\right] \quad \text{and} \quad \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}} = \mathbf{I} - \mathbf{P}_{\widehat{\mathbf{Q}}^{(2)}}.$$

Step 6: Re-estimate the spillovers and conduct final inference. With $\ddot{\mathbf{A}}^{(2)} = \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}}\mathbf{A}$, repeat plain LASSO to obtain $\widehat{\boldsymbol{\gamma}}_{i:}^{(2,L)}$. Construct the debiased estimator

$$\widehat{\boldsymbol{\gamma}}_{i:}^{(2,db)} = \widehat{\boldsymbol{\gamma}}_{i:}^{(2,L)} + \widehat{\boldsymbol{\Theta}}_{-i}^{(2)} \frac{1}{T} \ddot{\mathbf{X}}_{-i}^{(2)\prime} \mathbf{M}_{\ddot{\mathbb{Z}}_i^{(2)}} \left(\ddot{\mathbf{Y}}_i^{(2)} - \ddot{\mathbf{X}}_{-i}^{(2)} \widehat{\boldsymbol{\gamma}}_{i:}^{(2,L)} \right).$$

The double-refinement pointwise confidence intervals and max-row test use $\widehat{\boldsymbol{\Gamma}}^{(2,db)}$ and

$$\widehat{\mathrm{se}}\left(\widehat{\gamma}_{ij}^{(2,db)}\right) = \sqrt{\frac{\widehat{\sigma}_{\varepsilon,i}^{2(2)} \widehat{\omega}_{ij}^{(2)}}{T}},$$

where $\widehat{\sigma}_{\varepsilon,i}^{2(2)}$ is the residual-only variance estimator. As in the main procedure, the variance estimator contains neither a LASSO-penalty adjustment nor a sandwich correction.

Finally, keep $\widehat{S}_i^{(1)}$ fixed and refit it by OLS:

$$\widehat{\boldsymbol{\gamma}}_{i:,\widehat{S}_i^{(1)}}^{(2,PS)} = \left(\ddot{\mathbf{X}}_{\widehat{S}_i^{(1)}}^{(2)\prime} \mathbf{M}_{\ddot{\mathbb{Z}}_i^{(2)}} \ddot{\mathbf{X}}_{\widehat{S}_i^{(1)}}^{(2)} \right)^{-1} \ddot{\mathbf{X}}_{\widehat{S}_i^{(1)}}^{(2)\prime} \mathbf{M}_{\ddot{\mathbb{Z}}_i^{(2)}} \ddot{\mathbf{Y}}_i^{(2)}, \qquad \widehat{\boldsymbol{\gamma}}_{i:,\widehat{S}_i^{(1)c}}^{(2,PS)} = 0.$$

The final private-effect estimator is

$$\widehat{\boldsymbol{\theta}}^{(2)} = \left(\sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}} \mathbb{Z}_i \right)^{-1} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}} \left(\mathbf{Y}_i - \mathbf{X}_{-i} \widehat{\boldsymbol{\gamma}}_{i:}^{(2,PS)} \right).$$

Its variance is estimated by the same residual-based estimator used for the private effects in the main paper, i.e.,

$$\widehat{\boldsymbol{\Sigma}}_{\mathbb{U}\boldsymbol{\varepsilon}}^{(2)} = \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}} \mathbb{Z}_i \right)^{-1} \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}} \mathbb{Z}_i \frac{1}{T} \widehat{\boldsymbol{\varepsilon}}_i^{(2)\prime} \widehat{\boldsymbol{\varepsilon}}_i^{(2)} \right) \left(\frac{1}{NT} \sum_{i=1}^{N} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}} \mathbb{Z}_i \right)^{-1}$$

The double-refinement algorithm removes the estimated outcome-specific factor from the residuals and therefore has a smaller asymptotic variance than the original algorithm. Its asymptotic theory follows the proof strategy used for the main results, supplemented by bounds for the estimation

error in $\widehat{\mathbf{F}}_{\mathbf{Y}}^{(1)}$ and $\widehat{\mathbf{F}}_{\mathbf{Y}}^{(2)}$. The details of proof are left to the future work.

The second refinement is designed to refine the estimation of outcome-specific factors. The first refinement provides a rough estimate of the outcome-specific factor that depends on the preliminary estimates of parameters $\mathbf{\Gamma}$ and $\boldsymbol{\theta}$ that have relatively large variance, and the second refinement improves this estimate. Specifically, $\widehat{\mathbf{F}}_{\mathbf{Y}}^{(2)}$ is constructed from the more accurate first-refinement parameters estimates, reducing biases caused by the interaction between the estimated coefficients and other variables in the model. The following theorem shows the asymptotic normality of the double-refinement estimator for the spillover coefficients and private effects.

Theorem C2. Let

$$d_N := \max_{\ell \leq N} \sum_{j=1}^{N} \mathbf{1}\{\gamma_{jl} \neq 0\},$$

where d_N is the maximum number of receivers affected by one sender. Under the same extended assumptions as Theorem 4.2, and if $\sqrt{T/N} \to 0$, $d_N = \mathbf{o}\left(\frac{T}{\log N} + \frac{N}{\log N}\right)$ as $N, T \to \infty$, and the adaptive LASSO penalties in the first refinement $\upsilon_i^{(1,AL)}$ satisfies

$$\upsilon_i^{(1,AL)} \to 0, \quad \left(\frac{T}{\log^2 N} + \frac{N^2}{\log^3 N}\right) \upsilon_i^{(1,AL)} \to \infty \quad \text{for all } i.$$

Then for each i and any $d \in \mathbb{R}^{N-1}$ satisfying $\|d\|_1 = O(1)$,

$$d'\sqrt{T}\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(2,db)} - \boldsymbol{\gamma}_{i:}\right) / \sqrt{\left(d'\widehat{\mathbf{\Theta}}_{-i} d' \widehat{\sigma}_{\varepsilon,i}^{2(2)}\right)} \xrightarrow{d} N(0,1),$$
$$\left(\widehat{\mathbf{\Sigma}}_{\mathbb{U}\boldsymbol{\varepsilon}}^{(2)}\right)^{-1/2} \sqrt{NT}\left(\widehat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}\right) \xrightarrow{d} N(0, \mathbf{I}_{K+1}).$$

Moreover, let $\widehat{\mathcal{I}}_i^{(2)} = \left\| \widehat{\sigma}_{\varepsilon,i}^{(2)} \widehat{\Theta}_{-i}^{1/2} \boldsymbol{\xi} \right\|_\infty$ where $\boldsymbol{\xi} \sim N(0, \mathbf{I}_{N-1})$. Then

$$\sup_{a \geq 0} \left| \mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}^{(2,db)} - \boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right) - \mathbb{P}\left(\widehat{\mathcal{I}}_i^{(2)} \leq a\right) \right| \to 0.$$

Theorem C2 imposes two additional rate restrictions. First, $T/N \to 0$, a condition commonly used in the CCE and interactive-fixed-effects literatures (see, e.g. Pesaran 2006; Bai 2009; Rücker et al. 2025, among others). Second, the sender degree d_N must grow sufficiently slowly. This restriction controls the error in the estimated outcome-specific factors that is propagated from the

first-refinement spillover estimates.

To see why sender sparsity matters, consider the following component of the spillover effects estimation error:

$$\max_{i\leq N}\frac{1}{N\sqrt{T}}\mathbf{U}_{\mathbf{X},i}'\mathbf{U}_{\mathbf{X}}\left(\mathbf{\Gamma}-\widehat{\mathbf{\Gamma}}^{(1,PS)}\right)'\mathbf{\Lambda}_{\mathbf{Y}}=\max_{i\leq N}\frac{1}{N}\sum_{j=1}^{N-1}\frac{1}{\sqrt{T}}\mathbf{U}_{\mathbf{X},i}'\mathbf{U}_{\mathbf{X},S_j}\left(\boldsymbol{\gamma}_{j:,S_j}-\widehat{\boldsymbol{\gamma}}_{j:,S_j}^{(1,PS)}\right)\mathbf{\Lambda}_{\mathbf{Y},j}'.$$

For every receiver j such that $i\in S_j$, the cross product $T^{-1/2}\mathbf{U}_{\mathbf{X},i}'\mathbf{U}_{\mathbf{X},S_j}$ contains the own-series component $T^{-1/2}\mathbf{U}_{\mathbf{X},i}'\mathbf{U}_{\mathbf{X},i}$, which is of order $\mathcal{O}_p\left(\sqrt{T}\right)$. The number of such terms is bounded by d_N. If $d_N\asymp N$, the available bound need not converge to zero. The restriction on d_N ensures that its aggregate contribution is asymptotically negligible.

Simulation study

I compare the original algorithm from the main text with the single- and double-refinement estimators. The factor, loading, and idiosyncratic-error designs follow the unobserved-spillover simulation in Section 5.1. I set $K=1$, $(r,r_x,r_w)=(3,1,1)$, and $\rho=\beta=1$. The spillover matrix has nearest-neighbor coefficients equal to one, except that its first row is zero.

To generate the additional factor, let $\mathbf{g}_t$ collect the stochastic factors in $\mathbf{F}$, $\mathbf{F}_{\mathbf{X}}$, and $\mathbf{F}_{\mathbf{W}}$ at date t. I draw $\mathbf{F}_{\mathbf{Y}}\in\mathbb{R}^{T\times 1}$ jointly with these factors according to

$$\begin{pmatrix}\mathbf{g}_t\\ F_{\mathbf{Y},t}\end{pmatrix}=\mathbf{1}+\mathbf{\Sigma}_{\mathbf{GY}}^{1/2}\mathbf{v}_t,\qquad \mathbf{v}_t\overset{\text{i.i.d.}}{\sim}N(\mathbf{0},\mathbf{I}),\qquad (\mathbf{\Sigma}_{\mathbf{GY}})_{jk}=0.5^{|j-k|}.$$

This construction allows $\mathbf{F}_{\mathbf{Y}}$ to be correlated with the factors in the regressor equations. It remains independent of $\mathbf{U}_{\mathbf{X}}$, $\mathbf{U}_{\mathbf{W}}$, and $\boldsymbol{\varepsilon}$, as required by the exogeneity condition. The outcome-specific loading vector is

$$\mathbf{\Lambda}_{\mathbf{Y}}=\mathbf{1}_N+\mathbf{\Sigma}_N^{1/2}\mathbf{v}_{\mathbf{Y}},\qquad \mathbf{v}_{\mathbf{Y}}\sim N(\mathbf{0},\mathbf{I}_N),\qquad (\mathbf{\Sigma}_N)_{ij}=0.5^{|i-j|}.$$

The single-refinement estimator stops at Step 4 and uses $\widehat{\mathbf{\Gamma}}^{(1,db)}$ for spillover inference. The double-refinement estimator re-estimates $\mathbf{F}_{\mathbf{Y}}$, recomputes the plain and debiased estimators under $\mathbf{M}_{\widehat{\mathbf{Q}}^{(2)}}$, and refits the Step-4 support by OLS. The simulations therefore compare debiased-LASSO confidence intervals and max-row tests from the original, single-refinement, and double-refinement procedures.

Parameter	Estimator	$N = 50,\ T = 100$			$N = 50,\ T = 200$			$N = 100,\ T = 200$		
		Bias	SD	Cov.	Bias	SD	Cov.	Bias	SD	Cov.
$\rho = 1$	Original	−0.0064	0.0205	0.911	−0.0040	0.0136	0.940	−0.0019	0.0091	0.953
	Single refinement	−0.0075	0.0145	0.876	−0.0051	0.0098	0.882	−0.0026	0.0062	0.930
	Double refinement	−0.0037	0.0142	0.913	−0.0022	0.0096	0.925	−0.0010	0.0062	0.936
$\beta = 1$	Original	−0.0020	0.0197	0.944	−0.0017	0.0128	0.962	−0.0009	0.0091	0.962
	Single refinement	−0.0012	0.0144	0.912	−0.0013	0.0093	0.949	−0.0007	0.0065	0.943
	Double refinement	−0.0015	0.0142	0.920	−0.0016	0.0092	0.944	−0.0009	0.0065	0.946

Table 8: Private-effect estimation with an outcome-specific factor. Bias and SD are the Monte Carlo bias and standard deviation, and Cov. is the coverage rate of the nominal 95% confidence interval. Each design uses 1,000 replications.

I consider $(N, T) = (50, 100)$, $(50, 200)$, and $(100, 200)$, with 1,000 Monte Carlo replications for each design. All three estimators use the same data draws and replication seeds, and the original and double-refinement max tests use matched Gaussian multiplier draws. Pointwise confidence intervals have nominal coverage of 95%. The max-row test uses 1,000 multiplier draws and a 5% nominal level.

Table 8 reports the unscaled Monte Carlo bias, standard deviation, and coverage rate of the private-effect estimators. For an estimator $\widehat{\vartheta}$, these statistics are

$$\text{Bias}(\widehat{\vartheta}) = \frac{1}{R}\sum_{s=1}^{R}(\widehat{\vartheta}_s - \vartheta), \qquad \text{SD}(\widehat{\vartheta}) = \left\{\frac{1}{R-1}\sum_{s=1}^{R}(\widehat{\vartheta}_s - \overline{\widehat{\vartheta}})^2\right\}^{1/2}.$$

The refinements substantially increase precision. Relative to the original algorithm, double refinement reduces the Monte Carlo variance by 51–53% for $\widehat{\rho}$ and by 48–49% for $\widehat{\beta}$. Most of this gain occurs in the first refinement. The second refinement further improves the centering of $\widehat{\rho}$: across the three designs, it reduces the magnitude of the single-refinement bias from 0.0075 to 0.0037, from 0.0051 to 0.0022, and from 0.0026 to 0.0010. The estimates of β change little between the two refinements.

Table 9 reports spillover inference for all three procedures. The single- and double-refinement rows use $\widehat{\mathbf{\Gamma}}^{(1,db)}$ and $\widehat{\mathbf{\Gamma}}^{(2,db)}$, respectively.

Using the second-refined factor space reduces the Monte Carlo variance of the two focal debiased spillover estimators by 50–61% relative to the original algorithm. Relative to single refinement, the second refinement raises coverage of the zero coefficient from 0.945, 0.940, and 0.940 to 0.948,

Target	Estimator or test	$N = 50,\ T = 100$ SD	$N = 50,\ T = 100$ Rate	$N = 50,\ T = 200$ SD	$N = 50,\ T = 200$ Rate	$N = 100,\ T = 200$ SD	$N = 100,\ T = 200$ Rate
$\gamma_{12} = 0$	Original: coverage	0.1197	0.944	0.0825	0.950	0.0814	0.940
	Single: coverage	0.0766	0.945	0.0545	0.940	0.0529	0.940
	Double: coverage	0.0749	0.948	0.0534	0.947	0.0525	0.945
$\gamma_{21} = 1$	Original: coverage	0.2146	0.938	0.1542	0.938	0.1491	0.944
	Single: coverage	0.1510	0.917	0.0992	0.921	0.1000	0.934
	Double: coverage	0.1514	0.920	0.0963	0.930	0.0987	0.941
$H_0 : \boldsymbol{\gamma}_{1:} = \mathbf{0}$	Original: rejection	—	0.079	—	0.076	—	0.068
	Single: rejection	—	0.105	—	0.095	—	0.077
	Double: rejection	—	0.102	—	0.085	—	0.071
$H_0 : \boldsymbol{\gamma}_{2:} = \mathbf{0}$	Original: rejection	—	0.924	—	0.992	—	0.979
	Single: rejection	—	1.000	—	1.000	—	1.000
	Double: rejection	—	1.000	—	1.000	—	1.000

Table 9: Spillover inference with an outcome-specific factor. SD is the unscaled Monte Carlo standard deviation. Rate is pointwise coverage for γ_{12} and γ_{21} and the rejection rate of the 5% max-row test for the last two hypotheses. The first hypothesis is true, whereas the second is false because $\gamma_{21} = \gamma_{23} = 1$. Each design uses 1,000 replications.

0.947, and 0.945. Coverage of the nonzero coefficient rises from 0.917, 0.921, and 0.934 to 0.920, 0.930, and 0.941. It also lowers the null rejection rate from 0.105, 0.095, and 0.077 to 0.102, 0.085, and 0.071. Both refinements attain power equal to one in every design.

For the original estimator, pointwise coverage ranges from 0.911 to 0.962 for the private effects and from 0.938 to 0.950 for the two spillover coefficients. Coverage is lowest for ρ in the smallest design and otherwise lies between 0.938 and 0.962. These results provide finite-sample evidence that the studentized estimators induced by the projected leading score are well approximated by a normal distribution.

Appendix D: Heterogeneous panel models

In this appendix, I present the estimation and the corresponding theories of the following heterogeneous model:

$$
\begin{aligned}
y_{it} &= \rho_i x_{it} + \sum_{j \neq i}^{N} \gamma_{ij} x_{jt} + \sum_{k=1}^{K} \beta_{k,i} w_{it,k} + \lambda_i' f_t + \varepsilon_{it}, \\
x_{it} &= \eta_{x,1,i}' f_t + b_{x,i}' f_{x,t} + u_{x,it}, \\
w_{it,k} &= \eta_{w,k,i}' f_t + b_{w,k,i}' f_{w,k,t} + u_{w,k,it}, \quad k = 1, \ldots, K.
\end{aligned} \tag{20}
$$

The estimation procedure for the heterogeneous model follows the same priinciples as the homogeneous model. The main difference is that the coefficients ρ_i, γ_{ij} and β_{ik} are estimated using OLS for each unit i.

The algorithm for the observed spillover channels is as follows:

Algorithm for observed spillovers (Heterogeneous coefficients).

1. (Factor Projection) Estimate the factor space by PCA on $\mathbf{Z} = \mathbf{X} + \sum_k \mathbf{W}_k$ and compute the projection matrix $\mathbf{M}_{\widehat{\mathbf{G}}}$.
2. (OLS) Let $\check{\mathbb{Z}}_i = [\mathbf{X}\boldsymbol{\Omega}_{1,i:}, \ldots, \mathbf{X}\boldsymbol{\Omega}_{L,i:}, \mathbf{X}_i, \mathbf{W}_{1,i}, \ldots, \mathbf{W}_{K,i}]$ collects all regressors and $\delta_i = (\rho_{1,i}, \ldots, \rho_{L,i}, \rho_i, \beta_{1,i}, \ldots, \beta_{K,i})'$ collects parameters. For each i, $\widehat{\delta}_i$ is given by

$$
\widehat{\delta}_i = \left(\check{\mathbb{Z}}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \check{\mathbb{Z}}_i\right)^{-1} \check{\mathbb{Z}}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{Y}_i. \tag{21}
$$

For the unobserved interaction channels, since the spillover effects are by design heterogeneous, the estimators for γ_{ij} are the same as the algorithm for the homogeneous model. For the private effects $\beta_{k,i}$, post-selection technique is not required for inference. Let $\theta_i = (\rho_i, \beta_i')'$. Given $\mathbf{M}_{\widehat{\mathbf{G}}}$ and $\widehat{\boldsymbol{\gamma}}_{i:}$, $\widehat{\theta}_i$ is given by

$$
\widehat{\theta}_i = \left(\mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \mathbb{Z}_i\right)^{-1} \mathbb{Z}_i' \mathbf{M}_{\widehat{\mathbf{G}}} \left(\mathbf{Y}_i - \sum_{j \neq i} \mathbf{X}_j \cdot \widehat{\gamma}_{ij}\right). \tag{22}
$$

The following theorem gives the asymptotic normality for the private effect estimators $\widehat{\delta}_i$ when the spillover channels are observed.

Theorem D1. *Under Assumption 1 to 6, the estimator $\widehat{\delta}_i$ defined in Equation* (21) *satisfies*

$$
\sqrt{T}\left(\widehat{\delta}_i - \delta_i\right) \xrightarrow{d} N\left(0, \boldsymbol{\Sigma}_{\check{\mathbb{U}}\boldsymbol{\varepsilon},i}\right), \tag{23}
$$

where

$$\Sigma_{\check{\mathbb{U}}\varepsilon,i} = \left(\mathbb{E}\left[\check{\mathbb{U}}_{it}\check{\mathbb{U}}_{it}'\right]\right)^{-1}\sigma_{\varepsilon,i}^2.$$

Under the same conditions, the covariance matrix can be consistently estimated by the sample estimate.

The following theorem gives the asymptotic normality for the private effect estimators $\widehat{\theta}_i$ when the spillover channels are unobserved.

Theorem D2. *Under Assumption 1 to 5, suppose that* $\upsilon_i \geq C\left(\sqrt{\frac{\log(N)}{T}} + \frac{1}{N}\right)$ *for some* $C > 0$. *Then the estimator* $\widehat{\theta}_i$ *satisfies*

$$\sqrt{T}\left(\widehat{\theta}_i - \theta_i\right) \xrightarrow{d} N\left(0, \Sigma_{\mathbb{U}\varepsilon,i}\right),$$

where

$$\Sigma_{\mathbb{U}\varepsilon,i} = \left(\mathbb{E}\left[\mathbb{U}_{it}\mathbb{U}_{it}'\right]\right)^{-1}\sigma_{\varepsilon,i}^2.$$

Simiarly, the covariance matrix can be consistently estimated by the sample estimate under the same assumptions.

Simulation. Simulation studies are conducted to examine the finite sample performance of the proposed inference procedures in the heterogeneous panel model. The DGP is the same as in 5 but with heterogeneous coefficients $(\rho_i, \rho_i, \beta_i)$.

For the observed interaction channels, the parameters ρ_i, ρ_i, β_i are generated independently from $Unif[0,1]$. I construct the 95% confidence intervals for $(\rho_1, \rho_1, \beta_1)$ using the proposed estimators for parameters. The simulation is repeated for 2000 times. The coverage rates of the confidence intervals are reported in Table 10. It is seen that the coverage rates are all close to the nominal level 95% across different designs, which verifies the validity of the proposed inference procedure.

For the unobserved interaction channels, DGP is the same as in Section 5.1, but with the heterogeneous coefficients ρ_i, ρ_i, β_i generated independently from $Unif[0,1]$. I construct the 95% confidence intervals for ρ_2 and β_2. Table 11 reports the coverage rates of the confidence intervals for (ρ_2, β_2). The results show that the coverage rates are all close to the nominal level 95% across different designs.

N		50			100			200			300	
T	50	100	200	50	100	200	50	100	200	50	100	200
Sparse $\mathbf{\Omega}$												
ρ_1	0.93	0.94	0.94	0.92	0.94	0.94	0.92	0.93	0.95	0.92	0.94	0.95
ρ_1	0.92	0.94	0.94	0.91	0.94	0.94	0.92	0.94	0.94	0.92	0.94	0.95
β_1	0.93	0.94	0.95	0.92	0.94	0.94	0.92	0.94	0.95	0.92	0.94	0.96
Dense $\mathbf{\Omega}$												
ρ_1	0.93	0.94	0.94	0.92	0.94	0.94	0.92	0.94	0.95	0.92	0.94	0.95
ρ_1	0.92	0.94	0.94	0.91	0.93	0.94	0.92	0.94	0.94	0.92	0.94	0.94
β_1	0.93	0.94	0.95	0.92	0.94	0.94	0.92	0.94	0.95	0.92	0.94	0.95

Table 10: Coverage rates of the confidence intervals for parameters in heterogeneous panel model with observed spillover channels.

(N,T)	(50, 100)	(50, 200)	(100, 100)	(100, 200)	(200, 100)	(200, 200)
ρ_2	0.93	0.96	0.92	0.95	0.92	0.95
β_2	0.92	0.96	0.93	0.95	0.92	0.96

Table 11: Coverage rates of the confidence intervals for (ρ_2, β_2) in heterogeneous panel model with unobserved spillover channels.

Appendix E: Additional simulation results

This appendix examines robustness under three alternative specifications: cross-sectional correlation in the outcome errors, partial overlap of the outcome factors across covariates, and approximate sparsity of the spillover network.

E.1 Correlated errors

The first design examines the robustness of the proposed algorithm to weak cross-sectional correlation in the outcome error $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$. The design follows Section 5.1, changing only the cross-sectional dependence of this error. Specifically, I generate

$$\varepsilon_t = D_\sigma \mathbf{\Sigma}_N^{1/2} z_t, \qquad z_t \overset{i.i.d.}{\sim} N(0, I_N), \qquad D_\sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_N),$$

where $(\mathbf{\Sigma}_N)_{ij} = 0.5^{|i-j|}$ and $\sigma_i^2 \sim \mathrm{Unif}[0.5, 2]$, as in the baseline design. Conditional on the variance draws,

$$\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_i \sigma_j 0.5^{|i-j|}.$$

The marginal error variances are therefore unchanged, but errors are now correlated across units, with correlations declining geometrically in $|i-j|$. This dependence is weak in the sense that the absolute covariance sums remain bounded as N increases. Errors remain independent over time and independent of the factor and regressor innovations.

All other features of the baseline design are unchanged, including the sparse spillover network, factor structure, and regressor idiosyncratic components. I use 2,000 repetitions for each $(N, T) \in \{50, 200\} \times \{100, 200, 300\}$. Estimation and inference follow the same procedures as in the main text, including PCA-Each, BIC penalty selection, and post-selection. The variance estimator is also unchanged; no additional correction for cross-sectional error covariances is introduced.

Table E.1 shows that spillover inference performs similarly to the baseline. Coverage ranges from 93.5% to 95.0% for the zero coefficient and from 91.2% to 95.2% for the nonzero coefficient. The row-wise test rejects the true null in 5.2–7.8% of repetitions, while power ranges from 78.1% to 100% and reaches 100% in every design with $T \geq 200$.

Figures E.1 and E.2 show network recovery for $N = 50$ at $T = 100$ and 300, matching the designs plotted in the main text. The multiple-testing tradeoff is similar: at $T = 100$, BH and BY have empirical FDRs of 10.6% and 2.2% and power of 80.1% and 67.2%, respectively. At $T = 300$, power

reaches 99.8% for BH and 99.2% for BY, while their FDRs are 7.3% and 1.2%. BY therefore keeps the empirical FDR below the 5% target in both designs, with little loss of power at $T = 300$.

For the common coefficients, Table E.2 again shows that post-selection improves coverage mainly for the private effect ρ. Coverage rises from 83.7–90.2% without post-selection to 88.3–93.8% with post-selection. Coverage for the control coefficient β remains close to 95% under both methods. The remaining undercoverage for ρ and the mild overrejection of the row-wise tests resemble the baseline patterns. Overall, these results support the robustness of the algorithm to the weak cross-sectional error correlation considered here.

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Coverage rates for γ_{12} and γ_{21}						
γ_{12}	0.94	0.94	0.95	0.94	0.94	0.94
γ_{21}	0.91	0.92	0.93	0.95	0.95	0.95
Rejection rates of the row-wise tests						
$H_0 : \|\boldsymbol{\gamma}_{1\cdot}\|_\infty = 0$	0.08	0.05	0.06	0.06	0.06	0.06
$H_0 : \|\boldsymbol{\gamma}_{2\cdot}\|_\infty = 0$	0.91	1.00	1.00	0.78	1.00	1.00

Table E.1: Coverage and rejection rates for spillover inference with cross-sectionally correlated outcome errors. Columns report (N, T), with 2,000 repetitions per design. The upper panel reports coverage of nominal 95% confidence intervals for $\gamma_{12} = 0$ and $\gamma_{21} = 0.5$. The lower panel reports rejection rates of nominal 5% row-wise tests: unit 1 measures size and unit 2 measures power. Estimation uses PCA-Each.

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Non-PS: ρ	0.84	0.84	0.84	0.88	0.90	0.90
PS: ρ	0.88	0.93	0.93	0.90	0.94	0.94
Non-PS: β	0.93	0.94	0.95	0.94	0.94	0.94
PS: β	0.93	0.94	0.95	0.93	0.93	0.95

Table E.2: Empirical coverage of nominal 95% confidence intervals for the private effect $\rho = 1$ and control coefficient $\beta = 1$ with cross-sectionally correlated outcome errors. Non-PS omits post-selection; PS implements it. Both estimators use PCA-Each and the baseline variance estimator. Each column reports 2,000 repetitions for (N, T).

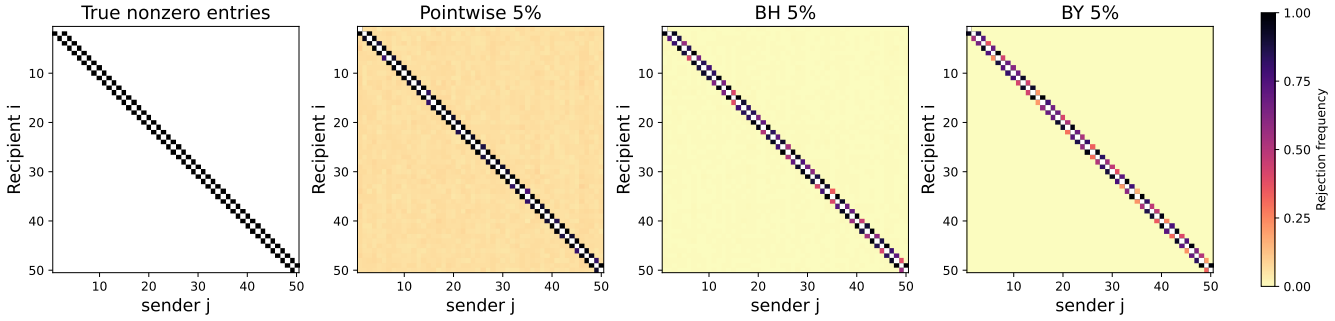


Figure E.1: Network recovery with cross-sectionally correlated outcome errors for $(N, T) = (50, 100)$ over 2,000 repetitions. Rows index recipients and columns index senders. The first panel shows the true support. The remaining panels show rejection frequencies for pointwise tests at the nominal 5% level and the BH and BY procedures, both targeting a 5% FDR. Darker cells indicate more frequent rejections on a common 0–1 scale. Diagonal entries are excluded from testing.

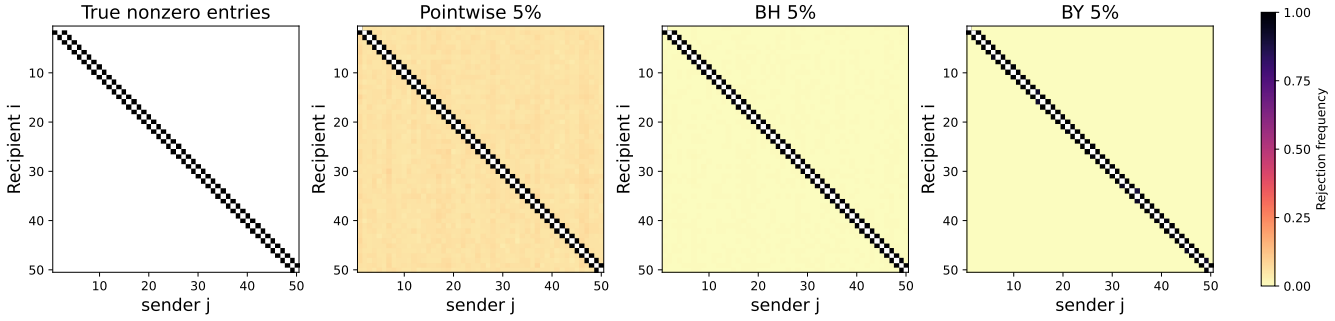


Figure E.2: Network recovery with cross-sectionally correlated outcome errors for $(N, T) = (50, 300)$ over 2,000 repetitions. Rows index recipients and columns index senders. The first panel shows the true support. The remaining panels show rejection frequencies for pointwise tests at the nominal 5% level and the BH and BY procedures, both targeting a 5% FDR. Darker cells indicate more frequent rejections on a common 0–1 scale. Diagonal entries are excluded from testing.

E.2 Partial factor overlap

The second design allows each covariate to contain only a subset of the outcome factors, while the two covariates jointly span the full outcome factor space. Let $\mathbf{F} = [\mathbf{1}_T, \mathbf{d}, \mathbf{f}]$ denote the three outcome factors in the baseline design. I retain the outcome equation and replace the regressor equations with

$$\mathbf{X} = \mathbf{1}_T \mathbf{a}_x' + \mathbf{d}\boldsymbol{\eta}_{d,x}' + \mathbf{f}_x \mathbf{b}_x' + \mathbf{U}_x,$$
$$\mathbf{W} = \mathbf{1}_T \mathbf{a}_w' + \mathbf{f}\boldsymbol{\eta}_{f,w}' + \mathbf{f}_w \mathbf{b}_w' + \mathbf{U}_w.$$

Thus, $\mathbf{X}$ contains the first two columns of $\mathbf{F}$, whereas $\mathbf{W}$ contains the first and third. Each also retains its own additional factor. Neither covariate alone spans all of $\mathbf{F}$, but combining their factor spaces captures every factor entering the outcome equation. PCA-Each therefore extracts three factors from each covariate, rather than four as in the baseline.

All other features follow Section 5.1, including the loading distributions, spillover network, panel dimensions, and 2,000 repetitions per design. In particular, outcome errors are independent across units, as in the baseline; this design changes factor overlap rather than error dependence. The remaining estimation and inference steps are unchanged.

Table E.3 shows that spillover inference remains similar to the baseline. Coverage is 94–95% for the zero coefficient and 91–96% for the nonzero coefficient. The row-wise test rejects the true null in about 6% of repetitions. Power is 93% at $(N, T) = (50, 100)$ and 78% at $(200, 100)$, reaching 100% whenever $T \geq 200$.

Figures E.3 and E.4 show the same pattern in network recovery. At $T = 100$, BH and BY have empirical FDRs of 9.7% and 2.0%, with power of 81.2% and 68.6%, respectively. At $T = 300$, power reaches 99.8% and 99.3%, while FDR falls to 7.4% and 1.2%. As in the main text, BY stays below the 5% FDR target and detects nearly all nonzero links when $T = 300$.

Table E.4 reports coverage for the common coefficients. Post-selection improves coverage for ρ in every design, from 86–92% without post-selection to 90–94% with it. Coverage for β under post-selection ranges from 90% to 94%, indicating some remaining undercoverage, particularly when $N = 50$. Overall, these results support the robustness of the algorithm when the covariates jointly span the outcome factor space even though neither does so individually.

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Coverage rates for γ_{12} and γ_{21}						
γ_{12}	0.94	0.95	0.94	0.94	0.94	0.95
γ_{21}	0.91	0.93	0.92	0.96	0.95	0.95
Rejection rates of the row-wise tests						
$H_0 : \|\boldsymbol{\gamma}_{1:}\|_\infty = 0$	0.07	0.06	0.06	0.06	0.06	0.06
$H_0 : \|\boldsymbol{\gamma}_{2:}\|_\infty = 0$	0.93	1.00	1.00	0.78	1.00	1.00

Table E.3: Coverage and rejection rates for spillover inference with partial factor overlap across covariates. Columns report (N, T), with 2,000 repetitions per design. The upper panel reports coverage of nominal 95% confidence intervals for $\gamma_{12} = 0$ and $\gamma_{21} = 0.5$. The lower panel reports rejection rates of nominal 5% row-wise tests: unit 1 measures size and unit 2 measures power. Estimation uses PCA-Each.

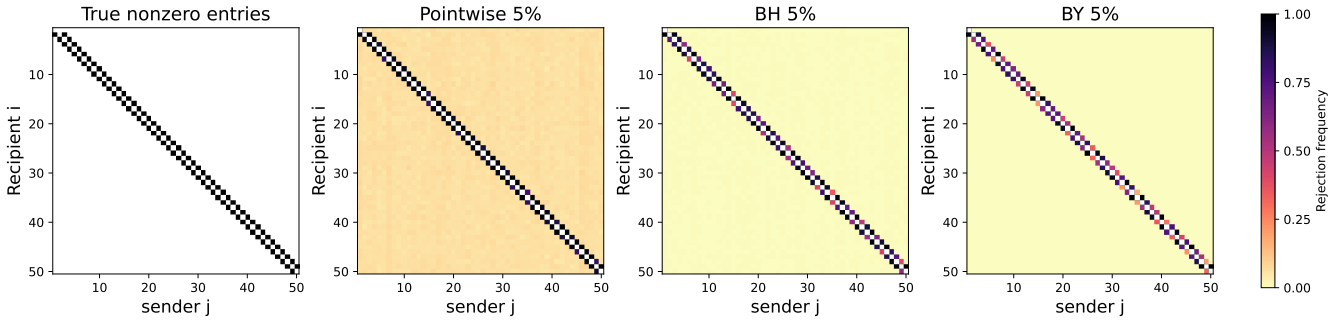


Figure E.3: Network recovery with partial factor overlap across covariates for $(N, T) = (50, 100)$ over 2,000 repetitions. Rows index recipients and columns index senders. The first panel shows the true support. The remaining panels show rejection frequencies for pointwise tests at the nominal 5% level and the BH and BY procedures, both targeting a 5% FDR. Darker cells indicate more frequent rejections on a common 0–1 scale. Diagonal entries are excluded from testing.

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Non-PS: ρ	0.86	0.87	0.87	0.88	0.92	0.92
PS: ρ	0.90	0.93	0.94	0.90	0.94	0.94
Non-PS: β	0.91	0.90	0.88	0.93	0.92	0.93
PS: β	0.90	0.91	0.90	0.93	0.93	0.94

Table E.4: Empirical coverage of nominal 95% confidence intervals for the private effect $\rho = 1$ and control coefficient $\beta = 1$ with partial factor overlap across covariates. Non-PS omits post-selection; PS implements it. Both estimators use PCA-Each and the baseline variance estimator. Each column reports 2,000 repetitions for (N, T).

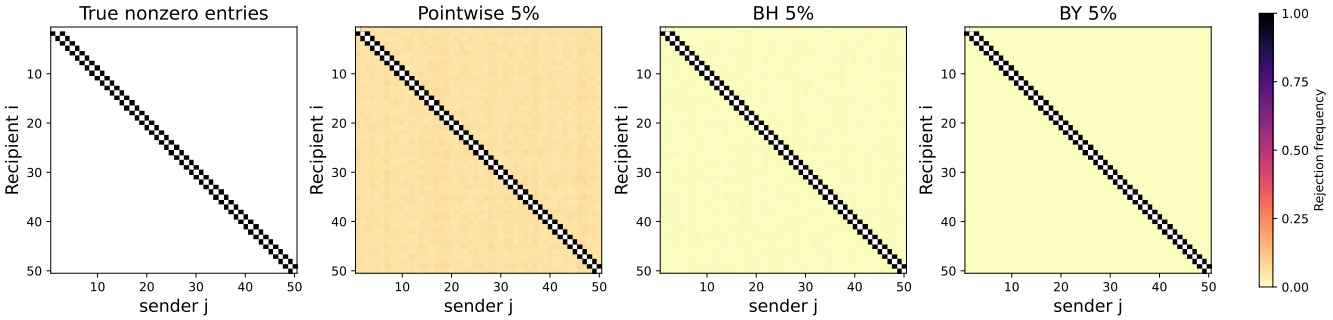


Figure E.4: Network recovery with partial factor overlap across covariates for $(N, T) = (50, 300)$ over 2,000 repetitions. Rows index recipients and columns index senders. The first panel shows the true support. The remaining panels show rejection frequencies for pointwise tests at the nominal 5% level and the BH and BY procedures, both targeting a 5% FDR. Darker cells indicate more frequent rejections on a common 0–1 scale. Diagonal entries are excluded from testing.

E.3 Approximate sparsity

The third design examines robustness to approximate sparsity. Let S_i contain the four units closest to i in index distance $|i - j|$, excluding i and breaking ties in favor of the smaller index. For $i \geq 2$ and $j \neq i$, set

$$\gamma_{ij} = \begin{cases} 0.5, & j \in S_i, \\ \tau_{NT} \dfrac{|i-j|^{-2}}{\sum_{k \notin S_i \cup \{i\}} |i-k|^{-2}}, & j \notin S_i, \end{cases} \qquad \tau_{NT} = \frac{1}{4} \min\left\{1, \frac{\log N}{T}\right\}.$$

Set $\gamma_{ii} = \gamma_{1j} = 0$ for all i, j. Each recipient other than unit 1 therefore has four strong links and many weak links whose coefficients sum to τ_{NT}. The weak links decay with distance, and their total weight decreases with T.

All other features follow Section 5.1, including the factor structure, error distributions, and common coefficients $\rho = \beta = 1$. I use 2,000 repetitions for each $(N, T) \in \{50, 200\} \times \{100, 200, 300\}$. Estimation uses PCA-Each with four factors per covariate, BIC penalty selection, and the same post-selection and inference procedures as in the baseline.

Table E.5 shows that network estimation becomes more accurate as the panel dimensions increase. For $N = 50$, post-selection RMSE falls from 0.08 at $T = 100$ to 0.03 at $T = 300$; for $N = 200$, it falls from 0.06 to 0.02. Post-selection also yields lower RMSE than the initial LASSO in every

design. Coverage for the zero spillover coefficient remains at 94–95%. For the strong coefficient, coverage increases from 87% to 91% as T rises when $N = 50$ and is close to 95% in all designs with $N = 200$.

Table E.6 shows that post-selection improves coverage for the private effect ρ in every design. Its coverage rises from 83–84% at $T = 100$ to about 94% at $T = 300$, compared with 67% and 85% without post-selection at $T = 300$ for $N = 50$ and 200, respectively. Post-selection coverage for β also approaches 95% as T increases. These results support the robustness of the algorithm to the approximate sparsity, with more accurate network estimates and improved inference in larger panels.

The empirical coverage rates for ρ and β are consistent with the conjecture that post-selection retains strong links while excluding weak links whose coefficients vanish faster than the LASSO penalty, drawing on the post-selection analysis of Belloni and Chernozhukov (2013). A formal treatment requires substantial additional technical analysis to account for error in the estimated factor space and is left for future work.

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Whole-network RMSE						
LASSO	0.09	0.06	0.05	0.06	0.04	0.03
Post-selection	0.08	0.04	0.03	0.06	0.03	0.02
Debiased	0.13	0.09	0.07	0.15	0.09	0.07
Coverage rates for γ_{12} and γ_{21}						
γ_{12}	0.94	0.94	0.95	0.94	0.94	0.95
γ_{21}	0.87	0.90	0.91	0.95	0.94	0.95

Table E.5: Network estimation and spillover inference under approximate sparsity. Columns report (N, T), with 2,000 repetitions per design. The upper panel reports whole-network RMSE, $\{R^{-1}\sum_{r=1}^{R} N^{-2}\|\widehat{\mathbf{\Gamma}}^{(r)} - \mathbf{\Gamma}\|_F^2\}^{1/2}$, where $R = 2{,}000$ and diagonal entries are fixed at zero. The lower panel reports coverage of nominal 95% confidence intervals for $\gamma_{12} = 0$ and $\gamma_{21} = 0.5$, using the debiased estimator. Estimation uses PCA-Each.

	(50, 100)	(50, 200)	(50, 300)	(200, 100)	(200, 200)	(200, 300)
Non-PS: ρ	0.68	0.69	0.67	0.79	0.82	0.85
PS: ρ	0.83	0.92	0.94	0.84	0.92	0.94
Non-PS: β	0.92	0.93	0.93	0.93	0.93	0.94
PS: β	0.91	0.93	0.94	0.92	0.93	0.94

Table E.6: Empirical coverage of nominal 95% confidence intervals for the private effect $\rho = 1$ and control coefficient $\beta = 1$ under approximate sparsity. Non-PS omits post-selection; PS implements it. Both estimators use PCA-Each and the baseline variance estimator. Each column reports 2,000 repetitions for (N, T).

Appendix F: Industry-level data construction

I construct a quarterly panel dataset by combining firm-level data from the Compustat Quarterly and Annual Fundamentals database. The main variables include research and development expenditures (R&D), sales, and property, plant, and equipment (PP&E), with employment obtained from the annual file. All nominal variables are converted into real terms using a quarterly GDP deflator from the Bureau of Economic Analysis (BEA).

I begin by harmonizing variable definitions across datasets and converting reporting dates into calendar year and quarter. Observations are indexed by firm identifier (gvkey), and duplicate entries within a given period are removed by retaining the most recent report. Firms are assigned to industries based on their two-digit SIC codes (sic2), which are treated as time-invariant.

To construct a consistent panel, I expand the data to include all firm–year–quarter combinations observed in the sample and merge the available quarterly data into this full grid. Because quarterly reporting is incomplete for many firms, I combine quarterly and annual information to improve coverage. For each firm-year, I first assess the number of available quarterly observations. If only one quarterly report is available, I treat the firm-year as effectively annual: flow variables such as R&D and sales are evenly allocated across quarters, while stock variables such as PP&E are carried forward to missing quarters. For firm-years with multiple quarterly observations, missing values are imputed within the year using nearest-neighbor interpolation, preserving the within-year variation in reported data.

After constructing firm-level quarterly variables, I deflate all nominal measures using the BEA price index to obtain real quantities. I then aggregate the data to the industry level using two-digit SIC codes, summing firm-level variables within each industry–year–quarter cell. Aggregation is performed allowing for missing firm-level observations, provided that at least one firm contributes data to a given cell.

I retain only industries with complete quarterly observations throughout the sample window. This yields a balanced industry-level panel for the empirical analysis.

After aggregation and taking logs for the main variables, I remove industry fixed effects and time trends ($\frac{t}{T}$) to isolate the variation of interest. The resulting panel is used for the estimation of spillover effects and common shocks in the main analysis.

Appendix G: Proofs of main theorems

In this appendix, I present the proofs of Theorem 4.1 and 4.2. Theroem 4.1 consists of two parts: one is the results for asymptotic normality of the debiased LASSO estimators and the others is the consistency of the variance estimator.

The following proposition shows the convergence rate of plain LASSO estimators $\widehat{\boldsymbol{\gamma}}_{i:}$.

Proposition G1. Under the same assumption of Theorem 4.1, for $i = 1, \ldots, N$, $\widehat{\boldsymbol{\gamma}}_{i:}$ defined in Step 2 of Algorithm 2 satisfies

$$\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\|_2 = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N}\right).$$

Proof of Prop. G1. Fix i. WLOG, assume $K = 1$, i.e. $\mathbb{Z}_i = (\mathbf{X}_i, \mathbf{W}_i) \in \mathbb{R}^{T\times 2}$. Rewrite the model that partial out $\mathbb{Z}_i$, we have

$$\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{Y}_i = \sum_{j\neq i}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_j \cdot \gamma_j + \mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{E}_i,$$

where $\mathbf{E}_i = \mathbf{F}\lambda_i + \boldsymbol{\varepsilon}_i$.

For the ease of notation, deonte $\widetilde{\mathbf{Y}}_i := \mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{Y}_i$, $\widetilde{\mathbf{X}}_{-i} := \mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{-i}$ and $\widetilde{\mathbf{E}}_i := \mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{E}_i$.

By Prop. J1, under the assumptions of Theorem 4.1, it is sufficient to show the following:

$$\frac{1}{T}\left\|\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N}\right), \tag{24}$$

$$\left\|\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{X}}_{-i} - \boldsymbol{\Sigma}_{\mathbf{U}_{x,-i}}\right\|_{\max} = \mathbf{o}_p(1), \tag{25}$$

where $\boldsymbol{\Sigma}_{\mathbf{U}_{x,-i}} = \mathbb{E}\left[\mathbf{U}_{x,-i,t}\mathbf{U}'_{x,-i,t}\right]$.

Write

$$\begin{aligned}\mathbb{Z}_i &= \mathbf{G}\mathbf{C}'_{\mathbb{Z},i} + \mathbf{U}_{\mathbb{Z},i}\\ \mathbf{X}_{-i} &= \mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X},-i} + \mathbf{U}_{\mathbf{X},-i}.\end{aligned}$$

Observe that $\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{P}_{\widehat{\mathbf{G}}} = \mathbf{P}_{\widehat{\mathbf{G}}}$, we have

$$\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}} = \mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i} = \mathbf{M}_{\widehat{\mathbb{Z}},i} - \mathbf{P}_{\widehat{\mathbf{G}}} = \mathbf{M}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbb{Z}},i}.$$

Therefore, to bound (24), we can decompose it as follows:

$$\begin{aligned}
\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i &= \frac{1}{T}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{E}_i \\
&= \frac{1}{T}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i + \frac{1}{T}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i \\
&= \frac{1}{T}\mathbf{U}'_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i - \frac{1}{T}\mathbf{U}'_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i + \frac{1}{T}\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i + \frac{1}{T}\mathbf{U}'_{\mathbf{X},-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i \\
&\quad + \frac{1}{T}\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i - \frac{1}{T}\mathbf{X}_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\boldsymbol{\varepsilon}_i - \frac{1}{T}\mathbf{X}_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{F}\lambda_i \\
&:= \sum_{i=1}^{7}\mathcal{A}_i.
\end{aligned} \tag{26}$$

By Lemma K1 and by Assumption 3 and Assumption 5, $\frac{1}{T}\left\|\mathbf{U}'_{\mathbf{X},-i}\mathbf{U}_{\mathbf{Z},i}\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{logN}{T}}\right)$. By Lemma I1, $\left\|\frac{1}{T}\mathbf{U}'_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty = \mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T} + \frac{\log N}{N}\right)$.

For $\frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty$,

$$\begin{aligned}
\frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty &= \frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\mathbf{H}'^{-1}_{\mathbf{Q}}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}} - \widehat{\mathbf{G}}_{\mathbf{X}}\right)'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty \\
&\leq \frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}} - \widehat{\mathbf{G}}_{\mathbf{X}}\right)'\boldsymbol{\varepsilon}_i\right\|_\infty + \frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}} - \widehat{\mathbf{G}}_{\mathbf{X}}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty
\end{aligned}$$

By Lemma K2 and Lemma H2,

$$\begin{aligned}
\frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}} - \widehat{\mathbf{G}}_{\mathbf{X}}\right)'\boldsymbol{\varepsilon}_i\right\|_\infty &\leq \max_{i\leq N}\left\|\mathbf{C}_{\mathbf{X},i}\right\|_2 \frac{1}{T}\left\|\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}} - \widehat{\mathbf{G}}_{\mathbf{X}}\right)\boldsymbol{\varepsilon}_i\right\|_F \\
&= \mathcal{O}_p\left(\frac{\sqrt{\log N}}{N} + \frac{\sqrt{\log N}}{T}\right).
\end{aligned}$$

For the second term, by Lemma I5,

$$\frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right)\leq\max_{j\leq N}\|\mathbf{C}_{\mathbf{X},j}\|_2\,\frac{1}{T}\left\|\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{Q}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_F$$
$$=\mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}\right).$$

Putting them together, we have

$$\|\mathcal{A}_3\|_\infty=\frac{1}{T}\left\|\mathbf{C}_{\mathbf{X},-i}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty=\mathcal{O}_p\left(\frac{\sqrt{\log N}}{N}+\frac{\sqrt{\log N}}{T}\right).$$

For $\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i$, expand λ_i so that $\widetilde{\lambda}_i=[\lambda_i',0]'\in\mathbb{R}^{r_{\mathbf{X}}}$. We have

$$\begin{aligned}\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i&=\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\widetilde{\lambda}_i\\&=\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\mathbf{M}_{\widehat{\mathbf{G}}}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i\\&=\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i-\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\mathbf{P}_{\widehat{\mathbf{G}}}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i.\end{aligned}$$

By Lemma K2 and Lemma H2,

$$\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},-i}'\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i\right\|_\infty\leq\max_{j\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},j}'\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\right\|_2\left\|\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i\right\|_F$$
$$=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}+\sqrt{\frac{\log N}{NT}}\right).$$

By Lemma I5 (ii),

$$\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},-i}'\mathbf{P}_{\widehat{\mathbf{G}}}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i\right\|_\infty\leq\max_{j\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\right\|_2\left\|\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\lambda}_i\right\|_F$$
$$=\mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}+\frac{\log N}{N\sqrt{N}}\right).$$

Hence,

$$\|\mathcal{A}_4\|_\infty=\left\|\frac{1}{T}\mathbf{U}_{\mathbf{X},-i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i\right\|_\infty=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}+\sqrt{\frac{\log N}{NT}}\right).$$

For $\mathcal{A}_5 = \frac{1}{T}\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i$,

$$\begin{aligned}\left\|\frac{1}{T}\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i\right\|_\infty &= \max_{j\neq i}\left\|\mathbf{C}_{\mathbf{X},j}\right\|_2\left\|\frac{1}{T}\mathbf{M}_{\widehat{\mathbf{G}}}\left(\widehat{\mathbf{G}}-\mathbf{G}\mathbf{H}\right)\right\|_F^2\left\|\mathbf{H}\right\|_2^2\left\|\lambda_i\right\|_2 \\ &= \mathcal{O}_p\left(\frac{\sqrt{\log N}}{N}+\frac{\sqrt{\log N}}{T^2}\right).\end{aligned}$$

Putting $\mathcal{A}_1$ to $\mathcal{A}_5$ together, we have

$$\frac{1}{T}\left\|\mathcal{A}_1-\mathcal{A}_2+\mathcal{A}_3+\mathcal{A}_4+\mathcal{A}_5\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right).$$

Denote the decomposition:

$$\begin{aligned}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i &= \mathbf{U}'_{\mathbf{X},-i}\mathbf{U}_{\mathbb{Z},i}-\mathbf{U}'_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbb{Z},i}+\mathbf{U}'_{\mathbf{X},-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbf{C}_{\mathbb{Z},i} \\ &\quad+\mathbf{C}'_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbb{Z},i}+\mathbf{C}_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbf{C}'_{\mathbb{Z},i} \\ &:= \mathcal{B}_1-\mathcal{B}_2+\mathcal{B}_3+\mathcal{B}_4+\mathcal{B}_5,\end{aligned}$$

and

$$\begin{aligned}\mathbb{Z}_i\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i &= \mathbf{U}'_{\mathbb{Z},i}\boldsymbol{\varepsilon}_i-\mathbf{U}'_{\mathbb{Z},i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i+\mathbf{C}_{\mathbb{Z},i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i \\ &= \mathcal{D}_1-\mathcal{D}_2+\mathcal{D}_3,\end{aligned}$$

and

$$\begin{aligned}\mathbb{Z}_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i &= \mathbf{U}'_{\mathbb{Z},i}\mathbf{F}\lambda_i-\mathbf{U}'_{\mathbb{Z},i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i+\mathbf{C}_{\mathbb{Z},i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\lambda_i \\ &= \mathcal{E}_1-\mathcal{E}_2+\mathcal{E}_3.\end{aligned}$$

With these notations, $\mathcal{A}_6$ can be written as

$$\frac{1}{T}\mathbf{X}_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\boldsymbol{\varepsilon}_i \leq \frac{1}{T}\left\|\mathcal{B}_1-\mathcal{B}_2+\mathcal{B}_3+\mathcal{B}_4+\mathcal{B}_5\right\|_\infty\left\|\left(\frac{1}{T}\mathbb{Z}'_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\right\|_F\frac{1}{T}\left\|\mathcal{D}_1-\mathcal{D}_2+\mathcal{D}_3\right\|_2.$$

It is easy to verify that $\mathcal{B}_1$ to $\mathcal{B}_5$ are analogous to $\mathcal{A}_1$ to $\mathcal{A}_5$, which gives

$$\frac{1}{T}\left\|\mathcal{B}_1-\mathcal{B}_2+\mathcal{B}_3+\mathcal{B}_4+\mathcal{B}_5\right\|_\infty=\mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right).$$

Note that $\frac{1}{T}\left\|\mathbf{U}_{\mathbb{Z},i}\mathbf{F}\lambda_i\right\|_2=\mathcal{O}_p\left(\sqrt{\frac{1}{T}}\right)$. By Lemma I5(iii), $\frac{1}{T}\left\|\mathcal{D}_2\right\|_2=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right)$. For $\frac{1}{T}\left\|\mathcal{D}_3\right\|_2$,

$$\begin{aligned}\frac{1}{T}\left\|\mathbf{C}_{\mathbb{Z},i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2 &\leq \left\|\mathbf{C}_{\mathbb{Z},i}\right\|_2\left\|\mathbf{H}\right\|_2\frac{1}{T}\left\|\left(\widehat{\mathbf{G}}-\mathbf{G}\mathbf{H}\right)'\boldsymbol{\varepsilon}_i\right\|_2+\left\|\mathbf{C}_{\mathbb{Z},i}\right\|_2\left\|\mathbf{H}\right\|_2\frac{1}{T}\left\|\left(\widehat{\mathbf{G}}-\mathbf{G}\mathbf{H}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\\ &=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right).\end{aligned}$$

Putting them together, we have

$$\frac{1}{T}\left\|\mathcal{D}_1-\mathcal{D}_2+\mathcal{D}_3\right\|_2=\mathcal{O}_p\left(\sqrt{\frac{1}{T}}+\frac{1}{N}\right).$$

By Lemma I5 (iv) and continuous mapping theorem, $\left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\right\|_F=\mathcal{O}_p\left(1\right)$. So

$$\left\|\mathcal{A}_6\right\|_\infty=\mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}+\frac{\log N}{N\sqrt{T}}\right).$$

Similarly, $\frac{1}{T}\left\|\mathcal{A}_7\right\|_\infty$ is bounded by

$$\frac{1}{T}\left\|\mathbf{X}_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{F}\lambda_i\right\|_\infty\leq\frac{1}{T}\left\|\mathcal{B}_1-\mathcal{B}_2+\mathcal{B}_3+\mathcal{B}_4+\mathcal{B}_5\right\|_\infty\left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\right\|_F\frac{1}{T}\left\|\mathcal{E}_1-\mathcal{E}_2+\mathcal{E}_3\right\|_2.$$

Following the same argument, we have $\frac{1}{T}\left\|\mathcal{E}_1\right\|_2=\mathcal{O}_p\left(\sqrt{\frac{1}{T}}\right)$, $\frac{1}{T}\left\|\mathcal{E}_2\right\|_2=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right)$ and $\frac{1}{T}\left\|\mathcal{E}_3\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T^2}\right)$. Hence,

$$\left\|\mathcal{A}_7\right\|_\infty=\mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}+\frac{\log N}{N\sqrt{T}}\right).$$

Putting $\mathcal{A}_1$ to $\mathcal{A}_7$ together, we have

$$\frac{1}{T}\left\|\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right).$$

This proves (24). Now we turn to (25). Write

$$\begin{aligned}
\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{X}}_{-i} - \mathbf{\Sigma}_{\mathbf{U}_\mathbf{X},-i} &= \frac{1}{T}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{-i} \\
&= \frac{1}{T}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{-i} - \frac{1}{T}\mathbf{X}'_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{X}_{-i} \\
&= \left(\frac{1}{T}\mathbf{U}_{\mathbf{X},-i'}\mathbf{U}_{\mathbf{X},-i} - \mathbf{\Sigma}_{\mathbf{U}_\mathbf{X},-i}\right) - \frac{1}{T}\mathbf{U}'_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},-i} \\
&\quad + \frac{1}{T}\mathbf{C}'_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},-i} + \frac{1}{T}\mathbf{U}'_{\mathbf{X},-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbf{C}_{\mathbf{X},-i} + \frac{1}{T}\mathbf{C}'_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbf{C}'_{\mathbf{X},-i} \\
&\quad - \frac{1}{T}\mathbf{X}'_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{X}_{-i} \\
&:= \mathcal{G}_1 - \mathcal{G}_2 + \mathcal{G}_3 + \mathcal{G}_4 + \mathcal{G}_5 + \mathcal{G}_6.
\end{aligned}$$

By Lemma K1, $\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},-i'}\mathbf{U}_{\mathbf{X},-i} - \mathbf{\Sigma}_{\mathbf{U}_\mathbf{X},-i}\right\|_{\max} = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}\right)$.

By Lemma I2, $\left\|\mathcal{G}_2\right\|_{\max} = \mathcal{O}_p\left(\frac{\log N}{N}+\frac{\log N}{T}\right)$.

For $\mathcal{G}_3$,

$$\begin{aligned}
&\left\|\frac{1}{T}\mathbf{C}'_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},-i}\right\|_{\max} \\
&\le \max_{j\neq i}\left\|\mathbf{C}_{\mathbf{X},j}\right\|_2\left\|\mathbf{H}\right\|_2\frac{1}{T}\left\|\left(\widehat{\mathbf{G}}-\mathbf{G}\mathbf{H}\right)'\mathbf{U}_{\mathbf{X},-i}\right\|_{\max} + \max_{j\neq i}\left\|\mathbf{C}_{\mathbf{X},j}\right\|_2\left\|\mathbf{H}\right\|_2\frac{1}{T}\left\|\left(\widehat{\mathbf{G}}-\mathbf{G}\mathbf{H}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},-i}\right\|_{\max} \\
&= \mathcal{O}_p\left(\frac{\sqrt{\log N}}{N}+\frac{\sqrt{\log N}}{T}+\frac{\log N}{\sqrt{NT}}\right)+\mathcal{O}_p\left(\frac{(\log N)^{3/2}}{\sqrt{NT}}\right) \\
&= \mathcal{O}_p\left(\frac{\sqrt{\log N}}{N}+\frac{\sqrt{\log N}}{T}+\frac{(\log N)^{3/2}}{\sqrt{NT}}\right)
\end{aligned}$$

$\mathcal{G}_4$ is equivalent to $\mathcal{G}_3$. For $\mathcal{G}_5$,

$$\left\|\frac{1}{T}\mathbf{C}'_{\mathbf{X},-i}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbf{C}'_{\mathbf{X},-i}\right\|_{\max} \leq \max_{j\neq i}\|\mathbf{C}_{\mathbf{X},j}\|_2^2\left\|\frac{1}{T}\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\right\|_F = \mathcal{O}_p\left(\frac{\log N}{N}+\frac{\log N}{T}\right).$$

For $\mathcal{G}_6$,

$$\begin{aligned}\left\|\frac{1}{T}\mathbf{X}'_{-i}\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{X}_{-i}\right\|_{\max} &= \left\|\frac{1}{T}\mathbf{X}'_{-i}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{-i}\right\|_{\max}\\ &\leq \max_{i\leq N}\|\mathcal{B}_{1i}-\mathcal{B}_{2i}+\mathcal{B}_{3i}+\mathcal{B}_{4i}+\mathcal{B}_{5i}\|_2^2\left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\right\|_F\\ &= \mathcal{O}_p\left(\frac{\log N}{T}+\frac{\log^2 N}{N^2}\right).\end{aligned}$$

Putting them together, we have

$$\left\|\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{X}}_{-i}-\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X}},-i}\right\|_{\max} = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right) = \mathbf{o}_p(1), \tag{27}$$

which concludes the proof. □

The following proposition shows the asymptotic normality of the debiased LASSO estimator with a fixed vector d.

Proposition G2. Under the same assumption of Theorem 4.1, for $i = 1,\ldots,N$, we have

$$\sqrt{T}d'\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}-\boldsymbol{\gamma}_{i:}\right)\xrightarrow{d} N(0,\sigma_{\varepsilon,i}^2 d'\mathbf{\Theta}_{-i}d). \tag{28}$$

Proof of Prop. G2 . Recall that

$$\widehat{\boldsymbol{\gamma}}_{i:} = \underset{\boldsymbol{\gamma}_{i:}}{\operatorname{argmin}}\left\|\widetilde{\mathbf{Y}}_i-\widetilde{\mathbf{X}}_{-i}\boldsymbol{\gamma}_{i:}\right\|_2^2+\upsilon_i\|\boldsymbol{\gamma}_{i:}\|_1.$$

By KKT condition of the above optimization problem, we have

$$\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\left(\widetilde{\mathbf{Y}}_i-\widetilde{\mathbf{X}}_{-i}\widehat{\boldsymbol{\gamma}}_{i:}\right) = \upsilon_i\widehat{\boldsymbol{\kappa}}_i, \tag{29}$$

where $\widehat{\boldsymbol{\kappa}}_i$ is the subgradient of $\|\widehat{\boldsymbol{\gamma}}_{i:}\|_1$. By plugging in the equation of $\widetilde{\mathbf{Y}}_i$ and rearranging, we have

$$\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right)+\upsilon_i\widehat{\boldsymbol{\kappa}}_i=\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i,$$

where $\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}=\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{X}}_{-i}$. Pre-multiply $\widehat{\boldsymbol{\Theta}}_{-i}$ on both sides and rearrange, we have

$$\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right)+\widehat{\boldsymbol{\Theta}}_{-i}\upsilon_i\widehat{\boldsymbol{\kappa}}_i=\widehat{\boldsymbol{\Theta}}_{-i}\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i-\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right).$$

Subsistuting $\upsilon_i\widehat{\boldsymbol{\kappa}}_i$ with the KKT condition (29) and by the definition of the proposed debiased LASSO estimators, we have

$$\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}-\boldsymbol{\gamma}_{i:}=\widehat{\boldsymbol{\Theta}}_{-i}\frac{1}{T}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i-\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right). \tag{30}$$

To show (28), note that

$$\sqrt{T}d'\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}-\boldsymbol{\gamma}_{i:}\right)=\frac{1}{\sqrt{T}}d'\widehat{\boldsymbol{\Theta}}_{-i}\widetilde{\mathbf{X}}'_{-i}\widetilde{\mathbf{E}}_i-\sqrt{T}d'\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right).$$

Decomposing the first term according to (26), we have

$$\begin{aligned}\sqrt{T}d'\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)}-\boldsymbol{\gamma}_{i:}\right)=&\frac{1}{\sqrt{T}}d'\boldsymbol{\Theta}_{-i}\mathbf{U}'_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i+d'\left(\widehat{\boldsymbol{\Theta}}_{-i}-\boldsymbol{\Theta}_{-i}\right)\frac{1}{\sqrt{T}}\mathbf{U}'_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\\&+d'\widehat{\boldsymbol{\Theta}}_{-i}\sqrt{T}\sum_{i=2}^{7}\mathcal{A}_i+\sqrt{T}d'\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right).\end{aligned}$$

Under the assumptions of Theorem 4.1 and by the standard CLT for martingale difference sequence, $\frac{1}{\sqrt{T}}d'\boldsymbol{\Theta}_{-i}\mathbf{U}'_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\xrightarrow{d}N(0,\sigma^2_{\varepsilon,i}d'\boldsymbol{\Theta}_{-i}d)$. It is sufficient to show the remaining terms are $\mathbf{o}_p(1)$.

To begin with, we need to bound the estimation error of $\widehat{\boldsymbol{\Theta}}_{-i}$. To ease the notation, denote $\sigma_j^2=\mathbb{E}\left[\frac{1}{T}\mathbf{U}'_{\mathbf{X},j}\mathbf{U}_{\mathbf{X},j}\right]$ and $\widehat{\sigma}_j^2=\frac{1}{T}\widetilde{\mathbf{X}}'_j\widetilde{\mathbf{X}}_j$. So $\widehat{\boldsymbol{\Theta}}_{-i}=\operatorname{diag}\left(\{\widehat{\sigma}_j^{-2}\}_{j\neq i}\right)$. Therefore, we have

$$\left\|\widehat{\boldsymbol{\Theta}}_{-i}-\boldsymbol{\Theta}_{-i}\right\|_\infty=\max_{j\neq i}\left|\widehat{\sigma}_j^{-2}-\sigma_j^{-2}\right|.$$

By Taylor expansion,

$$\left|\widehat{\sigma}_j^{-2}-\sigma_j^{-2}\right|=\frac{1}{\sigma_j^2}\left|\widehat{\sigma}_j^2-\sigma_j^2\right|+\mathcal{O}\left(\left|\widehat{\sigma}_j^2-\sigma_j^2\right|^2\right).$$

Since σ_j^2 is bounded away from zero, it suffices to bound $\max_{j\neq i}\left|\widehat{\sigma}_j^2-\sigma_j^2\right|$. Observe that

$$\max_{j\neq i}\left|\widehat{\sigma}_j^2-\sigma_j^2\right| \leq \left\|\frac{1}{T}\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{X}}_{-i}-\mathbf{\Sigma}_{\mathbf{U}_\mathbf{X},-i}\right\|_{\max} = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right)$$

by (27), which implies that

$$\left\|\widehat{\mathbf{\Theta}}_{-i}-\mathbf{\Theta}_{-i}\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right) \quad \text{and} \quad \left\|\widehat{\mathbf{\Theta}}_{-i}\right\|_\infty = \mathcal{O}_p\left(1\right). \tag{31}$$

So we have

$$\begin{aligned} d'\left(\widehat{\mathbf{\Theta}}_{-i}-\mathbf{\Theta}_{-i}\right)\frac{1}{\sqrt{T}}\mathbf{U}_{\mathbf{X},-i}'\boldsymbol{\varepsilon}_i &\leq \left\|d\right\|_1\left\|\widehat{\mathbf{\Theta}}_{-i}-\mathbf{\Theta}_{-i}\right\|_\infty \max_{j\neq i}\left|\frac{1}{\sqrt{T}}\mathbf{U}_{\mathbf{X},j}'\boldsymbol{\varepsilon}_i\right| \\ &= \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}+\frac{\log N}{N}\right)\mathcal{O}_p\left(\sqrt{\log N}\right) \\ &= \mathcal{O}_p\left(\frac{\log N}{\sqrt{T}}+\frac{\log^{3/2} N}{N}\right) = \mathbf{o}_p\left(1\right). \end{aligned}$$

Similarly,

$$\begin{aligned} d'\widehat{\mathbf{\Theta}}_{-i}\sqrt{T}\sum_{i=2}^{7}\mathcal{A}_i &\leq \left\|d\right\|_1\left\|\widehat{\mathbf{\Theta}}_{-i}\right\|_\infty \sqrt{T}\sum_{i=2}^{7}\left\|\mathcal{A}_i\right\|_\infty \\ &= \mathcal{O}_p\left(1\right)\mathcal{O}_p\left(\frac{(\log N)^{3/4}}{\sqrt{T}}+\frac{\sqrt{T}\log N}{N}\right) = \mathbf{o}_p\left(1\right), \end{aligned}$$

by the bounds we have established for $\mathcal{A}_2$ to $\mathcal{A}_7$ in the proof of Prop. G1. For the last term,

$$\sqrt{T}d'\left(\widehat{\mathbf{\Theta}}_{-i}\widehat{\mathbf{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\left(\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right) \leq \left\|d\right\|_1\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_1\left\|\left(\widehat{\mathbf{\Theta}}_{-i}\widehat{\mathbf{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\right\|_{\max}.$$

By Prop. G1, $\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_1 = \mathcal{O}_p\left(\sqrt{\log N}+\frac{\sqrt{T}\log N}{N}\right)$. Note that

$$\begin{aligned} \left\|\left(\widehat{\mathbf{\Theta}}_{-i}\widehat{\mathbf{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\right\|_{\max} &= \left\|\widehat{\mathbf{\Theta}}_{-i}\left(\widehat{\mathbf{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\widehat{\mathbf{\Theta}}_{-i}^{-1}\right)\right\|_{\max} \\ &\leq \left\|\mathbf{\Theta}_{-i}\right\|_{\max}\left\|\widehat{\mathbf{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\widehat{\mathbf{\Theta}}_{-i}^{-1}\right\|_{\max}+\left\|\widehat{\mathbf{\Theta}}_{-i}-\mathbf{\Theta}_{-i}\right\|_{\max}\left\|\widehat{\mathbf{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}}-\widehat{\mathbf{\Theta}}_{-i}^{-1}\right\|_{\max} \end{aligned}$$

$$= \mathcal{O}_p(1)\,\mathcal{O}_p\left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N}\right).$$

Therefore,

$$\sqrt{T}d'\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{X}}_{-i}} - \mathbf{I}\right)(\widehat{\boldsymbol{\gamma}}_{-i} - \boldsymbol{\gamma}_{-i}) = \mathcal{O}_p\left(\sqrt{\log N} + \frac{\sqrt{T}\log N}{N}\right)\mathcal{O}_p\left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N}\right) = \mathbf{o}_p(1).$$

This proves (28). □

The following proposition shows the Gaussian approximation of the debiased LASSO estimator in the ℓ_∞ norm.

Proposition G3. Under the same assumption of Theorem 4.1, for $i = 1, \ldots, N$, we have

$$\sup_{a\geq 0}\left|\mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right) - \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right| \to 0, \tag{32}$$

where $\boldsymbol{\zeta} \sim N(0, \sigma_{\varepsilon,i}^2\boldsymbol{\Theta}_{-i})$.

Proof of Prop. G3. Observe that

$$\begin{aligned}
&\sup_{a\geq 0}\left|\mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right) - \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right| \\
&\leq \sup_{a\geq 0}\left|\mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right) - \mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a\right)\right| \\
&\quad + \sup_{a\geq 0}\left|\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a\right) - \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right|.
\end{aligned}$$

By Lemma K3, for any $\Delta > 0$, the first term is bounded by

$$\begin{aligned}
&\sup_{a\geq 0}\left|\mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right) - \mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a\right)\right| \\
&\leq \mathbb{P}\left(\left|\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right\|_\infty - \left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty\right| > \Delta\right) + \sup_{a\geq 0}\mathbb{P}\left(a \leq \left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a + \Delta\right) \\
&\leq \mathbb{P}\left(\left|\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right\|_\infty - \left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty\right| > \Delta\right) \\
&\quad + \sup_{a\geq 0}\left|\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a + \Delta\right) - \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a + \Delta\right)\right|
\end{aligned}$$

$$+\sup_{a\geq 0}\left|\mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a+\Delta\right)-\mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right|$$
$$+\sup_{a\geq 0}\left|\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a\right)-\mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right|.$$

So, putting all together,

$$\begin{aligned}
&\sup_{a\geq 0}\left|\mathbb{P}\left(\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_\infty \leq a\right)-\mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right| \\
&\leq \mathbb{P}\left(\left|\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_\infty-\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty\right|>\Delta\right) \\
&\quad+3\sup_{a\geq 0}\left|\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a\right)-\mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right| \\
&\quad+\sup_{a\geq 0}\left|\mathbb{P}\left(a\leq\|\boldsymbol{\zeta}\|_\infty \leq a+\Delta\right)\right|.
\end{aligned}$$

By Theorem 1 of Chang et al. (2024) and under the assumptions of Theorem 4.1, we have

$$\sup_{a\geq 0}\left|\mathbb{P}\left(\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \leq a\right)-\mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a\right)\right|\to 0.$$

For the first term, by the decomposition in (30),

$$\begin{aligned}
\left|\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_\infty-\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty\right| &\leq \left\|\widehat{\boldsymbol{\Theta}}_{-i}-\boldsymbol{\Theta}_{-i}\right\|_\infty\left\|\frac{1}{\sqrt{T}}\mathbf{U}'_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty \\
&\quad+\left\|\widehat{\boldsymbol{\Theta}}_{-i}\right\|_\infty\sqrt{T}\sum_{i=2}^{7}\|\mathcal{A}_i\|_\infty \\
&\quad+\sqrt{T}\left\|\left(\widehat{\boldsymbol{\Theta}}_{-i}\widehat{\boldsymbol{\Sigma}}_{\tilde{\mathbf{X}}_{-i}}-\mathbf{I}\right)\right\|_{\max}\left\|\widehat{\boldsymbol{\gamma}}_{-i}-\boldsymbol{\gamma}_{-i}\right\|_1.
\end{aligned}$$

Therefore, by the same bounds established in the proof of Prop. G2, we have

$$\left|\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_\infty-\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty\right|=\mathcal{O}_p\left(\frac{\log N}{\sqrt{T}}+\frac{\sqrt{T}\log N}{N}\right).$$

So let $\Delta=\left(\frac{\log^{1+\delta}N}{\sqrt{T}}+\frac{\sqrt{T}\log^{1+\delta}N}{N}\right)$ for some small $\delta>0$. Then we have

$$\mathbb{P}\left(\left|\sqrt{T}\left\|\widehat{\boldsymbol{\gamma}}_{i:}-\boldsymbol{\gamma}_{i:}\right\|_\infty-\left\|\frac{1}{\sqrt{T}}\boldsymbol{\Theta}_{-i}\mathbf{U}_{\mathbf{X},-i}\boldsymbol{\varepsilon}_i\right\|_\infty\right|>\Delta\right)\to 0.$$

By Nazarov's anti-concentration inequality for Gaussian random vectors (see, e.g., Lemma A.1. in Chernozhukov et al. (2017)), we have

$$\mathbb{P}\left(a \leq \|\boldsymbol{\zeta}\|_{\infty} \leq a + \Delta\right) = \mathcal{O}\left(\sqrt{\log N}\Delta\right) \to 0.$$

So putting all together yields (32). □

The following proposition shows the consistency of the variance estimator $\widehat{\sigma}^2_{\varepsilon,i}$.

Proposition G4. Under the same assumption of Theorem 4.1, for $i = 1, \ldots, N$, we have

$$\widehat{\sigma}^2_{\varepsilon,i} - \sigma^2_{\varepsilon,i} = \mathcal{O}_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right).$$

Proof of Prop. G4. Decomposing $\widehat{\sigma}^2_{\varepsilon,i} - \sigma^2_{\varepsilon,i}$ according to the definition, we have

$$\left|\frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i - \sigma^2_{\varepsilon,i}\right| \leq \left|\frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right| + \left|\frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i - \sigma^2_{\varepsilon,i}\right|.$$

The second term is $\mathcal{O}_p\left(\sqrt{\frac{1}{T}}\right)$ by the standard CLT for martingale difference sequence. For the first term,

$$\begin{aligned}
&\frac{1}{T}\widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i \\
&= \frac{1}{T}\left(\widetilde{\mathbf{Y}}_i - \widetilde{\mathbf{X}}_{-i}\widehat{\boldsymbol{\gamma}}_{i:}\right)'\left(\widetilde{\mathbf{Y}}_i - \widetilde{\mathbf{X}}_{-i}\widehat{\boldsymbol{\gamma}}_{i:}\right) - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i \\
&= \left(\widetilde{\mathbf{X}}_{-i}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right) + \widetilde{\mathbf{E}}_i\right)'\left(\widetilde{\mathbf{X}}_{-i}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right) + \widetilde{\mathbf{E}}_i\right) - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i \\
&= \left(\frac{1}{T}\widetilde{\mathbf{E}}_i'\widetilde{\mathbf{E}}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right) + \frac{2}{T}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right)'\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{E}}_i + \frac{1}{T}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right)'\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{X}}_{-i}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right).
\end{aligned}$$

By (24) and (27) and Prop. G1,

$$\frac{2}{T}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right)'\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{E}}_i \leq \|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\|_1 \frac{2}{T}\left\|\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{E}}_i\right\|_{\infty} = \mathcal{O}_p\left(\frac{\log N}{T} + \frac{\log^2 N}{N^2}\right),$$

and

$$\frac{1}{T}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right)'\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{X}}_{-i}\left(\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\right) \leq \|\widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:}\|_1^2 \left\|\frac{1}{T}\widetilde{\mathbf{X}}_{-i}'\widetilde{\mathbf{X}}_{-i}\right\|_{\max} = \mathcal{O}_p\left(\frac{\log N}{T} + \frac{\log^2 N}{N^2}\right).$$

The first term can be decomposed as

$$\begin{aligned}
&\frac{1}{T}\widetilde{\mathbf{E}}_i'\widetilde{\mathbf{E}}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i \\
&= \left(\frac{1}{T}\widetilde{\boldsymbol{\varepsilon}}_i'\widetilde{\boldsymbol{\varepsilon}}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i\right) + \frac{2}{T}\boldsymbol{\Lambda}_i'\widetilde{\mathbf{F}}'\widetilde{\boldsymbol{\varepsilon}}_i + \frac{1}{T}\boldsymbol{\Lambda}_i'\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\boldsymbol{\Lambda}_i \\
&:= \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{H}_3.
\end{aligned}$$

Bound $\mathcal{H}_2$ first, observe that

$$\begin{aligned}
\frac{1}{T}\boldsymbol{\Lambda}_i'\widetilde{\mathbf{F}}'\widetilde{\boldsymbol{\varepsilon}}_i &= \frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\widehat{\mathbb{Z}},i}\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i \\
&= \frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i - \frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{P}_{\widehat{\mathbb{Z}},i}\boldsymbol{\varepsilon}_i
\end{aligned}$$

Since $\mathbf{F}$ is contained in $\mathbf{G}_{\mathbf{X}}$, let $\mathbf{C}_{\mathbf{Y},i} = [\boldsymbol{\Lambda}_i', \mathbf{0}]'$ expand the loading with zeros to match the dimension of $\mathbf{G}_{\mathbf{X}}$, then we have

$$\frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \frac{1}{T}\mathbf{C}_{\mathbf{Y},i}'\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \mathcal{O}_p\left(\frac{1}{N} + \frac{1}{T}\right),$$

by the same argument for $\mathcal{D}_3$ in the proof of Prop. G1. Also, we have

$$\frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{P}_{\widehat{\mathbb{Z}},i}\boldsymbol{\varepsilon}_i \leq \left\|\frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right\|_2 \left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\mathbf{G}}\mathbb{Z}_i\right)^{-1}\right\|_2 \left\|\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2.$$

By bounds on $\frac{1}{T}\left\|\mathcal{E}_1 - \mathcal{E}_2 + \mathcal{E}_3\right\|_2$ in the proof of Prop. G1, we have

$$\left\|\frac{1}{T}\boldsymbol{\Lambda}_i'\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right).$$

By Lemma I5(iv),

$$\left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\mathbf{G}}\mathbb{Z}_i\right)^{-1}\right\|_2 = \mathcal{O}_p\left(1\right).$$

By the bounds on $\frac{1}{T}\left\|\mathcal{D}_1 - \mathcal{D}_2 + \mathcal{D}_3\right\|_2$, we have

$$\left\|\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{T}} + \frac{1}{N}\right).$$

So we have

$$\mathcal{H}_2 = \frac{1}{T}\mathbf{\Lambda}_i'\widetilde{\mathbf{F}}'\widetilde{\boldsymbol{\varepsilon}}_i = \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right).$$

For $\mathcal{H}_3$,

$$\frac{1}{T}\mathbf{\Lambda}_i'\widetilde{\mathbf{F}}'\widetilde{\mathbf{F}}\mathbf{\Lambda}_i \leq \|\mathbf{\Lambda}_i\|_2^2 \frac{1}{T}\left\|\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\right\|_2 + \|\mathbf{\Lambda}_i\|_2^2 \frac{1}{T}\left\|\mathbf{F}'\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{F}\right\|_2.$$

$\frac{1}{T}\left\|\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\right\|_2 \leq \frac{1}{T}\left\|\left(\widehat{\mathbf{G}}_{\mathbf{X}} - \mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\mathbf{H}_{\mathbf{X}}^{-1}\mathbf{M}_{\widehat{\mathbf{G}}}\right\|_2^2 = \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{N}\right)$. And

$$\frac{1}{T}\left\|\mathbf{F}'\mathbf{P}_{\widehat{\mathbb{Z}},i}\mathbf{F}\right\|_2 \leq \left\|\frac{1}{T}\mathbf{F}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right\|_2^2 \left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\right\|_2 = \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{N^2}\right).$$

So we have $\mathcal{H}_3 = \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{N}\right)$. For $\mathcal{H}_1$,

$$\begin{aligned}\frac{1}{T}\widetilde{\boldsymbol{\varepsilon}}_i'\widetilde{\boldsymbol{\varepsilon}}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i &= \frac{1}{T}\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\mathbf{P}_{\mathbb{Z},i}\boldsymbol{\varepsilon}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\boldsymbol{\varepsilon}_i \\ &= \frac{1}{T}\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i - \frac{1}{T}\boldsymbol{\varepsilon}_i'\mathbf{P}_{\mathbb{Z},i}\boldsymbol{\varepsilon}_i\end{aligned}$$

The first term $\frac{1}{T}\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \mathcal{O}_p\left(\frac{1}{N}\right)$ by Lemma I4. The second term is bounded by

$$\frac{1}{T}\boldsymbol{\varepsilon}_i'\mathbf{P}_{\mathbb{Z},i}\boldsymbol{\varepsilon}_i \leq \left\|\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2^2 \left\|\left(\frac{1}{T}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i\right)^{-1}\right\|_2 = \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{N^2}\right).$$

The result follows by putting all the bounds together. □

Now we are ready to prove Theorem 4.1.

Proof of Theorem 4.1. To show the asymptotic normality in equation (11), by Prop. G2, we aleady have

$$\sqrt{T}d'\left(\widehat{\boldsymbol{\gamma}}_{i:}^{(db)} - \boldsymbol{\gamma}_{i:}\right)/\sqrt{d'\mathbf{\Theta}_{-i}d\sigma_{\varepsilon,i}^2} \xrightarrow{d} N(0,1).$$

It is sufficient to show that

$$d'\widehat{\mathbf{\Theta}}_{-i}d\widehat{\sigma}_{\varepsilon,i}^2 - d'\mathbf{\Theta}_{-i}d\sigma_{\varepsilon,i}^2 = \mathbf{o}_p(1),$$

and the result follows by Slutsky's theorem. By Prop. G4 and equation (31),

$$\begin{aligned}d'\widehat{\mathbf{\Theta}}_{-i}d\widehat{\sigma}_{\varepsilon,i}^2 - d'\mathbf{\Theta}_{-i}d\sigma_{\varepsilon,i}^2 &= d'\left(\widehat{\mathbf{\Theta}}_{-i}\widehat{\sigma}_{\varepsilon,i}^2 - \mathbf{\Theta}_{-i}\sigma_{\varepsilon,i}^2\right)d \\ &\leq \|d\|_1^2\left\|\widehat{\mathbf{\Theta}}_{-i}\widehat{\sigma}_{\varepsilon,i}^2 - \mathbf{\Theta}_{-i}\sigma_{\varepsilon,i}^2\right\|_{\max}\end{aligned}$$

$$
\begin{aligned}
&\leq \|d\|_1^2 \left\| \widehat{\boldsymbol{\Theta}}_{-i} \widehat{\sigma}^2_{\varepsilon,i} - \widehat{\boldsymbol{\Theta}}_{-i} \sigma^2_{\varepsilon,i} + \widehat{\boldsymbol{\Theta}}_{-i} \sigma^2_{\varepsilon,i} - \boldsymbol{\Theta}_{-i} \sigma^2_{\varepsilon,i} \right\|_{\max} \\
&\leq \|d\|_1^2 \left\| \widehat{\boldsymbol{\Theta}}_{-i} \right\|_{\max} \left| \widehat{\sigma}^2_{\varepsilon,i} - \sigma^2_{\varepsilon,i} \right| + \|d\|_1^2 \left\| \widehat{\boldsymbol{\Theta}}_{-i} - \boldsymbol{\Theta}_{-i} \right\|_{\max} \sigma^2_{\varepsilon,i} \\
&= \mathcal{O}_p \left(\sqrt{\frac{\log N}{T}} + \frac{\log N}{N} \right) = \mathbf{o}_p(1).
\end{aligned}
$$

So we prove (11). To show the Gaussian approximation in the ℓ_∞ norm (12), observe that

$$
\begin{aligned}
&\sup_{a \geq 0} \left| \mathbb{P}\left(\sqrt{T} \left\| \widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:} \right\|_\infty \leq a \right) - \mathbb{P}\left(\widehat{\mathcal{I}}_i \leq a \right) \right| \\
&\leq \sup_{a \geq 0} \left| \mathbb{P}\left(\sqrt{T} \left\| \widehat{\boldsymbol{\gamma}}_{i:} - \boldsymbol{\gamma}_{i:} \right\|_\infty \leq a \right) - \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a \right) \right| + \sup_{a \geq 0} \left| \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a \right) - \mathbb{P}\left(\widehat{\mathcal{I}}_i \leq a \right) \right|,
\end{aligned}
$$

where $\boldsymbol{\zeta}$ is defined in Prop. G3. By Prop. G3, the first term converges to zero. For the second term, by Lemma 2.1 in Chernozhukov et al. (2023),

$$
\sup_{a \geq 0} \left| \mathbb{P}\left(\|\boldsymbol{\zeta}\|_\infty \leq a \right) - \mathbb{P}\left(\widehat{\mathcal{I}}_i \leq a \right) \right| \lesssim \log N \left\| \widehat{\boldsymbol{\Theta}}_{-i} \widehat{\sigma}^2_{\varepsilon,i} - \boldsymbol{\Theta}_{-i} \sigma^2_{\varepsilon,i} \right\|_{\max} = \mathcal{O}_p \left(\frac{\log^{3/2} N}{T} + \frac{\log^2 N}{N} \right) = \mathbf{o}_p(1),
$$

which proves (12). □

Next, we proceed to prove Theorem 4.2. The following proposition shows the uniform model selection consistency of the adaptive LASSO estimators defined in Algorithm 3. Denote $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_{1:}, \ldots, \boldsymbol{\gamma}_{N:})' \in \mathbb{R}^{N(N-1)}$ as the vector of all spillover parameters and denote $\widehat{\boldsymbol{\gamma}}^{(al)}$ analogously. Define the variable selection consistency notation:

$$
\widehat{\boldsymbol{\gamma}}^{(al)} =_s \boldsymbol{\gamma} \quad \text{if} \quad \operatorname{sign}\left(\widehat{\boldsymbol{\gamma}}^{(al)} \right)_j = \operatorname{sign}(\boldsymbol{\gamma})_j \text{ for } \boldsymbol{\gamma}_j \neq 0 \text{ and } \widehat{\boldsymbol{\gamma}}^{(al)}_j = 0 \text{ for } \boldsymbol{\gamma}_j = 0,
$$

where, abusing the notation, $\boldsymbol{\gamma}_j$ denotes the j^{th} element of $\boldsymbol{\gamma}$.

Proposition G5. Under the assumptions of Theorem 4.2, we have

$$
\mathbb{P}\left(\widehat{\boldsymbol{\gamma}}^{(al)} =_s \boldsymbol{\gamma} \right) \to 1.
$$

Proof of Proposition G5. Observe that

$$
\mathbb{P}\left(\widehat{\boldsymbol{\gamma}}^{(al)} =_s \boldsymbol{\gamma} \right) = \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} =_s \boldsymbol{\gamma}_{i:} \text{ for all } i \right) = 1 - \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} \neq_s \boldsymbol{\gamma}_{i:} \text{ for some } i \right) \geq 1 - \sum_{i=1}^{N} \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} \neq_s \boldsymbol{\gamma}_{i:} \right).
$$

Therefore, it is sufficient to show that

$$\sum_i \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} \neq_s \boldsymbol{\gamma}_{i:}\right) \to 0.$$

For each i, without loss of generality, rearrange $\boldsymbol{\gamma}_{i:}$ and $\widehat{\boldsymbol{\gamma}}_{i:}$ so that the non-zero elements of $\boldsymbol{\gamma}_{i:}$ appear first. Let the subscript s denote the subset corresponding to the non-zero elements of $\boldsymbol{\gamma}_{i:}$ and s^C denote its complement. Denote

$$\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s} = \frac{1}{T}\widetilde{\mathbf{X}}'_{-i,s}\widetilde{\mathbf{X}}_{-i,s}, \quad \widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s,s^C} = \frac{1}{T}\widetilde{\mathbf{X}}'_{-i,s}\widetilde{\mathbf{X}}_{-i,s^C}, \quad \mathbf{M}_{\widetilde{\mathbf{X}}_{-i},s} = \mathbf{I}_T - \frac{1}{T}\widetilde{\mathbf{X}}_{-i,s}\left(\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s}\right)^{-1}\widetilde{\mathbf{X}}'_{-i,s}.$$

And denote

$$\widehat{\boldsymbol{\Psi}}_{i,s} = \operatorname{diag}\left(\left\{\frac{1}{\left|\widehat{\gamma}_{ij}^{(db)}\right|}\right\}_{j\in S_i}\right), \quad \widehat{\boldsymbol{\psi}}_{i,s} = \left(\left\{\frac{1}{\left|\widehat{\gamma}_{ij}^{(db)}\right|}\right\}_{j\in S_i}\right)' \quad \widehat{\boldsymbol{\psi}}_{i,s^C} = \left(\left\{\frac{1}{\left|\widehat{\gamma}_{ij}^{(db)}\right|}\right\}_{j\in S_i^C}\right)'.$$

By Prop. J2,

$$\mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} \neq_s \boldsymbol{\gamma}_{i:}\right) \leq \mathbb{P}\left(A_{i,T}^C\right) + \mathbb{P}\left(B_{i,T}^C\right),$$

where

$$A_{i,T}^C = \left\{2\frac{1}{T}\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s}^{-1}\widetilde{\mathbf{X}}'_{-i,s}\widetilde{\boldsymbol{\varepsilon}}_i\right| \geq 2\left|\boldsymbol{\gamma}_{i:,s}\right| - v_i\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s}^{-1}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right|\right\}$$
$$B_{i,T}^C = \left\{2\frac{1}{T}\left|\widetilde{\mathbf{X}}_{-i,s^C}\mathbf{M}_{\widetilde{\mathbf{X}}_{-i},s}\widetilde{\boldsymbol{\varepsilon}}_i\right| > v_i\widehat{\boldsymbol{\psi}}_{i,s^C} - v_i\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s,s^C}\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s}^{-1}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right|\right\}$$

Denote

$$\boldsymbol{\psi}_{i,s} = \boldsymbol{\psi}_{i,s} = \left(\left\{\frac{1}{|\gamma_{ij}|}\right\}_{j\in S_i}\right)', \quad \boldsymbol{\Sigma}_{\mathbf{U}_{\mathbf{X},-i,s}} = \mathbb{E}\left[\frac{1}{T}\mathbf{U}'_{\mathbf{X},-i,s}\mathbf{U}_{\mathbf{X},-i,s}\right].$$

$\underline{\phi}_i = \left|v_i\boldsymbol{\Sigma}_{\mathbf{U}_{\mathbf{X},-i},s}^{-1}\min_j\left|\boldsymbol{\psi}_{i,s}\right|_j\right|$, where $|\cdot|_j$ denotes the j^{th} element of absolute values of the vector. Then we bound $\mathbb{P}\left(A_{i,T}^C\right)$ by

$$\mathbb{P}\left(A_{i,T}^C\right) \leq \mathbb{P}\left(\frac{1}{T}\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s}^{-1}\widetilde{\mathbf{X}}'_{-i,s}\widetilde{\boldsymbol{\varepsilon}}_i\right| \geq \left|\boldsymbol{\gamma}_{i:,s}\right| - \underline{\phi}_i/2\right) + \mathbb{P}\left(v_i\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s}^{-1}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| < \underline{\phi}_i\right).$$

For the first probability term, following the same argument as in the proof of Prop. G2, the

left-hand side is dominated by

$$\frac{1}{T}\left|\boldsymbol{\Sigma}_{\mathbf{U}_{\mathbf{X},-i,s}}\mathbf{U}'_{\mathbf{X},-i,s}\boldsymbol{\varepsilon}_i\right|.$$

By Theroem 1, Remark 1 to 3 and Equation (1.11) in Merlevède et al. (2011), we have

$$\begin{aligned}\mathbb{P}\left(\frac{1}{T}\left|\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widetilde{\mathbf{X}}'_{-i,s}\widetilde{\boldsymbol{\varepsilon}}_i\right| \geq \left|\boldsymbol{\gamma}_{i:,s}\right| - \underline{\phi}_i/2\right) \lesssim & s_i(T+1)\exp\left(-T^{\iota}\left|\min_j\left|\boldsymbol{\gamma}_{i:,s}\right|_j - \underline{\phi}_i/2\right|^{\iota}\right) \\ & + s_i\exp\left(-T^2\left|\min_j\left|\boldsymbol{\gamma}_{i:,s}\right|_j - \underline{\phi}_i/2\right|^2\right)\end{aligned}$$

where $\iota = \frac{1}{1/\nu+1/2}$. As $v_i \to 0$ and $\iota \leq 2$, we have

$$\mathbb{P}\left(\frac{1}{T}\left|\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widetilde{\mathbf{X}}'_{-i,s}\widetilde{\boldsymbol{\varepsilon}}_i\right| \geq \left|\boldsymbol{\gamma}_{i:,s}\right| - \underline{\phi}_i/2\right) = \mathcal{O}\left(T\exp\left(-T^{\iota}\right)\right).$$

For the second probability term, observe that

$$\begin{aligned}\mathbb{P}\left(v_i\left|\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| < \underline{\phi}_i\right) &= \mathbb{P}\left(\left|\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| < \left|\boldsymbol{\Sigma}^{-1}_{\mathbf{U}_{\mathbf{X},-i},s}\min_j\left|\boldsymbol{\psi}_{i,s}\right|_j\right|\right) \\ &= \mathbb{P}\left(\left|\boldsymbol{\Sigma}_{\mathbf{U}_{\mathbf{X},-i},s}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| < \min_j\left|\boldsymbol{\psi}_{i,s}\right|_j\right).\end{aligned}$$

By a similar argument, The left-hand side is dominated by

$$\left|\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| = \left|\widehat{\boldsymbol{\psi}}_{i,s}\right|.$$

So we have

$$\begin{aligned}\mathbb{P}\left(\left|\boldsymbol{\Sigma}_{\mathbf{U}_{\mathbf{X},-i},s}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| < \min_j\left|\boldsymbol{\psi}_{i,s}\right|_j\right) &\lesssim \mathbb{P}\left(\left|\widehat{\boldsymbol{\psi}}_{i,s}\right| < \min_j\frac{1}{2}\left|\boldsymbol{\psi}_{i,s}\right|_j\right) \\ &= \mathbb{P}\left(\left|\widehat{\boldsymbol{\gamma}}^{(db)}_{i,s}\right| > 2\max_j\left|\boldsymbol{\gamma}_{i,s}\right|_j\right) \\ &\leq \mathbb{P}\left(\left|\widehat{\boldsymbol{\gamma}}^{(db)}_{i,s} - \boldsymbol{\gamma}_{i,s}\right| > \max_j\left|\boldsymbol{\gamma}_{i,s}\right|_j\right).\end{aligned}$$

By the proof of Prop. G2, the left-hand side is dominated by

$$\left|\frac{1}{T}\boldsymbol{\Theta}_{-i}\mathbf{U}'_{\mathbf{X},-i,s}\boldsymbol{\varepsilon}_i\right|.$$

So we have, by Merlevède et al. (2011) again,

$$\begin{aligned}\mathbb{P}\left(\left|\widehat{\boldsymbol{\gamma}}_{i,s}^{(db)}-\boldsymbol{\gamma}_{i,s}\right|>\max_j\left|\boldsymbol{\gamma}_{i,s}\right|_j\right) &\lesssim \mathbb{P}\left(\left|\frac{1}{T}\boldsymbol{\Theta}_{-i}\mathbf{U}'_{\mathbf{X},-i,s}\boldsymbol{\varepsilon}_i\right|>\frac{1}{2}\max_j\left|\boldsymbol{\gamma}_{i,s}\right|_j\right)\\ &\lesssim s_i(T+1)\exp\left(-T^{\iota}\left(\frac{1}{2}\max_j\left|\boldsymbol{\gamma}_{i,s}\right|_j\right)^{\iota}\right)+s_i\exp\left(-T^2\left(\frac{1}{2}\max_j\left|\boldsymbol{\gamma}_{i,s}\right|_j\right)^2\right)\\ &=\mathcal{O}\left(T\exp\left(-T^{\iota}\right)\right).\end{aligned}$$

Putting two pieces together gives the bound for $\mathbb{P}\left(A_{i,T}^C\right)$:

$$\mathbb{P}\left(A_{i,T}^C\right)=\mathcal{O}\left(T\exp\left(-T^{\iota}\right)\right).$$

For $\mathbb{P}\left(B_{i,T}^C\right)$, define $\underline{\phi}_i=\left|v_i\boldsymbol{\Sigma}^{-1}_{\mathbf{U}_{\mathbf{X},-i,s}}\min_j\left|\boldsymbol{\psi}_{i,s}\right|_j\right|$ and event $D_{i,T}$ as

$$D_{i,T}=\left\{v_i\widehat{\boldsymbol{\psi}}_{i,s^C}>3c_0, v_i\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s,s^C}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right|<c_0\right\},$$

for some choice of $c_0>0$. Then we have

$$\begin{aligned}\mathbb{P}\left(B_{i,T}^C\right)&\leq\mathbb{P}\left(B_{i,T}^C\cap D_{i,T}\right)+\mathbb{P}\left(D_{i,T}^C\right)\\ &\leq\mathbb{P}\left(\frac{1}{T}\left|\widetilde{\mathbf{X}}_{-i,s^C}\mathbf{M}_{\widetilde{\mathbf{X}}_{-i},s}\widetilde{\boldsymbol{\varepsilon}}_i\right|>c_0\right)+\mathbb{P}\left(v_i\widehat{\boldsymbol{\psi}}_{i,s^C}<3c_0\right)\\ &\quad+\mathbb{P}\left(v_i\left|\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s,s^C}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right|>c_0\right).\end{aligned}$$

By a similar argument from the first term of $\mathbb{P}\left(A_{i,T}^C\right)$, we have

$$\mathbb{P}\left(\frac{1}{T}\left|\widetilde{\mathbf{X}}_{-i,s^C}\mathbf{M}_{\widetilde{\mathbf{X}}_{-i},s}\widetilde{\boldsymbol{\varepsilon}}_i\right|>c_0\right)=\mathcal{O}\left(NT\exp\left(-T^{\iota}\right)\right),$$

since the dimension of s^C is at most N. For the third term, notice that

$$\mathbb{E}\left[\frac{1}{T}\mathbf{U}'_{\mathbf{X},-i,s}\mathbf{U}_{\mathbf{X},-i,s^C}\right]=0,$$

by Assumption 3. So $\widetilde{\boldsymbol{\Sigma}}_{\mathbf{X}_{-i},s,s^C}$ is going to zero in probability by the same argument as in the proof of Prop. G1. The remaining term $\widetilde{\boldsymbol{\Sigma}}^{-1}_{\mathbf{X}_{-i},s}\widehat{\boldsymbol{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)$ coincides with the second probability term

in $\mathbb{P}\left(A_{i,T}^{C}\right)$. Therefore, with some suitable choice of c_0, we have

$$\mathbb{P}\left(v_i \left|\widetilde{\mathbf{\Sigma}}_{\mathbf{X}_{-i},s,s^C}\widetilde{\mathbf{\Sigma}}_{\mathbf{X}_{-i},s}^{-1}\widehat{\mathbf{\Psi}}_{i,s}\operatorname{sign}\left(\boldsymbol{\gamma}_{i:,s}\right)\right| > c_0\right) = \mathbf{o}\left(NT\exp\left(-T^{\iota}\right)\right).$$

For the second term, note that

$$\begin{aligned}
\mathbb{P}\left(v_i\widehat{\boldsymbol{\psi}}_{i,s^C} < 3c_0\right) &= \mathbb{P}\left(\min_{j\in s^C}\frac{v_i}{\left|\widehat{\boldsymbol{\gamma}}_{ij}^{(db)}\right|} < 3c_0\right) \\
&= \mathbb{P}\left(\max_{j\in s^C}\left|\widehat{\boldsymbol{\gamma}}_{ij}^{(db)}\right| > \frac{v_i}{3c_0}\right) \\
&= \mathbb{P}\left(\max_{j\in s^C}\left|\widehat{\boldsymbol{\gamma}}_{ij}^{(db)} - \boldsymbol{\gamma}_{ij}\right| > \frac{v_i}{3c_0}\right) \\
&\lesssim N(T+1)\exp\left(-(Tv_i)^{\iota}\right),
\end{aligned}$$

by the same argument as in the first term of $\mathbb{P}\left(A_{i,T}^{C}\right)$. Putting all $A_{i,T}^{C}$ and $B_{i,T}^{C}$ together, we have

$$\sum_i \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} \neq_s \boldsymbol{\gamma}_{i:}\right) \lesssim N^2T\exp\left(-(Tv_i)^{\iota}\right).$$

By the condition on v_i in Theorem 4.2, we have $(Tv_i)^{\iota} \geq T^c$ for some $c > 0$. So

$$\sum_i \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}_{i:} \neq_s \boldsymbol{\gamma}_{i:}\right) \lesssim N^2T\exp\left(-(Tv_i)^{\iota}\right) \leq \exp\left(-\frac{2\log N\log T}{T^c}\right) \to 0,$$

which completes the proof. □

Proof of Theorem 4.2. Since Proposition J2 shows the selection consistency uniformly over i, we could derive the asymptotic normality of the private effects estimators conditional on the model selection consistency. Rigorouly, let $\widehat{\mathbf{S}}$ denote the selected set of spillover variables and $\mathbf{S}$ the true set. The post-selection OLS estimator $\widehat{\theta}$ depend on the $\widehat{\mathbf{S}}$ so I use $\widehat{\theta}(\widehat{\mathbf{S}})$ to denote the dependency. We have

$$\begin{aligned}
\mathbb{P}\left(\sqrt{NT}\left(\widehat{\theta}(\widehat{\mathbf{S}}) - \theta\right) \leq \delta\right) &\leq \mathbb{P}\left(\sqrt{NT}\left(\widehat{\theta}(\widehat{\mathbf{S}}) - \theta\right) \leq \delta, \widehat{\boldsymbol{\gamma}}^{(al)} =_s \boldsymbol{\gamma}\right) + \mathbb{P}\left(\widehat{\boldsymbol{\gamma}}^{(al)} \neq_s \boldsymbol{\gamma}\right) \\
&\leq \mathbb{P}\left(\sqrt{NT}\left(\widehat{\theta}(\widehat{\mathbf{S}}) - \theta\right) \leq \delta, \widehat{\boldsymbol{\gamma}}^{(al)} =_s \boldsymbol{\gamma}\right) + \mathbf{o}\left(1\right) \\
&= \mathbb{P}\left(\sqrt{NT}\left(\widehat{\theta}(\mathbf{S}) - \theta\right) \leq \delta\right) + \mathbf{o}\left(1\right)
\end{aligned}$$

by Prop. J2. Therefore, it is sufficient to show

$$\sqrt{NT}\left(\widehat{\theta}(\mathbf{S})-\theta\right)\xrightarrow{d}N(0,\mathbf{\Sigma}_{\mathbb{U}\varepsilon}).$$

Recall that $\mathbf{Z}_i$ stacks $(\mathbf{X}_i,\mathbf{W}_{1,i},\ldots,\mathbf{W}_{K,i})$ and therefore admits the following multi-variate factor model structure:

$$\mathbb{Z}_i=\mathbb{B}_i\mathbf{G}'+\mathbb{U}_i,$$

where $\mathbb{B}_i$ stacks the factor loadings of $\mathbf{X}_i$ and $\mathbf{W}_{k,i}$, and $\mathbb{U}_i$ stacks the idiosyncratic errors of $\mathbf{X}_i$ and $\mathbf{W}_{k,i}$. Decompose the post-selection OLS estimator:

$$\begin{aligned}\sqrt{NT}\left(\widehat{\theta}(\mathbf{S})-\theta\right)&=\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{U}_i'\boldsymbol{\varepsilon}_i+\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{U}_i'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i+\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{B}_i\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\\&\quad+\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S}_i}\left(\widehat{\boldsymbol{\gamma}}_{\mathbf{S}_i}^{(ps)}-\boldsymbol{\gamma}_{\mathbf{S}_i}\right)\\&\quad+\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{F}\mathbf{\Lambda}_i\\&:=\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{U}_i'\boldsymbol{\varepsilon}_i+\mathcal{H}_1+\mathcal{H}_2+\mathcal{H}_3+\mathcal{H}_4,\end{aligned}$$

where

$$\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}=\frac{1}{NT}\sum_i\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i.$$

By Lemma I5(iv), we have

$$\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}=\frac{1}{N}\sum_i\mathbb{E}\left[\frac{1}{T}\mathbb{U}_i'\mathbb{U}_i\right]+\mathbf{o}_p\left(1\right).$$

Therefore, under the assumptions of Theorem 4.2, we have

$$\widehat{\mathbf{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\mathbb{U}_i'\boldsymbol{\varepsilon}_i\xrightarrow{d}N(0,\mathbf{\Sigma}_{\mathbb{U}\varepsilon}).$$

It is sufficient to show the remaining terms are $\mathbf{o}_p\left(1\right)$. By Lemma I6(i), $\mathcal{H}_1=\mathcal{O}_p\left(\frac{1}{\sqrt{T}}\right)=\mathbf{o}_p\left(1\right)$.

For $\mathcal{H}_2$, we have

$$\begin{aligned}
\mathbb{E}\left\|\mathcal{H}_2\right\|_2^2 &= \mathbb{E}\left\|\frac{1}{NT}\sum_i\sum_j \mathbb{B}_i\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbb{B}_j'\right\|_2 \\
&= \mathbb{E}\left\|\frac{1}{NT}\sum_i \mathbb{B}_i\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbb{B}_i'\right\|_2 \\
&\leq \left\|\boldsymbol{\Sigma}_{\varepsilon,T}\right\|_2 \mathbb{E}\left\|\frac{1}{NT}\sum_i \mathbb{B}_i\mathbf{G}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}\mathbb{B}_i'\right\|_2 \\
&\leq \left\|\boldsymbol{\Sigma}_{\varepsilon,T}\right\|_2 \frac{1}{T}\mathbb{E}\left\|\widehat{\mathbf{G}}-\mathbf{G}\mathbf{H}\right\|_2^2 \mathbb{E}\left\|\frac{1}{N}\sum_i \mathbb{B}_i\mathbb{B}_i'\right\|_2 \\
&= \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{N^2}\right) = \mathbf{o}_p\left(1\right),
\end{aligned}$$

which implies $\mathcal{H}_2 = \mathbf{o}_p\left(1\right)$ by Chebyshev's inequality.

For $\mathcal{H}_3$, observe that

$$\begin{aligned}
\mathcal{H}_3 &= \widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S}_i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{F}}\boldsymbol{\Lambda}_i\right) \\
&\quad + \widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N \mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S}_i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\boldsymbol{\varepsilon}}_i\right) \\
&:= \mathcal{H}_{3,1}+\mathcal{H}_{3,2}.
\end{aligned}$$

Analogous to the argument for $\mathcal{A}_4$, $\mathcal{A}_5$ and $\mathcal{A}_7$ in the proof of Prop. J1 (no-max version so logarithmic factors are dropped), we have

$$\begin{aligned}
\left\|\mathcal{H}_{3,1}\right\|_2 &\leq \sqrt{N}\left\|\widehat{\boldsymbol{\Sigma}}_{\widehat{\mathbb{Z}}}^{-1}\right\|_2 \frac{1}{N}\sum_{i=1}^N \frac{1}{\sqrt{T}}\left\|\mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S}_i}\right\|_2 \left\|\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\right\|_2 \left\|\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{F}}\boldsymbol{\Lambda}_i\right\|_2 \\
&= \sqrt{N}\mathcal{O}_p\left(1\right)\frac{1}{N}\sum_{i=1}^N \mathcal{O}_p\left(1\right)\mathcal{O}_p\left(1\right)\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right) = \mathbf{o}_p\left(1\right).
\end{aligned}$$

For $\mathcal{H}_{3,2}$, abusing notations, let $\mathbf{M}_{\widehat{\mathbf{G}}\mathbb{Z}\widehat{\mathbf{G}},i} = \mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{M}_{\mathbb{Z},i}\mathbf{M}_{\widehat{\mathbf{G}}}$, WLOG, consider one entry of $\mathcal{H}_{3,2}$,

replacing $\mathbb{Z}_i$ with $\mathbf{X}_i$, we have

$$\begin{aligned}
\mathbb{E}\left\|\mathcal{H}_{3,2}\right\|_2^2 &= \frac{1}{NT}\mathbb{E}\left\|\sum_i \mathbb{Z}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S},i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\mathbf{X}_{\mathbf{S},i}\right)^{-1}\frac{1}{T}\mathbf{X}_{\mathbf{S},i}'\mathbf{M}_{\widehat{\mathbf{G}}\mathbb{Z}\widehat{\mathbf{G}},i}\boldsymbol{\varepsilon}_i\right\|_2^2 \\
&\leq \left\|\boldsymbol{\Sigma}_{\varepsilon,T}\right\|_2\frac{1}{N}\sum_i\mathbb{E}\left[\frac{1}{T}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S},i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\frac{1}{T}\mathbf{X}_{\mathbf{S},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_i\right] \\
&\lesssim \frac{1}{N}\sum_i\mathbb{E}\left[\frac{1}{T}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S},i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\frac{1}{T}\mathbf{X}_{\mathbf{S},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_i\right]
\end{aligned}$$

By Lemma H1(i)(ii) and Lemma I5(iii), we have

$$\frac{1}{T}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S},i} = \frac{1}{T}\mathbf{U}_{\mathbf{X},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}_\mathbf{S},i} + \mathbf{o}_p\left(1\right).$$

And by the argument similar to (27), we have

$$\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1} - \left(\frac{1}{T}\mathbf{U}_{\mathbf{X}_\mathbf{S},i}'\mathbf{U}_{\mathbf{X}_\mathbf{S},i}\right)^{-1} = \mathbf{o}_p\left(1\right).$$

So we have

$$\begin{aligned}
&\frac{1}{N}\sum_i\mathbb{E}\left[\frac{1}{T}\mathbf{X}_i'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_{\mathbf{S},i}\left(\frac{1}{T}\widetilde{\mathbf{X}}_{\mathbf{S},i}'\widetilde{\mathbf{X}}_{\mathbf{S},i}\right)^{-1}\frac{1}{T}\mathbf{X}_{\mathbf{S},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_i\right] \\
&= \frac{1}{N}\sum_i\mathbb{E}\left[\frac{1}{T}\mathbf{U}_{\mathbf{X},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}_\mathbf{S},i}\left(\frac{1}{T}\mathbf{U}_{\mathbf{X}_\mathbf{S},i}'\mathbf{U}_{\mathbf{X}_\mathbf{S},i}\right)^{-1}\frac{1}{T}\mathbf{U}_{\mathbf{X}_\mathbf{S},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},i}\right] + \mathbf{o}_p\left(1\right) \\
&= \frac{1}{T}\frac{1}{N}\sum_i\mathbb{E}\left[\mathbf{U}_{\mathbf{X},i}'\mathbf{P}_{\mathbf{U}_{\mathbf{X}_\mathbf{S},i}}\mathbf{U}_{\mathbf{X},i}\right] + \mathbf{o}_p\left(1\right),
\end{aligned}$$

where the last inequality is by the non-max version of Lemma I2. For each i, conditional on $\mathbf{U}_{\mathbf{X}_\mathbf{S},i}$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathbf{U}_{\mathbf{X},i}'\mathbf{P}_{\mathbf{U}_{\mathbf{X}_\mathbf{S},i}}\mathbf{U}_{\mathbf{X},i} \mid \mathbf{U}_{\mathbf{X}_\mathbf{S},i}\right] &= \mathbb{E}\left[\operatorname{tr}\left(\mathbf{P}_{\mathbf{U}_{\mathbf{X}_\mathbf{S},i}}\mathbf{U}_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}'\right) \mid \mathbf{U}_{\mathbf{X}_\mathbf{S},i}\right] \\
&\leq \left\|\mathbf{P}_{\mathbf{U}_{\mathbf{X}_\mathbf{S},i}}\right\|_*\left\|\boldsymbol{\Sigma}_{\mathbf{U}_{\mathbf{X}_\mathbf{S},i},T}\right\|_2 = \mathcal{O}_p\left(1\right),
\end{aligned}$$

which implies

$$\frac{1}{T}\frac{1}{N}\sum_i \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\mathbf{U}_{\mathbf{X}_\mathbf{S},i}}\mathbf{U}_{\mathbf{X},i}\right] = \mathcal{O}_p\left(\frac{1}{T}\right) + \mathbf{o}_p(1) = \mathbf{o}_p(1).$$

So $\mathcal{H}_3 = \mathbf{o}_p(1)$.

For $\mathcal{H}_4$, note that $\mathbf{F}$ is contained $\mathbf{G}_\mathbf{Q}$ for any $\mathbf{Q} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K\}$, it is sufficient to show

$$\frac{1}{\sqrt{NT}}\sum_i \mathbf{X}'_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_\mathbf{X}\widetilde{\mathbf{\Lambda}}_i = \mathbf{o}_p(1),$$

where, abusing the notation, $\widetilde{\mathbf{\Lambda}}_i$ expands $\mathbf{\Lambda}_i$ to the same dimension as $\mathbf{G}_\mathbf{X}$ with zeros.

By (37), we have

$$\begin{aligned}\left(\widehat{\mathbf{G}}_\mathbf{X} - \mathbf{G}_\mathbf{X}\mathbf{H}_\mathbf{X}\right)\mathbf{H}_\mathbf{X}^{-1} &= \frac{1}{NT}\mathbf{G}_\mathbf{X}\mathbf{C}'_\mathbf{X}\mathbf{U}'_\mathbf{X}\widehat{\mathbf{G}}_\mathbf{X}\widehat{\mathbf{V}}_\mathbf{X}^{-2}\mathbf{H}_\mathbf{X}^{-1} \\ &\quad + \frac{1}{NT}\mathbf{U}_\mathbf{X}\mathbf{U}'_\mathbf{X}\widehat{\mathbf{G}}_\mathbf{X}\widehat{\mathbf{V}}_\mathbf{X}^{-2}\mathbf{H}_\mathbf{X}^{-1} \\ &\quad + \frac{1}{N}\mathbf{U}_\mathbf{X}\mathbf{C}_\mathbf{X}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_\mathbf{X}}^{-1} \\ &:= \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3,\end{aligned}$$

where $\widehat{\mathbf{\Sigma}}_{\mathbf{C}_\mathbf{X}} = \frac{1}{N}\mathbf{C}'_\mathbf{X}\mathbf{C}_\mathbf{X}$. Then we have

$$\begin{aligned}\frac{1}{\sqrt{NT}}\sum_i \mathbf{X}'_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_\mathbf{X}\widetilde{\mathbf{\Lambda}}_i &= -\frac{1}{\sqrt{NT}}\sum_i \mathbf{U}'_{\mathbf{X},i}\mathbf{M}_{\widehat{\mathbf{G}}}\left(\mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3\right)\widetilde{\mathbf{\Lambda}}_i \\ &\quad + \frac{1}{\sqrt{NT}}\sum_i \mathbf{C}'_{\mathbf{X},i}\left(\mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3\right)'\mathbf{M}_{\widehat{\mathbf{G}}}\left(\mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3\right)\widetilde{\mathbf{\Lambda}}_i \\ &:= \sum_{p=1}^3 \mathcal{H}_{4,1p} + \sum_{p=1}^3\sum_{q=1}^3 \mathcal{H}_{4,2pq}\end{aligned}$$

For $\mathcal{H}_{4,11}$, observe that

$$\begin{aligned}\mathcal{H}_{4,11} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{U}'_{\mathbf{X},i}\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{G}_\mathbf{X}\mathbf{C}'_\mathbf{X}\mathbf{U}'_\mathbf{X}\widehat{\mathbf{G}}_\mathbf{X}\widehat{\mathbf{V}}_\mathbf{X}^{-2}\mathbf{H}_\mathbf{X}^{-1}\widetilde{\mathbf{\Lambda}}_i \\ &= \frac{1}{NT\sqrt{NT}}\operatorname{tr}\left(\mathbf{U}'_\mathbf{X}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_\mathbf{X}\mathbf{C}'_\mathbf{X}\mathbf{U}'_\mathbf{X}\widehat{\mathbf{G}}_\mathbf{X}\widehat{\mathbf{V}}_\mathbf{X}^{-2}\mathbf{H}_\mathbf{X}^{-1}\widetilde{\mathbf{\Lambda}}'\right)\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\left(\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\frac{1}{\sqrt{NT}}\widetilde{\mathbf{\Lambda}}'\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\frac{1}{NT}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right) \\
&\lesssim \sqrt{T}\left\|\frac{1}{\sqrt{NT}}\widetilde{\mathbf{\Lambda}}'\mathbf{U}_{\mathbf{X}}'\right\|_2\left\|\frac{1}{\sqrt{T}}\left(\widehat{\mathbf{G}}_{\mathbf{X}}-\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\right\|_2\left\|\frac{1}{NT}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2 \\
&= \mathcal{O}_p\left(\sqrt{\frac{T}{N}}+\frac{1}{\sqrt{T}}\right)\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)=\mathbf{o}_p\left(1\right),
\end{aligned}
$$

by Prop. H1, Lemma H1 and Lemma H2.

Similarly, for $\mathcal{H}_{4,12}$, we have

$$
\begin{aligned}
\mathcal{H}_{4,12} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{U}_{\mathbf{X},i}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}_i \\
&= \frac{1}{NT\sqrt{NT}}\operatorname{tr}\left(\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}'\right) \\
&\lesssim \sqrt{T}\left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\right\|_2\left\|\frac{1}{\sqrt{T}}\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2\left\|\frac{1}{\sqrt{NT}}\widetilde{\mathbf{\Lambda}}'\mathbf{U}_{\mathbf{X}}'\right\|_2 \\
&= \mathcal{O}_p\left(\frac{\sqrt{T}}{N}+\frac{1}{\sqrt{T}}\right)=\mathbf{o}_p\left(1\right).
\end{aligned}
$$

For $\mathcal{H}_{4,211}$,

$$
\begin{aligned}
\mathcal{H}_{4,211} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{C}_{\mathbf{X},i}'\frac{1}{NT}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}_i \\
&= \frac{1}{N^2T^2\sqrt{NT}}\operatorname{tr}\left(\mathbf{C}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}'\right) \\
&= \sqrt{NT}\operatorname{tr}\left(\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\frac{1}{NT}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\frac{1}{T}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\frac{1}{NT}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\frac{1}{N}\widetilde{\mathbf{\Lambda}}'\mathbf{C}_{\mathbf{X}}\right) \\
&= \sqrt{NT}\mathcal{O}_p\left(\frac{1}{N^2}+\frac{1}{NT}\right)\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T^2}\right)=\mathbf{o}_p\left(1\right).
\end{aligned}
$$

For $\mathcal{H}_{4,212}$,

$$
\begin{aligned}
\mathcal{H}_{4,212} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{C}_{\mathbf{X},i}'\frac{1}{NT}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}_i \\
&= \frac{1}{N^2T^2\sqrt{NT}}\operatorname{tr}\left(\mathbf{C}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}'\right)
\end{aligned}
$$

$$= \sqrt{NT}\,\mathrm{tr}\left(\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\frac{1}{NT}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\frac{1}{\sqrt{T}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\frac{1}{\sqrt{T}}\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\frac{1}{N}\widetilde{\mathbf{\Lambda}}'\mathbf{C}_{\mathbf{X}}\right)$$
$$= \sqrt{NT}\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)\mathcal{O}_p\left(\frac{1}{\sqrt{N}}+\frac{1}{T}\right)\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right) = \mathbf{o}_p(1).$$

For $\mathcal{H}_{4,213}$,

$$\begin{aligned}\mathcal{H}_{4,213} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{C}_{\mathbf{X},i}'\frac{1}{NT}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{N}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_{\mathbf{X}}}^{-1}\widetilde{\mathbf{\Lambda}}_i \\ &= \frac{1}{N^2T\sqrt{NT}}\mathrm{tr}\left(\mathbf{C}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_{\mathbf{X}}}^{-1}\widetilde{\mathbf{\Lambda}}'\right) \\ &= \sqrt{T}\,\mathrm{tr}\left(\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\frac{1}{NT}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\frac{1}{\sqrt{T}}\mathbf{G}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{\sqrt{NT}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_{\mathbf{X}}}^{-1}\frac{1}{N}\widetilde{\mathbf{\Lambda}}'\mathbf{C}_{\mathbf{X}}\right) \\ &= \sqrt{T}\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)\mathcal{O}_p\left(\frac{1}{\sqrt{N}}+\frac{1}{T}\right) = \mathbf{o}_p(1).\end{aligned}$$

$\mathcal{H}_{4,221}$ is analogous to $\mathcal{H}_{4,212}$ so the details are omitted. For $\mathcal{H}_{4,222}$,

$$\begin{aligned}\mathcal{H}_{4,222} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{C}_{\mathbf{X},i}'\frac{1}{NT}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}_i \\ &= \frac{1}{N^2T^2\sqrt{NT}}\mathrm{tr}\left(\mathbf{C}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\widetilde{\mathbf{\Lambda}}'\right) \\ &= \sqrt{NT}\,\mathrm{tr}\left(\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\frac{1}{\sqrt{T}}\widehat{\mathbf{G}}_{\mathbf{X}}'\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\frac{1}{\sqrt{T}}\widehat{\mathbf{G}}_{\mathbf{X}}\widehat{\mathbf{V}}_{\mathbf{X}}^{-2}\mathbf{H}_{\mathbf{X}}^{-1}\frac{1}{N}\widetilde{\mathbf{\Lambda}}'\mathbf{C}_{\mathbf{X}}\right) \\ &= \sqrt{NT}\mathcal{O}_p\left(\frac{1}{N^2}+\frac{1}{T^2}\right) = \mathbf{o}_p(1).\end{aligned}$$

For $\mathcal{H}_{4,223}$,

$$\begin{aligned}\mathcal{H}_{4,223} &= \frac{1}{\sqrt{NT}}\sum_i \mathbf{C}_{\mathbf{X},i}'\frac{1}{NT}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{N}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_{\mathbf{X}}}^{-1}\widetilde{\mathbf{\Lambda}}_i \\ &= \frac{1}{N^2T\sqrt{NT}}\mathrm{tr}\left(\mathbf{C}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_{\mathbf{X}}}^{-1}\widetilde{\mathbf{\Lambda}}'\right) \\ &= \sqrt{T}\,\mathrm{tr}\left(\mathbf{H}_{\mathbf{X}}^{-1\prime}\widehat{\mathbf{V}}_{\mathbf{X}}\frac{1}{\sqrt{T}}\widehat{\mathbf{G}}_{\mathbf{X}}'\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\mathbf{M}_{\widehat{\mathbf{G}}}\frac{1}{\sqrt{NT}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\widehat{\mathbf{\Sigma}}_{\mathbf{C}_{\mathbf{X}}}^{-1}\frac{1}{N}\widetilde{\mathbf{\Lambda}}'\mathbf{C}_{\mathbf{X}}\right) \\ &= \sqrt{T}\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right) = \mathbf{o}_p(1).\end{aligned}$$

$\mathcal{H}_{4,231}$ and $\mathcal{H}_{4,232}$ are analogous to $\mathcal{H}_{4,213}$ and $\mathcal{H}_{4,223}$ respectively, so the details are omitted.

For $\mathcal{H}_{4,13}$ and $\mathcal{H}_{4,233}$, observe that

$$\begin{aligned}
\mathcal{H}_{4,233} - \mathcal{H}_{4,13} &= \frac{1}{\sqrt{NT}} \sum_i \left(\frac{1}{N} \mathbf{C}'_{\mathbf{X},i} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \mathbf{C}'_{\mathbf{X}} \mathbf{U}'_{\mathbf{X}} \mathbf{M}_{\widehat{\mathbf{G}}} \frac{1}{N} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}_i - \mathbf{U}'_{\mathbf{X},i} \mathbf{M}_{\widehat{\mathbf{G}}} \frac{1}{N} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}_i \right) \\
&= -\frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{M_{C_X}} \mathbf{U}'_{\mathbf{X}} \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \right) \\
&= -\frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{U}'_{\mathbf{X}} \mathbf{M}_{\widehat{\mathbf{G}}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right) \\
&= -\frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{U}'_{\mathbf{X}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right) + \frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{U}'_{\mathbf{X}} \mathbf{P}_{\widehat{\mathbf{G}}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right)
\end{aligned} \tag{33}$$

For the first term, conditional on $\mathbf{C_X}$, by Assumption 3, we have

$$\begin{aligned}
\mathbb{E} \left[\frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{U}'_{\mathbf{X}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right) \mid \mathbf{C_X} \right] &= \frac{\sqrt{T}}{N\sqrt{N}} \operatorname{tr} \left(\mathbb{E} \left[\sigma_{\mathbf{U_X}} \mathbf{I}_N \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right] \right) \\
&= \frac{\sqrt{T}}{N\sqrt{N}} \sigma_{\mathbf{U_X}} \operatorname{tr} \left(\mathbf{M_{C_X}} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \right) = 0.
\end{aligned}$$

Let $\mathbf{A} = \frac{1}{N} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}}$. Note that

$$\|A\|_F = \left\| \frac{1}{N} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right\|_F = \mathcal{O}_p(1).$$

Conditonal on $\mathbf{C_X}$, we have the first term equal to

$$\begin{aligned}
\frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{U}'_{\mathbf{X}} \mathbf{U_X} \mathbf{A} \right) &= \sqrt{\frac{T}{N}} \frac{1}{T} \sum_t \mathbf{U}'_{\mathbf{X},t} \mathbf{A} \mathbf{U}_{\mathbf{X},t} - \mathbb{E} \left[\mathbf{U}'_{\mathbf{X},t} \mathbf{A} \mathbf{U}_{\mathbf{X},t} \mid \mathbf{C_X} \right] \\
&= \mathcal{O}_p \left(\frac{1}{\sqrt{N}} \right)
\end{aligned} \tag{34}$$

by the Bernstein-type inequality by Merlevède et al. (2011).

For the second term, we have

$$\begin{aligned}
\frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{U}'_{\mathbf{X}} \mathbf{P}_{\widehat{\mathbf{G}}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \right) &= \frac{1}{N\sqrt{NT}} \operatorname{tr} \left(\mathbf{P}_{\widehat{\mathbf{G}}} \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \mathbf{U}'_{\mathbf{X}} \right) \\
&\leq \left\| \mathbf{P}_{\widehat{\mathbf{G}}} \right\|_* \frac{1}{N\sqrt{NT}} \left\| \mathbf{U_X} \mathbf{C_X} \widehat{\mathbf{\Sigma}}_{\mathbf{C_X}}^{-1} \widetilde{\mathbf{\Lambda}}' \mathbf{M_{C_X}} \mathbf{U}'_{\mathbf{X}} \right\|_2.
\end{aligned}$$

$\left\|\mathbf{P}_{\widehat{\mathbf{G}}}\right\|_* = \mathcal{O}(1)$ and the second component is bounded by the first term, so we have $\mathcal{H}_{4,233} - \mathcal{H}_{4,13} = \mathbf{o}_p(1)$.

Putting them together, we have $\mathcal{H}_4 = \mathbf{o}_p(1)$. The results follow. □

The proof of Corollary 4.1 follows by repeated arguments in the proof of Theorem 4.1 and 4.2 so the details are omitted.

Appendix H: Factor-model-related results

In this section, I present the factor-model-related propositions and lemmas that are used to prove the main theorems.

Rewrite the factor structure (2) and the factor structure for the sum of covariates in $T \times N$ matrix form:

$$\mathbf{Q} = \mathbf{G}_\mathbf{Q}\mathbf{C}_\mathbf{Q}' + \mathbf{U}_\mathbf{Q}, \quad \text{for } \mathbf{Q} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{Z}\},$$

where $\mathbf{Z} = \mathbf{X} + \sum_{k=1}^K \mathbf{W}_k$, $\mathbf{G}_\mathbf{Q} = [\mathbf{F}, \mathbf{F}_\mathbf{Q}]$ and $\mathbf{C}_\mathbf{Q} = [\boldsymbol{\eta}_\mathbf{Q}, \mathbf{B}_\mathbf{Q}]$.

Proposition H1. *Let $\widehat{\mathbf{G}}_\mathbf{Q}$ be the estimated factors for each $\mathbf{Q} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{Z}\}$ from Algorithm 1, and $\widehat{\mathbf{G}}$ be the estimated factors from Algorithm 4. Under Assumption 1 to 5, there exists matrix $\mathbf{H}$ and $\mathbf{H}_\mathbf{Q}$ for $\mathbf{Q} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{Z}\}$ such that*

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_\mathbf{Q} - \mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_2 = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right), \quad \textit{for } \mathbf{Q} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{Z}\}, \tag{35}$$

$$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}} - \mathbf{G}\mathbf{H}\right\|_2 = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right). \tag{36}$$

Proof of Proposition H1. To prove (35), it is sufficient to show that $\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_\mathbf{Q} - \mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_2$ has the desired rate, since the same argument applies to $\widehat{\mathbf{G}}_{\mathbf{W}_k}$ for $k = 1, \ldots, K$. In addition, the proof of (36) is similar to that of$\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_\mathbf{Q} - \mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_2$, so I only present the proof of (35) here.

By the definition of $\widehat{\mathbf{G}}_\mathbf{Q}$, we have

$$\frac{1}{NT}\mathbf{Q}\mathbf{Q}' = \widehat{\mathbf{G}}_\mathbf{Q}\widehat{\mathbf{V}}_\mathbf{Q}^2,$$

where $\widehat{\mathbf{V}}_\mathbf{Q}^2$ is the diagonal matrix of the $r_\mathbf{Q}$ largest eigenvalues of $\mathbf{Q}\mathbf{Q}'/(NT)$. Expanding LHS and

rearranging terms yield:

$$\widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}} = \frac{1}{NT}\mathbf{G}_{\mathbf{Q}}\mathbf{C}'_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2} + \frac{1}{NT}\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2} + \frac{1}{NT}\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}, \quad (37)$$

which implies

$$\begin{aligned}\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2 =& \frac{1}{NT\sqrt{T}}\left\|\mathbf{G}_{\mathbf{Q}}\mathbf{C}'_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 + \frac{1}{NT\sqrt{T}}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &+ \frac{1}{NT\sqrt{T}}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2,\end{aligned}$$

By Lemma H1,

$$\begin{aligned}\frac{1}{NT\sqrt{T}}\left\|\mathbf{G}_{\mathbf{Q}}\mathbf{C}'_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 &\leq \frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{Q}}\right\|_2 \frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2 \frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_2 \left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \cdot \frac{1}{\sqrt{N}} \\ &= \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right)\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{1}{NT\sqrt{T}}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 &\leq \frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2 \frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{Q}}\right\|_2 \frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_2 \left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \cdot \frac{1}{\sqrt{N}} \\ &= \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right)\end{aligned}$$

By Lemma H1,

$$\begin{aligned}\frac{1}{NT\sqrt{T}}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 &\leq \frac{1}{NT}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\right\|_2 \frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_2 \left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &= \mathcal{O}_p\left(\frac{1}{N} + \frac{1}{T}\right).\end{aligned}$$

Putting pieces together, we have $\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2 = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{T}\right)$. □

The following lemma provides some bounds for the terms that are used in proofs.

Lemma H1. *Under Assumption 1 to 5, for* $\mathbf{Q}, \mathbf{R} \in \{\mathbf{X}, \mathbf{W}_1, \ldots, \mathbf{W}_K, \mathbf{Z}\}$,

(i) $\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 = \mathcal{O}_p(1)$.

(ii) $\left\|\mathbf{H}_{\mathbf{Q}}^{-1}\right\|_2 = \mathcal{O}_p(1)$.

(iii) $\frac{1}{\sqrt{NT}} \|\mathbf{U_Q C_R}\|_2 = \mathcal{O}_p(1)$.

(iv) $\frac{1}{\sqrt{NT}} \left\|\mathbf{G}'_{\mathbf{Q}}\mathbf{U_R}\right\|_2 = \mathcal{O}_p(1)$.

(v) $\frac{1}{NT} \left\|\mathbf{U_Q U}'_{\mathbf{Q}}\right\|_2 = \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$.

(vi) $\frac{1}{\sqrt{NT}} \left\|\mathbf{G}'_Q\mathbf{U}_Q\mathbf{C}_Q\right\|_2 = \mathcal{O}_p(1)$.

(vii) $\frac{1}{T} \left\|\mathbf{U}'_{\mathbf{Q},-i}\boldsymbol{\varepsilon}_i\right\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}\right)$ *for* $i = 1,\ldots,N$.

(viii) $\frac{1}{T} \|\mathbf{U}'_{\mathbf{R}}\mathbf{G_Q}\|_\infty = \mathcal{O}_p\left(\sqrt{\frac{\log N}{T}}\right)$ *for* $i = 1,\ldots,N$.

(ix) $\frac{1}{NT^2} \left\|\mathbf{U}'_{\mathbf{Q},i}\mathbf{U_R U}'_{\mathbf{R}}\mathbf{G_Q}\right\|_2 = \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{\sqrt{NT}}\right)$.

(x) $\max_{i\le N} \frac{1}{NT^2} \left\|\mathbf{U}'_{\mathbf{Q},i}\mathbf{U_R U}'_{\mathbf{R}}\mathbf{G_Q}\right\|_2 = \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{\sqrt{NT}}\right)$.

(xi) $\frac{1}{NT} \left\|\mathbf{U}'_{\mathbf{Q},i}\mathbf{U_Q C_R}\right\|_2 = \mathcal{O}_p\left(\sqrt{\frac{1}{NT}}+\frac{1}{N}\right)$.

(xii) $\max_{i\le N} \frac{1}{NT} \left\|\mathbf{U}'_{\mathbf{Q},i}\mathbf{U_Q C_R}\right\|_2 = \mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}+\frac{1}{N}\right)$.

(xiii) $\max_{i\le N} \frac{1}{T} \left|\mathbf{U}'_{\mathbf{Q},i}\mathbf{P}_{\mathbf{G_R}}, \boldsymbol{\varepsilon}_j\right| = \mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}\right)$.

(xiv) $\frac{1}{NT\sqrt{N}} \left\|\mathbf{U}'_{\mathbf{Q}}\mathbf{U_Q C_R}\right\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{NT}}+\frac{1}{T}\right)$.

(xv) $\frac{1}{NT}\frac{1}{T\sqrt{N}} \left\|\mathbf{U}'_{\mathbf{Q}}\mathbf{U_Q U}'_{\mathbf{Q}}\mathbf{G_Q}\right\|_2 = \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)$.

Proof of Lemma H1. The proof of (i) to (ii) can be found in Bai (2003). For (iii), it is sufficient to show $\frac{1}{\sqrt{NT}}\mathbf{U_Q C}_{\mathbf{R},k}$ for $k = 1,\ldots,r_{\mathbf{R}}$ has the desired rate, where $\mathbf{C}_{\mathbf{R},k}$ is the k-th column of $\mathbf{C_R}$. Observe that

$$\frac{1}{NT}\|\mathbf{U_Q C}_{\mathbf{R},i}\|_2^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{C}_{\mathbf{R},ik}\mathbf{U}_{\mathbf{Q},it}\right)^2 = \mathcal{O}_p(1).$$

The proof of (iv) is similar to that of (iii). (v) is by Theorem 4.7.1 in Vershynin (2024). For (vi), it is sufficient to show $\frac{1}{\sqrt{NT}}\left\|\mathbf{G}'_{\mathbf{Q},l}\mathbf{U_Q C}_{\mathbf{Q},k}\right\|_2$ for any k,l has the desired rate. Note that

$$\frac{1}{NT}\mathbb{E}\left[\left(\mathbf{G}'_{\mathbf{Q},l}\mathbf{U_Q C}_{\mathbf{Q},k}\right)^2\right] = \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{C}_{\mathbf{Q},ik}\mathbf{C}_{\mathbf{Q},jk}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{U}_{\mathbf{Q},is}\mathbf{U}_{\mathbf{Q},js}\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{C}_{\mathbf{Q},ik}^{2}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right]\right] \\
&\leq \mathbb{E}\left[\frac{1}{T}\max_{i\leq N}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right]\frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{\mathbf{Q},ik}^{2}\right] \\
&\leq \mathbb{E}\left[\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right]\right)^{2}\right]^{1/2}\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{\mathbf{Q},ik}^{2}\right)^{2}\right]^{1/2}
\end{aligned}
$$

Since $\mathbf{C}_{\mathbf{Q},ik}$ is sub-Gaussian and $\frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{\mathbf{Q},ik}^{2}=\mathcal{O}_{p}\left(1\right)$ by Assumption 2, $\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{C}_{\mathbf{Q},ik}^{2}\right)^{2}\right]=\mathcal{O}\left(1\right)$.

By Assumption 3, $\mathbf{U}_{\mathbf{Q},it}$ is stationary and α-mixing. Let $\tau\left(h\right)=\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},i(t+h)}\right]$. We have

$$
\begin{aligned}
\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right] &= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\tau(t-s) \\
&\leq \frac{1}{2T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}^{2}\tau(t-s)+\frac{1}{2T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},lt}^{2}\tau(t-s) \\
&= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},lt}^{2}\tau(t-s) \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{h=t-T}^{t-1}\mathbf{G}_{\mathbf{Q},lt}^{2}\tau(h) \\
&\leq \frac{1}{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}^{2}\sum_{h=-\infty}^{\infty}\left|\tau(h)\right| \\
&\lesssim \frac{1}{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}^{2}=\mathcal{O}_{p}\left(1\right).
\end{aligned}
$$

Since $\mathbf{G}_{\mathbf{Q},t}$ is sub-Gaussian, stationary and α-mixing by Assumption 1,

$$
\mathbb{E}\left[\left(\frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}\mathbf{G}_{\mathbf{Q},lt}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right]\right)^{2}\right]\lesssim\mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{G}_{\mathbf{Q},ls}^{2}\right)^{2}\right]=\mathcal{O}\left(1\right).
$$

The result follows by Chebyshev's inequality.

(vii) and (viii) are standard results by Lemma K1. For (ix), observe that, for each $k = 1, \ldots, r_{\mathbf{Q}}$,

$$\begin{aligned}\frac{1}{NT^2}\left|\mathbf{U}_{\mathbf{Q},i}\mathbf{U}_{\mathbf{R}}\mathbf{U}_{\mathbf{R}}'\mathbf{G}_{\mathbf{Q},k}\right| &= \left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\mathbf{U}_{\mathbf{R},t}'\mathbf{U}_{\mathbf{R},s}\mathbf{G}_{\mathbf{Q},sk}\right| \\ &\leq \left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\left(\mathbf{U}_{\mathbf{R},t}'\mathbf{U}_{\mathbf{R},s} - \sigma_{ts}\right)\mathbf{G}_{\mathbf{Q},sk}\right| + \left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\frac{1}{T}\sum_{s=1}^{T}\sigma_{ts}\mathbf{G}_{\mathbf{Q},sk}\right|,\end{aligned}$$

where $\sigma_{ts} = \mathbb{E}\left[\frac{1}{N}\mathbf{U}_{\mathbf{R},t}'\mathbf{U}_{\mathbf{R},s}\right]$. For the second term,

$$\left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\frac{1}{T}\sum_{s=1}^{T}\sigma_{ts}\mathbf{G}_{\mathbf{Q},sk}\right| \leq \frac{1}{T}\left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\right|^{1/2}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{T}\sigma_{ts}\mathbf{G}_{\mathbf{Q},sk}\right)^2\right|^{1/2}$$

The first part $\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2 = \mathcal{O}_p(1)$. For the second component,

$$\sum_{s=1}^{T}\sigma_{ts}\mathbf{G}_{\mathbf{Q},sk} = \sum_{s=1}^{T}\sigma_{ts}(\mathbf{G}_{\mathbf{Q},sk} - \mathbb{E}\left[\mathbf{G}_{\mathbf{Q},sk}\right]) + \sum_{s=1}^{T}\sigma_{ts}\mathbb{E}\left[\mathbf{G}_{\mathbf{Q},sk}\right].$$

By Assumption 3, $\sum_{s=1}^{T}\sigma_{ts}\mathbb{E}\left[\mathbf{G}_{\mathbf{Q},sk}\right] \lesssim \sum_{s=1}^{T}|\sigma_{ts}| = \mathcal{O}(1)$. And

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{s=1}^{T}\sigma_{ts}(\mathbf{G}_{\mathbf{Q},sk} - \mathbb{E}\left[\mathbf{G}_{\mathbf{Q},sk}\right])\right)^2\right] &= \sum_{s=1}^{T}\sum_{l=1}^{T}\sigma_{ts}\sigma_{tl}\operatorname{cov}\left(\mathbf{G}_{\mathbf{Q},sk}, \mathbf{G}_{\mathbf{Q},lk}\right) \\ &\leq \sum_{s=1}^{T}\sigma_{ts}\left(\sum_{l=1}^{T}\sigma_{tl}^2\right)^{1/2}\left(\sum_{l=1}^{T}\operatorname{cov}\left(\mathbf{G}_{\mathbf{Q},sk}, \mathbf{G}_{\mathbf{Q},lk}\right)^2\right)^{1/2} \\ &\lesssim \sum_{s=1}^{T}|\sigma_{ts}| = \mathcal{O}(1),\end{aligned}$$

by Assumption 3 and 1. So $\left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\frac{1}{T}\sum_{s=1}^{T}\sigma_{ts}\mathbf{G}_{\mathbf{Q},sk}\right| = \mathcal{O}_p\left(\frac{1}{T}\right)$ by Chebyshev's inequality. For the first term,

$$\left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\frac{1}{T}\sum_{s=1}^{T}\frac{1}{N}\left(\mathbf{U}_{\mathbf{R},t}'\mathbf{U}_{\mathbf{R},s} - \sigma_{ts}\right)\mathbf{G}_{\mathbf{Q},sk}\right|$$

$$\leq \left(\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\left(\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}-\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right]\right)\right)^2\right)^{1/2}\cdot\frac{1}{\sqrt{NT}}$$

Abusing the notation, let $\tau(t,s)=\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}\right]$. Note that

$$\mathbb{E}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\left(\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}-\tau(t,s)\right)\right]^2=\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\frac{1}{T}\left(\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\left(\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}-\tau(t,s)\right)\right)^2\right]$$

by Assumption 3. Therefore, it is sufficient to show $\mathbb{E}\left[\frac{1}{T}\left(\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\left(\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}-\tau(t,s)\right)\right)^2\right]=\mathcal{O}\left(1\right)$. Observe that

$$\frac{1}{T}\left(\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\left(\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},is}-\tau(t,s)\right)\right)^2=\left(\mathbf{U}_{\mathbf{Q},it}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\mathbf{U}_{\mathbf{Q},is}-\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\tau(t,s)\right)^2$$

It is easy to verify that $\mathbb{E}\left[\left(\mathbf{U}_{\mathbf{Q},it}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\mathbf{U}_{\mathbf{Q},is}\right)^2\right]=\mathcal{O}\left(1\right)$ and $\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\mathbf{G}_{\mathbf{Q},sk}\tau(t,s)\right)^2\right]=\mathcal{O}\left(1\right)$. The result follows by Chebyshev's inequality.

The proof of (x) is similar to that of (ix). The rate is the same since $\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2=\mathcal{O}_p\left(1+\sqrt{\frac{\log(N)}{T}}\right)=\mathcal{O}_p\left(1\right)=\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2$.

For (xi), for each $k=1,\ldots,r_{\mathbf{R}}$,

$$\begin{aligned}\frac{1}{NT}\mathbf{U}_{\mathbf{Q},i}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{R},k}&=\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}'\mathbf{C}_{\mathbf{R},jk}\\&=\frac{1}{NT}\sum_{t=1}^{T}\sum_{j\neq i}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}'\mathbf{C}_{\mathbf{R},jk}+\frac{1}{NT}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\mathbf{C}_{\mathbf{R},ik}\end{aligned}$$

It is easy to verify that the second term is $\mathcal{O}_p\left(\frac{1}{N}\right)$. For the first term,

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{j\neq i}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}'\mathbf{C}_{\mathbf{R},jk}=\frac{1}{\sqrt{NT}}\cdot\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\cdot a_{it},$$

where $a_{it}=\frac{1}{\sqrt{N}}\sum_{j\neq i}^{N}\mathbf{U}_{\mathbf{Q},jt}'\mathbf{C}_{\mathbf{R},jk}$ is a $\mathcal{O}_p\left(1\right)$ mean-zero random variable independent of $\mathbf{U}_{\mathbf{Q},it}$ by Assumption 3. Therefore, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\cdot a_{it}=\mathcal{O}_p\left(1\right)$ and the result follows.

For (xii), following the same decomposition as in (xi), the second term of RHS is then

$$\begin{aligned}
&\frac{1}{N}\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\mathbf{C}_{\mathbf{R},ik}\\
&=\frac{1}{N}\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{U}_{\mathbf{Q},it}^2\mathbf{C}_{\mathbf{R},ik}-\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\mathbf{C}_{\mathbf{R},ik}\right]\right)+\frac{1}{N}\max_{i\leq N}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\mathbf{C}_{\mathbf{R},ik}\right]\\
&=\mathcal{O}_p\left(\frac{1}{N}\sqrt{\frac{\log N}{T}}\right)+\mathcal{O}_p\left(\frac{1}{N}\right)=\mathcal{O}_p\left(\frac{1}{N}\right),
\end{aligned}$$

by Lemma K1 and the fact that $\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}^2\mathbf{C}_{\mathbf{R},ik}\right]=\mathcal{O}_p(1)$ by Assumption 3 and 2. For the first term of RHS, following the same argument as in (xi), we have

$$\frac{1}{\sqrt{N}}\max_{i\leq N}\frac{1}{T}\sum_{t=1}^{T}\mathbf{U}_{\mathbf{Q},it}\cdot a_{it}=\mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}\right),$$

by Assumption 2, 3 and Lemma K1. The result follows.

For (xiii), observe that

$$\begin{aligned}
\max_{i\leq N}\frac{1}{T}\left|\mathbf{U}_{\mathbf{Q},i}'\mathbf{P}_{\mathbf{G}_R},\boldsymbol{\varepsilon}_i\right|&\leq\max_{i\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},i}'\mathbf{G}_{\mathbf{R}}\right\|_2\left\|\left(\frac{1}{T}\mathbf{G}_{\mathbf{R}}'\mathbf{G}_{\mathbf{R}}\right)^{-1}\right\|_2\frac{1}{T}\left\|\mathbf{G}_{\mathbf{R}}'\boldsymbol{\varepsilon}_j\right\|_2\\
&=\mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}\right),
\end{aligned}$$

by Lemma K1.

For (xiv), WLOG assume $r_{\mathbf{R}}=1$,

$$\frac{1}{N^3T^2}\left\|\mathbf{U}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{R}}\right\|_2^2=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R},j}\right)^2,$$

where

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R},j}\leq\frac{1}{\sqrt{NT}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R}.j}-\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R}.j}\right]\right)$$

$$+\frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R}.j}\right].$$

Let $a_{it}=\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R}.j}-\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R}.j}\right]$, observe that a_{it} is a mean-zero α-mixing random variable with bounded variance under assumptions, so $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}a_{it}=\mathcal{O}_p\left(1\right)$, which gives

$$\frac{1}{\sqrt{NT}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}a_{it}=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right).$$

And

$$\frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}\mathbf{U}_{\mathbf{Q},jt}\mathbf{C}_{\mathbf{R}.j}\right]=\frac{1}{N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\mathbf{U}_{\mathbf{Q},it}^{2}\right]\mathbb{E}\left[\mathbf{C}_{\mathbf{R}.j}\right]=\mathcal{O}_p\left(\frac{1}{N}\right).$$

Hence,

$$\frac{1}{NT\sqrt{N}}\left\|\mathbf{U}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{R}}\right\|_2=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}+\frac{1}{N}\right).$$

The proof of (xv) is similar to (ix) and (x), so the details are omitted. □

The following lemma provides bounds for terms that involve the estimated factors.

Lemma H2. *Under Assumption 1 to 5, for* $\mathbf{Q},\mathbf{R}\in\{\mathbf{X},\mathbf{W}_1,\ldots,\mathbf{W}_K,\mathbf{Z}\}$,

(i) $\frac{1}{T}\left\|\mathbf{G}_R'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$.

(ii) $\frac{1}{T}\left\|\mathbf{U}_{\mathbf{R},i}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$.

(iii) $\max_{i\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{R},i}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}+\sqrt{\frac{\log N}{NT}}\right)$.

(iv) $\left\|\mathbf{I}_{r_{\mathbf{Q}}}-\left(\frac{1}{T}\mathbf{H}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)^{-1}\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$.

(v) $\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},-i}\mathbf{P}_{\widehat{\mathbf{G}}_{\mathbf{R}}}\boldsymbol{\varepsilon}_i\right\|_\infty=\mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}\right)$.

(vi) $\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{R}}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$.

(vii) $\frac{1}{NT}\left\|\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)'\mathbf{U}_{\mathbf{R}}\mathbf{C}_{\mathbf{R}}\right\|_2=\mathcal{O}_p\left(\frac{1}{N}\right)$.

Proof of Lemma H2. For (i), by (37),

$$\begin{aligned}\frac{1}{T}\mathbf{G}_{\mathbf{R}}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)=&\frac{1}{NT^2}\mathbf{G}_{\mathbf{R}}'\mathbf{G}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}+\mathbf{G}_{\mathbf{R}}'\frac{1}{NT^2}\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\\&+\frac{1}{NT^2}\mathbf{G}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}.\end{aligned}$$

Bounding the first term:

$$\begin{aligned}\frac{1}{NT^2}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{G}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2&\leq\frac{1}{T}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\frac{1}{NT}\left\|\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&\leq\frac{1}{\sqrt{NT}}\frac{1}{T}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&\quad+\frac{1}{\sqrt{N}}\frac{1}{T}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right)+\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right)=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

by Lemma H1. For the second term:

$$\begin{aligned}\frac{1}{NT^2}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2&\leq\frac{1}{\sqrt{NT}}\frac{1}{\sqrt{NT}}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2\frac{1}{T}\left\|\mathbf{G}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&\quad+\frac{1}{\sqrt{NT}}\frac{1}{\sqrt{NT}}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{Q}}'\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

by Lemma H1. For the third term,

$$\begin{aligned}\frac{1}{NT^2}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2&\leq\frac{1}{T}\frac{1}{NT}\left\|\mathbf{G}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&\quad+\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{R}}\right\|_2\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2\\&=\mathcal{O}_p\left(\frac{1}{T}\right)+\mathcal{O}_p\left(\frac{1}{N\sqrt{N}}+\frac{1}{NT}+\frac{1}{T\sqrt{N}}+\frac{1}{T^2}\right)=\mathcal{O}_p\left(\frac{1}{T}+\frac{1}{N\sqrt{N}}\right).\end{aligned}$$

by Lemma H1. Putting all pieces together yields $\frac{1}{T}\mathbf{G}_{\mathbf{R}}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$.

For (ii), again by (37),

$$\begin{aligned}\frac{1}{T}\mathbf{U}_{\mathbf{R},i}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)=&\frac{1}{NT^2}\mathbf{U}_{\mathbf{R},i}'\mathbf{G}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}+\frac{1}{NT^2}\mathbf{U}_{\mathbf{R},i}'\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\end{aligned}$$

$$+ \frac{1}{NT^2}\mathbf{U}'_{\mathbf{R},i}\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}.$$

Bounding the first term:

$$\begin{aligned}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{G}_{\mathbf{Q}}\mathbf{C}'_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 &\leq \frac{1}{T\sqrt{N}}\frac{1}{\sqrt{T}}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{G}_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}'_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &\quad + \frac{1}{\sqrt{NT}}\frac{1}{\sqrt{T}}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{G}_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}'_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &= \mathcal{O}_p\left(\frac{1}{T\sqrt{N}}\right)+\mathcal{O}_p\left(\frac{1}{N\sqrt{T}}+\frac{1}{T\sqrt{NT}}\right)=\mathcal{O}_p\left(\frac{1}{N\sqrt{T}}+\frac{1}{T\sqrt{N}}\right),\end{aligned}$$

by Lemma H1. For the second term:

$$\begin{aligned}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\mathbf{G}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 &\leq \frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2\frac{1}{T}\left\|\mathbf{G}'_{\mathbf{Q}}\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &\quad + \frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2\frac{1}{T}\left\|\mathbf{G}'_{\mathbf{Q}}(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}})\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &= \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

by Lemma H1. For the third term,

$$\begin{aligned}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\widehat{\mathbf{G}}_{\mathbf{Q}}\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 &\leq \frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{R},i}\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &\quad + \frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{R},i}\right\|_2\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{U}'_{\mathbf{Q}}\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2\left\|\widehat{\mathbf{V}}_{\mathbf{Q}}^{-2}\right\|_2 \\ &= \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{\sqrt{NT}}\right)+\mathcal{O}_p\left(\frac{1}{T^2}+\frac{1}{T\sqrt{N}}+\frac{1}{N\sqrt{N}}+\frac{1}{NT}\right) \\ &= \mathcal{O}_p\left(\frac{1}{T}+\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

by Lemma H1. Putting all pieces together yields $\frac{1}{T}\mathbf{U}'_{\mathbf{R},i}\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$. For (iii), the proof is similar to that of (ii) with the first component of each term replaced by its maximum over $i \leq N$. The second term gives $\mathcal{O}_p\left(\frac{1}{N}+\sqrt{\frac{\log N}{NT}}\right)$, while the rate of the third term remains the same by Lemma H1. The result follows.

For (iv), observe that

$$\begin{aligned}\left\|\mathbf{I}_{r_\mathbf{Q}}-\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)^{-1}\right\|_2 &= \left\|\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}-\mathbf{I}_{r_\mathbf{Q}}\right)\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)^{-1}\right\|_2 \\ &= \left\|\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}-\frac{1}{T}\widehat{\mathbf{G}}_\mathbf{Q}'\widehat{\mathbf{G}}_\mathbf{Q}\right)\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)^{-1}\right\|_2 \\ &= \left\|\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}-\frac{1}{T}\widehat{\mathbf{G}}_\mathbf{Q}'\widehat{\mathbf{G}}_\mathbf{Q}\right\|_2\left\|\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)^{-1}\right\|_2.\end{aligned}$$

Note that $\left\|\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)^{-1}\right\|_2=\mathcal{O}_p(1)$ and

$$\begin{aligned}\left\|\frac{1}{T}\widehat{\mathbf{G}}_\mathbf{Q}'\widehat{\mathbf{G}}_\mathbf{Q}-\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_2 &\lesssim \left\|\frac{2}{T}\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_2+\frac{1}{T}\left\|\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)'\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)\right\|_2 \\ &= \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)\end{aligned}$$

by (i). The result follows. For (v), observe that

$$\begin{aligned}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},-i}\mathbf{P}_{\widehat{\mathbf{G}}_\mathbf{R}}\boldsymbol{\varepsilon}_j\right\|_\infty &= \frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},-i}\mathbf{P}_{\mathbf{G}_\mathbf{R}}\boldsymbol{\varepsilon}_j\right\|_\infty+\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},-i}\left(\mathbf{P}_{\widehat{\mathbf{G}}_\mathbf{R}}-\mathbf{P}_{\mathbf{G}_\mathbf{R}}\right)\boldsymbol{\varepsilon}_j\right\|_\infty \\ &= \mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}\right)++\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},-i}\left(\mathbf{P}_{\widehat{\mathbf{G}}_\mathbf{R}}-\mathbf{P}_{\mathbf{G}_\mathbf{R}}\right)\boldsymbol{\varepsilon}_j\right\|_\infty,\end{aligned}$$

by Lemma H1. For the second term, first decompose $\mathbf{P}_{\widehat{\mathbf{G}}_\mathbf{R}}-\mathbf{P}_{\mathbf{G}_\mathbf{R}}$ as

$$\begin{aligned}&\frac{1}{T}\widehat{\mathbf{G}}_\mathbf{R}'\widehat{\mathbf{G}}_\mathbf{R}-\frac{1}{T}\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{G}\mathbf{H}_\mathbf{Q}\right)^{-1}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}' \\ &=\frac{1}{T}\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)'+\frac{1}{T}\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}' \\ &\quad+\frac{1}{T}\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)'\left(\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)+\frac{1}{T}\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\left(\mathbf{I}_{r_\mathbf{Q}}-\left(\frac{1}{T}\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right)^{-1}\right)\mathbf{H}_\mathbf{Q}'\mathbf{G}_\mathbf{Q}'\end{aligned}$$

so, by Lemma K2,

$$\begin{aligned}&\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},-i}\left(\mathbf{P}_{\widehat{\mathbf{G}}_\mathbf{R}}-\mathbf{P}_{\mathbf{G}_\mathbf{R}}\right)\boldsymbol{\varepsilon}_i\right\|_\infty \\ &\leq\frac{1}{\sqrt{T}}\max_{i\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{Q},i}'\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_\mathbf{Q}-\mathbf{G}_\mathbf{Q}\mathbf{H}_\mathbf{Q}\right\|_F\end{aligned}$$

$$
\begin{aligned}
&+ \frac{1}{\sqrt{T}} \max_{i \leq N} \frac{1}{T} \left\| \mathbf{U}'_{\mathbf{Q},i} \left(\widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right) \right\|_2 \frac{1}{\sqrt{T}} \left\| \mathbf{G}_{\mathbf{Q}} \right\|_F \left\| \mathbf{H}_{\mathbf{Q}} \right\|_2 \\
&+ \frac{1}{\sqrt{T}} \max_{i \leq N} \frac{1}{T} \left\| \mathbf{U}'_{\mathbf{Q},i} \left(\widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right) \right\|_2 \frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right\|_F \\
&+ \frac{1}{\sqrt{T}} \max_{i \leq N} \frac{1}{T} \left\| \mathbf{U}'_{\mathbf{Q},i} \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right\|_2 \left\| \mathbf{I}_{r_{\mathbf{Q}}} - \left(\frac{1}{T} \mathbf{H}'_{\mathbf{Q}} \mathbf{G}'_{\mathbf{Q}} \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right)^{-1} \right\|_2 \frac{1}{\sqrt{T}} \frac{1}{\sqrt{T}} \left\| \mathbf{G}_{\mathbf{Q}} \right\|_F \left\| \mathbf{H}_{\mathbf{Q}} \right\|_2 \\
&= \mathcal{O}_p \left(\frac{\sqrt{\log N}}{T\sqrt{N}} + \frac{\sqrt{\log N}}{T^2} \right) + \mathcal{O}_p \left(\frac{1}{N\sqrt{T}} + \frac{1}{T\sqrt{T}} + \frac{\sqrt{\log N}}{T\sqrt{N}} \right) \\
&= \mathcal{O}_p \left(\frac{1}{N\sqrt{T}} + \frac{1}{T\sqrt{T}} + \frac{\sqrt{\log N}}{T\sqrt{N}} \right),
\end{aligned}
$$

by (i), (iii) and (iv). Since $\sqrt{T}/N \to 0$ as $T, N \to \infty$, these rates are dominated by $\mathcal{O}_p\left(\frac{\sqrt{\log N}}{T}\right)$. The result follows.

For (vi), observe that

$$
\begin{aligned}
\frac{1}{\sqrt{N}T} \left\| \mathbf{U}'_{\mathbf{R}} \left(\widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right) \right\|_2 \lesssim{}& \frac{1}{T^2 N \sqrt{N}} \left\| \mathbf{U}'_{\mathbf{R}} \mathbf{G}_{\mathbf{Q}} \mathbf{C}'_{\mathbf{Q}} \mathbf{U}'_{\mathbf{Q}} \widehat{\mathbf{G}}_{\mathbf{Q}} \right\|_2 + \frac{1}{T^2 N \sqrt{N}} \left\| \mathbf{U}'_{\mathbf{Q}} \mathbf{U}_{\mathbf{R}} \mathbf{C}_{\mathbf{Q}} \mathbf{G}'_{\mathbf{Q}} \widehat{\mathbf{G}}_{\mathbf{Q}} \right\|_2 \\
&+ \frac{1}{T^2 N \sqrt{N}} \left\| \mathbf{U}'_{\mathbf{R}} \mathbf{U}_{\mathbf{Q}} \mathbf{U}'_{\mathbf{Q}} \widehat{\mathbf{G}}_{\mathbf{Q}} \right\|_2.
\end{aligned}
$$

Bound them separately, for the first term,

$$
\begin{aligned}
\frac{1}{T^2 N \sqrt{N}} \left\| \mathbf{U}'_{\mathbf{R}} \mathbf{G}_{\mathbf{Q}} \mathbf{C}'_{\mathbf{Q}} \mathbf{U}'_{\mathbf{Q}} \widehat{\mathbf{G}}_{\mathbf{Q}} \right\|_2 \leq{}& \frac{1}{T\sqrt{N}} \frac{1}{\sqrt{NT}} \left\| \mathbf{U}'_{\mathbf{R}} \mathbf{G}_{\mathbf{Q}} \right\|_2 \frac{1}{\sqrt{NT}} \left\| \mathbf{C}'_{\mathbf{Q}} \mathbf{U}'_{\mathbf{Q}} \mathbf{G}_{\mathbf{Q}} \right\|_2 \left\| \mathbf{H}_{\mathbf{Q}} \right\|_2 \\
&+ \frac{1}{\sqrt{NT}} \frac{1}{\sqrt{NT}} \left\| \mathbf{U}'_{\mathbf{R}} \mathbf{G}_{\mathbf{Q}} \right\|_2 \frac{1}{\sqrt{NT}} \left\| \mathbf{C}'_{\mathbf{Q}} \mathbf{U}'_{\mathbf{Q}} \right\|_2 \frac{1}{\sqrt{T}} \left\| \widehat{\mathbf{G}}_{\mathbf{Q}} - \mathbf{G}_{\mathbf{Q}} \mathbf{H}_{\mathbf{Q}} \right\|_2 \\
={}& \mathcal{O}_p \left(\frac{1}{T\sqrt{N}} \right) + \frac{1}{\sqrt{NT}} \mathcal{O}_p \left(\frac{1}{\sqrt{N}} + \frac{1}{T} \right) \\
={}& \mathcal{O}_p \left(\frac{1}{T\sqrt{N}} + \frac{1}{N\sqrt{T}} \right).
\end{aligned}
$$

The second term:

$$
\begin{aligned}
\frac{1}{T^2 N \sqrt{N}} \left\| \mathbf{U}'_{\mathbf{Q}} \mathbf{U}_{\mathbf{R}} \mathbf{C}_{\mathbf{Q}} \mathbf{G}'_{\mathbf{Q}} \widehat{\mathbf{G}}_{\mathbf{Q}} \right\|_2 &\leq \frac{1}{NT\sqrt{N}} \left\| \mathbf{U}'_{\mathbf{Q}} \mathbf{U}_{\mathbf{R}} \mathbf{C}_{\mathbf{Q}} \right\|_2 \frac{1}{T} \left\| \mathbf{G}'_{\mathbf{Q}} \widehat{\mathbf{G}}_{\mathbf{Q}} \right\|_2 \\
&\leq \mathcal{O}_p \left(\frac{1}{N} + \frac{1}{\sqrt{NT}} \right)
\end{aligned}
$$

by Lemma H1(*xiv*).

The third term:

$$\begin{aligned}\frac{1}{T^2N\sqrt{N}}\left\|\mathbf{U}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_2 &\leq \frac{1}{T^2N\sqrt{N}}\left\|\mathbf{U}_{\mathbf{R}}'\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}\right\|_2\left\|\mathbf{H}_{\mathbf{Q}}\right\|_2 \\ &\quad+\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{R}}\right\|_2\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right\|_2 \\ &=\mathcal{O}_p\left(\frac{1}{T}+\frac{1}{\sqrt{NT}}\right)+\mathbf{o}_p\left(\frac{1}{N}+\frac{1}{T}\right)\mathcal{O}_p\left(\frac{1}{\sqrt{N}}+\frac{1}{T}\right) \\ &=\mathcal{O}_p\left(\frac{1}{T}+\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

Putting all pieces we have,

$$\frac{1}{\sqrt{N}T}\left\|\mathbf{U}_{\mathbf{R}}'\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)\right\|_2=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right).$$

by Lemma H1(*xiv*). The result follows.

For (vii), by (37) and by Lemma H1

$$\begin{aligned}\frac{1}{NT}\left\|\left(\widehat{\mathbf{G}}_{\mathbf{Q}}-\mathbf{G}_{\mathbf{Q}}\mathbf{H}_{\mathbf{Q}}\right)'\mathbf{U}_{\mathbf{R}}\mathbf{C}_{\mathbf{R}}\right\|_2 &\lesssim \frac{1}{N\sqrt{T}}\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_F\left\|\frac{1}{\sqrt{NT}}\mathbf{U}_{\mathbf{Q}}\mathbf{C}_{\mathbf{Q}}\right\|_2\left\|\frac{1}{\sqrt{NT}}\mathbf{Q}_{\mathbf{Q}}\mathbf{U}_{\mathbf{R}}\mathbf{C}_{\mathbf{R}}\right\|_2 \\ &\quad+\frac{1}{N}\left\|\frac{1}{T}\widehat{\mathbf{G}}_{\mathbf{Q}}'\mathbf{G}_{\mathbf{Q}}\right\|_F\left\|\frac{1}{\sqrt{NT}}\mathbf{C}_{\mathbf{Q}}'\mathbf{U}_{\mathbf{Q}}'\right\|_2\left\|\frac{1}{\sqrt{NT}}\mathbf{U}_{\mathbf{R}}\mathbf{C}_{\mathbf{R}}\right\|_2 \\ &\quad+\frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{T}}\widehat{\mathbf{G}}_{\mathbf{Q}}\right\|_F\left\|\frac{1}{NT}\mathbf{U}_{\mathbf{Q}}\mathbf{U}_{\mathbf{Q}}'\right\|_2\left\|\frac{1}{\sqrt{NT}}\mathbf{U}_{\mathbf{R}}\mathbf{C}_{\mathbf{R}}\right\|_2 \\ &=\mathcal{O}_p\left(\frac{1}{N\sqrt{T}}\right)+\mathcal{O}_p\left(\frac{1}{N}\right)+\mathcal{O}_p\left(\frac{1}{N\sqrt{N}}+\frac{1}{T\sqrt{N}}\right) \\ &=\mathcal{O}_p\left(\frac{1}{N}\right).\end{aligned}$$

□

Appendix I: Strategy to tackle the rank-degeneracy in $\mathbf{P}_{\widehat{\mathbf{G}}}$

Since $\mathbf{X}$ and $\mathbf{W}_k$ are allowed to share common factors $\mathbf{F}$, $\widehat{\mathbf{G}}$ obtained from Algorithm 1 is a rank-degenerate matrix as N and T grow large. Therefore, one need to be careful when bounding terms that involves $\mathbf{P}_{\widehat{\mathbf{G}}}$ due to the presence of pseudo-inverse in $\mathbf{P}_{\widehat{\mathbf{G}}}$. To be specific, the term being taken the pseudo-inverse in $\widehat{\mathbf{P}}_{\mathbf{G}}$ is $\frac{1}{T}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}$, which converges to $\frac{1}{T}\mathbf{H}'\mathbf{G}'\mathbf{G}\mathbf{H}$ at the rate of

$\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$ following the same argument in Lemma H2 (v). However, continuous mapping theorem does not apply to obtain $\left(\frac{1}{T}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}\right)^{-}-\left(\frac{1}{T}\mathbf{H}'\mathbf{G}'\mathbf{G}\mathbf{H}\right)^{-}=\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$ since the rank of $\frac{1}{T}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}$ is larger than that of $\frac{1}{T}\mathbf{H}'\mathbf{G}'\mathbf{G}\mathbf{H}$ with probability approaching one and pseudo-inverse is not continuous on the set of matrices with different ranks. The intuition is straightforward: when one takes the pseudo-inverse of a symmetric matrix, the eigenvectors are preserved, and the non-zero eigenvalues are inverted while the zero eigenvalues remain unchanged. Analogous to the scalar function $f(x):\mathbb{R}\to\mathbb{R}$ defined as $f(x)=1/x$ for $x\neq 0$ and $f(0)=0$, the pseudo-inverse of a matrix is not continuous at points where the eigenvalues change from non-zero to zero. For the sake of it, one can only obtain $\left(\frac{1}{T}\widehat{\mathbf{G}}'\widehat{\mathbf{G}}\right)^{-}-\left(\frac{1}{T}\mathbf{H}'\mathbf{G}'\mathbf{G}\mathbf{H}\right)^{-}=\mathcal{O}_p\left(N+T\right)$. Similar issue arises in CCE estimators by Pesaran (2006) when the number of variables to proxy factors is larger than the number of factors as discussed by Karabiyik et al. (2017).

The solution given by Karabiyik et al. (2017) is to amplify the degenerate component of the matrix with a suitable diverging sequence so that it converges to another matrix with the same rank. This approach, however, is relatively conservative in the sense that it amplifies the noise component. Although it gives the same convergence rate for CCE estimator, it leads to a slower rate in my setting due to a more complicated error of the factor estimation from PCA. Instead, I adopt the leave-one-out technique by Abbe et al. (2022) and discussed in Section 4.1 in Chen et al. (2021), which utilizes the cross-sectional independence in the idiosyncratic components in Assumption 3. I demonstrate this technique in the proof of the following lemma.

Lemma I1. Under the assumptions of Theorem 4.1, we have

$$\frac{1}{T}\|\mathbf{U}'_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\|_\infty=\mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T}+\frac{\log N}{N}\right),$$

and same rate holds if $\boldsymbol{\varepsilon}_i$ is replaced by $\mathbf{U}_{\mathbf{Q},i}$ for $\mathbf{Q}=\mathbf{X},\mathbf{W}_1,\ldots,\mathbf{W}_K$. Similarly, we have

$$\frac{1}{T}\|\mathbf{U}'_{\mathbf{Q},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{R}}\|_\infty=\mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T}+\frac{\log N}{N}\right),$$

with $\mathbf{Q},\mathbf{R}\in\{\mathbf{X},\mathbf{W}_1,\ldots,\mathbf{W}_K\}$.

Proof of Lemma I1. To show the first part of the lemma, let $\widehat{\mathbf{G}}^{-(ij)}$ denote the estimated factors by applying Algorithm 1 to data with the i-th and j-th units removed. The objective term can be

decomposed as

$$\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty \leq \max_{j\neq i}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right| + \max_{j\neq i}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},j}'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right)\boldsymbol{\varepsilon}_i\right|.$$

To bound the second term, write

$$\max_{j\neq i}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},j}'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right)\boldsymbol{\varepsilon}_i\right| \leq \max_{j\neq i}\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},j}\right\|_2 \cdot \max_{j\neq i}\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right\|_2 \frac{1}{\sqrt{T}}\left\|\boldsymbol{\varepsilon}_i\right\|_2.$$

By Assumption 3, $\frac{1}{\sqrt{T}}\left\|\boldsymbol{\varepsilon}_i\right\|_2 = \mathcal{O}_p(1)$. By Lemma K1, $\max_{j\neq i}\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},j}\right\|_2 = \mathcal{O}_p(1)$. To bound $\max_{j\neq i}\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right\|_2$, observe that the ranks of $\mathbf{P}_{\widehat{\mathbf{G}}}$ and $\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}$ are the same with probability approaching one. By applying Davis-Kahan $\sin\Theta$ Theorem twice, we have

$$\begin{aligned}
\max_{j\neq i}\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right\|_2 &= \max_{j\neq i}\left\|U_r(\widehat{\mathbf{G}})U_r(\widehat{\mathbf{G}})' - U_r(\widehat{\mathbf{G}}^{-(ij)})U_r(\widehat{\mathbf{G}}^{-(ij)})'\right\|_2 \\
&\leq \max_{j\neq i}\left\|\widehat{\mathbf{G}}-\widehat{\mathbf{G}}^{-(ij)}\right\|_2 \\
&\lesssim \max_{j\neq i}\left\|\widehat{\mathbf{G}}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}^{-(ij)}\right\|_2 \\
&\lesssim \max_{j\neq i}\frac{\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}'-\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right\|_2}{\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right)-\lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right)\right|} \\
&\leq \frac{\max_{j\neq i}\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}'-\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right\|_2}{\min_{j\neq i}\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right)-\lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right)\right|} \\
&\leq \frac{1}{\lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right)-\mathbf{o}_p(1)}\max_{j\neq i}\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}'-\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right\|_2 \\
&= \mathcal{O}_p\left(\frac{\log N}{N}\right),
\end{aligned}$$

where $U_r(\cdot)$ denotes the top r left singular vectors of the corresponding matrix, $\mathbf{X}^{-(ij)}$ is the data matrix with the i-th and j-th columns removed and $\lambda_r(\cdot)$ and $\lambda_{r+1}(\cdot)$ are the r-th and $(r+1)$-th largest eigenvalues of the corresponding matrix. The last two steps follow from Lemma I8 (c) and (d). Therefore

$$\max_{j\neq i}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},j}'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right)\boldsymbol{\varepsilon}_i\right| = \mathcal{O}_p\left(\frac{\log N}{N}\right).$$

To bound the first term $\max_{j\neq i}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|$, let $\mathcal{F}_{-j}$ denote the σ-algebra generated by

$\{\mathbf{U}_{\mathbf{X},k}\}_{k\neq j}$, $\{\mathbf{C}_{\mathbf{X},k}\}_{k\neq j}$ and $\{\boldsymbol{\varepsilon}_k\}_{k\neq j}$. For each j,

$$\begin{aligned}\mathbb{P}\left(\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\right)&=\mathbb{E}\left[\mathbb{1}\left\{\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\right\}\right]\\&=\mathbb{E}\left[\mathbb{E}\left[\mathbb{1}\left\{\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\right\}\mid\mathcal{F}_{-j}\right]\right]\\&=\mathbb{E}\left[\mathbb{P}\left(\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\mid\mathcal{F}_{-j}\right)\right]\end{aligned}$$

By Bonnferoni's inequality and Theorem 2.6.3 in Vershynin (2024), we have

$$\begin{aligned}\mathbb{P}\left(\max_{j\neq i}\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\right)&\leq\sum_{j\neq i}\mathbb{P}\left(\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\right)\\&\leq\sum_{j\neq i}\mathbb{E}\left[\mathbb{P}\left(\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\mid\mathcal{F}_{-j}\right)\right]\\&\leq\sum_{j\neq i}2\mathbb{E}\left[\exp\left(-c\frac{t^2}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2}\right)\right]\end{aligned}$$

for some constant $c>0$. Observe that, for some sequence M_N,

$$\mathbb{E}\left[\exp\left(-c\frac{t^2}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2}\right)\right]\leq\mathbb{E}\left[\exp\left(-c\frac{t^2}{M_N}\right)\right]\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2\leq M_N\right)+\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2>M_N\right).$$

Plug in back to the previous display, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{j\neq i}\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right|>t\right)\\&\quad\leq2\mathbb{E}\left[\exp\left(-c\frac{t^2}{M_N}\right)\right]\sum_{j\neq i}\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2\leq M_N\right)+2\sum_{j\neq i}\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2>M_N\right).\end{aligned}$$

By Hanson-Wright inequality in Rudelson and Vershynin (2013) or Theorem 6.2.1 in Vershynin (2024), and the idempotent property of projection matrices, and the fact that $\mathbb{E}\left[\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2|\mathcal{F}_{-i}\right]=\mathcal{O}(1)$, we have

$$\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2^2>M_N^2\right)\leq2\exp\left(-c_2\min\left(\frac{M_N^4}{r^4},\frac{M_N^2}{r^2}\right)\right)$$

for some $M_N>\mathbb{E}\left[\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2|\mathcal{F}_{-i}\right]$, where r is the rank of $\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}$. Pick $M_N^2=C\log N$ for some sufficiently large constant C such that $1-Cc_2<0$, we have $\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2>\sqrt{\log N}\right)\leq2N^{-Cc_2}$

and

$$\sum_{j\neq i}\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2 > M_N\right) \leq 2N^{1-Cc_2} = \mathbf{o}_p(1) \quad \text{and} \quad \sum_{j\neq i}\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2 \leq M_N\right) \leq N.$$

Let $t = \sqrt{C}(\log N)^{3/4}$ for some C large enough such that $1 - Cc < 0$, we have

$$\begin{aligned} 2\mathbb{E}\left[\exp\left(-c\frac{t^2}{M_N}\right)\right]\sum_{j\neq i}\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right\|_2 \leq M_N\right) &\leq 2N\mathbb{E}\left[\exp\left(-c\frac{t^2}{M_N}\right)\right] \\ &= 2N^{1-Cc} = \mathbf{o}_p(1). \end{aligned}$$

Putting them together, we have

$$\mathbb{P}\left(\max_{j\neq i}\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right| > \sqrt{C}(\log N)^{3/4}\right) = \mathbf{o}(1),$$

which implies $\max_{j\neq i}\frac{1}{T}\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\boldsymbol{\varepsilon}_i\right| = \mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T}\right)$. Therefore,

$$\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},-i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_\infty = \mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T} + \frac{\log N}{N}\right),$$

which completes the proof of the first part of the lemma. The same argument applies to the second statement with leave-one-out projection matrix $\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}$ and the fact that $\frac{1}{\sqrt{T}}\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{G}_{\mathbf{X}}\right\|_2 = \mathcal{O}_p(1)$. □

Lemma I2. Under the assumptions of Theorem 4.1, we have

$$\frac{1}{T}\left\|\mathbf{U}'_{\mathbf{X}}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\right\|_{\max} = \mathcal{O}_p\left(\frac{\log N}{T} + \frac{\log N}{N}\right).$$

Proof of Lemma I2. Observe that

$$\begin{aligned} \frac{1}{T}\left\|\mathbf{U}'_{\mathbf{X}}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\right\|_{\max} &= \max_{i\leq N}\max_{j\leq N}\left|\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\right| \\ &\leq \max_{i\leq N}\max_{j\leq N}\frac{1}{T}\left|\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right| + \max_{i\leq N}\max_{j\leq N}\frac{1}{T}\left|\mathbf{U}'_{\mathbf{X},i}\left(\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right)\mathbf{U}_{\mathbf{X},j}\right|. \end{aligned}$$

The second term is bounded by

$$\begin{aligned}\frac{1}{T}\max_{i\leq N}\max_{j\leq N}\left|\mathbf{U}'_{\mathbf{X},i}\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right)\mathbf{U}_{\mathbf{X},j}\right| &\leq \max_{i\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},i}\right\|_2^2\cdot\max_{i,j\leq N}\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right\|_2\\ &=\mathcal{O}_p\left(\frac{\log N}{N}\right)\end{aligned}$$

The first term can be decomposed as

$$\max_{i\leq N}\max_{j\leq N}\frac{1}{T}\left|\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right|\leq\max_{i\leq N}\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}+\max_{i\leq N}\max_{j\neq i}\frac{1}{T}\left|\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right|.$$

By Hanson-Wright inequality (Theorem 1 in Rudelson and Vershynin (2013) or Theorem 6.2.1 in Vershynin (2024)), we have

$$\begin{aligned}\mathbb{P}\left(\left(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}-\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}\right]\right)>t\right)&\leq 2\exp\left(-c\min\left(\frac{t^2}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right\|_F^4},\frac{t}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right\|_F^2}\right)\right)\\ &\leq 2\exp\left(-c\min\left(\frac{t^2}{r^4},\frac{t}{r^2}\right)\right)\end{aligned}$$

for some constant $c>0$, where r is the rank of $\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}$. Pick $t=Cr^2\log N$ for some sufficiently large constant C such that $1-Cc<0$, we have

$$\begin{aligned}\mathbb{P}\left(\max_{i\leq N}\left(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}-\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}\right]\right)>Cr^2\log N\right)&\leq 2N\exp\left(-cC\log N\right)\\ &=2N^{1-Cc}=\mathbf{o}_p\left(1\right),\end{aligned}$$

which implies $\max_{i\leq N}\left(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}-\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}\right]\right)=\mathcal{O}_p\left(\log N\right)$. For the expectation term, we have

$$\begin{aligned}\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}\right]&=\mathbb{E}\left[\operatorname{tr}\left(\mathbf{P}_{\widehat{\mathbf{G}}^{(-i)}}\mathbf{U}_{\mathbf{X},i}\mathbf{U}'_{\mathbf{X},i}\right)\right]\\ &=\mathbb{E}\left[\operatorname{tr}\left(\mathbf{P}_{\widehat{\mathbf{G}}^{(-i)}}\mathbb{E}\left[\mathbf{U}_{\mathbf{X},i}\mathbf{U}'_{\mathbf{X},i}\mid\mathcal{F}_{-i}\right]\right)\right]\\ &=\mathbb{E}\left[\operatorname{tr}\left(\mathbf{P}_{\widehat{\mathbf{G}}^{(-i)}}\Sigma_{\mathbf{X},i}\right)\right]\\ &\leq\mathbb{E}\left[\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{(-i)}}\right\|_*\right]\cdot\left\|\mathbf{\Sigma}_{\mathbf{X},i}\right\|_2\\ &=r\cdot\left\|\mathbf{\Sigma}_{\mathbf{X},i}\right\|_2,\end{aligned}$$

therefore $\max_{i\leq N}\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}=\mathcal{O}_p\left(\frac{\log N}{T}\right)$.

For the second term, fix i and $j\neq i$, following a similar argument as in the proof of Lemma I1, we have

$$\begin{aligned}
&\mathbb{P}\left(\left|(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2-\mathbb{E}\left[(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2\right]\right|>t\right)\\
&\leq\mathbb{E}\left[\exp\left(-c\min\left(\frac{t^2}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^4},\frac{t}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^2}\right)\right)\right]\\
&\leq\mathbb{E}\left[\exp\left(-c\min\left(\frac{t^2}{M_{NT}^2},\frac{t}{M_{NT}}\right)\right)\right]\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^2\leq M_{NT}\right)+\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^2>M_{NT}\right)\\
&\leq\mathbb{E}\left[\exp\left(-c\min\left(\frac{t^2}{M_{NT}^2},\frac{t}{M_{NT}}\right)\right)\right]+\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^2>M_{NT}\right).
\end{aligned}$$

So, by union bound,

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\leq N}\max_{j\neq i}\left|(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2-\mathbb{E}\left[(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2\right]\right|>t\right)\\
&\leq 2N^2\mathbb{E}\left[\exp\left(-c\min\left(\frac{t^2}{M_{NT}^2},\frac{t}{M_{NT}}\right)\right)\right]+2N^2\mathbb{P}\left(\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^2>M_{NT}\right).
\end{aligned}$$

Note that

$$\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right\|_F^2=\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}.$$

So, by the same argument as above

$$\mathbb{P}\left(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}>s+\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}\right]\right)\leq 2\exp\left(-c\min\left(\frac{s^2}{r^4},\frac{s}{r^2}\right)\right).$$

So, let $M_{NT}=Cr^2\log N+\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i}\right]$ for some sufficiently large constant C such that $2-Cc<0$ and $t=C'M_{NT}\log N$ for some sufficiently large constant C' such that $2-C'c<0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{i\leq N}\max_{j\neq i}\left|(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2-\mathbb{E}\left[(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2\right]\right|>C'M_{NT}\log N\right)\\
&\leq 2N^{(2-Cc)}+2N^{(2-C'c)}=\mathbf{o}_p\left(1\right),
\end{aligned}$$

which implies

$$\max_{i\leq N}\max_{j\neq i}\left|(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2-\mathbb{E}\left[(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2\right]\right|=\mathcal{O}_p\left((\log N)^2\right).$$

For the expectation term, by a similar argument as above, we have

$$\mathbb{E}\left[(\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j})^2\right]\leq\left\|\mathbf{\Sigma}_{\mathbf{X},i}\right\|_2\left\|\mathbf{\Sigma}_{\mathbf{X},j}\right\|_2\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right\|_*=\mathcal{O}(1).$$

Putting them together, we have $\max_{i,j\leq N}\frac{1}{T}\left|\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\mathbf{U}_{\mathbf{X},j}\right|=\mathcal{O}_p\left(\frac{\log N}{T}+\frac{\log N}{N}\right)$. The result follows.

□

The following lemma provides the bound for the 2-norm of a general term.

Lemma I3. Under the assumptions of Theorem 4.1, let $\mathbf{V}$ be any conformable random matrix that is independent of $\boldsymbol{\varepsilon}_i$ and satisfies $\|\mathbf{V}\|_2=\mathcal{O}_p(a_{NT})$ for some deterministic sequence a_{NT}, we have

$$\left\|\mathbf{V}'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2=\mathcal{O}_p(a_{NT}).$$

for any i.

Proof of Lemma I3. The leave-one-out technique is applied in a smilar way as in the proof of Lemma I1. Let $\widehat{\mathbf{G}}^{-(i)}$ denote the estimated factors by applying Algorithm 1 to data with the i-th units removed. We have

$$\left\|\mathbf{V}'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\leq\left\|\mathbf{V}'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i\right\|_2+\left\|\mathbf{V}'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right)\boldsymbol{\varepsilon}_i\right\|_2.$$

Let $\mathcal{F}$ denotes the σ-algebra generated by $\mathbf{V}$ and data excluding the i-th unit. By the same argument as in the proof of Lemma I1, we have

$$\mathbb{P}\left(\left\|\mathbf{V}'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i\right\|_2^2\geq t\mid\mathcal{F}\right)\lesssim\exp\left(-\min\left(\frac{t}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{V}\right\|_2^2},\frac{t^2}{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{V}\right\|_2^4}\right)\right).$$

Pick $t=Ca_{NT}^2$, we have

$$\left\|\mathbf{V}'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i\right\|_2=\mathcal{O}_p(a_{NT}).$$

The second term is bounded by

$$\left\|\mathbf{V}'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right)\boldsymbol{\varepsilon}_i\right\|_2 \leq \|\mathbf{V}\|_2\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right\|_2\|\boldsymbol{\varepsilon}_i\|_2 = \mathcal{O}_p\left(a_{NT}\cdot\frac{\sqrt{T}}{N}\right) = \mathbf{o}_p\left(a_{NT}\right),$$

by Lemma I7. □

Lemma I4. Under the assumptions of Theorem 4.1, for all i, we have

$$\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \mathcal{O}_p\left(1+\frac{T}{N}\right).$$

Proof of Lemma I4. The target term can be decomposed as

$$\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i + \boldsymbol{\varepsilon}_i'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right)\boldsymbol{\varepsilon}_i.$$

The first term can be bounded by Markov's inequality conditional on the data excluding the i-th unit,

$$\begin{aligned}
\mathbb{P}\left(\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i|\mathcal{F}_{-i}>t\right) &\leq \frac{\mathbb{E}\left[\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i|\mathcal{F}_{-i}\right]}{t} \\
&= \frac{\operatorname{tr}\left(\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\Sigma_{\varepsilon,i}\right)}{t} \\
&\leq \frac{\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right\|_*\cdot\|\Sigma_{\varepsilon,i}\|_2}{t} \\
&= \frac{r\cdot\|\Sigma_{\varepsilon,i}\|_2}{t},
\end{aligned}$$

which implies $\boldsymbol{\varepsilon}_i'\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\boldsymbol{\varepsilon}_i = \mathcal{O}_p\left(1\right)$. The second term is bounded by

$$\boldsymbol{\varepsilon}_i'\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right)\boldsymbol{\varepsilon}_i \leq \|\boldsymbol{\varepsilon}_i\|_2^2\cdot\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right\|_2 = \mathcal{O}_p\left(\frac{T}{N}\right).$$

The result follows. □

The following lemma provides the bound for terms that involve $\mathbf{P}_{\widehat{\mathbf{G}}}$ in the main proof of Theorem 4.1.

Lemma I5. Under the assumptions of Theorem 4.1, we have

(i) $\frac{1}{T}\left\|\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right)$ and same rate holds for $\mathbf{W}_k$ replacing $\mathbf{X}$;

(ii) $\max_{j\leq N}\frac{1}{T}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\left(\widehat{\mathbf{G}}_{\mathbf{X}}-\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\right\|_2=\mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}+\frac{\log N}{N\sqrt{N}}\right)=\mathcal{O}_p\left(\frac{\log N}{\sqrt{NT}}\right)$;

(iii) $\frac{1}{T}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\left(\widehat{\mathbf{G}}_{\mathbf{X}}-\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\right\|_2=\mathcal{O}_p\left(\sqrt{\frac{1}{NT}}+\frac{1}{N\sqrt{N}}\right)=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right)$.

(iv) $\frac{1}{T}\mathbb{Z}'_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbb{Z}_i=\mathbb{E}\left[\frac{1}{T}\mathbb{U}'_i\mathbb{U}_i\right]+\mathbf{o}_p(1)$.

Proof of Lemma I5. For (i), decomposing it yields

$$\begin{aligned}&\frac{1}{T}\left\|\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\\&\lesssim\frac{1}{NT^2}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2+\frac{1}{NT^2}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}\mathbf{C}'_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2+\frac{1}{NT^2}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\end{aligned}$$

The first term is bounded by

$$\begin{aligned}\frac{1}{NT^2}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\mathbf{G}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2&\leq\frac{1}{T\sqrt{N}}\left\|\mathbf{H}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{G}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{G}'_{\mathbf{X}}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\\&\quad+\frac{1}{\sqrt{NT}}\left\|\mathbf{H}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{X}}-\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{G}'_{\mathbf{X}}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\\&=\mathcal{O}_p\left(\frac{1}{T\sqrt{N}}\right),\end{aligned}$$

by Lemma H1, Lemma H2 and Lemma I3 with $\left\|\mathbf{G}_{\mathbf{X}}\right\|_2=\mathcal{O}_p\left(\sqrt{T}\right)$. The second term is bounded by

$$\begin{aligned}\frac{1}{NT^2}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}\mathbf{C}'_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2&\leq\frac{1}{\sqrt{NT}}\frac{1}{T}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}'_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\\&=\mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right),\end{aligned}$$

by Lemma I3 with $\frac{1}{\sqrt{NT}}\left\|\mathbf{C}'_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\right\|_2=\mathcal{O}_p(1)$ by Lemma H1. The last term is bounded by

$$\begin{aligned}\frac{1}{NT^2}\left\|\widehat{\mathbf{G}}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2&\leq\frac{1}{\sqrt{T}}\left\|\mathbf{H}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X}}\mathbf{U}'_{\mathbf{X}}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2\\&=\frac{1}{\sqrt{T}}\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right),\end{aligned}$$

by Lemma I3 with $\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\right\|_2 = \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$ by Lemma H1. Therefore,

$$\frac{1}{T}\left\|\left(\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}-\widehat{\mathbf{G}}_{\mathbf{X}}\right)\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{NT}}\right).$$

For (ii), decompose it as

$$\begin{aligned}\max_{j\le N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\left(\widehat{\mathbf{G}}_{\mathbf{X}}-\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\right\|_2 \lesssim{}& \max_{j\le N}\frac{1}{NT^2}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2\\ &+\max_{j\le N}\frac{1}{NT^2}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2\\ &+\max_{j\le N}\frac{1}{NT^2}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2\end{aligned}$$

Observe that the first term is bounded by

$$\begin{aligned}\max_{j\le N}\frac{1}{NT^2}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2 \le{}& \frac{1}{\sqrt{NT}}\max_{j\le N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}\right\|_2\left\|\mathbf{H}_{\mathbf{X}}\right\|_2\\ &+\frac{1}{\sqrt{N}}\max_{j\le N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\right\|_2\frac{1}{\sqrt{T}}\left\|\widehat{\mathbf{G}}_{\mathbf{X}}-\mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right\|_2\\ ={}&\mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T\sqrt{NT}}+\frac{\log N}{N\sqrt{NT}}\right)\\ &+\mathcal{O}_p\left(\frac{(\log N)^{3/4}}{NT}+\frac{\log N}{N^2}+\frac{(\log N)^{3/4}}{T^2\sqrt{N}}+\frac{\log N}{NT\sqrt{N}}\right)\\ ={}&\mathcal{O}_p\left(\frac{(\log N)^{3/4}}{T^2}+\frac{\log N}{N^2}\right),\end{aligned}$$

by Lemma I1.

For the second term,

$$\max_{j\le N}\frac{1}{NT^2}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2 \le \max_{j\le N}\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\right\|_2\frac{1}{T}\left\|\mathbf{G}_{\mathbf{X}}'\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2.$$

So the rate is determined by $\max_{j\le N}\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\right\|_2$. Decompose it as

$$\begin{aligned}\max_{j\le N}\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\right\|_2 \le{}& \max_{j\le N}\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X},j}'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2\\ &+\frac{1}{NT}\max_{j\le N}\left\|\mathbf{U}_{\mathbf{X},j}'\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\mathbf{C}_{\mathbf{X},j}\right\|_2.\end{aligned}$$

Examine the second term, it is bounded by

$$\frac{1}{NT}\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\mathbf{C}_{\mathbf{X},j}\right\|_2 \leq \frac{1}{NT}\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\right\|_2 \max_{j\leq N}\|\mathbf{C}_{\mathbf{X},j}\|_2$$

It is easy to verify that $\max_{j\leq N}\|\mathbf{C}_{\mathbf{X},j}\|_2 = \mathcal{O}_p\left(\sqrt{\log N}\right)$. The first component is proceeded by the leave-one-out analysis. Bound it by

$$\frac{1}{NT}\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\right\|_2 \leq \frac{1}{NT}\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},j}\right\|_2 + \frac{1}{NT}\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}-\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X},j}\right\|_2,$$

where

$$\begin{aligned}\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}-\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X},j}\right\|_2 &\leq r_{\mathbf{X}}\frac{1}{N}\max_{j\leq N}\frac{1}{T}\|\mathbf{U}_{\mathbf{X},j}\|_2^2\max_{j\leq N}\left\|\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}-\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2 \\ &= \mathcal{O}_p\left(\frac{\log N}{N^2}\right),\end{aligned}$$

by Lemma I7. And by Theorem 6.2.1 in Vershynin (2024), we have

$$\mathbb{P}\left(\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},j}\right\|_2 > t \mid \mathcal{F}_{-j}\right) \leq \mathbb{P}\exp\left(-c_x t\right)$$

for some $c_x > 0$ depends on $r_{\mathbf{X}}$. Since the right hand side does not depend on $\mathcal{F}_{-j}$, we have

$$\mathbb{P}\left(\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},j}\right\|_2 > t \mid \mathcal{F}_{-j}\right) \leq N\exp\left(-c_x t\right),$$

which implies $\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},j}\right\|_2 = \mathcal{O}_p\left(\log N\right)$. Therefore,

$$\frac{1}{NT}\max_{j\leq N}\left\|\mathbf{U}'_{\mathbf{X},j}\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\mathbf{C}_{\mathbf{X},j}\right\|_2 = \mathcal{O}_p\left(\frac{(\log N)^{3/2}}{N^2}+\frac{(\log N)^{3/2}}{NT}\right).$$

Now we bound the first term. Again, applying the leave-one-out analysis, decompose it as

$$\begin{aligned}\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2 \leq &\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2 \\ &+\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2.\end{aligned}$$

Examine the second term,

$$\begin{aligned}
&\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2 \\
&\quad\leq\frac{1}{\sqrt{N}}\max_{j\leq N}\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},j}\right\|_2\max_{j\leq N}\left\|\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}-\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2\max_{j\leq N}\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2
\end{aligned}$$

$\max_{j\leq N}\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},j}\right\|_2=\mathcal{O}_p\left(1\right)$ and $\max_{j\leq N}\left\|\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}}-\widehat{\mathbf{P}}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2=\mathcal{O}_p\left(\frac{\log N}{N}\right)$ by Lemma I7. In the following I will show $\max_{j\leq N}\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2=\mathcal{O}_p\left(1\right)$, which gives us

$$\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2=\mathcal{O}_p\left(\frac{\log N}{N\sqrt{N}}\right).$$

To bound $\max_{j\leq N}\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2$, WLOG assume $r_{\mathbf{X}}=1$, observe that

$$\begin{aligned}
\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2^2\leq\max_{j\leq N}\frac{1}{NT}&\left|\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}-\mathbb{E}\left[\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right]\right| \\
&+\max_{j\leq N}\frac{1}{NT}\mathbb{E}\left[\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right].
\end{aligned}$$

It is easy to verify that $\max_{j\leq N}\frac{1}{NT}\mathbb{E}\left[\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right]=\mathcal{O}\left(1\right)$. By Theorem 6.2.1 in Vershynin (2024), conditional on $\mathbf{U}_{\mathbf{X},-j}$, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{NT}\left|\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}-\mathbb{E}\left[\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right]\right|>t\mid\mathbf{U}_{\mathbf{X},-j}\right) \\
&<\exp\left(-c\min\left(\frac{t^2}{\left\|\frac{1}{NT}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\right\|_F^2},\ \frac{t}{\left\|\frac{1}{NT}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\right\|_F}\right)\right).
\end{aligned}$$

So, for some sequence M_{NT}, we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{j\leq N}\frac{1}{NT}\left|\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right|>t+\mathbb{E}\left[\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right]\right) \\
&<\exp\left(-c\min\left(\frac{t^2}{M_{NT}^2},\ \frac{t}{M_{NT}}\right)\right)\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\right\|_F\leq M_{NT}\right) \\
&\quad+\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\right\|_F>M_{NT}\right).
\end{aligned}$$

Note that

$$\begin{aligned}\left\|\frac{1}{NT}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\right\|_F &= \frac{1}{NT}\left\|\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_F \\ &\leq \frac{\min\left(\sqrt{N},\sqrt{T}\right)}{NT}\left\|\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_2\end{aligned}$$

Following the argument of Theorem 4.6.1 in Vershynin (2024), we have

$$\mathbb{P}\left(\left\|\frac{1}{N}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_2 > t + \left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},-j},T}\right\|_2\right) \leq 9^T \exp\left(-c\min\left(t,t^2\right)N\right),$$

where $\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},-j},T} = \mathbb{E}\left[\frac{1}{N}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right] = \mathcal{O}(1)$. This implies

$$\begin{aligned}\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{N}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_2 > t + \left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},-j},T}\right\|_2\right) &\leq N\cdot 9^T\exp\left(-c\min\left(t,t^2\right)N\right) \\ &\leq \exp\left(\log N + (\log 9)T - c\min\left(t,t^2\right)N\right).\end{aligned}$$

Let $\delta = \sqrt{\frac{T}{N}}$ and let $t = C\max\left(\delta,\delta^2\right)$ for some sufficiently large C such that $\log 9 - Cc < 0$, we have

$$\begin{aligned}\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{N}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_2 > C\max\left(\sqrt{\frac{T}{N}},\frac{T}{N}\right) + \left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},-j},T}\right\|_2\right) &\leq \exp\left(\log N + (\log 9 - Cc)T\right) \\ &= \mathbf{o}(1).\end{aligned} \tag{38}$$

It implies that, for some large enough $C > \left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},-j},T}\right\|_2$,

$$\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\right\|_F > C\left(\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{T}}\right)\right) \leq \exp\left(\log N + (\log 9 - Cc)T\right) = \mathbf{o}(1).$$

Back to the bound for the first term, pick $M_{NT} = C\left(\frac{1}{\sqrt{N}}+\frac{1}{\sqrt{T}}\right)$ and $t = C\log NM_{NT}$ for some large enough $C > \mathbb{E}\left[\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right]$, we have

$$\begin{aligned}&\mathbb{P}\left(\max_{j\leq N}\frac{1}{NT}\left|\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right| > C(\log NM_{NT}+1)\right) \\ &\leq N^{1-Cc} + \mathbf{o}(1) = \mathbf{o}(1).\end{aligned} \tag{39}$$

So, we have

$$\max_{j\leq N}\frac{1}{NT}\left|\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right|=\mathcal{O}_p\left(\frac{\log N}{\sqrt{N}}+\frac{\log N}{\sqrt{T}}+1\right)=\mathcal{O}_p\left(1\right).$$

Now bound the first term $\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2$, following a similar argument but with Hoeffding's inequality (Theorem 2.6.3 in Vershynin (2024)), we have

$$\begin{aligned}
&\mathbb{P}\left(\max_{j\leq N}\frac{1}{\sqrt{NT}}\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right|>t\right)\\
&\leq\exp\left(-c\frac{t^2}{M_{NT}^2}\right)\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{\sqrt{NT}}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2\leq M_{NT}\right)\\
&\quad+\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{\sqrt{NT}}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2>M_{NT}\right)
\end{aligned}$$

Note that $\left\|\frac{1}{\sqrt{NT}}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2\lesssim\left\|\frac{1}{\sqrt{NT}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2$ and by the analysis (39),

$$\sum_{j\leq N}\mathbb{P}\left(\frac{1}{\sqrt{NT}}\sqrt{\left|\mathbf{C}'_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right|}>C(\sqrt{\log NM_{NT}^2}+1)\right)=\mathbf{o}_p\left(1\right),$$

where $M_{NT}^2=C\left(\frac{1}{N}+\frac{1}{T}\right)$. So, let $t=C(\log NM_{NT}+1)$ for some large enough $C>0$, we have

$$\max_{j\leq N}\frac{1}{\sqrt{NT}}\left|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right|=\mathcal{O}_p\left(\sqrt{\log N}\right).$$

It follows that

$$\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{C}_{\mathbf{X},-j}\right\|_2=\mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}+\frac{\log N}{N\sqrt{N}}\right).$$

Lastly, we bound the third term of (ii) $\max_{j\leq N}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2$, observe that

$$\begin{aligned}
\max_{j\leq N}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2&\leq\max_{j\leq N}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\right\|_2\left\|\mathbf{U}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2\\
&\leq\max_{j\leq N}\frac{1}{\sqrt{T}}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}\right\|_2\left\|\mathbf{H}_{\mathbf{X}}\right\|_2
\end{aligned}$$

$$+ \max_{j\leq N} \frac{1}{T\sqrt{N}} \left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\right\|_2 \frac{1}{T\sqrt{N}} \left\|\mathbf{U}_{\mathbf{X}}\left(\widehat{\mathbf{G}}_{\mathbf{X}} - \mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\right\|_2$$

By Lemma H1 and H2, $\frac{1}{\sqrt{NT}}\left\|\mathbf{U}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}\right\|_2 = \mathcal{O}_p(1)$ and $\frac{1}{T\sqrt{N}}\left\|\mathbf{U}_{\mathbf{X}}\left(\widehat{\mathbf{G}}_{\mathbf{X}} - \mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)\right\|_2 = \mathcal{O}_p\left(\frac{1}{N} + \frac{1}{T}\right)$. So the rate is determined by $\max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\right\|_2$. Again, applying the leave-one-out analysis, we have

$$\max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X}}\right\|_2 \leq \max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X}}\right\|_2 + \max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X}}\right\|_2.$$

The second term is bounded by

$$\begin{aligned}\max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\left(\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right)\mathbf{U}_{\mathbf{X}}\right\|_2 &\leq \max_{j\leq N}\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},j}\right\|_2 \max_{j\leq N}\left\|\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2 \frac{1}{\sqrt{NT}}\left\|\mathbf{U}'_{\mathbf{X}}\right\|_2 \\ &= \mathcal{O}_p\left(\frac{\log N}{N\sqrt{N}} + \frac{\log N}{N\sqrt{T}}\right).\end{aligned}$$

Examinng the first term,

$$\max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}'_{\mathbf{X}}\right\|_2 \leq \max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},j}\right\|_2 + \max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\right\|_2.$$

By applying union bound and Theorem 6.2.1 in Vershynin (2024), we have

$$\max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},j}\right\|_2 = \mathcal{O}_p\left(\frac{\log N}{T\sqrt{N}}\right).$$

For the second term,

$$\begin{aligned}&\mathbb{P}\left(\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\right\|_2^2 > t\right) \\ &\leq \exp\left(-c\min\left(\frac{t^2}{M_{NT}}, \frac{t}{M_{NT}}\right)\right)\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_F \leq M_{NT}\right) \\ &\quad + \sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_F > M_{NT}\right).\end{aligned}$$

By Lemma K2 and the idempotent property of projection matrix, we have

$$\left\|\frac{1}{NT}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_F \lesssim \left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_2.$$

So, by (38), let $M_{NT} = C\left(\frac{1}{T}+\frac{1}{N}\right)$ for some large enough $C > \mathbb{E}\left[\left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},-j},T}\right\|_2\right]$, we have

$$\begin{aligned}&\sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_F > C\left(\frac{1}{T}+\frac{1}{N}\right)\right)\\ &\lesssim \sum_{j\leq N}\mathbb{P}\left(\left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X},-j}\mathbf{U}'_{\mathbf{X},-j}\right\|_2 > C\left(\frac{1}{T}+\frac{1}{N}\right)\right) = \mathbf{o}(1).\end{aligned}$$

Pick $t = C\log NM_{NT}$ for some large enough $C > 0$, we have

$$\max_{j\leq N}\frac{1}{NT}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\right\|_2^2 = \mathcal{O}_p\left(\frac{\log N}{N}+\frac{\log N}{T}\right),$$

which gives

$$\max_{j\leq N}\frac{1}{T\sqrt{N}}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\mathbf{U}_{\mathbf{X},-j}\right\|_2 = \mathcal{O}_p\left(\sqrt{\frac{\log N}{NT}}+\frac{\sqrt{\log N}}{T}\right).$$

Putting all pieces together, we have

$$\max_{j\leq N}\frac{1}{NT^2}\left\|\mathbf{U}'_{\mathbf{X},j}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}'_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}\widehat{\mathbf{G}}_{\mathbf{X}}\right\|_2 = \mathcal{O}_p\left(\frac{\sqrt{\log N}}{T\sqrt{N}}+\frac{\sqrt{\log N}}{T\sqrt{T}}\right).$$

Putting three terms for (ii) together, we have the desired rate. For (iii), the result follows by combining the argument for (i) and (ii) so the details are omitted here.

For (iv), it is sufficient to show that

$$\frac{1}{T}\mathbf{X}'_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_i = \mathbb{E}\left[\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}\right]+\mathbf{o}_p(1).$$

Decomposing the left hand side, we have

$$\frac{1}{T}\mathbf{X}'_i\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{X}_i = \frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i} - \frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},i} + \frac{2}{T}\mathbf{C}'_{\mathbf{X},i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},i} + \frac{1}{T}\mathbf{C}'_{\mathbf{X},i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X},i},$$

it is easy to verify that the first term $\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i} = \mathbb{E}\left[\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}\right]+\mathcal{O}_p\left(\frac{1}{\sqrt{T}}\right)$ so it suffices to show

the remaining three terms are $\mathbf{o}_p(1)$. Applying the leave-one-out analysis, we have

$$\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},i} = \frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i} + \frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\left(\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right)\mathbf{U}_{\mathbf{X},i}.$$

By Lemma I7, $\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\left(\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right)\mathbf{U}_{\mathbf{X},i} = \mathcal{O}_p\left(\frac{1}{N}\right)$. For the second term, by Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i} > t \mid \mathcal{F}_{-i}\right) &\leq \frac{\mathbb{E}\left[\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i} \mid \mathcal{F}_{-i}\right]}{t} \\
&\leq \frac{1}{T}\mathbb{E}\left[\operatorname{tr}\left(\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbb{E}\left[\mathbf{U}_{\mathbf{X},i}\mathbf{U}'_{\mathbf{X},i} \mid \mathcal{F}_{-i}\right]\right)\right]/t \\
&\leq \frac{1}{T}\mathbb{E}\left[\left\|\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\right\|_*\right]\left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},i},T}\right\|_2/t \\
&= \frac{r}{T}\left\|\mathbf{\Sigma}_{\mathbf{U}_{\mathbf{X},i},T}\right\|_2/t,
\end{aligned}$$

which implies

$$\frac{1}{T}\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{-(i)}}\mathbf{U}_{\mathbf{X},i} = \mathcal{O}_p\left(\frac{1}{T}\right).$$

The third term is bounded by

$$\begin{aligned}
\frac{1}{T}\left\|\mathbf{C}'_{\mathbf{X},i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},i}\right\|_2 &\leq \|\mathbf{C}_{\mathbf{X},i}\|_2\|\mathbf{H}_{\mathbf{X}}\|_2\left\|\frac{1}{T}\left(\widehat{\mathbf{G}}_{\mathbf{X}} - \mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)'\mathbf{U}_{\mathbf{X},i}\right\|_2 \\
&\quad + \|\mathbf{C}_{\mathbf{X},i}\|_2\|\mathbf{H}_{\mathbf{X}}\|_2\left\|\frac{1}{T}\left(\widehat{\mathbf{G}}_{\mathbf{X}} - \mathbf{G}_{\mathbf{X}}\mathbf{H}_{\mathbf{X}}\right)'\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},i}\right\|_2 \\
&= \mathcal{O}_p\left(\frac{1}{N} + \frac{1}{T}\right).
\end{aligned}$$

The fourth term:

$$\begin{aligned}
\frac{1}{T}\left\|\mathbf{C}'_{\mathbf{X},i}\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X},i}\right\|_2 &\leq \|\mathbf{C}_{\mathbf{X},i}\|_2^2\frac{1}{T}\left\|\mathbf{G}'_{\mathbf{X}}\mathbf{M}_{\widehat{\mathbf{G}}}\mathbf{G}_{\mathbf{X}}\right\|_2 \\
&= \mathcal{O}_p\left(\frac{1}{N} + \frac{1}{T^2}\right).
\end{aligned}$$

Putting all four terms together, we have the desired result. □

The following lemma provides the bounds for the terms involving $\mathbf{P}_{\widehat{\mathbf{G}}}$ in the proof of Theorem 4.2.

Lemma I6. Under the assumptions of Theorem 4.2,

(i) $\frac{1}{\sqrt{NT}}\sum_i \mathbb{U}_i \mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \mathcal{O}_p\left(\frac{1}{\sqrt{T}}\right)$.

Proof of Lemma I6. For (i), it is sufficient to show the rate of term replacing $\mathbb{U}_i$ with $\mathbf{U}_{\mathbf{X},i}$.

Observe that

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\sqrt{NT}}\sum_i \mathbf{U}_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\right]^2 &= \frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}}\mathbf{U}_{\mathbf{X},j}\right] \\
&= \frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right] \\
&\quad + \frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right] \\
&\quad + \frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\mathbf{U}_{\mathbf{X},j}\right] \\
&\quad + \frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\mathbf{U}_{\mathbf{X},j}\right].
\end{aligned}$$

For the first term, by the independence assumption in Theorem 4.2, it is easy to see that

$$\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right] = 0, \quad \text{for } i \neq j.$$

For $i = j$, conditional on $\mathcal{F}_{(-ij)}$,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},i} \mid \mathcal{F}_{(-ij)}\right] &\leq \|\boldsymbol{\Sigma}_{\varepsilon,T}\|_2\, \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},i} \mid \mathcal{F}_{(-ij)}\right] \\
&\leq \|\boldsymbol{\Sigma}_{\varepsilon,T}\|_2\, \|\boldsymbol{\Sigma}_{\mathbf{U}_\mathbf{X},T}\|_2\, \left\|\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\right\|_* \\
&= \mathcal{O}(1),
\end{aligned}$$

which gives

$$\frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right] = \frac{1}{NT}\sum_i \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},i}\right] = \mathcal{O}\left(\frac{1}{T}\right).$$

For the second term,

$$\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right]^2 \leq \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\boldsymbol{\varepsilon}_i\right]^2 \mathbb{E}\left[\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right]^2.$$

By the same argument,

$$\mathbb{E}\left[\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right]^2 = \mathcal{O}(1).$$

For the first component, observe that

$$\begin{aligned}\mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\mathbf{U}_{\mathbf{X},i}\right] &\leq \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\mathbf{U}_{\mathbf{X},i}\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\right)\right] \\ &\leq \|\boldsymbol{\Sigma}_{\varepsilon,T}\|_2\,\mathbb{E}\left[\operatorname{tr}\left((\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\mathbf{U}_{\mathbf{X},i}\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\right)\right] \\ &= \mathcal{O}_p\left(\frac{T}{N^2}\right),\end{aligned}$$

where the second inequality is due to the fact that $\widehat{\mathbf{G}}$ is calculated using only $\mathbf{X}$ and $\mathbf{W}$ and therefore independent of $\boldsymbol{\varepsilon}_i$ by assumptions. So

$$\frac{1}{NT}\sum_i\sum_j \mathbb{E}\left[\mathbf{U}'_{\mathbf{X},i}(\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}})\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_j\mathbf{P}_{\widehat{\mathbf{G}}^{(-ij)}}\mathbf{U}_{\mathbf{X},j}\right] = \mathcal{O}_p\left(\frac{1}{N\sqrt{T}}\right).$$

Following similar argument, the third term is $\mathcal{O}_p\left(\frac{1}{N\sqrt{T}}\right)$ and the fourth term is $\mathcal{O}_p\left(\frac{1}{N^2}\right)$. Putting all terms together and by Chebyshev's inequality, we have

$$\frac{1}{\sqrt{NT}}\sum_i \mathbb{U}_i\mathbf{P}_{\widehat{\mathbf{G}}}\boldsymbol{\varepsilon}_i = \mathcal{O}_p\left(\frac{1}{\sqrt{T}}\right).$$

□

The following lemma provides the bound for the distance between $\mathbf{P}_{\widehat{\mathbf{G}}}$ and leave-one-out projection matrix $\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}$, which is a key component in the proof of Lemma I1.

Lemma I7. Let $\widehat{\mathbf{G}}^{-(j)}$ denote the estimated factors by applying Algorithm 1 to data with the j-th unit removed. Under the assumptions of Lemma I1, we have

$$\begin{aligned}\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2 &= \mathcal{O}_p\left(\frac{1}{N}\right) \\ \max_{j\leq N}\left\|\mathbf{P}_{\widehat{\mathbf{G}}}-\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2 &= \mathcal{O}_p\left(\frac{\log N}{N}\right).\end{aligned}$$

The results hold trivially for leave-two-out projection matrix $\mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}$ for $j \neq i$.

Proof of Lemma I7. Observe that the ranks of $\mathbf{P}_{\widehat{\mathbf{G}}}$ and $\mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}$ are the same with probability approaching one. By applying Davis-Kahan $\sin\Theta$ Theorem twice, we have

$$\begin{aligned}
\left\|\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(j)}}\right\|_2 &= \left\|U(\widehat{\mathbf{G}})U(\widehat{\mathbf{G}})' - U(\widehat{\mathbf{G}}^{-(j)})U(\widehat{\mathbf{G}}^{-(j)})'\right\|_2 \\
&\leq \left\|\widehat{\mathbf{G}} - \widehat{\mathbf{G}}^{-(j)}\right\|_2 \\
&\lesssim \left\|\widehat{\mathbf{G}}_{\mathbf{X}} - \widehat{\mathbf{G}}_{\mathbf{X}}^{-(j)}\right\|_2 \\
&\lesssim \frac{\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(j)}\mathbf{X}^{-(j)'}\right\|_2}{\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) - \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(j)}\mathbf{X}^{-(j)'}\right)\right|} \\
&\leq \frac{\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(j)}\mathbf{X}^{-(j)'}\right\|_2}{\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) - \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(j)}\mathbf{X}^{-(j)'}\right)\right|} \\
&\leq \frac{1}{\lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right) - \mathbf{o}_p(1)}\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(j)}\mathbf{X}^{-(j)'}\right\|_2 \\
&= \mathcal{O}_p\left(\frac{1}{N}\right),
\end{aligned}$$

by Lemma I8 (a) and (b). The second result follows from the same argument and by Lemma I8 (c) and (d):

$$\begin{aligned}
\max_{j\leq N}\left\|\mathbf{P}_{\widehat{\mathbf{G}}} - \mathbf{P}_{\widehat{\mathbf{G}}^{-(ij)}}\right\|_2 &= \max_{j\leq N}\left\|U(\widehat{\mathbf{G}})U(\widehat{\mathbf{G}})' - U(\widehat{\mathbf{G}}^{-(ij)})U(\widehat{\mathbf{G}}^{-(ij)})'\right\|_2 \\
&\leq \max_{j\leq N}\left\|\widehat{\mathbf{G}} - \widehat{\mathbf{G}}^{-(ij)}\right\|_2 \\
&\lesssim \max_{j\leq N}\left\|\widehat{\mathbf{G}}_{\mathbf{X}} - \widehat{\mathbf{G}}_{\mathbf{X}}^{-(ij)}\right\|_2 \\
&\lesssim \max_{j\leq N}\frac{\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right\|_2}{\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) - \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right)\right|} \\
&\leq \frac{\max_{j\leq N}\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right\|_2}{\min_{j\leq N}\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) - \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right)\right|} \\
&\leq \frac{1}{\lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right) - \mathbf{o}_p(1)}\max_{j\leq N}\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(ij)}\mathbf{X}^{-(ij)'}\right\|_2 \\
&= \mathcal{O}_p\left(\frac{\log N}{N}\right).
\end{aligned}$$

□

The following are helper lemmas for the proof of Lemma I1.

Lemma I8. Under the assumptions of Lemma I1, we have

(a) $\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right\|_2 = \mathcal{O}_p\left(\frac{1}{N}\right)$;

(b) $\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) - \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right)\right| \geq \lambda_r\left(\frac{\mathbf{G}_\mathbf{X}'\mathbf{G}_\mathbf{X}}{T}\right)\lambda_r\left(\frac{\mathbf{C}_\mathbf{X}'\mathbf{C}_\mathbf{X}}{N}\right) - \mathbf{o}_p(1)$,

and

(c) $\max_{i\leq N}\left\|\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right\|_2 = \mathcal{O}_p\left(\frac{\log N}{N}\right)$;

(d) $\min_{i\leq N}\left|\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) - \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right)\right| \geq \lambda_r\left(\frac{\mathbf{G}_\mathbf{X}'\mathbf{G}_\mathbf{X}}{T}\right)\lambda_r\left(\frac{\mathbf{C}_\mathbf{X}'\mathbf{C}_\mathbf{X}}{N}\right) - \mathbf{o}_p(1)$,

for $r = r_\mathbf{X}$. Same results hold trivially for $\mathbf{W}_k$ for $k = 1, \ldots, K$.

Proof of Lemma I8. To show (a) and (c), observe that

$$\begin{aligned}
&\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'} \\
&= \frac{1}{NT}\left(\mathbf{X}_i\mathbf{X}_i' + \sum_{j\neq i}\mathbf{X}_j\mathbf{X}_j' - \sum_{j\neq i}\mathbf{X}_j\mathbf{X}_j'\right) \\
&= \frac{1}{NT}\mathbf{X}_i\mathbf{X}_i' \\
&= \frac{1}{NT}\mathbf{G}_\mathbf{X}\mathbf{C}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\mathbf{G}_\mathbf{X}' + \frac{1}{NT}\mathbf{G}_\mathbf{X}\mathbf{C}_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}' + \frac{1}{NT}\mathbf{U}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\mathbf{G}_\mathbf{X}' + \frac{1}{NT}\mathbf{U}_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}'.
\end{aligned}$$

It is easy to verify that $\frac{1}{N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}'\right\|_2 = \frac{1}{N}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},i}'\mathbf{U}_{\mathbf{X},i}\right| = \mathcal{O}_p\left(\frac{1}{N}\right)$ and $\frac{1}{N}\max_{i\leq N}\frac{1}{T}\left|\mathbf{U}_{\mathbf{X},i}'\mathbf{U}_{\mathbf{X},i}\right| = \mathcal{O}_p\left(\frac{1}{N}\right)$ by Lemma K1.

Also, note that $\frac{1}{NT}\left\|\mathbf{G}_\mathbf{X}\mathbf{C}_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}'\right\|_2 = \frac{1}{N}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_\mathbf{X}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},i}\right\|_2\left\|\mathbf{C}_{\mathbf{X},i}\right\|_2 = \mathcal{O}_p\left(\frac{1}{N}\right)$, and

$$\begin{aligned}
\max_{i\leq N}\frac{1}{NT}\left\|\mathbf{G}_\mathbf{X}\mathbf{C}_{\mathbf{X},i}\mathbf{U}_{\mathbf{X},i}'\right\|_2 &= \frac{1}{N}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_\mathbf{X}\right\|_2\max_{i\leq N}\frac{1}{\sqrt{T}}\left\|\mathbf{U}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\right\|_2 \\
&= \frac{1}{N}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_\mathbf{X}\right\|_2\sqrt{\max_{i\leq N}\frac{1}{T}\left\|\mathbf{U}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\right\|_2^2} \\
&= \mathcal{O}_p\left(\frac{1}{N}\right).
\end{aligned}$$

For the last term,

$$\begin{aligned}\frac{1}{NT}\left\|\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\mathbf{G}_{\mathbf{X}}'\right\|_2 &\leq \frac{1}{N}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2\left\|\mathbf{C}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\right\|_2\\ &= \frac{1}{N}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2\left|\mathbf{C}_{\mathbf{X},i}'\mathbf{C}_{\mathbf{X},i}\right|\\ &= \mathcal{O}_p\left(\frac{1}{N}\right),\end{aligned}$$

and

$$\begin{aligned}\frac{1}{NT}\max_{i\leq N}\left\|\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X},i}\mathbf{C}_{\mathbf{X},i}'\mathbf{G}_{\mathbf{X}}'\right\|_2 &\leq \frac{1}{N}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2\max_{i\leq N}\left|\mathbf{C}_{\mathbf{X},i}'\mathbf{C}_{\mathbf{X},i}\right|\\ &= \mathcal{O}_p\left(\frac{\log N}{N}\right).\end{aligned}$$

Putting pieces together gives the desired results.

To show (b) and (d), it is sufficient to show:

- $\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) \geq \lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right) - \mathbf{o}_p\left(1\right)$;
- $\lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right) = \mathbf{o}_p\left(1\right)$, and
- $\max_{i\leq N}\lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right) = \mathbf{o}_p\left(1\right)$.

To show the first claim, by Weyl's inequality,

$$\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) \geq \lambda_r\left(\frac{1}{NT}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\right) - \left\|\frac{1}{NT}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\right\|_2 - \left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}'\right\|_2 - \left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\right\|_2.$$

For the first term,

$$\begin{aligned}\lambda_r\left(\frac{1}{NT}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\right) &= \lambda_r\left(\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right)^{1/2}\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right)^{1/2}\right)\\ &\geq \lambda_r\left(\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right)^{1/2}\right)\lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right)^{1/2}\right)\\ &= \lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right),\end{aligned}$$

where the second inequality follows the fact that the rank of each matrix in the first equality is r, so its r-th largest eigenvalue is in fact the smallest eigenvalue, which is lower-bounded by the product of the smallest eigenvalues of the three matrices.

The second and third term share the same bound. By Lemma H1,

$$\frac{1}{NT}\left\|\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\right\|_2 \leq \frac{1}{\sqrt{N}}\frac{1}{\sqrt{T}}\left\|\mathbf{G}_{\mathbf{X}}\right\|_2 \frac{1}{\sqrt{NT}}\left\|\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\right\|_2 = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

The last term is $\mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right)$ by Lemma H1. Therefore, $\lambda_r\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) \geq \lambda_r\left(\mathbf{G}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}/T\right)\lambda_r\left(\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}/N\right) - \mathbf{o}_p(1)$.

To show the second claim, by Weyl's inequality,

$$\begin{aligned}\lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}^{-(i)}\mathbf{X}^{-(i)'}\right) &= \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}\mathbf{X}' - \frac{1}{NT}\mathbf{X}^{(i)}\mathbf{X}^{(i)'}\right) \\ &\leq \lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) + \left\|\frac{1}{NT}\mathbf{X}^{(i)}\mathbf{X}^{(i)'}\right\|_2.\end{aligned}$$

By (a), $\left\|\frac{1}{NT}\mathbf{X}^{(i)}\mathbf{X}^{(i)'}\right\|_2 = \mathbf{o}_p(1)$. Also, by Weyl's inequality again,

$$\begin{aligned}\lambda_{r+1}\left(\frac{1}{NT}\mathbf{X}\mathbf{X}'\right) &\leq \lambda_{r+1}\left(\frac{1}{NT}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{C}_{\mathbf{X}}\mathbf{G}_{\mathbf{X}}'\right) + \left\|\frac{1}{NT}\mathbf{G}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{U}_{\mathbf{X}}'\right\|_2 + \left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{C}_{\mathbf{X}}'\mathbf{G}_{\mathbf{X}}'\right\|_2 + \left\|\frac{1}{NT}\mathbf{U}_{\mathbf{X}}\mathbf{U}_{\mathbf{X}}'\right\|_2 \\ &\leq 0 + 2\cdot\mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right) + \mathcal{O}_p\left(\frac{1}{N}+\frac{1}{T}\right) = \mathbf{o}_p(1).\end{aligned}$$

So the second claim follows. The third claim can be shown by the same argument and by (c), which completes the proof. □

Appendix J: LASSO-related results

In this section, I present the LASSO-related lemmas and propositions that are used in the main proofs.

Consider a simple linear regression model with a high-dimensional parameter vector $\beta \in \mathbb{R}^p$:

$$Y = X'\beta + E, \quad Y, X, E \in \mathbb{R}^T, \tag{40}$$

and the LASSO estimator is defined as

$$\widehat{\beta} = \arg\min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{2T}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1\right\}, \tag{41}$$

where $\lambda > 0$ is the regularization parameter.

Given an index set S, let S^C be its complement and $\beta_S \in \mathbb{R}^{|S|}$ be the subvector of β indexed by S. Let s_0 be the number of non-zero elements in β, i.e. $s_0 = |\{j : \beta_j \neq 0\}|$. Let $\Sigma_X = \mathbb{E}[X'X/T]$ be the covariance matrix of the regressors X and $\widehat{\Sigma}_X = \frac{1}{T}X'X$ be the sample covariance matrix of the regressors.

The following proposition provides the estimation error bound for the LASSO estimator under the restricted eigenvalue condition.

Proposition J1. Given a set $S \subset \{1, \ldots, p\}$, define set $\mathcal{C}(S, 3) = \{\beta \in \mathbb{R}^p : \|\beta_{S^C}\|_1 \leq 3\|\beta_S\|_1\}$. Assume that $\|\beta\|_2^2 \leq C\beta'\Sigma_X\beta$ for some constant $C > 0$ and all $\beta \in \mathcal{C}(S, 3)$. Define two events:

$$\mathcal{E}_1 = \left\{ \left\| \frac{X'E}{T} \right\|_\infty \leq \frac{\lambda}{2} \right\},$$
$$\mathcal{E}_2 = \left\{ \left\| \widehat{\Sigma}_X - \Sigma_X \right\|_{max} \leq \frac{c}{s_0} \right\}, \quad \text{for some constant } c > 0.$$

Then, on the event $\mathcal{E}_1 \cap \mathcal{E}_2$, we have $\widehat{\beta}$ defined in (41) satisfies

$$\|\widehat{\beta} - \beta\|_2 \leq \frac{12(1 - 16Cc)}{C}\lambda\sqrt{s_0}.$$

The proof of this proposition relies on the following lemma:

Lemma J1. Assume that $\|\beta\|_2^2 \leq C\beta'\Sigma_X\beta$ for some constant $C > 0$ and all $\beta \in \mathcal{C}(S, 3)$, then under Event $\mathcal{E}_2$, we have

$$\|\beta\|_2^2 \leq \frac{C}{1 - 16Cc}\beta'\widehat{\Sigma}_X\beta$$

Proof of Lemma J1. This lemma is modified from Lemma 6.17 and Corollary 6.8 in Bühlmann and Van De Geer (2011). The proof is given here for completeness. Let $\beta \in \mathcal{C}(S, 3)$, then we have

$$\|\beta\|_1 = \|\beta_S\|_1 + \|\beta_{S^C}\|_1 \leq 4\|\beta_S\|_1 \leq 4\sqrt{s_0}\|\beta_S\|_2.$$

So, for c small enough such that $1 - 16Cc > 0$, we have,

$$\begin{aligned}
\left|\beta'\left(\widehat{\Sigma}_X - \Sigma_X\right)\beta\right| &\leq \left\|\widehat{\Sigma} - \Sigma\right\|_{\max} \|\beta\|_1^2 \\
&\leq \frac{c}{s_0} 16 s_0 \|\beta_S\|_2^2 && \text{by } \mathcal{E}_2 \\
&\leq 16c \|\beta\|_2^2 \\
&\leq 16Cc\beta'\Sigma_X\beta. && \text{by } \beta \in \mathcal{C}(S,3)
\end{aligned}$$

It follows that

$$(1 - 16Cc)\,\beta'\Sigma_X\beta \leq \beta'\widehat{\Sigma}_X\beta \leq (1 + 16Cc)\beta'\Sigma_X\beta.$$

By $\beta \in \mathcal{C}(S,3)$,

$$\frac{1}{C}(1 - 16Cc)\,\|\beta\|_2^2 \leq (1 - 16Cc)\beta'\Sigma_X\beta \leq \beta'\widehat{\Sigma}_X\beta.$$

We have

$$\|\beta\|_2^2 \leq \frac{C}{1 - 16Cc}\beta'\widehat{\Sigma}_X\beta.$$

□

Proof of Prop. J1. By the definition of $\widehat{\beta}$, we have

$$\frac{1}{2T}\|Y - X\widehat{\beta}\|_2^2 + \lambda\|\widehat{\beta}\|_1 \leq \frac{1}{2T}\|Y - X\beta\|_2^2 + \lambda\|\beta\|_1.$$

By rearranging the terms, we have

$$\frac{1}{2T}\left\|X\left(\widehat{\beta} - \beta\right)\right\|_2 \leq \frac{1}{T}E'X\left(\widehat{\beta} - \beta\right) + \lambda\left(\|\beta\|_1 - \left\|\widehat{\beta}\right\|_1\right).$$

Let S denote the support of β. Deonte $\Delta = \widehat{\beta} - \beta$, then we have

$$\|\beta\|_1 - \left\|\widehat{\beta}\right\|_1 \leq \|\beta_S\|_1 - \|\beta_S + \Delta_S\|_1 - \|\Delta_{S^C}\|_1 \leq \|\Delta_S\|_1 - \|\Delta_{S^C}\|_1,$$

where the last inequality is by reverse triangle inequality. So, we have

$$\begin{aligned}
\frac{1}{T}\|X\Delta\|_2^2 &\leq \frac{2}{T}E'X\Delta + 2\lambda\left(\|\Delta_S\|_1 - \|\Delta_{S^C}\|_1\right) \\
&\leq \left\|\frac{2}{T}X'E\right\|_\infty \|\Delta\|_1 + 2\lambda\left(\|\Delta_S\|_1 - \|\Delta_{S^C}\|_1\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \lambda \|\Delta\|_1 + 2\lambda \left(\|\Delta_S\|_1 - \|\Delta_{S^C}\|_1\right) && \text{by } \mathcal{E}_1 \\
&= \lambda \left(\|\Delta_S\|_1 + \|\Delta_{S^C}\|_1\right) + 2\lambda \left(\|\Delta_S\|_1 - \|\Delta_{S^C}\|_1\right) \\
&= \lambda \left(3\|\Delta_S\|_1 - \|\Delta_{S^C}\|_1\right) \\
&\leq 3\lambda \left(\|\Delta_S\|_1 - \frac{1}{3}\|\Delta_{S^C}\|_1\right) \\
&\leq 3\lambda \left(\|\Delta_S\|_1 + \|\Delta_{S^C}\|_1\right) = 3\lambda \|\Delta\|_1 .
\end{aligned}
$$

By the second equality, since LHS is non-negative, we have $\|\Delta_{S^C}\|_1 \leq 3\|\Delta_S\|_1$, so $\Delta \in \mathcal{C}(S,3)$. Under Event $\mathcal{E}_2$, by Lemma J1, we have

$$
\frac{C}{1-16Cc}\|\Delta\|_2^2 \leq \frac{1}{T}\|X\Delta\|_2^2 \leq 3\lambda\|\Delta\|_1 \leq 12\lambda\sqrt{s_0}\|\Delta\|_2 ,
$$

which implies

$$
\|\Delta\|_2 \leq \frac{12(1-16Cc)}{C}\lambda\sqrt{s_0}.
$$

□

Consider the adaptive LASSO estimator defined as

$$
\widehat{\beta}^{(al)} = \arg\min_{\beta\in\mathbb{R}^p} \left\{ \frac{1}{2T}\|Y - X\beta\|_2^2 + \lambda^{(al)} \sum_{j=1}^{p} w_j|\beta_j| \right\}, \tag{42}
$$

where $w_j > 0$ is the weight for the j-th coefficient. Define the model selection consistency notation:

$$
\widehat{\beta}^{(al)} =_s \beta \quad \text{if} \quad \text{sign}\left(\widehat{\beta}_j^{(al)}\right) = \text{sign}\left(\beta_j\right) \text{ for all } \beta_j \neq 0 \text{ and } \widehat{\beta}_j^{(al)} = 0 \text{ for all } \beta_j = 0.
$$

The following proposition is proposition 1 in Huang et al. (2008), which provides the events under which the model selection consistency of the adaptive LASSO estimator holds. Without the loss of generality, we assume that the first s_0 elements of β are non-zero and the rest are zero. Denote X_s the submatrix of X consisting of the first s_0 columns and X_{sC} the submatrix of X consisting of the last $p - s_0$ columns. Denote β_s the subvector of β consisting of the first s_0 elements and β_{sC} the subvector of β consisting of the last $p - s_0$ elements. Define

$$
\widehat{\Sigma}_s = \frac{1}{T}X_s'X_s, \quad \widehat{\Sigma}_{s^C,s} = \frac{1}{T}X_{s^C}'X_s,
$$

and

$$W_s = \text{diag}\,(w_1, \ldots, w_{s_0}), \quad \mathbf{w}_{s^C} = (w_{s_0+1}, \ldots, w_p)'.$$

Proposition J2. Under the event $A_T \cap B_T$, where

$$A_T = \left\{ 2T^{-1/2} \left| \widehat{\Sigma}_s^{-1} X_s' E \right| < 2\sqrt{T}\,|\beta_s| - T^{-1/2}\lambda^{(al)} \left| \widehat{\Sigma}_s^{-1} W_s \,\text{sign}\,(\beta_s) \right| \right\}$$
$$B_T = \left\{ 2T^{-1/2}\,|X_{s^C} M_{X_s} E| \leq T^{-1/2}\lambda^{(al)} \mathbf{w}_{s^C} - T^{-1/2}\lambda^{(al)} \left| \widehat{\Sigma}_{s^C,s} \widehat{\Sigma}_s^{-1} W_s \,\text{sign}\,(\beta_s) \right| \right\},$$

with $M_{X_s} = I - \frac{1}{T} X_s \widehat{\Sigma}_s^{-1} X_s'$, we have $\widehat{\beta}^{(al)} =_s \beta$.

The proof of this proposition directly follows from the proof of Proposition 1 in Zhao and Yu (2006) so the details are omitted.

Appendix K: Some useful lemmas

The following lemma is derived based on the Bernstein-type inequality for weakly dependent sequences in Merlevède et al. (2011).

Lemma K1. Suppose x_{it} is a mean-zero general sub-exponential such that

$$\mathbb{E}\left[\exp\left(c_1 |x_{it}|^{\gamma_1}\right)\right] \leq c_2 \quad \text{for some constants } c_1, c_2 > 0 \text{ and } \gamma_1 > 0,$$

and stationary α-mixing process with mixing coefficients $\alpha(k) \leq \exp(-Ck^{\gamma_2})$ for some constant $C_\alpha > 0$ and $\gamma_2 > 0$, for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$. Define $\gamma := \frac{1}{1/\gamma_1 + 1/\gamma_2}$. Assume that $\gamma < 1$ and $(\log N)^{2/\gamma - 1} (\log T)^{2/\gamma} = o(T)$. Then we have

$$\max_{1 \leq i \leq N} \left| \frac{1}{T} \sum_{t=1}^{T} x_{it} \right| = O_p\left(\sqrt{\frac{\log N}{T}} \right).$$

Proof of Lemma K1. By Bonferroni's inequality, we have

$$\mathbb{P}\left(\max_{i \leq N} \left| \frac{1}{T} \sum_{t=1}^{T} x_{it} \right| > \varepsilon \right) \leq \sum_{i=1}^{N} \mathbb{P}\left(\left| \frac{1}{T} \sum_{t=1}^{T} x_{it} \right| > \varepsilon \right).$$

By Theorem 1, Remark 1∼3 and Equation (1.11) in Merlevède et al. (2011), we have, for $\varepsilon \gtrsim$

$\log(T)^\eta/T$ for some $\eta > 0$,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} x_{it}\right| > \varepsilon\right) \lesssim (T+1)\exp\left(-(T\varepsilon)^\gamma\right) + \exp\left(-(T\varepsilon)^2\right).$$

So we have

$$\mathbb{P}\left(\max_{i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T} x_{it}\right| > \varepsilon\right) \lesssim N(T+1)\exp\left(-(T\varepsilon)^\gamma\right) + N\exp\left(-(T\varepsilon)^2\right).$$

Let the first term on RHS go to zero, we have $\varepsilon \gtrsim \frac{(\log NT)^{1/\gamma}}{T}$. Let the second term go to zero, we have $\varepsilon \gtrsim \sqrt{\frac{\log N}{T}}$. Since $(\log N)^{2/\gamma-1}(\log T)^{2/\gamma} = o(T)$, we have $\frac{(\log NT)^{1/\gamma}}{T} = o\left(\sqrt{\frac{\log N}{T}}\right)$. Therefore, we have $\varepsilon \gtrsim \sqrt{\frac{\log N}{T}}$. Also, it is easy to verify that $\frac{(\log T)^\eta}{T} = o\left(\sqrt{\frac{logN}{T}}\right)$ for any $\eta > 0$. Hence,

$$\mathbb{P}\left(\max_{i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T} x_{it}\right| > c\sqrt{\frac{\log N}{T}}\right) = o(1) \quad \text{for some constant } c > 0,$$

which implies that $\max_{i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T} x_{it}\right| = O_p\left(\sqrt{\frac{\log N}{T}}\right)$. □

The following lemma provides some useful bounds for matrix norms. The proofs of these bounds are omitted.

Lemma K2. For two matrices $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$,

(i) $\operatorname{tr}(AB) \leq \|A\|_2 \|B\|_*$;

(ii) $\|AB\|_F \leq \|A\|_2 \|B\|_F$;

(iii) $\|AB\|_\infty \leq \max_{1\leq i\leq m} \|A_{i:}\|_2 \|B\|_F$, where $A_{i:}$ is the i-th row of A;

The following lemma provides useful results for the derivation of the Gaussian approximation of the debiased LASSO estimator.

Lemma K3. Given two random variables X and Y, for any $a \in \mathbb{R}$ and $\Delta \in \mathbb{R}$, we have

(1) $\mathbb{P}(X \leq a) \leq \mathbb{P}(Y \leq a+\Delta) + \mathbb{P}(|X-Y| > \Delta)$;

(2) $\mathbb{P}(X \leq a) \geq \mathbb{P}(Y \leq a-\Delta) - \mathbb{P}(|X-Y| > \Delta)$;

(3) $\sup_{a\in\mathbb{R}} |\mathbb{P}(X \leq a) - \mathbb{P}(Y \leq a)| \leq \mathbb{P}(|X-Y| > \Delta) + \sup_{a\in\mathbb{R}} \mathbb{P}(a \leq Y \leq a+\Delta)$.

Proof of Lemma K3. To show (1), note that

$$\begin{aligned}\mathbb{P}(X \leq a) &= \mathbb{P}(X \leq a, |X - Y| \leq \Delta) + \mathbb{P}(x \leq a, |X - Y| > \Delta) \\ &\leq \mathbb{P}(X \leq a, |X - Y| \leq \Delta) + \mathbb{P}(|X - Y| > \Delta).\end{aligned}$$

Define two events: $A = \{Y \leq a + \Delta\}$ and $B = \{X \leq a, |X - Y| \leq \Delta\}$. if $\omega \in B$, then we have

$$|X - Y| \leq \Delta \Rightarrow Y \leq \Delta + X \Rightarrow Y \leq a + \Delta \Rightarrow \omega \in A.$$

Therefore, we have $\mathbb{P}(B) \leq \mathbb{P}(A)$, which implies (1).

The proof of (2) is similar. Transforming the inequality in (2), we have

$$\mathbb{P}(Y \leq a - \Delta) \leq \mathbb{P}(X \leq a) + \mathbb{P}(|X - Y| > \Delta).$$

Then (2) follows from (1) by replacing Y with X and a with $a - \Delta$.

To show (3), observe that, by (1),

$$\begin{aligned}\mathbb{P}(X \leq a) - \mathbb{P}(Y \leq a) &\leq \mathbb{P}(Y \leq a + \Delta) - \mathbb{P}(Y \leq a) + \mathbb{P}(|X - Y| > \Delta) \\ &= \mathbb{P}(a \leq Y \leq a + \Delta) + \mathbb{P}(|X - Y| > \Delta).\end{aligned}$$

Similarly, by (2), we have

$$\begin{aligned}\mathbb{P}(X \leq a) - \mathbb{P}(Y \leq a) &\geq \mathbb{P}(Y \leq a - \Delta) - \mathbb{P}(Y \leq a) - \mathbb{P}(|X - Y| > \Delta) \\ &= -\mathbb{P}(a - \Delta \leq Y \leq a) - \mathbb{P}(|X - Y| > \Delta).\end{aligned}$$

Putting the above two inequalities together yields the inequality in (3). □