

Forecasting Using Supervised Factors and Idiosyncratic Elements

Daanish Padha* and Tae-Hwy Lee†

October 1, 2026

Abstract

We extend the Three-Pass Regression Filter (3PRF) in two dimensions. First, we allow the factors driving the predictor panel to be weak. Second, we allow the idiosyncratic components of the predictors to contain additional predictive information for the forecasting target. We establish conditions under which 3PRF consistently recovers the target-relevant factor space under weak factor structures and characterize how its convergence rate depends on the relative strength of relevant and irrelevant factors. Stronger relevant factors improve the rate at which 3PRF approaches the infeasible factor forecast, whereas stronger irrelevant factors have the opposite effect. We use these results to compare 3PRF with Principal Component Regression (PCR) and identify settings in which either procedure has a relative advantage. To accommodate predictive information outside the common factor space, we augment 3PRF with a Least Absolute Shrinkage and Selection Operator (LASSO) regression on estimated idiosyncratic components, yielding the 3PRF LASSO estimator. We derive its prediction-error rate while accounting for the fact that the idiosyncratic regressors are generated from a first-stage factor estimator. Monte Carlo simulations support the asymptotic results and indicate strong finite-sample forecasting performance of 3PRF LASSO. In an empirical application involving forecasts of U.S. headline CPI inflation using FRED-QD data, 3PRF LASSO delivers competitive forecast performance.

Keywords: Weak Factors, Forecasting, High-dimensional data, Supervision, Three pass regression filter, LASSO.

JEL Classification: C18, C22, C53, C55, E27

*Economics, Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ, UK. E-mail: Daanish.Padha@glasgow.ac.uk

†Department of Economics, University of California, Riverside, CA 92521, USA. E-mail: tae.lee@ucr.edu

1 Introduction

Factor models are ubiquitous in the econometric analysis of high-dimensional data. Starting from the seminal work of [Forni *et al.* \[2000\]](#), [Stock & Watson \[2002\]](#), and [Bai \[2003\]](#), the utility of these models has been increasingly acknowledged in high-dimensional multivariate analysis. Notably, they have found extensive use in two key areas: high-dimensional covariance estimation and forecasting. This paper delves into the latter domain.

The literature on forecasting with factor models is extensive. Some prominent papers include [Ludvigson & Ng \[2016\]](#), who highlight their effectiveness in financial forecasting; [Stock & Watson \[2002\]](#) and [Stock & Watson \[2003\]](#), who demonstrate their importance in macroeconomic prediction, among others. The efficacy of these models is well-documented in the literature. [Stock & Watson \[2012\]](#) find that forecasts derived from factor models outperform those generated by several shrinkage-based techniques. [Kim & Swanson \[2018\]](#) corroborate these findings by demonstrating superior predictive performance of factor-augmented models compared to a wide array of machine learning methods.

The framework for forecasting with factor models is well-established. Consider a scenario where we have a large number of predictors organized in a vector, $\mathbf{x}_t$, and we aim to forecast a single target variable h -periods ahead, y_{t+h} . In this context, a standard factor-based forecasting model can be expressed as follows:

$$y_{t+h} = \beta_0 + \boldsymbol{\beta}' \mathbf{F}_t + u_{t+h}, \quad (1.1)$$

$$\mathbf{x}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi} \mathbf{F}_t + \boldsymbol{\varepsilon}_t. \quad (1.2)$$

The N -dimensional vector of covariates $\mathbf{x}_t$ is decomposed into three latent components: an intercept term $\boldsymbol{\phi}_0$, a low-rank component $\boldsymbol{\Phi} \mathbf{F}_t$, and a vector of idiosyncratic components $\boldsymbol{\varepsilon}_t$. The low-rank component captures the systematic variation across the covariates and is driven by the K -dimensional vector of latent factors $\mathbf{F}_t$, where K remains fixed asymptotically. The $N \times K$ matrix $\boldsymbol{\Phi}$ represents the temporally invariant matrix of factor loadings. These loadings quantify the influence of factors on the observed covariates. The vector of idiosyncratic components $\boldsymbol{\varepsilon}_t$, as the name suggests, represents variation unique to individual covariates, which is non-systematic and cannot be further decomposed into a lower-rank structure.

Since these factors are latent, they must be estimated using an appropriate method. The benchmark approach in the literature is the method of principal components, an unsupervised technique that derives factors exclusively from the predictor matrix $\mathbf{X}$. This method does not incorporate information about the target variable $\mathbf{y}$ during factor estimation, which limits its predictive utility. To address this limitation, [Kelly & Pruitt \[2015\]](#) introduced 3PRF, a supervised framework that leverages additional information from ‘proxy’

variables, denoted as $\mathbf{Z}$. These proxies, either pre-specified or constructed based on $\mathbf{y}$, enable the estimation of factors relevant to $\mathbf{y}$, which often constitute a strict subset of the factors driving $\mathbf{X}$. By aligning factor estimation with the goal of forecasting the target variable, 3PRF improves forecasting efficiency compared to unsupervised methods.

In formulating their theoretical framework, Kelly & Pruitt [2015] rely on foundational assumptions inspired by earlier works such as Stock & Watson [2002] and Bai & Ng [2006]. These assumptions facilitate the identification of latent factors and provide the basis for deriving the properties of their estimator. In this study, we revisit these assumptions and relax two critical ones, leveraging recent advancements in literature. Below, we outline these assumptions and discuss the rationale for their relaxation.

The first assumption pertains to the proportion of variation in the predictors that is explained by the factors, a concept commonly referred to as the “strength” of the factors. Like most earlier studies in the literature, Kelly & Pruitt [2015] assume that all factors driving the predictors are strong. Specifically, with reference to Equation 1.2, this assumption is formalized as $\frac{\Phi'\Phi}{N}$ converging to a non-zero limit, indicating that the factors explain a substantial share of the variance in the predictors. However, evidence from studies such as Bailey *et al.* [2021] and Freyaldenhoven [2022] indicates that this strong factors assumption often fails in practice. Factors may instead be weak, i.e., $\frac{\Phi'\Phi}{N^\psi}$, for some $\psi < 1$, may converge to a non-zero limit. Bailey *et al.* [2021] provide a framework for estimating factor strength and apply their method to key macroeconomic and financial datasets, demonstrating that the strong factor assumption is frequently violated. Weakness in factors can primarily be attributed to two reasons, or a combination thereof: (i) when factors influence only a subset of predictors, commonly referred to as local factors (see Freyaldenhoven [2022]), or (ii) when idiosyncratic variances are large (see Bai & Ng [2023] and references therein). In such scenarios, the theoretical properties of factor-based forecasting methods warrant closer examination.

Recent studies on the principal components method for factor estimation have sought to relax the strong factor assumption; see Bai & Ng [2023] and Freyaldenhoven [2022]. These papers primarily examine the implications of a weak factor structure on estimating the factors using the principal components method. In contrast, our focus is on evaluating the effect of weak factors on forecasting. We extend the theory of 3PRF to accommodate settings where predictors follow a weak factor structure. We allow target-relevant factors to have a different strength compared to target-irrelevant factors. Our theoretical results provide bounds on how weak the target-relevant factors can be. When developing the asymptotic theory under the assumption of a strong factor structure, it is sufficient for the sample size (T) and the number of predictors (N) to approach infinity, with no restriction on their relative growth rates. However, in the weak factor setting, the derivation of the

theoretical properties of 3PRF reveals that T must grow at a sufficiently fast rate to ensure consistency, a requirement absent under the strong factor assumption. This condition is formalized in Assumption 6 of the paper. Furthermore, we show that if irrelevant factors are too strong relative to relevant factors, the convergence rate of 3PRF is severely reduced, and beyond a specific limit, we encounter inconsistency.

We establish that 3PRF converges to the infeasible best forecast at a faster rate than PCR when the relevant factors are stronger than the irrelevant ones. This finding provides a rationale for 3PRF’s strong performance in many empirical settings, as this assumption is likely to hold for a wide range of economic target variables. Conversely, when all irrelevant factors are stronger than the relevant ones, 3PRF converges to the infeasible best forecast at a slower rate compared to PCR. In cases where some irrelevant factors are weaker and others stronger than the relevant ones, this comparison is complicated.

While Kelly & Pruitt [2015] demonstrate the consistency and establish convergence rates of 3PRF, they do not theoretically establish any superiority over PCR. Our contribution fills this gap by explicitly identifying cases where 3PRF has a clear advantage in terms of faster convergence rates.

The second assumption in Kelly & Pruitt [2015], which we address in this paper, pertains to the orthogonality of the idiosyncratic components to the target variable. In Equation 1.1, the forecast $\beta_0 + \beta'F_t$ is optimal only if $\mathbb{E}(u_{t+h} \mid \varepsilon_t) = 0$, as assumed by Kelly & Pruitt [2015]. However, in high-dimensional settings, it is unlikely that all idiosyncratic components ($\varepsilon_i \mid i \in \{1, \dots, N\}$) are uncorrelated with the target variable. If even a small number of these components are correlated with the target, the forecast constructed using Equation 1.1 becomes suboptimal, as it fails to fully exploit the predictive information contained in $\mathbf{X}$.

Allowing idiosyncratic components to correlate with the target variable reflects the reality of many economic forecasting scenarios, particularly when dealing with a large number of predictors. Fan *et al.* [2023] provide several examples demonstrating this. Beyhum & Striaukas [2024] introduce a framework for testing a dense model specification, specifically factor regression, against a hybrid model that combines dense and sparse components; a factor model augmented with sparse idiosyncratic components. They apply this test to various macroeconomic and financial datasets, frequently rejecting the null hypothesis of the dense factor regression specification. They note, “This suggests the presence of sparsity — alongside a dense component — in widely studied economic applications.”

Few studies have addressed this limitation by exploiting predictive information in the idiosyncratic components. Notable examples include Kneip & Sarda [2011], Fan *et al.* [2020], and Fan *et al.* [2023], which augment factor based models with a sparse subset of idiosyncratic components relevant to the target.

Motivated by this line of work, our paper extends the 3PRF framework by introducing an additional step to incorporate the predictive content of idiosyncratic components. Specifically, we employ LASSO (Tibshirani [1996]) to select relevant idiosyncratic components as additional predictors. We refer to this augmented estimator as 3PRF LASSO.

Augmenting a principal components based factor model to account for ignored idiosyncratic dependence is relatively straightforward, as the unsupervised nature of factor estimation via the principal components method ensures that the factor estimation process remains unaffected by the data-generating process of the target $\mathbf{y}$. However, in the 3PRF framework, allowing idiosyncratic elements to possess predictive content for the target $\mathbf{y}$, and thus for its proxies $\mathbf{Z}$, introduces a form of ‘corruption’ in the supervision process, as clarified in Section 2.¹ The principal components method is unaffected by the correlation between idiosyncratic components and the target, as neither the target nor its proxies are used during factor estimation. In contrast, within the 3PRF framework, where proxies are utilized for supervision, such correlations can potentially undermine the benefits of this supervised approach. The 3PRF methodology is designed to extract relevant factor loadings from the proxies to estimate factors pertinent to the target. However, when idiosyncratic dependence is present, this extraction process loses precision, capturing information unrelated to the factors but stemming from the correlation between the proxies and the idiosyncratic components. To address this issue, we outline assumptions in Section 3 that prevent this ‘corruption’ from adversely impacting the asymptotic convergence rates, ensuring that 3PRF retains its robustness even in the presence of idiosyncratic dependence.

In addition to relaxing these two assumptions within the 3PRF framework, it is noteworthy that, unlike the literature on PCR, which addresses the challenges posed by weak factors and idiosyncratic dependence independently without considering their combined impact, we extend the theoretical framework of 3PRF to account for both phenomena simultaneously. Specifically, we derive the asymptotic rate at which the regularization parameter in the proposed Stage 2 of 3PRF LASSO must approach zero, with this rate determined by the strength of the factors. Crucially, this rate governs the convergence rate of the LASSO step, as established in the LASSO literature. This analysis provides a deeper understanding of the interplay between the estimation of weak factors and the handling of idiosyncratic dependence, and how these interactions influence the overall performance of the model.

Our simulation results indicate that, under these general assumptions, the performance of 3PRF frequently surpasses that of its closest competitor, PCR. The augmented method, 3PRF LASSO, which combines 3PRF with a LASSO regression involving the idiosyncratic components, consistently outperforms both 3PRF and PCR. In many cases, it also exceeds

¹ $\mathbf{Z}$ mimics the data generating process (DGP) of $\mathbf{y}$ in that it depends on the same set of factors and idiosyncratic components as $\mathbf{y}$. This is clarified in Assumptions 1 and 9.

the performance of PCR augmented with a LASSO step. When N is large relative to T , this advantage becomes almost universal, extending across various factor strengths and serial/cross-sectional correlations in factors and idiosyncratic components. Furthermore, 3PRF LASSO frequently outperforms LASSO when relevant factors are relatively strong compared to irrelevant factors.

Our empirical application evaluates the 3PRF LASSO approach by forecasting U.S. headline CPI inflation using the FRED-QD dataset by [Clark & McCracken \[2023\]](#) across several forecast horizons and evaluation samples. The 3PRF LASSO method demonstrates competitive performance compared to established methods.

The paper is structured as follows. Section 2 introduces the proposed estimator, detailing its formulation and operational mechanics. In Section 3, we outline a series of assumptions necessary to establish the theoretical results presented in Section 4. This theoretical framework is then put to the test in Section 5, where we explore the numerical properties of our estimator through comprehensive Monte Carlo simulations. Section 6 focuses on empirical applications, demonstrating the estimator's performance with real-world data. Finally, Section 7 summarizes the contributions and offers concluding remarks.

1.1 Definitions and Notations

Let $\mathbf{y}$ denote the $T \times 1$ vector of the target variable, i.e., $\mathbf{y} = (y_{1+h}, \dots, y_{T+h})$. We have N predictors, each with T observations. The cross-section of predictors at time t is given by the $N \times 1$ vector $\mathbf{x}_t$. The temporal observations of predictor i form the $T \times 1$ vector $\mathbf{x}_i$. The predictors are stacked in a $T \times N$ matrix $\mathbf{X}$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)' = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$. We have L proxies stacked in a $T \times L$ matrix $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_T)'$.

Define $\mathbf{J}_T \equiv \mathbb{I}_T - \frac{1}{T} \iota_T \iota_T'$, where $\mathbb{I}_T$ is the $T \times T$ identity matrix and ι_T is the $T \times 1$ vector of ones. $\mathbf{J}_N$ is defined analogously. For matrices $\mathbf{U}$ and $\mathbf{V}$ of conformable dimensions, let $\mathbf{W}_{UV} \equiv \mathbf{J}_N \mathbf{U}' \mathbf{J}_T \mathbf{V}$ and $\mathbf{S}_{UV} \equiv \mathbf{U}' \mathbf{J}_T \mathbf{V}$.

Given an index set $S \subset \{1, \dots, N\}$ and a vector $\mathcal{X}$ with i -th component $\mathcal{X}_i$, define $\mathcal{X}_{i,S} = \mathcal{X}_i \mathbb{1}\{i \in S\}$, where $\mathbb{1}$ is the indicator function. For a set A , $|A|$ denotes its cardinality. For a vector $\mathbf{v}$, $v(m)$ denotes its m -th component. The norms we use in the paper are:

$$\|\mathbf{v}\|_1 = \sum_i |v_i|, \quad \|\mathbf{v}\|_2 = \left(\sum_i v_i^2 \right)^{1/2}, \quad \|\mathbf{v}\|_\infty = \max_i |v_i|.$$

For an $m \times n$ matrix $A = [a_{ij}]$, the following norms are used:

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|, \quad \|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|, \quad \|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 \right)^{1/2}.$$

Stochastic orders are denoted by O_p and o_p , while deterministic orders are O and o . We say $a \asymp b$, if $a = O(b)$ and $b = O(a)$. For matrices, $\mathbf{O}_p$ and $\mathbf{o}_p$ denote element-wise stochastic orders. A matrix $\mathbf{A}$ is $\mathbf{O}_p(1)$ if all elements are $O_p(1)$, and $\mathbf{o}_p(1)$ if all elements are $o_p(1)$. The notation $O_p(a \vee b)$ denotes $O_p(\max(a, b))$ and $O_p(a \wedge b)$ denotes $O_p(\min(a, b))$. $\max_i \{\mathbf{A}_i\}_{i \in \{1, \dots, N\}}$ denotes the element-wise maximum of matrices $\{\mathbf{A}_i\}_{i \in \{1, \dots, N\}}$. The abbreviation ‘w.p.’ stands for ‘with probability’ and ‘w.r.t.’ stands for with respect to. For a real number x , $\lfloor x \rfloor$ denotes the floor of x , defined as the greatest integer less than or equal to x .

2 The 3PRF LASSO Estimator

We predict the target $\mathbf{y}$ using a two-stage process which we call 3PRF LASSO. Stage 1 of this process is the 3PRF procedure by [Kelly & Pruitt \[2015\]](#). 3PRF is essentially a sequence of linear regressions aimed at consolidating information from a large set of predictors in a small set of factors. The procedure relies on a set of ‘proxies’ $\mathbf{Z}$, which, as in [Kelly & Pruitt \[2015\]](#), are required to be driven by target-relevant factors while remaining unaffected by target-irrelevant factors. This requirement, explained in greater detail in the next section, is a crucial element in identifying the target-relevant factors. Once we obtain the target relevant factors from Stage 1, we regress each predictor $\mathbf{x}_i$ on them and estimate the residuals. Thereafter, we perform a LASSO regression to extract any predictive content in these residuals for our target $\mathbf{y}$. Detailed procedure is outlined below.

Algorithm 1 : 3PRF LASSO Procedure

Stage 1 (Three Pass regression Filter, 3PRF)	
Pass	Description
1.	Run time series regression of $\mathbf{x}_i$ on $\mathbf{Z}$ for $i = 1, \dots, N$, $x_{i,t} = \tilde{\phi}_{0,i} + \tilde{\phi}_i' \mathbf{z}_t + \hat{v}_{it}$, retain slope estimate $\tilde{\phi}_i$.
2.	Run cross-section regression of $\mathbf{x}_t$ on $\tilde{\phi}_i$ for $t = 1, \dots, T$, $x_{i,t} = \tilde{\phi}_{0,t} + \tilde{\phi}_i' \hat{\mathbf{F}}_t + \tilde{\varepsilon}_{it}$, retain slope estimate $\hat{\mathbf{F}}_t$.
3.	Run time series regression of y_{t+h} on predictive factors $\hat{\mathbf{F}}_t$, $y_{t+h} = \hat{\beta}_0 + \hat{\beta}' \hat{\mathbf{F}}_t + \hat{u}_{t+h}$, delivers initial forecast $\hat{y}_{t+h,f} = \hat{\beta}_0 + \hat{\mathbf{F}}_t' \hat{\beta}$. Retain the residual $\hat{u}_{t+h}$ and the Stage 1 forecast $\hat{y}_{t+h,f}$.
Stage 2 (Three-Pass Regression Filter augmented with LASSO, 3PRF LASSO)	
Pass	Description
4.	Run time series regression of $\mathbf{x}_i$ on $\hat{\mathbf{F}}$ for $i = 1, \dots, N$, $x_{i,t} = \hat{\phi}_{0,i} + \hat{\phi}_i' \hat{\mathbf{F}}_t + \hat{\varepsilon}_{it}$, retain the residual $\hat{\varepsilon}_{it}$
5.	Run LASSO regression of $\hat{u}_{t+h}$ obtained from Pass 3 in Stage 1 on the estimated residuals $\hat{\varepsilon}_{it}$, $\hat{u}_{t+h} = \hat{\gamma}' \hat{\varepsilon}_t + \hat{\eta}_{t+h}$.
The final forecast is given by	
6.	$\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\beta}' \hat{\mathbf{F}}_t + \hat{\gamma}' \hat{\varepsilon}_t$

From Stage 1 and Stage 2, the final forecast is obtained in Pass 6. We can rewrite the final forecast as

$$\hat{\mathbf{y}} = \underbrace{\iota_T \bar{y} + \mathbf{J}_T \hat{\mathbf{F}} \hat{\beta}}_{\hat{\mathbf{y}}_f} + \underbrace{\hat{\varepsilon} \hat{\gamma}}_{\hat{\mathbf{y}}_\varepsilon},$$

where the Stage 1 forecast is given by

$$\begin{aligned} \hat{\mathbf{y}}_f &= \iota_T \bar{y} + \mathbf{J}_T \hat{\mathbf{F}} \hat{\beta} \\ &= \iota_T \bar{y} + \mathbf{J}_T \mathbf{X} \mathbf{W}_{XZ} (\mathbf{W}_{XZ}' \mathbf{S}_{XX} \mathbf{W}_{XZ})^{-1} \mathbf{W}_{XZ}' \mathbf{S}_{Xy}. \end{aligned}$$

The estimated factors are given by

$$\hat{\mathbf{F}}' = \mathbf{S}_{ZZ} (\mathbf{W}_{XZ}' \mathbf{S}_{XZ})^{-1} \mathbf{W}_{XZ}' \mathbf{X}',$$

and the estimated coefficients of the factors are given by

$$\hat{\beta} = \mathbf{S}_{ZZ} \mathbf{W}_{XZ} \mathbf{S}_{XZ} (\mathbf{W}_{XZ}' \mathbf{S}_{XX} \mathbf{W}_{XZ})^{-1} \mathbf{W}_{XZ}' \mathbf{S}_{Xy}.$$

Alternatively, we can rewrite the Stage 1 forecast as

$$\hat{\mathbf{y}}_f = \nu \bar{\mathbf{y}} + \mathbf{J}_T \mathbf{X} \hat{\boldsymbol{\alpha}},$$

where $\hat{\boldsymbol{\alpha}}$ is the implied predictive coefficient for $\mathbf{X}$ and is given by

$$\hat{\boldsymbol{\alpha}} = \mathbf{W}_{XZ} (\mathbf{W}'_{XZ} \mathbf{S}_{XX} \mathbf{W}'_{XZ})^{-1} \mathbf{W}'_{XZ} \mathbf{S}_{Xy}.$$

The procedure described above relies on the availability of suitable proxies, which can be obtained through relationships established in economic theory or constructed using the target variable in a sequential manner. Proxies constructed using $\mathbf{y}$ are referred to as automatic proxies (auto-proxies for short). Kelly & Pruitt [2015] explains how such auto-proxies can always be constructed. The process to obtain L proxies is laid out below. Theorem 7 of Kelly & Pruitt [2015] proves that such proxies are valid; in the sense that they adhere to the assumptions of the model outlined in Section 3.

Algorithm 2: Auto-Proxy Algorithm

-
-
0. Initialize $\mathbf{r}_0 = \mathbf{y}$. For $k = 1, \dots, L$ (where L is the total number of proxies):
 1. Define the k^{th} automatic proxy to be $\mathbf{r}_{k-1}$. Stop if $k = L$; otherwise proceed.
 2. Compute 3PRF for target $\mathbf{y}$ using cross-section $\mathbf{X}$ and statistical proxies 1 through k . Denote the resulting forecast $\hat{\mathbf{y}}_k$.
 3. Calculate $\mathbf{r}_k = \mathbf{y} - \hat{\mathbf{y}}_k$, advance k , and go to step 1.
-
-

To understand the functioning of this three-pass procedure, it is instructive to look at the DGP for the proxies.

$$\mathbf{Z} = \nu_T \boldsymbol{\lambda}'_0 + \mathbf{F} \boldsymbol{\Lambda}' + \boldsymbol{\varepsilon} \boldsymbol{\zeta}' + \boldsymbol{\omega}. \quad (2.1)$$

Kelly & Pruitt [2015] provide a detailed explanation of how the supervision process operates. Pass 1 of 3PRF constitutes the supervision step, where the relevant factor loadings across predictors are estimated, up to a rotation, while those associated with irrelevant factors are filtered out. Kelly & Pruitt [2015] mention that “Fluctuations in the latent factors cause the cross-section of predictors to fan out and compress over time. First-pass coefficient estimates map the cross-sectional distribution of predictors to the latent factors.” This statement holds true only if the composite error term $\boldsymbol{\varepsilon} \boldsymbol{\zeta}' + \boldsymbol{\omega}$ is uncorrelated with the idiosyncratic components of the predictors. This is trivially true in the framework of Kelly & Pruitt [2015], since they assume $\boldsymbol{\zeta} = \mathbf{0}$ and $\boldsymbol{\omega}$ is uncorrelated with $\boldsymbol{\varepsilon}$. When $\boldsymbol{\zeta} \neq \mathbf{0}$, we encounter what we refer to as the ‘corrupted’ supervisor problem; the supervisor is imperfect

in the sense that it cannot estimate some of the loadings, upto a rotation, consistently in Pass 1 of 3PRF. To illustrate this, consider a simplified case with only one factor. It can be easily verified that when $\zeta = \mathbf{0}$ and ω is uncorrelated with ε , then, through Pass 1 of Stage 1, we obtain $\tilde{\phi}_i = c\phi_i + O_p(T^{-1/2})$, where c is a constant not dependent on i . This implies that Pass 1 of 3PRF, in this setting, can estimate all loadings up to a constant of proportionality. However, this convenient feature is lost when $\zeta \neq 0$, since in that case, for all j such that $\zeta_j \neq 0$, we would have $\tilde{\phi}_j = c\phi_j + d_j + O_p(T^{-1/2})$. This d_j term is $O_p(1)$ and arises from the correlation between $\{\varepsilon_j | \zeta_j \neq 0\}$ in the data generating process of $\mathbf{Z}$ and predictors in the set $\Delta_j = \{\mathbf{x}_i | \varepsilon_i \text{ is correlated with } \varepsilon_j\}$. Through a set of assumptions, we restrict the extent of this corruption. Once Pass 1 is ensured to function adequately, that is, the extent of ‘corruption’ is negligible asymptotically, Kelly & Pruitt [2015] explain that “second-pass cross-sectional regressions use the estimated mapping in Pass 1 to back out estimates of the factors at each point in time.”, enabling consistent estimation in Stage 1 of 3PRF. Stage 2 simply proceeds by using consistent estimates of the factors in Stage 1.

Remark 1. *One practical issue is choosing the number of factors when we are using the Auto-Proxy algorithm. Kelly & Pruitt [2015] adopt a method initially presented by Krämer & Sugiyama [2011] to calculate the number of factors. We do not delve into the question of estimating the number of factors in this paper. One may use an information criterion as mentioned or divide the data into a training and validation set and estimate the number of relevant factors using a cross validation technique. Using a single 3PRF factor is a prudent choice as highlighted in Appendix 7.2 of Kelly & Pruitt [2015]. They demonstrate that there are situations where the DGP of $\mathbf{y}$, which involves multiple relevant factors, can be reformulated as a DGP with a single relevant factor. Moreover, in cases where a single-factor representation is not feasible, the variation in the target explained by the first estimated factor typically far exceeds that explained by the factors estimated subsequently, as demonstrated in Appendix 7.3 of Kelly & Pruitt [2015]. This is due to the fact that 3PRF estimates a rotation of underlying factors, with the first estimated factor explaining the maximal variation of the target.*

3 Setup

Below, we delineate our data generating process and the associated assumptions.

Assumption 1. *(Data Generating process). The data are generated as follows:*

$$\begin{aligned} x_t &= \phi_0 + \Phi \mathbf{F}_t + \varepsilon_t, & y_{t+h} &= \beta_0 + \beta' \mathbf{F}_t + \gamma' \varepsilon_t + \eta_{t+h}, & z_t &= \lambda_0 + \Lambda \mathbf{F}_t + \zeta \varepsilon_t + \omega_t, \\ \mathbf{X} &= \iota_T \phi'_0 + \mathbf{F} \Phi' + \varepsilon, & \mathbf{y} &= \iota_T \beta_0 + \mathbf{F} \beta + \varepsilon \gamma + \eta, & \mathbf{Z} &= \iota_T \lambda'_0 + \mathbf{F} \Lambda' + \varepsilon \zeta' + \omega, \end{aligned}$$

where $\mathbf{F}_t = (\mathbf{f}'_t, \mathbf{g}'_t)'$, $\mathbf{\Phi} = (\mathbf{\Phi}_f, \mathbf{\Phi}_g)$, $\mathbf{\Lambda} = (\mathbf{\Lambda}_f, \mathbf{\Lambda}_g)$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_f, \mathbf{0}')'$ with $\beta_f(m) \neq 0$ for all $m = 1, \dots, K_f$. $K_f > 0$ is the dimension of vector $\mathbf{f}_t$, $K_g \geq 0$ is the dimension of vector $\mathbf{g}_t$, $L > 0$ is the dimension of vector $\mathbf{z}_t$, and $K = K_f + K_g$. Furthermore, for all $m = 1, \dots, K$ and all $j \in \{i | \gamma_i \neq 0\}$, $\beta(m) = 0$ implies $\phi_j(m) = 0$.

Assumption 1 characterizes the factor structure of the predictors and the data-generating process for both the target and proxies. The target is driven by a subset of factors that drive variation in the predictors. In addition, we allow the target to be correlated with the idiosyncratic components, a modification from Kelly & Pruitt [2015]. In the usual framework, factors act as a convenient conduit relating $\mathbf{X}$ to $\mathbf{y}$. This involves an implicit assumption that $\mathbf{X}$ has no explanatory power for the target after accounting for the latent factors, which may be unrealistic in various settings. The proxies are driven by factors and idiosyncratic components.

The idea of allowing the predictors to retain explanatory power for the target after accounting for latent factors has been explored in other studies as well. Examples include Kneip & Sarda [2011], Kapetanios & Marcellino [2010], and Fan *et al.* [2024]. The latter two papers assume a DGP for the target, which takes the following form,

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\nu}_T \boldsymbol{\beta}_0^* + \mathbf{F} \boldsymbol{\beta}^* + \mathbf{X} \boldsymbol{\gamma}^* + \boldsymbol{\eta} \\ &= \boldsymbol{\nu}_T (\boldsymbol{\beta}_0^* + \boldsymbol{\phi}'_0 \boldsymbol{\gamma}^*) + \mathbf{F} (\boldsymbol{\beta}^* + \boldsymbol{\phi}' \boldsymbol{\gamma}^*) + \boldsymbol{\varepsilon} \boldsymbol{\gamma}^* + \boldsymbol{\eta}. \end{aligned}$$

Comparing it with the DGP of $\mathbf{y}$ given in Assumption 1, one can clearly see that, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ and $\boldsymbol{\beta}(m) = \boldsymbol{\beta}^*(m) + \sum_i \phi_i(m) \gamma_i^* = 0$. We assume that $\boldsymbol{\beta}^*(m) + \sum_i \phi_i(m) \gamma_i^* = 0$ only if $\boldsymbol{\beta}^*(m) = 0$ and for all $j \in \{i | \gamma_i^* \neq 0\}$ we have $\phi_j(m) = 0$. We are ruling out the pathological cases where both these aforementioned quantities are not zero but the sum $\boldsymbol{\beta}^*(m) + \sum_i \phi_i(m) \gamma_i^* = 0$. This assumption is succinctly expressed by stating that for all $j \in \{i | \gamma_i \neq 0\}$, $\beta(m) = 0$ implies $\phi_j(m) = 0$. This assumption allows us to consistently recover the true idiosyncratic components for the relevant $\mathbf{x}_i$ (i.e., $\{\mathbf{x}_i | \gamma_i \neq 0\}$) in Stage 2 Pass 4 and subsequently implement Pass 5 in Stage 2.

Assumption 2. (*Factors, Loadings and Residuals*). Let $M < \infty$. For any i, s, t , $0 < \psi_f \leq 1$ and $0 < \psi_g \leq 1$,

1. $\mathbb{E} \|\mathbf{F}_t\|^4 < M, T^{-1} \sum_{s=1}^T \mathbf{F}_s \xrightarrow[T \rightarrow \infty]{p} \boldsymbol{\mu}$ and $T^{1/2} \left(\frac{\mathbf{F}' \mathbf{J}_T \mathbf{F}}{T} - \boldsymbol{\Delta}_F \right) = O_p(1)$.
2. $\mathbb{E} \|\phi_i\|^4 \leq M$. For $v = f, g$, $N^{-\psi_v} \sum_{j=1}^N \phi_{vj} \xrightarrow[N \rightarrow \infty]{p} \bar{\phi}_v < \infty$, $N^{\psi_v/2} \left(\frac{\mathbf{\Phi}_v' \mathbf{J}_N \mathbf{\Phi}_v}{N^{\psi_v}} - \mathcal{P}_v \right) = O_p(1)$ and for $\psi_s = \min(\psi_f, \psi_g)$, $N^{\psi_s/2} \left(\frac{\mathbf{\Phi}_f' \mathbf{J}_N \mathbf{\Phi}_g}{N^{\psi_s}} - \mathcal{P}_{fg} \right) = O_p(1)$.

3. $\mathbb{E}(\varepsilon_{it}) = 0, \mathbb{E} \|\varepsilon_{it}\|^8 \leq M$.
4. $\mathbb{E}(\omega_t) = \mathbf{0}, \mathbb{E} \|\omega_t\|^4 \leq M, T^{-1/2} \sum_{s=1}^T \omega_s = \mathcal{O}_p(1)$ and $T^{1/2} \left(\frac{\omega' \mathbf{J}_T \omega}{T} - \Delta_\omega \right) = \mathcal{O}_p(1)$.
5. Let $\mathcal{I}_t$ denoting the information set containing all information available through time t . For any $h > 0$, $\mathbb{E}_t(\eta_{t+h}) = \mathbb{E}(\eta_{t+h} | \mathcal{I}_t) = 0$, $\mathbb{E}(\eta_{t+h}^2) = \delta_\eta < \infty$, and η_{t+h} is independent of the collection $\{\mathbf{F}_t, \omega_t, \phi_i, \varepsilon_{i,t}\}$.

The vector $\boldsymbol{\mu}$ is non-stochastic. The vectors $\bar{\phi}_v$ for $v = f, g$ are non-stochastic. The matrix Δ_F , the matrices $\mathcal{P}_v$ for $v = f, g$, and $\mathcal{P}_{fg}$ are non-stochastic and are further characterized in Assumption 4. Similarly, the matrix Δ_ω is non-stochastic.

If $\psi_f = \psi_g = 1$ in Assumption 2.2, this corresponds to the strong factor assumption. Combined with Assumptions 2.1, 2.3, and 2.5, it characterizes the typical structure of forecasting models based on strong factors. These assumptions are standard in the literature; see Stock & Watson [2002] and Bai & Ng [2006] in the context of PCR and Kelly & Pruitt [2015] in the context of 3PRF.

We allow for weak factors in the 3PRF framework by relaxing the strong factor assumption, considering a broader range of factor strengths with $0 < \psi_f \leq 1$ and $0 < \psi_g \leq 1$, where ψ_f and ψ_g may differ. Similar approaches have been explored in the context of factor estimation using the principal components method, as in Freyaldenhoven [2022] and Bai & Ng [2023].

Assumption 2.4 is similar to the assumption in Kelly & Pruitt [2015] which ensures that the proxy noise is well-behaved. The fact that conditional expectation of η_{t+h} with respect to information set in time t is zero implies that $\beta_0 + \beta'_f \mathbf{f}_t + \gamma' \varepsilon_t$ provides the optimal forecast of the target at time t . However, this forecast is infeasible as the factors and idiosyncratic components are not known.

Assumption 3. (Dependence). For $M < \infty$ and any i, j, t, s, m, m_1, m_2 and $v = f, g$

1. Let $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$, $|\sigma_{ij,ts}| \leq \tau_{ts}$, and
 - (a) $N^{-1} \sum_{i,j=1}^N \bar{\sigma}_{ij} \leq M$,
 - (b) $T^{-1} \sum_{t,s=1}^T \tau_{ts} \leq M$,
 - (c) $N^{-1} \sum_{i,s} |\sigma_{ii,ts}| \leq M$,
 - (d) $T^{-1} \sum_{i,t} |\sigma_{ij,tt}| \leq M$,
 - (e) $N^{-1} T^{-1} \sum_{i,j,t,s} |\sigma_{ij,ts}| \leq M$.
2. (a) $\mathbb{E} \left| N^{-1/2} T^{-1/2} \sum_{s=1}^T \sum_{i=1}^N [\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}] \right|^4 \leq M$,
- (b) $\mathbb{E} \left| N^{-1/2} T^{-1/2} \sum_{s=1}^T \sum_{i=1}^N [\varepsilon_{is}\varepsilon_{js} - \sigma_{ij,ss}] \right|^4 \leq M$.

3. $\mathbb{E} \left| N^{-\psi_v/2} T^{-1/2} \sum_{t=1}^T \sum_{i=1}^N \phi_{iv}(m) [\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij,tt}] \right|^2 \leq M.$ ²
4. $\mathbb{E} \left| T^{-1/2} \sum_{t=1}^T F_t(m_1) \omega_t(m_2) \right|^2 \leq M.$
5. $\mathbb{E} \left| T^{-1/2} \sum_{t=1}^T \omega_t(m) \varepsilon_{it} \right|^2 \leq M.$
6. $\mathbb{E} \left| T^{-1/2} \sum_{t=1}^T F_t(m) \varepsilon_{it} \right|^2 \leq M.$
7. $\mathbb{E} \left| N^{-\psi_v/2} \sum_{i=1}^N \phi_{iv}(m_1) \varepsilon_{it} \right|^2 \leq M.$

Assumption 3.1-3.2 allow weak cross-sectional and temporal dependence in the idiosyncratic components; this set of assumptions characterize an approximate factor model (Chamberlain & Rothschild [1983]). Essentially, we require that the idiosyncratic components lack an underlying factor structure, as the presence of such a structure would render the true factor space unidentifiable. Assumptions 3.1-3.4 with $\psi_f = \psi_g = 1$ are common in literature, see Bai [2003], Stock & Watson [2002], Kelly & Pruitt [2015] among others. Bai & Ng [2023] extends Bai [2003] to accommodate weak factors by making similar adjustments to the assumptions as we have done above.

Assumptions 3.4-3.7 are reasonable since they involve products of orthogonal series. We can specify lower-level conditions (several mixing conditions) which guarantee 3.1-3.7, but for the sake of simplicity, we instead state these high-level assumptions, as done in other papers, see Kelly & Pruitt [2015] and Bai [2003] among others.

Assumption 4. (*Uncorrelated loadings and Factors*). For matrices $\mathcal{P}_f$, $\mathcal{P}_{fg}$ and Δ_F , featuring in Assumption 2, we require the following:

1. $\mathcal{P}_f$ is positive definite and $\mathcal{P}_{fg} = \mathbf{0}$.
2. $\Delta_F \equiv \begin{pmatrix} \Delta_f & \Delta_{fg} \\ \Delta'_{fg} & \Delta_g \end{pmatrix}$ is positive definite and $\Delta_{fg} = \mathbf{0}$.

We require that the relevant factors be uncorrelated with the irrelevant factors and that the associated relevant factor loadings also be uncorrelated with the irrelevant factor loadings. This condition is less stringent than the assumption in Kelly & Pruitt [2015], where all loadings are assumed to be orthogonal to each other, and all factors are also assumed to be mutually orthogonal.³

²If weakness in loadings is induced by sparsity, i.e., the factors are local, then we can use Assumption 3.2 to prove 3.3 by slightly modifying the argument in Kelly & Pruitt [2015] Lemma 1.1 and Lemma 1.2. However, we consider a more general setting where factors may not be local; instead, all loadings are weak, and hence we introduce this assumption.

³See Assumption 5 (Normalization) in Kelly & Pruitt [2015].

Assumption 5. (*Relevant Proxies*).

1. $L = K_f$ and $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$.
2. $\mathbf{\Lambda}_f$ is non-singular.

Assumption 5 is borrowed from Kelly & Pruitt [2015]. We require proxies to mimic the target in terms of their dependence on factors. The assumption asserts that proxies must meet three criteria: (i) they should not load on irrelevant factors, (ii) their loadings on relevant factors should be linearly independent, and (iii) their number should match the number of relevant factors. When combined with Assumption 4, this implies that the common components of proxies span the relevant factor space and that none of the variation in proxies stems from irrelevant factors.

Assumption 6. For $v = f, g$, define $\Gamma_{N_v, T} \equiv \min(\sqrt{N^{\psi_v}}, \sqrt{T})$ and $\delta_{N, T} \equiv \min(\sqrt{N}, \sqrt{T})$. We need the following:

1. $\lim_{N \rightarrow \infty, T \rightarrow \infty} \frac{N^{1-\psi_f}}{\sqrt{T} \delta_{N, T}} = \max\left(\frac{N^{1-\psi_f}}{T}, \frac{N^{1/2-\psi_f}}{T^{1/2}}\right) = 0$.
2. $\lim_{N \rightarrow \infty, T \rightarrow \infty} \frac{N^{\psi_g-\psi_f}}{\Gamma_{N_g, T}} = \max\left(\frac{N^{\psi_g-\psi_f}}{\sqrt{T}}, N^{-\psi_f+\frac{\psi_g}{2}}\right) = 0$.

Assumption 6 specifies the necessary growth rate of T relative to N under a weak factor structure. It is automatically satisfied when $\psi_f = \psi_g = 1$. When all factors have same strength, $\psi_f = \psi_g = \psi$, Assumption 6.1 implies that a small ψ necessitates a larger T relative to N for consistent estimation. This requirement embodies an implicit cost imposed by a weak factor structure. When $\psi_f \geq \psi_g$, Assumption 6.2 is automatically satisfied. Conversely, when $\psi_f < \psi_g$, Assumption 6.2 reflects the cost of having higher noise (irrelevant factors) relative to the signal (relevant factors). Additionally, Assumption 6.2 imposes a limit on the weakness of relevant factors relative to irrelevant factors for the consistency of 3PRF, requiring that $\psi_f > \frac{\psi_g}{2}$.

Assumption 7. (*Uniform bounds*). For all m , N and T , $v = f, g$ and some positive constants $r_1, \dots, r_5$,

1. $\max_i |\phi_i(m)| = O_p(1)$.
2. $\max_{it} |\varepsilon_{it}| = O_p((\log N)^{r_1}) + O_p((\log T)^{r_1})$.
3. $\max_i \left| \sum_{t=1}^T \frac{1}{\sqrt{T}} \varepsilon_{it} \right| = O_p((\log N)^{r_2})$ and $\max_t \left| \sum_{i=1}^N \frac{1}{\sqrt{N}} \varepsilon_{it} \right| = O_p((\log T)^{r_2})$.
4. $\max_i \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t(m) \varepsilon_{it} \right| = O_p((\log N)^{r_3})$.

$$5. \max_i \left| N^{-\psi_f/2} T^{-1/2} \sum_{j,t} \phi_{jf}(m) \varepsilon_{jt} \varepsilon_{it} \right| = O_p((\log N)^{r_4}).$$

$$6. \max_i \left| N^{-1/2} T^{-1/2} \sum_{j,t} \varepsilon_{it} \varepsilon_{jt} \right| = O_p((\log N)^{r_5}).$$

We impose some high-level assumptions. We require uniform bounds on certain empirical processes to ensure that the prediction error in the Stage 1 does not adversely affect the theoretical results in Stage 2 LASSO regression. Such assumptions are prevalent in the literature. Specifically, Assumption 7.1 featured in Fan *et al.* [2020] and Giglio *et al.* [2026] and references therein. In fact, Fan *et al.* [2020] assumes that the maxima of factors, loadings, and idiosyncratic terms do not scale with N and T and are uniformly bounded by some constant.⁴ Similarly, Giglio *et al.* [2026] incorporates Assumption 7.2 with $r_1 = 1/2$. The partial sums (after centering) in Assumptions 7.4–7.6 (without taking the maximum over i) are bounded by Assumptions 3.2, 3.3, and 3.6. We assume that the maximum of these empirical processes scales as some power of the logarithm of N . This assumption is equivalent to imposing a uniform tail bound on these partial sums (see Remark 2).

Remark 2. *The scaling properties outlined in Assumption 7 are often associated with Weibull distributions, which are widely employed in economics. Their flexibility and ability to capture various shapes of distributions make them a general and versatile assumption in economic analyses. A more general assumption could be to impose moment bounds on the random variables and scaled partial sums in Assumption 7. For a random variable $\mathcal{X}_i$, where $i \in \{1, \dots, N\}$, if $\mathbb{E}(\mathcal{X}_i^k) < M$ for some finite k , then by Markov Inequality we have*

$$\mathbb{P}(\mathcal{X}_i > N^{1/k}) < \frac{\mathbb{E}(\mathcal{X}_i^k)}{(N^{1/k})^k} \leq \frac{M}{N}.$$

Hence, $\max_i \mathbb{P}(\mathcal{X}_i > N^{1/k}) \leq \frac{MN}{N} \leq M$; the highest ordered statistic of $\mathcal{X}$ scales with the rate at most $N^{1/k}$. Therefore, we can instead state Assumption 7 by assuming the existence of a sufficiently large k that allows the theoretical properties of Stage 2 involving LASSO as stated in Theorem 5 to hold.

Assumption 8. *(Weak cross-sectional dependence). For each i , let*

$$\Delta_{i,\varepsilon} \equiv \left\{ j : \mathbb{E} \left| T^{-1/2} \sum_{t=1}^T \varepsilon_{it} \varepsilon_{jt} \right|^2 \leq M < \infty \right\}. \text{ Then } N - |\Delta_{i,\varepsilon}| \leq M_2 < \infty.$$

Assumption 8 strengthens Assumption 3 by imposing a truncated form of cross-sectional dependence. Such a truncation in temporal dependence is commonly assumed in the literature; for instance, see Gonçalves *et al.* [2017]. A non-zero ζ introduces ‘corruption’ in the supervisor, which necessitates assuming weaker dependence in the idiosyncratic terms to enable consistent estimation.

⁴See Assumption 4.6 regarding $\mathbf{W}_{\max}$ in Fan *et al.* [2020].

Assumption 9. (*‘Relevant’ idiosyncratic terms and mimicking proxies*).

1. Let $S \equiv \{i | \gamma_i \neq 0\}$. The cardinality of set S is bounded, i.e., $|S| \leq M < \infty$.
2. $\gamma_i = 0$ if and only if $\zeta'_i = 0$, where ζ'_i denotes the i^{th} row of matrix ζ' .

We require the set of ‘relevant’ idiosyncratic terms to be finite. If this set were allowed to grow in size, we would need to adjust the rates in various ensuing Theorems. The dependence of the target (and thereby proxies) on idiosyncratic terms leads to noisier estimation in Stage 1 of the 3PRF LASSO procedure. If many idiosyncratic terms have explanatory power for $\mathbf{y}$, the extent of corruption in the Stage 1 Pass 1 (supervising step) is greater. The assumption that $\gamma_i = 0$ if and only if $\zeta'_i = 0$ can be relaxed. We only require that $|\{i | \zeta'_i \neq 0\}| < \infty$ to establish our theoretical results. However, since $\mathbf{Z}$ is used as a proxy for $\mathbf{y}$, it is reasonable to assume that their data-generating processes are similar.

Remark 3. *The dependence of the target on idiosyncratic terms can be ‘dense’ in the sense that a lot of idiosyncratic terms have non-zero small coefficients, see He [2024]. In such cases, as shown in his paper, ridge regression is asymptotically efficient in capturing both factor and idiosyncratic information among the entire class of spectral regularized estimators. Our model, unlike his paper, assumes that the dependence of the target on idiosyncratic terms is sparse, more akin to the setting in Fan et al. [2020] and Kneip & Sarda [2011].*

Assumption 10. For $r_1, \dots, r_5$ as defined in Assumption 7, we require:

1. (a) For all $r \in \{r_1, \dots, r_5\}$, $\frac{(\log N)^r}{\sqrt{T}} = O(1)$.
 (b) $\frac{N^{1-\psi_f}}{\sqrt{T}} [(\log N)^{r_1} + (\log T)^{r_1}] = O(1)$.
2. There exists a sufficiently large constant $\kappa > 0$ such that for all $i \in \{1, \dots, N\}$, all T , and all $s > 0$, we have,

$$\mathbb{P} \left(\frac{|\sum_{t=1}^T \varepsilon_{it} \eta_{t+h}|}{\sqrt{T}} > s \right) \leq \exp \left(\frac{-s^2}{\kappa} \right).$$

3. Define $\Delta_{\varepsilon,g} := (\varepsilon + \mathbf{J}_T \mathbf{g} \Phi'_g)' (\varepsilon + \mathbf{J}_T \mathbf{g} \Phi'_g) / T$. For the $N \times N$ matrix $\Delta_{\varepsilon,g}$, we say that, with respect to $\Delta_{\varepsilon,g}$, the compatibility condition is met for some set $A \subset \{1, \dots, N\}$, if for some compatibility constant $\nu > 0$, and for all $N \times 1$ vectors Θ satisfying $\|\Theta_{A^c}\|_1 \leq 3 \|\Theta_A\|_1$, it holds that

$$\|\Theta_A\|_1^2 \leq (\Theta' \Delta_{\varepsilon,g} \Theta) |A| / \nu^2.$$

We assume that, with probability approaching one, the compatibility condition is met for set S defined in Assumption 9, with respect to $\Delta_{\varepsilon,g}$ and the associated compatibility constant is ν_0 .

Assumption 10.1 is required to bound the impact of estimation errors in Stage 1 on Stage 2, as generated regressors are used in Stage 2. Assumption 10.2 is naturally satisfied for i.i.d. sub-Gaussian processes, which are commonly employed in deriving the properties of LASSO, as seen in Bühlmann & Van De Geer [2011]. Although the current setup does not assume i.i.d., we assume a weak dependence structure that obeys a similar sub-Gaussian-type bound on partial sums. Assumption 10.3, the compatibility condition, is standard in the LASSO literature (see Bühlmann & Van De Geer [2011]). This assumption essentially restricts the correlation among the relevant idiosyncratic components, ensuring they are not excessively interdependent.

4 Theoretical Results

We present each of the theorems 1-4 in two parts, labeled as parts (a) and (b). The first part (a) accommodates the idea of weak factors, while the second part (b) focuses on cases where idiosyncratic terms have predictive content for the target. We, therefore, cater to diverse readerships; certain readers may find one part more interesting than the other, while some may be interested in both.

Asymptotic results throughout the paper are based on simultaneous N and T limits, as in Bai [2003] and Kelly & Pruitt [2015]. As explained in Bai [2003], a simultaneous limit implies the existence of coinciding sequential and path-wise limits, but not vice-versa. Proofs for all the theorems are provided in Online Appendix A.

Define $\Xi_{N,T}^{-1} \equiv \max \left(T^{-1/2}, N^{-\psi_f/2}, \frac{N^{\psi_g - \psi_f}}{\Gamma_{N_g,T}} \right) = \max \left(T^{-1/2}, N^{-\psi_f/2}, N^{-\psi_f + \psi_g/2}, \frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \right)$, where the equality follows from substituting the value of $\Gamma_{N_g,T}$ as defined in Assumption 6.

Theorem 1. (a) Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, then we have,

$$\hat{y}_{t+h,f} - \mathbb{E}_t y_{t+h} = O_p(\Xi_{N,T}^{-1}),$$

where $\mathbb{E}_t y_{t+h}$ denotes the conditional expectation of the target variable at $t+h$ given the information set at t .

(b) Let Assumptions 1-6 and 8-9 hold, $\gamma \neq 0$ and $\zeta \neq 0$. Furthermore, $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$, then

$$\hat{y}_{t+h,f} - \mathbb{E}(y_{t+h} | \mathbf{F}_t) = O_p(\Xi_{N,T}^{-1}).$$

Theorem 1 (a) specifies the rate of convergence of the Stage 1 forecast when $\gamma = 0$ and $\zeta = 0$ but $0 < \psi_f \leq 1$ and $0 < \psi_g \leq 1$, hence generalizing Kelly & Pruitt [2015] by accommodating weak factors. When $\psi_f = \psi_g = 1$, the rate is $\delta_{N,T}^{-1}$ ($\delta_{N,T} \equiv \min(\sqrt{N}, \sqrt{T})$).

Theorem 1 (b) establishes that the Stage 1 forecast converges to the conditional expectation of the target w.r.t. true relevant factors. Unlike Theorem 1 (a), this factor-based forecast is no longer optimal because γ is allowed to take a value different from 0, indicating that the idiosyncratic components contain predictive information for the target. This predictive content in idiosyncratic components is harnessed in subsequent Stage 2. Theorem 1 (b) generalizes Kelly & Pruitt [2015] along 2 dimensions, i.e., accommodating weak factors and abstracting away from the assumption that $y_{t+h} - \mathbb{E}(y_{t+h}|\mathbf{F}_t)$ has a conditional expectation of zero with respect to information set in period t .

Remark 4. *If factors are strong, Theorem 1(a) would imply that $\hat{y}_{t+h,f} - \mathbb{E}_t y_{t+h} = O_p(\delta_{N,T}^{-1})$. This is different from the result in Kelly & Pruitt [2015] where the rate is $O_p(T^{-1/2})$, (see Theorem 4 in their paper). Their proof follows two steps. First they show that $\hat{y}_{t+h} - \tilde{y}_{t+h} = O_p(T^{-1/2})$ and then they argue that $\sqrt{T}\tilde{y}_{t+h} \xrightarrow{T, N \rightarrow \infty} \mathbb{E}_t y_{t+h}$. Since $\tilde{y}_{t+h}$ is $O_p(1)$, $\sqrt{T}\tilde{y}_{t+h}$ would diverge to infinity and their statement would be false. If they erroneously wrote this and instead wanted to imply that $\sqrt{T}(\tilde{y}_{t+h} - \mathbb{E}_t y_{t+h}) \xrightarrow{T, N \rightarrow \infty} 0$, then, again this statement is false because $\tilde{y}_{t+h} - \mathbb{E}_t y_{t+h}$ has random elements which converge to 0 at a rate which is $O_p(N^{-1/2}) + O_p(T^{-1/2}) = O_p(\delta_{N,T}^{-1})$.*

Theorem 2. *Let $\hat{\alpha}_i$ denote the i^{th} element of $\hat{\alpha}$. If Assumptions 1-6 hold, $\gamma = 0$ and $\zeta = 0$. If $\mathcal{P}_f = \mathbb{I}_{K_f}$, then for any i ,*

$$N^{\psi_f} \hat{\alpha}_i - \left(\phi_{if} - N^{\psi_f-1} \overline{\phi}_f \right)' \beta_f \xrightarrow[T, N \rightarrow \infty]{p} 0$$

Theorem 2 establishes consistency of the implied predictive coefficient $\hat{\alpha}$ in a scenario with possibly weak factors. This generalizes Theorem 2 of Kelly & Pruitt [2015], where this result is stated for $\psi_f = \psi_g = 1$. As argued in Kelly & Pruitt [2015], as N grows, the predictive information in $\mathbf{f}$ is spread across a larger number of predictors so each predictor's contribution approaches zero at the rate of $\frac{1}{N}$. That is the case when the number of predictors which load on the factors is proportionate to N , i.e., strong factors. When the factors are weak, they may either be local or have weak (local to zero) loadings or an amalgam thereof. If the factors are not pervasive, the predictive information contained within the vector $\mathbf{f}$ is dispersed across a few variables. The standardization term N^{ψ_f} illustrates that the predictive information is distributed across a subset of predictors; where the size of this subset is proportional to N^{ψ_f} . Hence, the contribution of each predictor goes to 0 at a slower rate compared to pervasive factors. When the factors are pervasive but

loadings are weak, in the sense that $\phi_{if} = c_N \tilde{\phi}_{if}$, where $\tilde{\phi}_{if}$ is a constant (not dependent on N), then Assumption 2.2, would imply that $c_n = O(N^{\psi_f - 1})$, which would imply that $N\hat{\alpha}_i \xrightarrow[T, N \rightarrow \infty]{p} (\tilde{\phi}_{if} - \bar{\phi}_f)' \beta_f$. Consequently, when factors are pervasive but all loadings are weak (local to zero), the predictive information in $\mathbf{f}$ is distributed across all predictors, and the relative contribution of each predictor diminishes at a rate of $\frac{1}{N}$, similar to the scenario with strong factors.

Define $\mathbf{G}_\beta \equiv \hat{\beta}_1^{-1} \hat{\beta}_2 (\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f')^{-1} (\Lambda_f \Delta_f \mathcal{P}_f \Delta_f)$, where $\hat{\beta}_1 = T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z}$ and $\hat{\beta}_2 = N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z}$.

Theorem 3. (a) Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, then we have,

$$\hat{\beta} - \mathbf{G}_\beta \beta_f = \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

(b) Let Assumptions 1-6 and 8-9 hold, $\gamma \neq 0$ and $\zeta \neq 0$. Furthermore, $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$, then,

$$\hat{\beta} - \mathbf{G}_\beta \beta_f = \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

Theorem 3 (a) specifies the convergence rate of the vector of predictive coefficients of the factors, i.e., $\hat{\beta}$ to a rotated version of the true coefficient vector β . This generalizes Theorem 5 of Kelly & Pruitt [2015] by accommodating weak factors. Just like Theorem 1 (a), when $\psi_f = \psi_g = 1$, the rate is $\delta_{N,T}^{-1}$, which is dissimilar to the $\sqrt{T}$ rate specified in Kelly & Pruitt [2015]. This difference stems from the definition of rotation matrix $\mathbf{G}_\beta$, see Remark 5. Theorem 3 (b) extends the scope of Theorem 3 (a) by allowing a more general DGP where idiosyncratic elements possess predictive capabilities for the target.

Define $\mathbf{H}_f \equiv \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \Lambda_f \Delta_f \mathcal{P}_f$ and $\mathbf{H}_0 \equiv \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} [N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0]$, where, $\hat{\mathbf{F}}_A = T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z}$ and $\hat{\mathbf{F}}_B = N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z}$.

Theorem 4. (a) Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, we have,

$$\hat{\mathbf{F}}_t - (\mathbf{H}_0 + \mathbf{H}_f \mathbf{f}_t) = \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

(b) Let Assumptions 1-6 and 8-9 hold, $\gamma \neq 0$ and $\zeta \neq 0$. Furthermore, $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$, we have,

$$\hat{\mathbf{F}}_t - (\mathbf{H}_0 + \mathbf{H}_f \mathbf{f}_t) = \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

Similar to the aforementioned theorems, both Theorem 4 (a) and Theorem 4 (b) extend the findings of Theorem 6 in Kelly & Pruitt [2015] by accommodating weak factors and permitting idiosyncratic terms to have predictive information for the target variable respectively.

Theorem 4 (a) and 4 (b) establish the convergence of the estimated factors to a rotation of the true relevant factors and provide the corresponding rate. Our convergence result diverges from the one presented in Kelly & Pruitt [2015]. They demonstrate the convergence of $\hat{\mathbf{F}}_t$ to a vector $\mathbf{H}\mathbf{F}_t$ ($\mathbf{H} \neq \mathbf{H}_f$) at a $\sqrt{N}$ rate. However, the matrix $\mathbf{H}$, as defined in their paper, does not satisfy certain desirable properties, which we highlight in Remark 5.

Remark 5. *As highlighted in Bai & Ng [2006] and also emphasized in Kelly & Pruitt [2015], the presence of matrices $\mathbf{H}_f$ and $\mathbf{G}_\beta$ in Theorem 3 and Theorem 4 stems from our estimation of a vector space. These theorems “pertain to the difference between $[\hat{\mathbf{F}}_t/\hat{\beta}]$ and the space spanned by $[\mathbf{F}_t/\beta]$ ”. The product $\mathbf{H}'_f\mathbf{G}_\beta$ equals an identity matrix, thereby nullifying the rotations in the predictive coefficients and relevant factors and preserving the direction spanned by $\beta'\mathbf{F}_t$. However, this characteristic is absent in theorems 5 and 6 of Kelly & Pruitt [2015]. The matrices $\mathbf{H}$ and $\mathbf{G}_\beta$ as defined in their paper do not necessarily yield a product that equals an identity matrix.*

Theorem 5. *Let the regularization parameter in the Stage 2 Pass 5 regression be given by $\lambda := 8 \frac{\sqrt{c + \kappa \log N}}{\Xi_{N,T}}$, where $c > 0$ and κ is defined in Assumption 10.2. Then, if Assumptions 1-10 hold and $\frac{T}{N} = O(1)$, with probability at least $1 - \exp\left(-\frac{c}{\kappa}\right) + o(1)$, we have*

$$\frac{1}{\sqrt{T}} \|\hat{\epsilon}\hat{\gamma} - \epsilon\gamma\|_2 = O_p\left(\frac{\sqrt{\log N}}{\Xi_{N,T}}\right).$$

Corollary 5.1. *From Theorem 5 and Theorem 1 (b), it follows that*

$$\frac{1}{\sqrt{T}} \|\hat{\mathbf{y}} - \mathbb{E}(\mathbf{y}|\mathbf{F}, \epsilon)\|_2 = O_p\left(\frac{\sqrt{\log N}}{\Xi_{N,T}}\right).$$

Theorem 5 establishes the rate at which the average prediction error of the Stage 2 LASSO regression converges to 0. The rate $\frac{\sqrt{\log N}}{\Xi_{N,T}}$, in general, is different from $\sqrt{\frac{\log N}{T}}$, which represents the optimal convergence rate for a high-dimensional M estimator, as indicated in Bickel *et al.* [2009]. This slower convergence rate is induced by the complexities associated with the generated regressor problem. To mitigate the maximum estimation error (across i) when generating the idiosyncratic components, it becomes imperative to adjust the rate at which the regularization parameter asymptotically converges to zero. This necessitated adjustment leads to a more gradual rate of convergence.

4.1 Discussion

The theoretical results derived in this paper are closely related to those in [Kelly & Pruitt \[2015\]](#), but differ in two fundamental respects. First, our framework accommodates potentially weak factors and potentially informative idiosyncratic components. Second, the convergence rates in our analysis depend on both the cross-sectional dimension N and the time dimension T , rather than on T alone as in their setting. Consequently, their results are not nested as special cases of ours, under strong-factor and uninformative-idiosyncratic scenarios (see [Remarks 4–5](#)).

Another recent and closely related paper that addresses weak factors in a forecasting context is [Giglio *et al.* \[2026\]](#). Their framework assumes the existence of a subset of predictors I_0 on which the factor structure is sufficiently strong for identification. Their method employs an iterative thresholding procedure to identify weak factors. In contrast, our framework does not require isolating such a subset; instead, we characterize conditions under which 3PRF recovers the relevant factors without thresholding. This allows both large idiosyncratic variances and local factor loadings to generate factor weakness. In their illustrative sparse loading model, let $N_0 = |I_0|$ denote the number of predictors in this subset. Their inconsistency condition for Partial Least Squares (PLS), under which $N/(N_0T)$ converges to a nonzero constant, corresponds under $N_0 \asymp N^{\psi_f}$ to $N^{1-\psi_f}/T$ converging to a nonzero constant, which we rule out to ensure consistent supervised factor recovery.

Another recent and conceptually proximate contribution is [Fan *et al.* \[2024\]](#). Both their framework and ours combine estimated factors with a sparse subset of idiosyncratic components selected via LASSO, yet the factor estimation and identification assumptions differ considerably. Their factors are extracted by unsupervised PCA under a strong-factor structure, rendering factor identification trivial and yielding the standard high-dimensional rate of convergence for the sparse regression that follows. Moreover, because the method of Principal Components is unsupervised, the factor estimation step in their paper remains unaffected by any potential dependence of the target variable on the idiosyncratic components, as discussed in the introduction section. By contrast, in our supervised framework, excessive dependence of the target on idiosyncratic terms could compromise the quality of the supervision step, a risk we control through our identification assumptions. Our estimator employs a supervised factor extraction step to recover target-relevant factors even under weak identification, followed by a LASSO stage that incorporates idiosyncratic predictive content. The effects of generated regressors are absent in their procedure due to strong identification, whereas in our framework, the properties of the downstream LASSO regression depend directly on the accuracy of the first-stage factor estimates. By combining supervision with flexible identification conditions, our framework retains the predictive advantages of target-guided factor estimation while remaining robust to weak factors and

idiosyncratic dependence.

4.2 Comparison with PCR

We now compare the convergence rates of 3PRF and PCR. To emphasize the role of factor weakness in determining the convergence rates, we assume that $N = O(T)$ and $\sqrt{T}/N^{\psi_{\min}} = o(1)$, where $\psi_{\min}$ denotes the strength of the weakest factor driving the predictor panel. Let $\mathbf{H}$ be the rotation matrix such that $\hat{\mathbf{F}}_t^{PC}$ converges to $\mathbf{H}\mathbf{F}_t$.

Consider first PCR using the first K principal components, where K is the number of factors driving the predictor panel. Following Bai & Ng [2006], the prediction error can be written as

$$\hat{y}_{t+h}^{PCR} - \mathbb{E}y_{t+h} = \left(\hat{\beta}^{PC'} - \beta' \mathbf{H}^{-1} \right) \hat{\mathbf{F}}_t^{PC} + \beta' \mathbf{H}^{-1} \left(\hat{\mathbf{F}}_t^{PC} - \mathbf{H}\mathbf{F}_t \right). \quad (4.1)$$

The first term reflects estimation error in the predictive coefficients. Adapting Lemma 4 of Bai & Ng [2023] to heterogeneous factor strengths, this term is $O_p(T^{-1/2})$ under $\sqrt{T}/N^{\psi_{\min}} = o(1)$. The second term reflects estimation error in the latent factors.

Although β assigns zero coefficients to target irrelevant factors, $\beta' \mathbf{H}^{-1}$ does not in general preserve this sparse structure unless $\mathbf{H}$ is the identity matrix. Freyaldenhoven [2022] shows that, under the population orthogonality and identifying normalization conditions imposed there, the rotation matrix converges to the identity matrix. Order the factor strengths as $1 \geq \psi_1 \geq \psi_2 \geq \dots \geq \psi_K > 0$. Using Lemma 3 of Freyaldenhoven [2022], the k th row of the rotation matrix satisfies

$$\mathbf{H}_{k\cdot} - \boldsymbol{\iota}'_k = O_p \left(N^{\psi_1/4 - \psi_k/2} \right) + O_p \left(N^{1/2 - \psi_k} \right),$$

where $\boldsymbol{\iota}'_k$ denotes the k th row of the identity matrix.⁵ Since K is fixed, the Frobenius norm of the rotation error is determined by the weakest factor,

$$\|\mathbf{H} - \mathbf{I}\|_F = O_p \left(N^{\psi_1/4 - \psi_K/2} \right) + O_p \left(N^{1/2 - \psi_K} \right).$$

By comparison, Bai & Ng [2023] show that estimation of the complete factor space is also governed by the weakest factor. In our notation, their result implies, under the corresponding conditions,

$$\hat{\mathbf{F}}_t^{PC} - \mathbf{H}\mathbf{F}_t = O_p \left(N^{-\psi_{\min}/2} \right).$$

⁵Freyaldenhoven [2022] uses the convention in which the k th estimated factor converges to $\mathbf{H}'_{\cdot k} \mathbf{F}_t$ and therefore states Lemma 3 in terms of the k th column $\mathbf{H}_{\cdot k}$. Under our convention, $\hat{\mathbf{F}}_t^{PC}$ converges to $\mathbf{H}\mathbf{F}_t$, so the corresponding object is the k th row $\mathbf{H}_{k\cdot}$.

Thus, the rate at which the estimated factors converge to the rotated factor space differs from the rate at which the rotation converges to the identity matrix. The former is $N^{-\psi_{\min}/2}$, whereas the latter is $N^{\psi_1/4-\psi_{\min}/2} + N^{1/2-\psi_{\min}}$, and is therefore slower. Consequently, convergence of the estimated principal components to the rotated factor space does not by itself imply equally fast recovery of the individual factors relevant for prediction.

Suppose now that the relevant factors have common strength ψ_f , the irrelevant factors have common strength ψ_g , and $\psi_f > \psi_g$. Then $\psi_{\min} = \psi_g$, and

$$\hat{\mathbf{F}}_t^{PC} - \mathbf{H}\mathbf{F}_t = O_p\left(N^{-\psi_g/2}\right).$$

When all K principal components are included, the PCR forecast therefore depends on estimation of the entire factor space, including the weaker target irrelevant factors. Under $N = O(T)$, PCR converges at rate $N^{\psi_g/2}$. By contrast, Theorem 1 shows that 3PRF targets the relevant factor space through supervision and converges faster when $\psi_f > \psi_g$.

Restricting PCR to the leading K_f principal components does not eliminate the rotation problem. Consider a two factor model in which F_{1t} is target relevant with strength ψ_f , F_{2t} is target irrelevant with strength $\psi_g < \psi_f$, and only the leading principal component is retained. Let h_{11} and h_{12} denote the $(1, 1)$ and $(1, 2)$ elements of $\mathbf{H}$, so that the first row of $\mathbf{H}$ is $\mathbf{H}_1 = (h_{11}, h_{12})$. Then

$$\hat{\mathbf{F}}_{1t}^{PC} = h_{11}F_{1t} + h_{12}F_{2t} + \left(\hat{\mathbf{F}}_{1t}^{PC} - \mathbf{H}_1\mathbf{F}_t\right).$$

The leading principal component is therefore an estimate of a linear combination of the relevant and irrelevant factors rather than of F_{1t} alone. For the leading factor, Lemma 3 of Freyaldenhoven [2022] implies that $\mathbf{H}_1 - (1, 0) = O_p(N^{-\psi_f/4}) + O_p(N^{1/2-\psi_f})$. Hence, although $\hat{\mathbf{F}}_{1t}^{PC}$ may converge relatively quickly to $\mathbf{H}_1\mathbf{F}_t$, its contamination by F_{2t} disappears only as $h_{11} \rightarrow 1$ and $h_{12} \rightarrow 0$. The leading principal component therefore remains a noisy estimate of the true relevant regressor at the slower rotation rate.

This gives the distinction between the two PCR specifications. When all K principal components are used, the convergence rate is governed by estimation of the entire factor space and therefore by the weakest factor. When only the leading K_f principal components are used, the weaker factors are omitted directly, but their effect remains through the rotation of the estimated leading factors. In either case, PCR does not generally attain the rate associated with observing the target relevant factors directly.

When target irrelevant factors are stronger than target relevant factors, the comparison is reversed. PCR converges faster than 3PRF when $\psi_f < \psi_g < 2\psi_f$, while Theorem 1 implies that 3PRF becomes inconsistent when $\psi_g \geq 2\psi_f$. Figure 1 summarizes these cases.

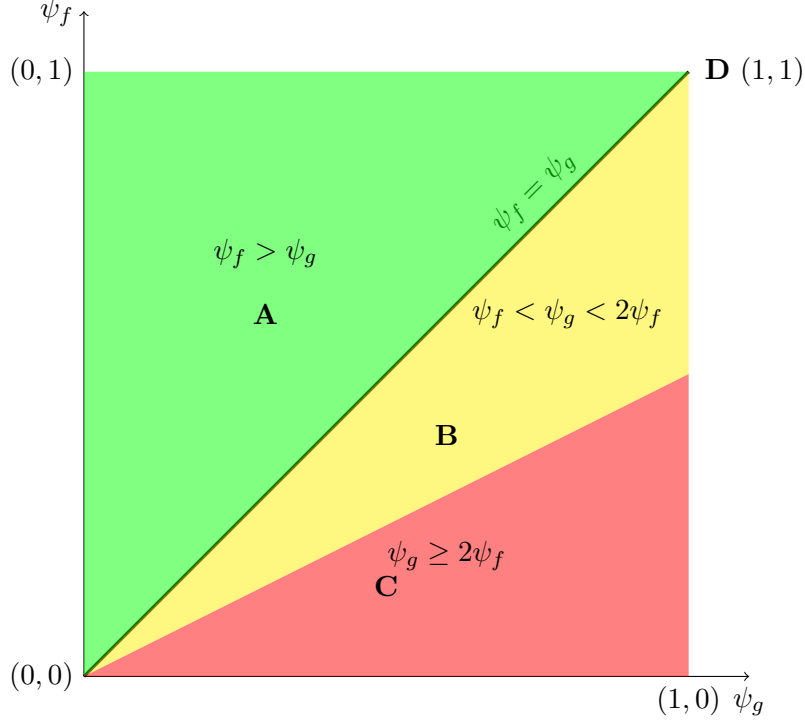


Figure 1: Convergence rate based on Factor strength. (Drawn for the case when $N \asymp T$).

Remark 6. *If target irrelevant factors have heterogeneous strengths, the convergence rate of 3PRF depends on the relevant factor strength and the strongest irrelevant factor. For PCR using the full factor space, the rate is determined by the weakest factor that must be estimated. The comparison is therefore more involved when irrelevant factors differ in strength.*

5 Simulation Analysis

Our Monte Carlo design closely resembles that in [Kelly & Pruitt \[2015\]](#). We consider one factor that is relevant to the target and four factors that are irrelevant to it. Let f_t denote the relevant factor and let $\mathbf{g}_t = (g_{1,t}, g_{2,t}, g_{3,t}, g_{4,t})'$ denote the vector of irrelevant factors. Thus, $\mathbf{F}_t = (f_t, \mathbf{g}_t)'$ denotes the vector of all five factors. The relevant factor is generated as $f_t = \rho_f f_{t-1} + u_t$, where $f_1 \sim N(0, 1)$ and $u_t \sim N(0, 1)$. For each $j = 1, \dots, 4$, the j th irrelevant factor is generated as $g_{j,t} = \rho_g g_{j,t-1} + w_{j,t}$, where $g_{j,1} \sim N(0, 1)$ and $w_{j,t} \sim N(0, 1)$. The innovations are independent across j and t and are independent of the other random variables in the data-generating process. Each factor series is demeaned and standardized after it is generated.

The idiosyncratic components allow for serial and cross-sectional dependence. We draw

$v_{i,t} \sim N(0, 1)$ independently across i and t and set $v_{0,t} = v_{N+1,t} = 0$. The innovation entering the i th idiosyncratic component is generated as $\tilde{\varepsilon}_{i,t} = (1 + d^2)v_{i,t} + dv_{i-1,t} + dv_{i+1,t}$ for $i = 1, \dots, N$. The idiosyncratic components are then generated recursively as $\varepsilon_{i,t} = a\varepsilon_{i,t-1} + \tilde{\varepsilon}_{i,t}$, where $\varepsilon_{i,1} \sim N(0, 1)$. Each idiosyncratic series is subsequently demeaned and standardized.

The idiosyncratic components of predictors 2 through 5 enter the target. We define $\tilde{\varepsilon}_{R,t} = (\varepsilon_{2,t} + \varepsilon_{3,t} + \varepsilon_{4,t} + \varepsilon_{5,t})/4$ and standardize this series to obtain $\varepsilon_{R,t}$. We also independently generate $\tilde{\eta}_t \sim N(0, 1)$ and standardize the resulting series to obtain η_t . The target is generated as $\tilde{y}_t = \sqrt{0.30}f_t + \sqrt{0.30}\varepsilon_{R,t} + \sqrt{0.40}\eta_t$ and is subsequently demeaned and standardized to obtain y_t . The factors, primitive shocks generating the idiosyncratic components, and the target noise are generated independently.

The predictors are generated as $\mathbf{x}_t = \mathbf{\Phi}\mathbf{F}_t + \boldsymbol{\varepsilon}_t$, where $\mathbf{\Phi} = (\boldsymbol{\phi}_f, \mathbf{\Phi}_g)$, $\boldsymbol{\phi}_f$ denotes the $N \times 1$ vector of loadings on the relevant factor, $\mathbf{\Phi}_g$ denotes the $N \times 4$ matrix of loadings on the irrelevant factors, and $\boldsymbol{\varepsilon}_t$ denotes the vector of idiosyncratic components. All potential factor loadings are drawn independently from the uniform distribution $U(-2, 2)$. For the relevant factor, $\lfloor N^{\psi_f} \rfloor$ loading locations are selected uniformly without replacement and retain their original uniform draws, while the remaining loadings are set equal to zero. The same procedure is applied separately to each irrelevant factor using $\lfloor N^{\psi_g} \rfloor$ loading locations. Factor weakness is therefore generated through limited pervasiveness. We consider $(\psi_f, \psi_g) \in \{(1, 1), (1, 0.8), (0.8, 1), (0.8, 0.8)\}$. Because the idiosyncratic components of predictors 2 through 5 enter the target, these predictors form the target-relevant idiosyncratic set. In accordance with Assumption 1, their loadings on all four target-irrelevant factors are set equal to zero.

We consider nine configurations of (ρ_f, ρ_g, a, d) . The first is $(0, 0, 0, 0)$, while the remaining eight include all combinations of $(\rho_f, \rho_g) \in \{(0.3, 0.9), (0.9, 0.3)\}$, $a \in \{0.3, 0.9\}$, and $d \in \{0, 1\}$.

We compare the out-of-sample performance of six methods: PCR, as described in [Stock & Watson \[2002\]](#); 3PRF, as proposed by [Kelly & Pruitt \[2015\]](#); LASSO, as proposed by [Tibshirani \[1996\]](#); PCR LASSO; 3PRF LASSO; and Elastic Net, as proposed by [Zou & Hastie \[2005\]](#). PCR uses the first $K = K_f + K_g = 5$ principal components, whereas 3PRF estimates a single target-relevant factor. For both 3PRF and 3PRF LASSO, we use the target $\mathbf{y}$ as the auto-proxy. PCR LASSO is implemented as a two-stage procedure. In the first stage, the target is regressed on the five leading principal components, while the predictors are regressed on the same components to obtain estimates of their idiosyncratic components as residuals. In the second stage, LASSO is used to regress the residual from the first-stage target regression on the estimated idiosyncratic components. 3PRF LASSO follows the same procedure, except that the factor and idiosyncratic components are estimated using

3PRF. The second-stage LASSO regressions are estimated without an intercept. LASSO and Elastic Net are applied directly to the predictor matrix $\mathbf{X}$ and include an intercept. The tuning parameters are selected using 10-fold cross-validation, following [Fan *et al.* \[2020\]](#).

The column labeled Oracle reports the out-of-sample R^2 obtained by regressing the target on an intercept, the true relevant factor, and the true relevant idiosyncratic components $\varepsilon_{2,t}, \dots, \varepsilon_{5,t}$. This Oracle regression is estimated using the training sample and evaluated over the testing sample.

For each replication, a total of $T' = 200$ observations are divided equally between a training sample of size $T = 100$ and a testing sample of size 100. We consider $N \in \{100, 200\}$ and use a fixed estimation window, as in [West & McCracken \[1998\]](#). We evaluate forecasting performance using the out of sample (OOS) R^2 , following the forecast evaluation framework of [West & McCracken \[1998\]](#), defined as

$$R_{\text{OOS}}^2 = 1 - \frac{\sum_{t \in \mathcal{T}_{\text{OOS}}} (y_t - \hat{y}_t)^2}{\sum_{t \in \mathcal{T}_{\text{OOS}}} (y_t - \bar{y}_{t,\text{train}})^2},$$

where $\bar{y}_{t,\text{train}}$ denotes the historical mean computed from the estimation sample available at forecast origin t . We report averages across 100 Monte Carlo replications.

The simulation results are reported in Tables 1–5. We focus the discussion mainly on the $N = 100, T = 100$ designs in Tables 1–4, and report the case $(\psi_f, \psi_g) = (1, 1)$ for $N = 200, T = 100$ in Table 5. The remaining $N = 200, T = 100$ factor strength configurations are reported in Online Appendix B and give the same broad qualitative patterns discussed below.

The strongest results for 3PRF LASSO arise when the target relevant common variation in the predictors is strong. Table 1 considers $(\psi_f, \psi_g) = (1, 0.8)$, where the relevant factor is stronger than the irrelevant factors. 3PRF LASSO has the highest out of sample R^2 in all nine dependence configurations. This is in line with the theory. When the relevant factor is stronger than the irrelevant factors, the convergence rate favors 3PRF relative to PCR, and the accuracy of the first stage carries over to the generated regressors used in the LASSO step. 3PRF LASSO therefore benefits from the better first stage while also using the predictive information contained in the idiosyncratic components.

The results remain favorable when the target relevant and target irrelevant factors have the same strength. Table 2 considers $(\psi_f, \psi_g) = (1, 1)$, where both types of common variation are fully pervasive. In this case, the theoretical convergence rates no longer favor 3PRF over PCR. Nevertheless, 3PRF performs better than PCR in almost all of the reported configurations, while 3PRF LASSO remains one of the leading methods and has the highest out of sample R^2 in five of the nine designs. Thus, the strong performance of 3PRF LASSO is not limited to cases in which the relevant common variation is stronger than the

irrelevant common variation. It also performs well when both are equally pervasive.

The dependence parameters add some qualification to these results. For instance, when d increases from 0 to 1, the purely sparse methods, LASSO and Elastic Net, generally gain more than 3PRF LASSO, so the gap between them narrows in some configurations. The case $d = 1$ implies strong neighborhood correlation among the idiosyncratic components. In many macroeconomic data sets, once the common variation in the predictors is removed, the remaining residual correlation is considerably weaker. The $d = 1$ design therefore represents a relatively demanding case of residual dependence. There is also some tendency for the 3PRF-based procedures to perform better when persistence is concentrated in the target-relevant common component rather than in the target-irrelevant common component.

The less favorable case for 3PRF arises when the target relevant common variation becomes weaker than the irrelevant common variation. Table 3 considers $(\psi_f, \psi_g) = (0.8, 1)$. Here 3PRF performs below PCR throughout the table, which is again in line with the theory: when the irrelevant factors are stronger than the relevant factor, the convergence of 3PRF deteriorates relative to PCR. Since the second stage of 3PRF LASSO uses regressors generated from the first stage, this weaker factor estimation also affects the LASSO step. LASSO and Elastic Net generally perform better in this configuration. The decline is, however, much smaller for 3PRF LASSO than for 3PRF itself. 3PRF LASSO remains close to the leading procedures in several cases and continues to outperform PCR LASSO throughout the table. The additional LASSO step can therefore somewhat offset part of the deterioration in the factor only procedure. We return below to why 3PRF LASSO can continue to outperform PCR LASSO even when the first stage advantage of 3PRF is weak.

A related pattern appears in Table 4, where both factor types have strength 0.8. Since the relevant factor is no longer weaker than the irrelevant factors, 3PRF again performs better than PCR throughout the dependence grid. However, the common factor signal itself is weaker than in the $(1, 1)$ case, and the purely sparse methods become more competitive. LASSO and Elastic Net outperform 3PRF LASSO in several configurations, although 3PRF LASSO remains close to them in the more favorable cases and is the best performing method under independent design. Across the four strength configurations, the results follow the theoretical comparison between 3PRF and PCR quite closely. When the relevant factor is stronger, the first stage favors 3PRF and 3PRF LASSO performs particularly well. When the irrelevant factor is stronger, 3PRF performs worse and this also carries over to 3PRF LASSO. When the two have the same strength, 3PRF remains competitive with PCR, often outperforming it. Weaker common variation gives LASSO and Elastic Net more scope to perform well relative to the factor-based procedures, since weaker factors are estimated at slower rates. Moreover, as discussed in Fan *et al.* [2020], weaker dependence among the regressors is favorable for sparse estimation.

One further feature of the results is that 3PRF LASSO often performs substantially better than PCR LASSO even when the difference between 3PRF and PCR is relatively small. In Online Appendix B, we report the true and false positive rates from the second stage regressions, which are informative about this difference. Across almost all of the simulation designs, 3PRF LASSO has a substantially lower false positive rate than PCR LASSO, while the true positive rates are generally similar. We conjecture that this is related to the different residual structures generated by the two procedures. 3PRF removes the target relevant common variation, but the common variation generated by the target irrelevant factors remains in the residualized predictors. The estimated idiosyncratic components corresponding to irrelevant predictors can therefore remain highly correlated. PCR estimates the idiosyncratic components after projecting the predictors onto the estimated space spanned by both the target-relevant and target-irrelevant factors. It therefore projects out the fitted contribution of both sets of common factors.

As discussed in [Fan *et al.* \[2020\]](#), strong dependence involving predictors that are relevant for the target can be detrimental to sparse estimation, leading to larger selected models and poorer LASSO performance. Our setting is different because the additional dependence generated by the 3PRF residualization is concentrated among predictors that are irrelevant for the target. For a group of highly correlated regressors, LASSO tends to select only one or a small number of variables and shrink the remaining coefficients toward zero; see, for example, [Zou & Hastie \[2005\]](#) and [Wang *et al.* \[2011\]](#). We conjecture that, when such correlation is concentrated among irrelevant predictors, this property can reduce false selections: if spurious variation leads LASSO to select a variable from a highly correlated group of irrelevant predictors, it is less likely to select many members of that group simultaneously. This provides a possible explanation for the lower false positive rates of 3PRF LASSO. In Online Appendix B, we consider a separate experiment in which only the correlation among irrelevant predictors is varied. The false positive rate generally decreases as this correlation increases, while the true positive rate remains broadly stable, providing additional support for this conjecture.

Finally, Table 5 reports the $(\psi_f, \psi_g) = (1, 1)$ design for $N = 200, T = 100$. The remaining three combinations of $\psi_f, \psi_g \in \{1, 0.8\}$ for $N = 200$ are reported in Online Appendix B and follow the same qualitative patterns as the $N = 100, T = 100$ results discussed above.

6 Empirical Application

We assess the out-of-sample forecasting performance of the proposed method using U.S. headline CPI inflation as the target variable. Let P_t denote the level of the headline Consumer Price Index. Following [Hamilton *et al.* \[2026\]](#), we define annualized headline CPI

inflation over horizon h as

$$y_t^{(h)} = \frac{400}{h} [\log(P_t) - \log(P_{t-h})]. \quad (6.1)$$

The predictor variables are taken from the quarterly FRED-QD dataset published by [Clark & McCracken \[2023\]](#). The raw dataset contains 246 series. To maintain a balanced predictor panel, we remove any series containing a missing observation over the full available sample, 1959Q1–2023Q3, leaving 176 predictor series. Many macroeconomic variables are non-stationary ([Beveridge & Nelson \[1981\]](#); [Nelson & Plosser \[1982\]](#)); we therefore transform the predictor series following [Hamilton *et al.* \[2026\]](#) before conducting the forecasting exercise.

To provide a direct empirical counterpart to the Monte Carlo analysis in Section 5, we compare the same six forecasting methods: PCR, 3PRF, PCR LASSO, 3PRF LASSO, LASSO, and Elastic Net. We consider forecast horizons of one quarter, two quarters, and four quarters, $h \in \{1, 2, 4\}$. The forecasting results are reported in Table 6.

We conduct the exercise over three samples. The first covers the Great Moderation, from 1984Q1 to 2007Q4. The second covers 2008Q1 to 2019Q4 and deliberately ends before the COVID-19 episode, which represents an unusually large disruption in inflation dynamics. As a robustness exercise, we also consider the full available post-2008 sample through 2023Q3, thereby including both the pandemic and the subsequent inflation episode.

We construct OOS forecasts using a rolling estimation window, following the rolling forecast evaluation framework discussed by [West & McCracken \[1998\]](#). For each forecast horizon h , the target y_t is aligned with predictors dated h periods earlier, $\mathbf{x}_{t-h}$. Within each subsample, the rolling training window contains 40 percent of the total aligned observations. Forecast performance is measured using OOS R^2 , defined as in Section 5. The corresponding evaluation periods are 1993Q3–2007Q4 for the Great Moderation sample, 2013Q1–2019Q4 for the 2008–2019 sample, and 2014Q2–2023Q3 for the sample extending through 2023Q3.

Prior to forecasting, both the target and the predictor variables are partialled with respect to a constant and four lags of the target, following [Kelly & Pruitt \[2015\]](#), [Bai & Ng \[2008\]](#), and [Stock & Watson \[2012\]](#). Letting $s = t - h$, we construct

$$\begin{aligned} \ddot{y}_t^{(h)} &= y_t^{(h)} - \widehat{\mathbb{E}}\left(y_t^{(h)} \mid y_s^{(h)}, y_{s-1}^{(h)}, y_{s-2}^{(h)}, y_{s-3}^{(h)}\right), \\ \ddot{\mathbf{x}}_s &= \mathbf{x}_s - \widehat{\mathbb{E}}\left(\mathbf{x}_s \mid y_s^{(h)}, y_{s-1}^{(h)}, y_{s-2}^{(h)}, y_{s-3}^{(h)}\right), \end{aligned} \quad (6.2)$$

where $\widehat{\mathbb{E}}(\cdot \mid \Omega)$ denotes linear projection on Ω and a constant. These projections are re-estimated using the current training sample at every forecast origin. The predictors are then standardized using the mean and standard deviation from the corresponding training

window.

For PCR, the number of principal components is selected separately in each rolling training sample using the eigenvalue ratio criterion of [Ahn & Horenstein \[2013\]](#). For 3PRF, we use a single supervised factor.⁶ For LASSO, Elastic Net, and the second stage regressions in PCR LASSO and 3PRF LASSO, tuning parameters are selected using ten-fold cross-validation.

The results in [Table 6](#) reveal a clear difference between short- and longer-horizon inflation predictability. At the one-quarter horizon, every method produces a negative OOS R^2 in each of the three samples. Thus, although the relative ranking of the methods varies at $h = 1$, none improves on the rolling historical-mean benchmark. Evidence of predictability becomes considerably stronger as the forecast horizon increases. At $h = 2$, 3PRF LASSO performs best during both the Great Moderation and the 2008–2019 sample, while Elastic Net performs best in the sample extending through 2023Q3.

The four-quarter-ahead prediction results display a particularly stable pattern across the three samples. 3PRF LASSO has the highest OOS R^2 in all three cases. Its improvement over 3PRF is also evident at the individual forecast-date level: 3PRF LASSO produces a lower squared forecast error at 72.4% of the Great Moderation evaluation dates, 50.0% of the 2008–2019 dates, and 60.5% of the dates in the sample through 2023Q3. Thus, in every sample, it performs at least as frequently better than plain 3PRF as worse, while also delivering a higher aggregate OOS R^2 . Relative to Elastic Net, the most competitive alternative to 3PRF LASSO at this horizon, the corresponding proportions are 55.2%, 67.9%, and 50.0% in the Great Moderation, 2008–2019, and 2008–2023 samples, respectively.

Overall, the empirical evidence suggests that combining the target-relevant supervised factor with sparse idiosyncratic predictive information is particularly useful for forecasting inflation at medium and longer horizons.

7 Conclusion

This paper extends the 3PRF framework in two directions. First, we allow the factors driving the predictor panel to be weak and characterize how factor strength affects the convergence of supervised factor forecasts. In particular, when target relevant factors are stronger than target irrelevant factors, 3PRF converges to the infeasible factor forecast faster than PCR because supervision targets the predictive factor space rather than the

⁶3PRF targets the factor space relevant for the target, whose dimension is no larger than the dimension of the factor space driving the predictor panel. Moreover, [Kelly & Pruitt \[2015\]](#) show that a target depending on multiple relevant factors can, under particular covariance structures, admit an equivalent single factor representation. In our application, the eigenvalue ratio criterion selects a very low dimensional PCR factor space, predominantly one or two factors across the rolling samples. We therefore use a single 3PRF factor.

entire factor space. When target irrelevant factors become sufficiently strong relative to the relevant factors, this advantage disappears and 3PRF can eventually become inconsistent.

Second, we allow the idiosyncratic components of the predictors to contain additional predictive information for the target. Because this information also enters the proxies used for supervision, it introduces an additional source of estimation error in the first stage. We establish conditions under which this additional dependence does not affect the convergence rate of 3PRF. We then augment 3PRF with a LASSO regression on estimated idiosyncratic components. The resulting 3PRF LASSO estimator explicitly accounts for the generated regressor nature of the second stage, and we derive the corresponding prediction error rate.

The Monte Carlo results support the theoretical analysis and show that 3PRF LASSO performs particularly well when the target relevant common variation is relatively strong. In the empirical application to U.S. headline CPI inflation using FRED-QD, 3PRF LASSO provides reliable forecasts, with particularly strong performance at longer horizons. Taken together, the theoretical results, Monte Carlo evidence, and empirical application support combining supervised factor extraction with sparse idiosyncratic information as a useful approach to high dimensional forecasting. These results also provide a basis for further theoretical developments and empirical applications.

Table 1: Out-of-sample R^2 values: $N = 100$, $T = 100$, $\psi_f = 1$, $\psi_g = 0.8$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.570	0.257	0.289	0.365	0.485 ^[1]	0.456 ^[2]	0.451 ^[3]
0.3	0.9	0.3	0	0.573	0.238	0.299	0.328	0.484 ^[1]	0.446 ^[2]	0.434 ^[3]
0.3	0.9	0.3	1	0.567	0.213	0.301	0.476	0.506 ^[1]	0.503 ^[3]	0.503 ^[2]
0.3	0.9	0.9	0	0.580	0.175	0.265	0.061	0.326 ^[1]	0.317 ^[3]	0.319 ^[2]
0.3	0.9	0.9	1	0.590	0.196	0.282	0.389	0.463 ^[1]	0.451 ^[3]	0.456 ^[2]
0.9	0.3	0.3	0	0.586	0.291	0.330	0.416	0.501 ^[1]	0.477 ^[2]	0.471 ^[3]
0.9	0.3	0.3	1	0.593	0.285	0.334	0.526	0.543 ^[1]	0.535 ^[2]	0.529 ^[3]
0.9	0.3	0.9	0	0.582	0.207	0.246	0.063	0.388 ^[1]	0.352 ^[3]	0.352 ^[2]
0.9	0.3	0.9	1	0.577	0.183	0.238	0.425	0.471 ^[1]	0.460 ^[3]	0.470 ^[2]

Notes: Entries report mean out-of-sample R^2 values across Monte Carlo replications. T denotes the training-sample size and is fixed at 100. The testing sample has the same size as the training sample. Superscripts ^[1], ^[2], and ^[3] denote, respectively, the highest, second-highest, and third-highest out-of-sample R^2 values among PCR, 3PRF, PCR LASSO, 3PRF LASSO, LASSO, and Elastic Net. The Oracle is excluded from these rankings.

Table 2: Out-of-sample R^2 values: $N = 100$, $T = 100$, $\psi_f = 1$, $\psi_g = 1$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.577	0.252	0.260	0.395	0.504 ^[1]	0.486 ^[3]	0.487 ^[2]
0.3	0.9	0.3	0	0.580	0.227	0.271	0.409	0.487 ^[1]	0.459 ^[3]	0.461 ^[2]
0.3	0.9	0.3	1	0.582	0.229	0.290	0.488	0.520 ^[3]	0.527 ^[1]	0.524 ^[2]
0.3	0.9	0.9	0	0.599	0.142	0.222	-0.289	0.357 ^[1]	0.349 ^[2]	0.348 ^[3]
0.3	0.9	0.9	1	0.589	0.160	0.223	0.358	0.453 ^[3]	0.484 ^[2]	0.484 ^[1]
0.9	0.3	0.3	0	0.596	0.288	0.297	0.432	0.526 ^[1]	0.502 ^[3]	0.506 ^[2]
0.9	0.3	0.3	1	0.592	0.284	0.312	0.514	0.542 ^[2]	0.541 ^[3]	0.544 ^[1]
0.9	0.3	0.9	0	0.594	0.242	0.260	-0.451	0.439 ^[1]	0.395 ^[3]	0.410 ^[2]
0.9	0.3	0.9	1	0.579	0.209	0.251	0.429	0.488 ^[3]	0.510 ^[1]	0.506 ^[2]

Notes: See the notes to Table 1.

Table 3: Out-of-sample R^2 values: $N = 100$, $T = 100$, $\psi_f = 0.8$, $\psi_g = 1$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.571	0.251	0.188	0.380	0.481 ^[1]	0.471 ^[3]	0.471 ^[2]
0.3	0.9	0.3	0	0.578	0.236	0.143	0.379	0.466 ^[3]	0.474 ^[1]	0.473 ^[2]
0.3	0.9	0.3	1	0.584	0.217	0.162	0.493	0.498 ^[3]	0.524 ^[2]	0.527 ^[1]
0.3	0.9	0.9	0	0.580	0.122	0.055	0.045	0.336 ^[3]	0.370 ^[2]	0.387 ^[1]
0.3	0.9	0.9	1	0.592	0.143	0.108	0.410	0.436 ^[3]	0.506 ^[1]	0.503 ^[2]
0.9	0.3	0.3	0	0.617	0.299	0.191	0.408	0.512 ^[3]	0.523 ^[2]	0.525 ^[1]
0.9	0.3	0.3	1	0.604	0.281	0.239	0.517	0.531 ^[3]	0.543 ^[2]	0.548 ^[1]
0.9	0.3	0.9	0	0.572	0.186	0.145	-0.001	0.364 ^[3]	0.391 ^[2]	0.421 ^[1]
0.9	0.3	0.9	1	0.594	0.219	0.181	0.398	0.446 ^[3]	0.474 ^[2]	0.489 ^[1]

Notes: See the notes to Table 1.

Table 4: Out-of-sample R^2 values: $N = 100$, $T = 100$, $\psi_f = 0.8$, $\psi_g = 0.8$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.570	0.238	0.263	0.371	0.444 ^[1]	0.437 ^[2]	0.436 ^[3]
0.3	0.9	0.3	0	0.584	0.227	0.260	0.361	0.432 ^[3]	0.438 ^[1]	0.437 ^[2]
0.3	0.9	0.3	1	0.569	0.199	0.296	0.373	0.480 ^[3]	0.496 ^[1]	0.494 ^[2]
0.3	0.9	0.9	0	0.586	0.175	0.204	0.144	0.284 ^[3]	0.292 ^[2]	0.325 ^[1]
0.3	0.9	0.9	1	0.577	0.192	0.251	0.372	0.410 ^[3]	0.447 ^[2]	0.461 ^[1]
0.9	0.3	0.3	0	0.593	0.277	0.292	0.344	0.462 ^[3]	0.468 ^[2]	0.474 ^[1]
0.9	0.3	0.3	1	0.579	0.247	0.314	0.494 ^[3]	0.494	0.502 ^[2]	0.504 ^[1]
0.9	0.3	0.9	0	0.580	0.225	0.276	0.014	0.341 ^[3]	0.347 ^[2]	0.356 ^[1]
0.9	0.3	0.9	1	0.586	0.198	0.287	0.382	0.406 ^[3]	0.471 ^[2]	0.484 ^[1]

Notes: See the notes to Table 1.

Table 5: Out-of-sample R^2 values: $N = 200$, $T = 100$, $\psi_f = 1$, $\psi_g = 1$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.570	0.248	0.250	0.133	0.486 ^[1]	0.460 ^[2]	0.459 ^[3]
0.3	0.9	0.3	0	0.569	0.230	0.259	0.104	0.465 ^[1]	0.459 ^[2]	0.451 ^[3]
0.3	0.9	0.3	1	0.572	0.212	0.259	0.325	0.492 ^[1]	0.486 ^[2]	0.485 ^[3]
0.3	0.9	0.9	0	0.576	0.149	0.202	-1.240	0.320 ^[3]	0.323 ^[2]	0.334 ^[1]
0.3	0.9	0.9	1	0.575	0.128	0.217	-0.341	0.407 ^[3]	0.434 ^[1]	0.425 ^[2]
0.9	0.3	0.3	0	0.602	0.307	0.300	0.149	0.521 ^[1]	0.515 ^[2]	0.513 ^[3]
0.9	0.3	0.3	1	0.584	0.270	0.280	0.362	0.529 ^[3]	0.534 ^[1]	0.532 ^[2]
0.9	0.3	0.9	0	0.587	0.224	0.234 ^[3]	-0.856	0.234	0.379 ^[2]	0.393 ^[1]
0.9	0.3	0.9	1	0.596	0.253	0.267	-0.121	0.492 ^[3]	0.496 ^[2]	0.505 ^[1]

Notes: See the notes to Table 1.

Table 6: Out-of-Sample Forecasting of Headline CPI Inflation

Method	$h = 1$	$h = 2$	$h = 4$
<i>Panel A: Great Moderation, 1984Q1–2007Q4</i>			
PCR	−0.063 ^[1]	−0.056	−0.137
3PRF	−0.337 ^[2]	0.051	0.168
PCR LASSO	−0.822	0.173 ^[2]	0.598 ^[4]
3PRF LASSO	−0.760 ^[4]	0.211 ^[1]	0.675 ^[1]
LASSO	−0.753 ^[3]	0.077 ^[4]	0.627 ^[3]
Elastic Net	−0.895	0.172 ^[3]	0.643 ^[2]
<i>Panel B: 2008Q1–2019Q4</i>			
PCR	−0.269 ^[3]	0.032	0.269
3PRF	−0.201 ^[2]	0.045	0.488
PCR LASSO	−0.745	0.241 ^[4]	0.717 ^[2]
3PRF LASSO	−0.485 ^[4]	0.578 ^[1]	0.749 ^[1]
LASSO	−0.770	0.337 ^[3]	0.550 ^[4]
Elastic Net	−0.156 ^[1]	0.559 ^[2]	0.607 ^[3]
<i>Panel C: 2008Q1–2023Q3</i>			
PCR	−0.884 ^[2]	0.298 ^[2]	0.477
3PRF	−1.695 ^[3]	0.176 ^[4]	0.353
PCR LASSO	−7.786	−9.019	0.561 ^[2]
3PRF LASSO	−5.842	0.171	0.676 ^[1]
LASSO	−0.787 ^[1]	0.184 ^[3]	0.495 ^[4]
Elastic Net	−1.695 ^[4]	0.395 ^[1]	0.553 ^[3]

Notes: This table reports out-of-sample (OOS) R^2 for the empirical exercise discussed in Section 6. Panel A considers the Great Moderation sample, 1984Q1–2007Q4; Panel B considers the post-2008 sample ending before the COVID-19 episode, 2008Q1–2019Q4; and Panel C considers the full available post-2008 sample, 2008Q1–2023Q3.

The corresponding OOS evaluation periods are 1993Q3–2007Q4, 2013Q1–2019Q4, and 2014Q2–2023Q3, respectively. The rolling training windows contain 38, 20, and 25 quarterly observations. For horizon h , an outcome dated t is forecast using predictors dated $t - h$. Thus, $h = 1, 2, 4$ correspond to one-quarter, two-quarter, and four-quarter forecast horizons.

OOS R^2 is computed relative to the rolling historical-mean benchmark. Within each panel and forecast horizon, superscripts ^[1], ^[2], ^[3], and ^[4] identify the first-, second-, third-, and fourth-highest OOS R^2 values among the six forecasting methods. The highest value is additionally reported in bold.

For Panel C at $h = 1$, 3PRF and Elastic Net both round to -1.695 ; their ranking is determined using the corresponding unrounded forecast errors.

References

- Ahn, Seung C, & Horenstein, Alex R. 2013. Eigenvalue ratio test for the number of factors. *Econometrica*, **81**(3), 1203–1227.
- Bai, Jushan. 2003. Inferential theory for factor models of large dimensions. *Econometrica*, **71**(1), 135–171.
- Bai, Jushan, & Ng, Serena. 2006. Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, **74**(4), 1133–1150.
- Bai, Jushan, & Ng, Serena. 2008. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, **146**(2), 304–317.
- Bai, Jushan, & Ng, Serena. 2023. Approximate factor models with weaker loadings. *Journal of Econometrics*, **235**(2), 1893–1916.
- Bailey, Natalia, Kapetanios, George, & Pesaran, M Hashem. 2021. Measurement of factor strength: Theory and practice. *Journal of Applied Econometrics*, **36**(5), 587–613.
- Beveridge, Stephen, & Nelson, Charles R. 1981. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the ‘business cycle’. *Journal of Monetary Economics*, **7**(2), 151–174.
- Beyhum, Jad, & Striaukas, Jonas. 2024. Testing for sparse idiosyncratic components in factor-augmented regression models. *Journal of Econometrics*, **244**(1), 105845.
- Bickel, Peter J, Ritov, Ya’Acov, & Tsybakov, Alexandre B. 2009. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, **37**(4), 1705–1732.
- Bühlmann, Peter, & Van De Geer, Sara. 2011. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.
- Chamberlain, Gary, & Rothschild, Michael. 1983. Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica: Journal of the Econometric Society*, 1281–1304.
- Clark, Todd E., & McCracken, Michael W. 2023. *FRED-QD: A Quarterly Database for Macroeconomic Research (September Ed.)*. Federal Reserve Bank of St. Louis.
- Fan, Jianqing, Ke, Yuan, & Wang, Kaizheng. 2020. Factor-adjusted regularized model selection. *Journal of Econometrics*, **216**(1), 71–85.

- Fan, Jianqing, Masini, Ricardo P, & Medeiros, Marcelo C. 2023. Bridging factor and sparse models. *The Annals of Statistics*, **51**(4), 1692–1717.
- Fan, Jianqing, Lou, Zhipeng, & Yu, Mengxin. 2024. Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, **119**(546), 1076–1088.
- Forni, Mario, Hallin, Marc, Lippi, Marco, & Reichlin, Lucrezia. 2000. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, **82**(4), 540–554.
- Freyaldenhoven, Simon. 2022. Factor models with local factors—determining the number of relevant factors. *Journal of Econometrics*, **229**(1), 80–102.
- Giglio, Stefano, Xiu, Dacheng, & Zhang, Dake. 2026. Prediction When Factors Are Weak. *Journal of the American Statistical Association*, **121**(554).
- Gonçalves, Sílvia, McCracken, Michael W, & Perron, Benoit. 2017. Tests of equal accuracy for nested models with estimated factors. *Journal of Econometrics*, **198**(2), 231–252.
- Hamilton, James D., Ma, Xinwei, & Xi, Jin. 2026. Principal Component Analysis for a Mix of Stationary and Nonstationary Variables. *Journal of Econometrics*, **258**.
- He, Yi. 2024. Ridge Regression Under Dense Factor Augmented Models. *Journal of the American Statistical Association*, **119**(546), 1566–1578.
- Kapetanios, George, & Marcellino, Massimiliano. 2010. Factor-GMM estimation with large sets of possibly weak instruments. *Computational Statistics & Data Analysis*, **54**(11), 2655–2675.
- Kelly, Bryan, & Pruitt, Seth. 2015. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, **186**(2), 294–316.
- Kim, Hyun Hak, & Swanson, Norman R. 2018. Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods. *International Journal of Forecasting*, **34**(2), 339–354.
- Kneip, Alois, & Sarda, Pascal. 2011. Factor models and variable selection in high-dimensional regression analysis. *The Annals of Statistics*, **39**(5), 2410–2447.
- Krämer, Nicole, & Sugiyama, Masashi. 2011. The degrees of freedom of partial least squares regression. *Journal of the American Statistical Association*, **106**(494), 697–705.

- Ludvigson, Sydney C, & Ng, Serena. 2016. A Factor Analysis of Bond Risk Premia. *Handbook of Empirical Economics and Finance*, 313.
- Nelson, Charles R, & Plosser, Charles R. 1982. Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics*, **10**(2), 139–162.
- Stock, James H, & Watson, Mark W. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, **97**(460), 1167–1179.
- Stock, James H, & Watson, Mark W. 2003. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, **41**(3), 788–829.
- Stock, James H, & Watson, Mark W. 2012. Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, **30**(4), 481–493.
- Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**(1), 267–288.
- Wang, Sijian, Nan, Bin, Rosset, Saharon, & Zhu, Ji. 2011. Random lasso. *The Annals of Applied Statistics*, **5**(1), 468.
- West, Kenneth, & McCracken, Michael. 1998. Regression-Based Tests of Predictive Ability. *International Economic Review*, **39**(4), 817–40.
- Zou, Hui, & Hastie, Trevor. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **67**(2), 301–320.

Online Appendix for Forecasting Using Supervised Factors and Idiosyncratic Elements

Daanish Padha* and Tae-Hwy Lee†

The Online Appendix is organized into two parts: Appendix A and Appendix B.

Appendix A provides the mathematical proofs for the theoretical results presented in the paper. Section A1 lists the assumptions, which are the same as those stated in the paper. Section A2 presents supplementary lemmas used throughout the proofs. Section A3 contains the proofs of Theorems 1(a), 2, 3(a), and 4(a), which establish the Stage 1 results of 3PRF LASSO for the cases where $\gamma = 0$ and $\zeta = 0$. The lemmas supporting these proofs are also included in Section A3. Section A4 covers Theorems 1(b), 3(b), and 4(b), focusing on the Stage 1 results of 3PRF LASSO when $\gamma \neq 0$ and $\zeta \neq 0$. Section A5 presents the Stage 2 results of 3PRF LASSO, including the proof of Theorem 5 and its associated supporting lemmas.

Appendix B presents additional simulation results not included in the main text. Section B1 reports supplementary simulation results that complement those in Section 5 of the paper, specifically tables comparing the relative performance of the six competing methods under various sample sizes and factor-strength combinations, using out-of-sample R^2 as the evaluation metric. Section B2 focuses on the true and false positive rates from Stage 2 of PCR LASSO and 3PRF LASSO and discusses the underlying mechanisms driving the observed differences.

*Economics, Adam Smith Business School, University of Glasgow, Glasgow G12 8QQ. E-mail: Daanish.Padha@glasgow.ac.uk

†Department of Economics, University of California, Riverside, CA 92521. E-mail: tae.lee@ucr.edu

Appendix A

A1 Assumptions

Assumption 1. (*Data-generating process*). The data are generated as follows:

$$\begin{aligned} \mathbf{x}_t &= \phi_0 + \Phi \mathbf{F}_t + \varepsilon_t, & y_{t+h} &= \beta_0 + \beta' \mathbf{F}_t + \gamma' \varepsilon_t + \eta_{t+h}, & \mathbf{z}_t &= \lambda_0 + \Lambda \mathbf{F}_t + \zeta \varepsilon_t + \omega_t, \\ \mathbf{X} &= \iota_T \phi_0' + \mathbf{F} \Phi' + \varepsilon, & \mathbf{y} &= \iota_T \beta_0 + \mathbf{F} \beta + \varepsilon \gamma + \eta, & \mathbf{Z} &= \iota_T \lambda_0' + \mathbf{F} \Lambda' + \varepsilon \zeta' + \omega, \end{aligned}$$

where $\mathbf{F}_t = (\mathbf{f}_t', \mathbf{g}_t')'$, $\Phi = (\Phi_f, \Phi_g)$, $\Lambda = (\Lambda_f, \Lambda_g)$, and $\beta = (\beta_f', \mathbf{0}')'$ with $\beta_f(m) \neq 0$ for all $m = 1, \dots, K_f$. $K_f > 0$ is the dimension of vector $\mathbf{f}_t$, $K_g \geq 0$ is the dimension of vector $\mathbf{g}_t$, $L > 0$ is the dimension of vector $\mathbf{z}_t$, and $K = K_f + K_g$. Furthermore, for all $m = 1, \dots, K$ and all $j \in \{i | \gamma_i \neq 0\}$, $\beta(m) = 0$ implies $\phi_j(m) = 0$.

Assumption 2. (*Factors, Loadings and Residuals*). Let $M < \infty$. For any i, s, t , $0 < \psi_f \leq 1$ and $0 < \psi_g \leq 1$,

1. $\mathbb{E} \|\mathbf{F}_t\|^4 < M$, $T^{-1} \sum_{s=1}^T \mathbf{F}_s \xrightarrow[T \rightarrow \infty]{p} \boldsymbol{\mu}$ and $T^{1/2} \left(\frac{\mathbf{F}' \mathbf{J}_T \mathbf{F}}{T} - \boldsymbol{\Delta}_F \right) = O_p(1)$.
2. $\mathbb{E} \|\phi_i\|^4 \leq M$. For $v = f, g$, $N^{-\psi_v} \sum_{j=1}^N \phi_{vj} \xrightarrow[N \rightarrow \infty]{p} \bar{\phi}_v < \infty$, $N^{\psi_v/2} \left(\frac{\Phi_v' \mathbf{J}_N \Phi_v}{N^{\psi_v}} - \mathcal{P}_v \right) = O_p(1)$ and for $\psi_s = \min(\psi_f, \psi_g)$, $N^{\psi_s/2} \left(\frac{\Phi \mathbf{f}' \mathbf{J}_N \Phi_g}{N^{\psi_s}} - \mathcal{P}_{fg} \right) = O_p(1)$.
3. $\mathbb{E}(\varepsilon_{it}) = 0$, $\mathbb{E} \|\varepsilon_{it}\|^8 \leq M$.
4. $\mathbb{E}(\omega_t) = \mathbf{0}$, $\mathbb{E} \|\omega_t\|^4 \leq M$, $T^{-1/2} \sum_{s=1}^T \omega_s = O_p(1)$ and $T^{1/2} \left(\frac{\omega' \mathbf{J}_T \omega}{T} - \boldsymbol{\Delta}_\omega \right) = O_p(1)$.
5. Let $\mathcal{I}_t$ denote the information set containing all information available through time t . For any $h > 0$, $\mathbb{E}_t(\eta_{t+h}) = \mathbb{E}(\eta_{t+h} | \mathcal{I}_t) = 0$, $\mathbb{E}(\eta_{t+h}^2) = \delta_\eta < \infty$, and η_{t+h} is independent of the collection $\{\mathbf{F}_t, \omega_t, \phi_i, \varepsilon_{i,t}\}$.

Assumption 3. (*Dependence*). For $M < \infty$ and any i, j, t, s, m, m_1, m_2 and $v = f, g$:

1. Let $\mathbb{E}(\varepsilon_{it} \varepsilon_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$, $|\sigma_{ij,ts}| \leq \tau_{ts}$, and
 - (a) $N^{-1} \sum_{i,j=1}^N \bar{\sigma}_{ij} \leq M$,
 - (b) $T^{-1} \sum_{t,s=1}^T \tau_{ts} \leq M$,
 - (c) $N^{-1} \sum_{i,s} |\sigma_{ii,ts}| \leq M$,
 - (d) $T^{-1} \sum_{i,t} |\sigma_{ij,tt}| \leq M$,

- (e) $N^{-1}T^{-1} \sum_{i,j,t,s} |\sigma_{ij,ts}| \leq M.$
2. (a) $\mathbb{E} \left| N^{-1/2}T^{-1/2} \sum_{s=1}^T \sum_{i=1}^N [\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}] \right|^4 \leq M,$
 (b) $\mathbb{E} \left| N^{-1/2}T^{-1/2} \sum_{s=1}^T \sum_{i=1}^N [\varepsilon_{is}\varepsilon_{js} - \sigma_{ij,ss}] \right|^4 \leq M.$
3. $\mathbb{E} \left| N^{-\psi_v/2}T^{-1/2} \sum_{t=1}^T \sum_{i=1}^N \phi_{iv}(m) [\varepsilon_{it}\varepsilon_{jt} - \sigma_{ij,tt}] \right|^2 \leq M.$
4. $\mathbb{E} \left| T^{-1/2} \sum_{t=1}^T F_t(m_1) \omega_t(m_2) \right|^2 \leq M.$
5. $\mathbb{E} \left| T^{-1/2} \sum_{t=1}^T \omega_t(m) \varepsilon_{it} \right|^2 \leq M.$
6. $\mathbb{E} \left| T^{-1/2} \sum_{t=1}^T F_t(m) \varepsilon_{it} \right|^2 \leq M.$
7. $\mathbb{E} \left| N^{-\psi_v/2} \sum_{i=1}^N \phi_{iv}(m_1) \varepsilon_{it} \right|^2 \leq M.$

Assumption 4. (*Uncorrelated loadings and Factors*). For matrices $\mathcal{P}_f$, $\mathcal{P}_{fg}$ and Δ_F , featuring in Assumption 2, we require the following:

1. $\mathcal{P}_f$ is positive definite and $\mathcal{P}_{fg} = \mathbf{0}.$
2. $\Delta_F \equiv \begin{pmatrix} \Delta_f & \Delta_{fg} \\ \Delta'_{fg} & \Delta_g \end{pmatrix}$ is positive definite and $\Delta_{fg} = \mathbf{0}.$

Assumption 5. (*Relevant Proxies*).

1. $L = K_f$ and $\Lambda = \begin{bmatrix} \Lambda_f & \mathbf{0} \end{bmatrix}.$
2. Λ_f is non-singular.

Assumption 6. For $v = f, g$, define $\Gamma_{N,T} \equiv \min(\sqrt{N^{\psi_v}}, \sqrt{T})$ and $\delta_{N,T} \equiv \min(\sqrt{N}, \sqrt{T})$. We need the following:

1. $\lim_{N \rightarrow \infty, T \rightarrow \infty} \frac{N^{1-\psi_f}}{\sqrt{T} \delta_{N,T}} = \max \left(\frac{N^{1-\psi_f}}{T}, \frac{N^{1/2-\psi_f}}{T^{1/2}} \right) = 0.$
2. $\lim_{N \rightarrow \infty, T \rightarrow \infty} \frac{N^{\psi_g-\psi_f}}{\Gamma_{N,T}} = \max \left(\frac{N^{\psi_g-\psi_f}}{\sqrt{T}}, N^{-\psi_f+\frac{\psi_g}{2}} \right) = 0.$

Assumption 7. (*Uniform bounds*). For all m , N and T , $v = f, g$ and some positive constants $r_1, \dots, r_5$,

1. $\max_i |\phi_i(m)| = O_p(1).$

2. $\max_{it} |\varepsilon_{it}| = O_p((\log N)^{r_1}) + O_p((\log T)^{r_1})$.
3. $\max_i \left| \sum_{t=1}^T \frac{1}{\sqrt{T}} \varepsilon_{it} \right| = O_p((\log N)^{r_2})$ and $\max_t \left| \sum_{i=1}^N \frac{1}{\sqrt{N}} \varepsilon_{it} \right| = O_p((\log T)^{r_2})$.
4. $\max_i \left| \frac{1}{\sqrt{T}} \sum_{t=1}^T F_t(m) \varepsilon_{it} \right| = O_p((\log N)^{r_3})$.
5. $\max_i \left| N^{-\psi_f/2} T^{-1/2} \sum_{j,t} \phi_{jf}(m) \varepsilon_{jt} \varepsilon_{it} \right| = O_p((\log N)^{r_4})$.
6. $\max_i \left| N^{-1/2} T^{-1/2} \sum_{j,t} \varepsilon_{it} \varepsilon_{jt} \right| = O_p((\log N)^{r_5})$.

Assumption 8. (*Weak cross-sectional dependence*). For each i , let

$$\Delta_{i,\varepsilon} \equiv \left\{ j : \mathbb{E} \left| T^{-1/2} \sum_{t=1}^T \varepsilon_{it} \varepsilon_{jt} \right|^2 \leq M < \infty \right\}. \text{ Then } N - |\Delta_{i,\varepsilon}| \leq M_2 < \infty.$$

Assumption 9. (*‘Relevant’ idiosyncratic terms and mimicking proxies*).

1. Let $S \equiv \{i | \gamma_i \neq 0\}$. The cardinality of set S is bounded, i.e., $|S| \leq M < \infty$.
2. $\gamma_i = 0$ if and only if $\zeta'_i = 0$, where ζ'_i denotes the i^{th} row of matrix ζ' .

Assumption 10. For $r_1, \dots, r_5$ as defined in Assumption 7, we require:

1. (a) For all $r \in \{r_1, \dots, r_5\}$, $\frac{(\log N)^r}{\sqrt{T}} = O(1)$.
 (b) $\frac{N^{1-\psi_f}}{\sqrt{T}} [(\log N)^{r_1} + (\log T)^{r_1}] = O(1)$.
2. There exists a sufficiently large constant $\kappa > 0$ such that for all $i \in \{1, \dots, N\}$, all T , and all $s > 0$, we have,

$$\mathbb{P} \left(\frac{|\sum_{t=1}^T \varepsilon_{it} \eta_{t+h}|}{\sqrt{T}} > s \right) \leq \exp \left(\frac{-s^2}{\kappa} \right).$$

3. Define $\Delta_{\varepsilon,g} := (\varepsilon + \mathbf{J}_T \mathbf{g} \Phi'_g)' (\varepsilon + \mathbf{J}_T \mathbf{g} \Phi'_g) / T$. For the $N \times N$ matrix $\Delta_{\varepsilon,g}$, we say that, with respect to $\Delta_{\varepsilon,g}$, the compatibility condition is met for some set $A \subset \{1, \dots, N\}$, if for some compatibility constant $\nu > 0$, and for all $N \times 1$ vectors Θ satisfying $\|\Theta_{A^c}\|_1 \leq 3 \|\Theta_A\|_1$, it holds that

$$\|\Theta_A\|_1^2 \leq (\Theta' \Delta_{\varepsilon,g} \Theta) |A| / \nu^2.$$

We assume that, with probability approaching one, the compatibility condition is met for set S defined in Assumption 9, with respect to $\Delta_{\varepsilon,g}$ and the associated compatibility constant is ν_0 .

A2 Supplementary Lemmas

Lemma 1. *Under Assumptions 2-3, for all s, t, i, m, m_1, m_2 and $v = f, g$ the following results hold.*

1. $\mathbb{E} \left| (NT)^{-1/2} \sum_{i,s} F_s(m) [\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}] \right|^2 \leq M.$
2. $\mathbb{E} \left| (NT)^{-1/2} \sum_{i,s} \omega_s(m) [\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}] \right|^2 \leq M.$
3. $N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} = O_p(1), \quad N^{-1/2} \sum_i \varepsilon_{it} = O_p(1) \quad \text{and} \quad T^{-1/2} \sum_t \varepsilon_{it} = O_p(1).$
4. $T^{-1/2} \sum_t \eta_{t+h} = O_p(1).$
5. $T^{-1/2} \sum_t F_t(m) \eta_{t+h} = O_p(1) \quad \text{and} \quad T^{-1/2} \sum_t \omega_t(m) \eta_{t+h} = O_p(1).$
6. $N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} \eta_{t+h} = O_p(1).$
7. $N^{-\psi_v/2}T^{-1} \sum_{i,t} \phi_{iv}(m_1) \varepsilon_{it} F_t(m_2) = O_p(1).$
8. $N^{-\psi_v/2}T^{-1} \sum_{i,t} \phi_{iv}(m_1) \varepsilon_{it} \omega_t(m_2) = O_p(1).$
9. $N^{-\psi_v/2}T^{-1/2} \sum_{i,t} \phi_{iv}(m) \varepsilon_{it} \eta_{t+h} = O_p(1).$
10. (a) $N^{-1}T^{-1/2} \sum_{i,s} \varepsilon_{is} \varepsilon_{it} = O_p(\delta_{N,T}^{-1}) \quad \text{and} \quad (b) \quad N^{-1/2}T^{-1} \sum_{i,t} \varepsilon_{it} \varepsilon_{jt} = O_p(\delta_{N,T}^{-1}).$
11. $N^{-1}T^{-3/2} \sum_{i,s,t} \varepsilon_{is} \varepsilon_{it} \eta_{t+h} = O_p(\delta_{N,T}^{-1}).$
12. $N^{-1}T^{-1/2} \sum_{i,s} F_s(m) \varepsilon_{is} \varepsilon_{it} = O_p(\delta_{N,T}^{-1}).$
13. $N^{-1}T^{-1/2} \sum_{i,s} \omega_s(m) \varepsilon_{is} \varepsilon_{it} = O_p(\delta_{N,T}^{-1}).$
14. $N^{-1}T^{-1} \sum_{i,s,t} F_s(m) \varepsilon_{is} \varepsilon_{it} \eta_{t+h} = O_p(1).$
15. $N^{-1}T^{-1} \sum_{i,s,t} \omega_s(m) \varepsilon_{is} \varepsilon_{it} \eta_{t+h} = O_p(1).$
16. $N^{-\psi_v/2}T^{-1} \sum_{i,t} \phi_{iv}(m) \varepsilon_{it} \varepsilon_{jt} = O_p(\Gamma_{N_v,T}^{-1}).$

The stochastic order is understood to hold as $N, T \rightarrow \infty$, $\delta_{N,T} \equiv \min(\sqrt{N}, \sqrt{T})$, $\Gamma_{N_v,T} \equiv \min(\sqrt{N^{\psi_v}}, \sqrt{T})$.

Proof: Items 1-4, 6, 10(a) and 11-15 have been proved in Kelly & Pruitt [2015], Auxiliary Lemma

1. We prove the rest below.

Item 5: Given Assumption 2.5, we have that

$$\begin{aligned}\mathbb{E} \left| T^{-1/2} \sum_t F_t(m) \eta_{t+h} \right|^2 &= T^{-1} \sum_t \mathbb{E} [\eta_{t+h}^2] \mathbb{E} [F_t(m)^2] \\ &\leq T^{-1} \sum_t \delta_\eta M \\ &= O(1)\end{aligned}$$

by Assumption 2.1 and 2.5. Therefore,

$$T^{-1/2} \sum_t F_t(m) \eta_{t+h} = O_p(1),$$

and similarly,

$$T^{-1/2} \sum_t \omega_t(m) \eta_{t+h} = O_p(1)$$

using Assumption 2.4, 2.5.

Item 7: For $v \in \{f, g\}$, using the Cauchy–Schwarz inequality,

$$\begin{aligned}N^{-\psi_v/2} T^{-1} \sum_{i,t} \phi_{iv}(m_1) \varepsilon_{it} F_t(m_2) &\leq \left(T^{-1} \sum_t F_t(m_2)^2 \right)^{1/2} \left(T^{-1} \sum_t \left[N^{-\psi_v/2} \sum_i \phi_{iv}(m_1) \varepsilon_{it} \right]^2 \right)^{1/2} \\ &= O_p(1) O_p(1)\end{aligned}$$

by Assumptions 3.7 and 2.1.

Item 8: For $v \in \{f, g\}$, using the Cauchy–Schwarz inequality,

$$\begin{aligned}N^{-\psi_v/2} T^{-1} \sum_{i,t} \phi_{iv}(m_1) \varepsilon_{it} \omega_t(m_2) &\leq \left(T^{-1} \sum_t \omega_t(m_2)^2 \right)^{1/2} \left(T^{-1} \sum_t \left[N^{-\psi_v/2} \sum_i \phi_{iv}(m_1) \varepsilon_{it} \right]^2 \right)^{1/2} \\ &= O_p(1) O_p(1)\end{aligned}$$

by Assumptions 3.7 and 2.4.

Item 9: Since Assumption 2.5 holds for any $h > 0$, taking $h = 1$ gives $\mathbb{E}(\eta_{r+1} \mid \mathcal{I}_r) = 0$ for every r .

Hence, by iterated expectations, $\mathbb{E}[\eta_{t+h}\eta_{s+h}] = 0$ for $t \neq s$. Therefore, by Assumption 2.5,

$$\begin{aligned} \mathbb{E} \left| N^{-\psi_v/2} T^{-1/2} \sum_{i,t} \phi_{iv}(m) \varepsilon_{it} \eta_{t+h} \right|^2 &= N^{-\psi_v} T^{-1} \sum_{i,j,t} \mathbb{E} [\phi_{iv}(m) \phi_{jv}(m) \varepsilon_{it} \varepsilon_{jt} \eta_{t+h}^2] \\ &= T^{-1} \sum_t \mathbb{E} [\eta_{t+h}^2] \mathbb{E} \left[\left(N^{-\psi_v/2} \sum_i \phi_{iv}(m) \varepsilon_{it} \right)^2 \right] \\ &= O(1). \end{aligned}$$

The final line follows from Assumptions 3.7 and 2.5. Hence,

$$N^{-\psi_v/2} T^{-1/2} \sum_{i,t} \phi_{iv}(m) \varepsilon_{it} \eta_{t+h} = O_p(1).$$

Item 10(b):

$$\begin{aligned} N^{-1/2} T^{-1} \sum_{i,t} \varepsilon_{it} \varepsilon_{jt} &= N^{-1/2} T^{-1} \sum_{i,t} [\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij,tt}] + N^{-1/2} T^{-1} \sum_{i,t} \sigma_{ij,tt} \\ &= O_p(T^{-1/2}) + O_p(N^{-1/2}) \end{aligned}$$

by Assumption 3.2(b) and 3.1(d).

Item 16 : We have,

$$\begin{aligned} N^{-\psi_v/2} T^{-1} \sum_{i,t} \phi_{iv}(m) \varepsilon_{it} \varepsilon_{jt} &= T^{-1/2} \left(N^{-\psi_v/2} T^{-1/2} \sum_{i,t} \phi_{iv}(m) [\varepsilon_{it} \varepsilon_{jt} - \sigma_{ij,tt}] \right) \\ &\quad + N^{-\psi_v/2} \left(T^{-1} \sum_{i,t} \phi_{iv}(m) \sigma_{ij,tt} \right) \\ &= 16.I + 16.II. \end{aligned}$$

16.I is $O_p(T^{-1/2})$ by Assumption 3.3.

16.II is $O_p(N^{-\psi_v/2})$ since $\mathbb{E} \left| T^{-1} \sum_{i,t} \phi_{iv}(m) \sigma_{ij,tt} \right| \leq \max_i \mathbb{E} |\phi_{iv}(m)| T^{-1} \sum_{i,t} |\sigma_{ij,tt}| = O_p(1)$ by Assumption 2.2 and 3.1(d). Hence, $N^{-\psi_v/2} T^{-1} \sum_{i,t} \phi_{iv}(m) \varepsilon_{it} \varepsilon_{jt} = O_p(\Gamma_{N_v, T}^{-1})$.

Lemma 2. *Under Assumptions 2-3, we have the following:*

1. $T^{-1/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega} = O_p(1).$
2. $T^{-1/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\eta} = O_p(1).$
3. $T^{-1/2} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta} = O_p(1).$
4. $N^{-\psi_f} \boldsymbol{\varepsilon}'_t \mathbf{J}_N \boldsymbol{\Phi} = O_p \left(N^{-\frac{\psi_f}{2}} \vee N^{-\psi_f + \frac{\psi_g}{2}} \right).$
5. $N^{-\psi_f} T^{-1} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = O_p \left(\Gamma_{N_f, T}^{-1} \vee \frac{N^{\psi_g - \psi_f}}{\Gamma_{N_g, T}} \right).$
6. $N^{-\psi_f} T^{-1} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} = O_p \left(\Gamma_{N_f, T}^{-1} \vee \frac{N^{\psi_g - \psi_f}}{\Gamma_{N_g, T}} \right).$
7. $N^{-\psi_f} T^{-1/2} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta} = O_p \left(N^{-\frac{\psi_f}{2}} \vee N^{-\psi_f + \frac{\psi_g}{2}} \right).$
8. $N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = O_p \left(\delta_{N, T}^{-1} \right).$
9. $N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = O_p \left(\delta_{N, T}^{-1} \right).$
10. $N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} = O_p \left(\delta_{N, T}^{-1} \right).$
11. $N^{-1} T^{-1/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}_t = O_p \left(\delta_{N, T}^{-1} \right).$
12. $N^{-1} T^{-1/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}_t = O_p \left(\delta_{N, T}^{-1} \right).$
13. $N^{-1} T^{-3/2} \boldsymbol{\eta}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = O_p \left(\delta_{N, T}^{-1} \right).$
14. $N^{-1} T^{-3/2} \boldsymbol{\eta}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} = O_p \left(\delta_{N, T}^{-1} \right).$

Proof: Items 1-3 and 8-14 have been proved in Kelly & Pruitt [2015], Auxiliary Lemma 2. We prove the rest below.

Expanding Item 4 we have,

$$N^{-\psi_f} \boldsymbol{\varepsilon}'_t \mathbf{J}_N \boldsymbol{\Phi} = \begin{bmatrix} N^{-\psi_f} \boldsymbol{\varepsilon}'_t \mathbf{J}_N \boldsymbol{\Phi}_f & N^{\psi_g - \psi_f} (N^{-\psi_g} \boldsymbol{\varepsilon}'_t \mathbf{J}_N \boldsymbol{\Phi}_g) \end{bmatrix}.$$

The m^{th} element of $N^{-\psi_f} \boldsymbol{\varepsilon}'_t \mathbf{J}_N \boldsymbol{\Phi}_f$ is given by,

$$N^{-\psi_f/2} \left(N^{-\psi_f/2} \sum_i \varepsilon_{it} \phi_{if}(m) - \left(N^{-1+\psi_f/2} \sum_i \varepsilon_{it} \right) \left(N^{-\psi_f} \sum_i \phi_{if}(m) \right) \right) = N^{-\psi_f/2} (1.\text{I} + 1.\text{II})$$

1.I : This term is $\mathbf{O}_p(1)$ by Assumption 3.7.

1.II : Since $N^{-1/2} \sum_i \varepsilon_{it} = \mathbf{O}_p(1)$ by Lemma 1.3, we have $N^{-1+\frac{\psi_f}{2}} \sum_i \varepsilon_{it} = \mathbf{O}_p(1)$ as $0 < \psi_f \leq 1$. By Assumption 2.2, $N^{-\psi_f} \sum_i \phi_{if}(m) = \mathbf{O}_p(1)$. Hence, (I + II) is $\mathbf{O}_p(1)$.

Therefore,

$$\begin{aligned} N^{-\psi_f} \varepsilon'_t \mathbf{J}_N \Phi_f &= N^{-\psi_f/2} \mathbf{O}_p(1) \\ &= \mathbf{O}_p\left(N^{-\psi_f/2}\right). \end{aligned}$$

Similarly,

$$N^{-\psi_g} \varepsilon'_t \mathbf{J}_N \Phi_g = \mathbf{O}_p\left(N^{-\psi_g/2}\right),$$

which implies

$$N^{-\psi_f} \varepsilon'_t \mathbf{J}_N \Phi_g = \mathbf{O}_p\left(N^{\psi_g-\psi_f} \times N^{-\psi_g/2}\right).$$

Hence, the whole matrix has the following order.

$$N^{-\psi_f} \varepsilon'_t \mathbf{J}_N \Phi = \mathbf{O}_p\left(N^{-\psi_f/2} \vee N^{-\psi_f+\psi_g/2}\right).$$

Item 5 can be expanded as

$$N^{-\psi_f} T^{-1} \Phi'_f \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F} = \begin{pmatrix} N^{-\psi_f} T^{-1} \Phi'_f \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F} \\ N^{-\psi_f+\psi_g} (N^{-\psi_g} T^{-1} \Phi'_g \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F}) \end{pmatrix}.$$

We show that $N^{-\psi_f} T^{-1} \Phi'_f \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F}$ is $\mathbf{O}_p\left(\Gamma_{N_f, T}^{-1}\right)$. By symmetry, $N^{-\psi_g} T^{-1} \Phi'_g \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F}$ will be $\mathbf{O}_p\left(\Gamma_{N_g, T}^{-1}\right)$. Hence $N^{-\psi_f+\psi_g} (N^{-\psi_g} T^{-1} \Phi'_g \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F})$ is $\mathbf{O}_p\left(\frac{N^{\psi_g-\psi_f}}{\Gamma_{N_g, T}}\right)$ and therefore the whole matrix is $\mathbf{O}_p\left(\Gamma_{N_f, T}^{-1} \vee \frac{N^{\psi_g-\psi_f}}{\Gamma_{N_g, T}}\right)$. Therefore, it is sufficient to show $N^{-\psi_f} T^{-1} \Phi'_f \mathbf{J}_N \varepsilon' \mathbf{J}_T \mathbf{F}$ is $\mathbf{O}_p\left(\Gamma_{N_f, T}^{-1}\right)$ and the stochastic order for the matrix follows. We show this below.

$N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \mathbf{F}$ is a $K_f \times K$ matrix with generic (m_1, m_2) element given by

$$\begin{aligned} & N^{-\psi_f} T^{-1} \sum_{i,t} \phi_{if}(m_1) F_t(m_2) \varepsilon_{it} - N^{-\psi_f-1} T^{-1} \sum_{i,j,t} \phi_{if}(m_1) F_t(m_2) \varepsilon_{jt} \\ & - N^{-\psi_f} T^{-2} \sum_{j,s,t} F_s(m_2) \phi_{jf}(m_1) \varepsilon_{jt} + N^{-\psi_f-1} T^{-2} \sum_{i,j,s,t} F_s(m_2) \phi_{if}(m_1) \varepsilon_{jt} \\ & = 5.I - 5.II - 5.III + 5.IV. \end{aligned}$$

5.I is $O_p(N^{-\psi_f/2})$ by Lemma 1.7.

5.II is $O_p(T^{-1/2})$ since $N^{-\psi_f} \sum_i \phi_{if}(m_1) = O_p(1)$ by Assumption 2.2 and

$N^{-1} \sum_j (T^{-1/2} \sum_t F_t(m_2) \varepsilon_{jt}) = O_p(1)$ by Assumption 3.6.

5.III is $O_p(N^{-\psi_f/2})$ since $T^{-1} \sum_s F_s(m_2) = O_p(1)$ by Assumption 2.1 and

$T^{-1} \sum_t (N^{-\psi_f/2} \sum_j \phi_{jf}(m_1) \varepsilon_{jt}) = O_p(1)$ by Assumption 3.7.

5.IV is $O_p(T^{-1/2} N^{-1/2})$ by Assumptions 2.1, 2.2 and Lemma 1.3.

Summing these terms, $N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \mathbf{F}$ is $O_p(\Gamma_{N_f, T}^{-1})$.

Item 6 can be expanded as

$$N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} = \begin{pmatrix} N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} \\ N^{-\psi_f+\psi_g} (N^{-\psi_g} T^{-1} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega}) \end{pmatrix}.$$

As in the case of Item 5, it suffices to show that $N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega}$ is $O_p(\Gamma_{N_f, T}^{-1})$.

$N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega}$ is a $K_f \times L$ matrix with generic (m_1, m_2) element given by

$$\begin{aligned} & N^{-\psi_f} T^{-1} \sum_{i,t} \phi_{if}(m_1) \omega_t(m_2) \varepsilon_{it} - N^{-\psi_f-1} T^{-1} \sum_{i,j,t} \phi_{if}(m_1) \omega_t(m_2) \varepsilon_{jt} \\ & - N^{-\psi_f} T^{-2} \sum_{j,s,t} \omega_s(m_2) \phi_{jf}(m_1) \varepsilon_{jt} + N^{-\psi_f-1} T^{-2} \sum_{i,j,s,t} \omega_s(m_2) \phi_{if}(m_1) \varepsilon_{jt} \\ & = 6.I - 6.II - 6.III + 6.IV. \end{aligned}$$

6.I is $O_p(N^{-\psi_f/2})$ by Lemma 1.8.

6.II is $O_p(T^{-1/2})$ since $N^{-\psi_f} \sum_i \phi_{if}(m_1) = O_p(1)$ by Assumption 2.2 and

$N^{-1} \sum_j (T^{-1/2} \sum_t \omega_t(m_2) \varepsilon_{jt}) = O_p(1)$ by Assumption 3.5.

6.III is $O_p(N^{-\psi_f/2} T^{-1/2})$ since $T^{-1/2} \sum_s \omega_s(m_2) = O_p(1)$ by Assumption 2.4

and $T^{-1} \sum_t \left(N^{-\psi_f/2} \sum_j \phi_{jf}(m_1) \varepsilon_{jt} \right) = O_p(1)$ by Assumption 3.7.

6.IV is $O_p(T^{-1}N^{-1/2})$ by Assumption 2.2, 2.4 and Lemma 1.3.

Summing these terms, we have that $N^{-\psi_f} T^{-1} \Phi_f' J_N \varepsilon' J_T \omega$ is $O_p(\Gamma_{N_f, T}^{-1})$.

Item 7: Similar to arguments presented in the case of Item 5 and 6, to show that $N^{-\psi_f} T^{-1/2} \Phi_f' J_N \varepsilon' J_T \eta$ is $O_p\left(N^{-\frac{\psi_f}{2}} \vee N^{-\psi_f + \frac{\psi_g}{2}}\right)$, it suffices to show that $N^{-\psi_f} T^{-1/2} \Phi_f' J_N \varepsilon' J_T \eta$ is $O_p\left(N^{-\frac{\psi_f}{2}}\right)$. We show this below.

The m^{th} element of $N^{-\psi_f} T^{-1/2} \Phi_f' J_N \varepsilon' J_T \eta$ is given by,

$$\begin{aligned} & N^{-\psi_f} T^{-1/2} \sum_{i,t} \phi_{if}(m) \varepsilon_{it} \eta_{t+h} - N^{-\psi_f} T^{-3/2} \sum_{i,s,t} \phi_{if}(m) \varepsilon_{it} \eta_{s+h} \\ & - N^{-\psi_f-1} T^{-1/2} \sum_{i,j,t} \phi_{if}(m) \varepsilon_{jt} \eta_{t+h} + N^{-\psi_f-1} T^{-3/2} \sum_{i,j,s,t} \phi_{if}(m) \varepsilon_{jt} \eta_{s+h} \\ & = 7.I - 7.II - 7.III + 7.IV. \end{aligned}$$

7.I is $O_p(N^{-\psi_f/2})$ by Lemma 1.9.

7.II can be written as

$$N^{-\psi_f/2} \left(T^{-1} \sum_t \left[N^{-\psi_f/2} \sum_i \phi_{if}(m) \varepsilon_{it} \right] \right) \left(T^{-1/2} \sum_s \eta_{s+h} \right).$$

This is $O_p(N^{-\psi_f/2})$ by Assumption 3.7 and Lemma 1.4.

7.III can be written as

$$\begin{aligned} & \left(N^{-\psi_f} \sum_i \phi_{if}(m) \right) \left(N^{-1/2} T^{-1/2} \sum_{j,t} \varepsilon_{jt} \eta_{t+h} \right) \left(N^{-1/2} \right) \\ & = O_p(1) O_p(1) \left(N^{-1/2} \right). \end{aligned}$$

This follows from Assumption 2.2 and Lemma 1.6.

7.IV can be written as

$$\begin{aligned}
& \left(N^{-1/2} T^{-1/2} \sum_{j,t} \varepsilon_{jt} \right) \left(T^{-1/2} \sum_s \eta_{s+h} \right) \left(N^{-\psi_f} \sum_i \phi_{if}(m) \right) \left(N^{-1/2} \times T^{-1/2} \right) \\
&= O_p(1) O_p(1) O_p(1) \left(N^{-1/2} \times T^{-1/2} \right) \\
&= O_p \left(N^{-1/2} \times T^{-1/2} \right).
\end{aligned}$$

This follows from Assumption 2.2 Lemma 1.3 and 1.4.

Summing these terms, $N^{-\psi_f} T^{-1/2} \Phi_f' J_N \varepsilon' J_T \eta$ is $O_p(N^{-\psi_f/2})$.

Lemma 3. *Under Assumptions 2-3, we have:*

1. $N^{-2\psi_f} T^{-1} \Phi' J_N \varepsilon' J_T \varepsilon J_N \Phi = O_p \left(N^{-\psi_f/2} \Gamma_{N_f, T}^{-1} \vee \frac{N^{2(\psi_g - \psi_f)}}{N^{\psi_g/2} \Gamma_{N_g, T}} \right).$
2. $N^{-2\psi_f} T^{-2} \Phi' J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F = O_p \left(T^{-1/2} N^{-\frac{3\psi_f}{2} + 1} \delta_{N, T}^{-1} \vee \frac{N^{2(\psi_g - \psi_f)}}{T^{1/2} N^{\frac{3\psi_g}{2} - 1} \delta_{N, T}} \right).$
3. $N^{-2\psi_f} T^{-2} \Phi' J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega = O_p \left(T^{-1/2} N^{-\frac{3\psi_f}{2} + 1} \delta_{N, T}^{-1} \vee \frac{N^{2(\psi_g - \psi_f)}}{T^{1/2} N^{\frac{3\psi_g}{2} - 1} \delta_{N, T}} \right).$
4. $N^{-2\psi_f} T^{-3} F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F = O_p \left(\left(\frac{N^{1-\psi_f}}{\sqrt{T} \delta_{N, T}} \right)^2 \right).$
5. $N^{-2\psi_f} T^{-3} F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega = O_p \left(\left(\frac{N^{1-\psi_f}}{\sqrt{T} \delta_{N, T}} \right)^2 \right).$
6. $N^{-2\psi_f} T^{-3} \omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega = O_p \left(\left(\frac{N^{1-\psi_f}}{\sqrt{T} \delta_{N, T}} \right)^2 \right).$

Proof: We prove 1-3 below. Items 4-6 have been proved in Kelly & Pruitt [2015], Online Appendix, Lemma Web. They show that $N^{-2} T^{-3} F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F$, $N^{-2} T^{-3} F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega$ and $N^{-2} T^{-3} \omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega$ are all $O_p \left(T^{-1} \delta_{N, T}^{-2} \right)$. Therefore changing N^{-2} in the normalization term by $N^{-2\psi_f}$ gives the stochastic orders as listed in Lemma 3 above.

Item 1 is a $K \times K$ matrix given by

$$N^{-2\psi_f} T^{-1} \Phi' J_N \epsilon' J_T \epsilon J_N \Phi = \begin{bmatrix} N^{-2\psi_f} T^{-1} \Phi'_f J_N \epsilon' J_T \epsilon J_N \Phi_f & N^{-2\psi_f} T^{-1} \Phi'_f J_N \epsilon' J_T \epsilon J_N \Phi_g \\ N^{-2\psi_f} T^{-1} \Phi'_g J_N \epsilon' J_T \epsilon J_N \Phi_f & N^{-2\psi_f} T^{-1} \Phi'_g J_N \epsilon' J_T \epsilon J_N \Phi_g \end{bmatrix}.$$

First, we show that $N^{-2\psi_f} T^{-1} \Phi'_f J_N \epsilon' J_T \epsilon J_N \Phi_f$ is $O_p \left(N^{-\psi_f/2} \Gamma_{N_f, T}^{-1} \right)$.

$N^{-2\psi_f} T^{-1} \Phi'_f J_N \epsilon' J_T \epsilon J_N \Phi_f$ has generic (m_1, m_2) element given as

$$\begin{aligned} & N^{-2\psi_f} T^{-1} \sum_{i,j,t} \phi_{fi}(m_1) \varepsilon_{it} \varepsilon_{jt} \phi_{fj}(m_2) - N^{-2\psi_f-1} T^{-1} \sum_{i,j,k,t} \phi_{fi}(m_1) \varepsilon_{it} \varepsilon_{jt} \phi_{fk}(m_2) \\ & - N^{-2\psi_f-1} T^{-1} \sum_{i,j,k,t} \phi_{fi}(m_1) \varepsilon_{jt} \varepsilon_{kt} \phi_{fk}(m_2) + N^{-2\psi_f-2} T^{-1} \sum_{i,j,k,l,t} \phi_{fi}(m_1) \varepsilon_{jt} \varepsilon_{kt} \phi_{fl}(m_2) \\ & - N^{-2\psi_f} T^{-2} \sum_{i,j,s,t} \phi_{fi}(m_1) \varepsilon_{is} \varepsilon_{jt} \phi_{fj}(m_2) + N^{-2\psi_f-1} T^{-2} \sum_{i,j,k,s,t} \phi_{fi}(m_1) \varepsilon_{is} \varepsilon_{jt} \phi_{fk}(m_2) \\ & + N^{-2\psi_f-1} T^{-2} \sum_{i,j,k,s,t} \phi_{fi}(m_1) \varepsilon_{js} \varepsilon_{kt} \phi_{fk}(m_2) - N^{-2\psi_f-2} T^{-2} \sum_{i,j,k,l,s,t} \phi_{fi}(m_1) \varepsilon_{js} \varepsilon_{kt} \phi_{fl}(m_2) \\ & = \text{1.I} - \dots - \text{1.VIII.} \end{aligned}$$

1.I can be written as

$$\begin{aligned} & N^{-\psi_f} \left(T^{-1} \sum_t \left(N^{-\psi_f/2} \sum_i \phi_{fi}(m_1) \varepsilon_{it} \right) \left(N^{-\psi_f/2} \sum_j \phi_{fj}(m_2) \varepsilon_{jt} \right) \right) \\ & = O_p \left(N^{-\psi_f} \right) \end{aligned}$$

by Assumption 3.7.

1.II can be written as

$$\begin{aligned} & N^{-\psi_f/2} \left(N^{-1} \sum_j N^{-\psi_f/2} T^{-1} \sum_{i,t} \phi_{fi}(m_1) \varepsilon_{it} \varepsilon_{jt} \right) \left(N^{-\psi_f} \sum_k \phi_{fk}(m_2) \right) \\ & = O_p \left(N^{-\psi_f/2} \Gamma_{N_f, T}^{-1} \right) \end{aligned}$$

by Lemma 1.16 and Assumption 2.2.

1.III is $O_p \left(N^{-\psi_f/2} \Gamma_{N_f, T}^{-1} \right)$, the proof is identical to 1.II.

1.IV can be written as

$$\begin{aligned}
& N^{-1/2} \left(N^{-\psi_f} \sum_i \phi_{fi}(m_1) \right) \left(N^{-\psi_f} \sum_l \phi_{fl}(m_2) \right) \left(N^{-1} \sum_k N^{-1/2} T^{-1} \sum_{j,t} \varepsilon_{jt} \varepsilon_{kt} \right) \\
&= N^{-1/2} O_p(1) O_p(1) O_p(\delta_{N,T}^{-1}) \\
&= O_p(\delta_{N,T}^{-1} N^{-1/2})
\end{aligned}$$

by Assumption 2.2 and Lemma 1.10.

1.V is $O_p(N^{-\psi_f})$. The proof is identical to 1.I.

1.VI can be written as

$$\begin{aligned}
& N^{-\frac{\psi_f-1}{2}} T^{-\frac{1}{2}} \left(T^{-1} \sum_s N^{-\psi_f/2} \sum_i \phi_{fi}(m_1) \varepsilon_{is} \right) \\
& \times \left(T^{-1/2} N^{-1/2} \sum_{j,t} \varepsilon_{jt} \right) \left(N^{-\psi_f} \sum_k \phi_{fk}(m_2) \right) \\
&= N^{-\frac{\psi_f-1}{2}} T^{-\frac{1}{2}} O_p(1) O_p(1) O_p(1) \\
&= O_p\left(N^{-\frac{\psi_f-1}{2}} T^{-\frac{1}{2}}\right)
\end{aligned}$$

by Assumption 2.2, 3.7 and Lemma 1.10.

1.VII is $O_p\left(N^{-\frac{\psi_f-1}{2}} T^{-\frac{1}{2}}\right)$. The proof is identical to 1.VI.

1.VIII can be written as

$$\begin{aligned}
& N^{-1} T^{-1} \left(N^{-\psi_f} \sum_i \phi_{fi}(m_1) \right) \left(N^{-1/2} T^{-1/2} \sum_{js} \varepsilon_{js} \right) \\
& \times \left(N^{-1/2} T^{-1/2} \sum_{kt} \varepsilon_{kt} \right) \left(N^{-\psi_f} \sum_l \phi_{fl}(m_2) \right) \\
&= N^{-1} T^{-1} O_p(1) O_p(1) O_p(1) \\
&= O_p(N^{-1} T^{-1})
\end{aligned}$$

by Assumptions 2.2 and Lemma 1.3.

Summing all these terms gives us $N^{-2\psi_f} T^{-1} \Phi_f' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \Phi_f$ is $O_p\left(N^{-\psi_f/2} \Gamma_{N_f, T}^{-1}\right)$.

By a symmetric argument $N^{-2\psi_g}T^{-1}\Phi'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\Phi_g$ is $O_p\left(N^{-\psi_g/2}\Gamma_{N_g,T}^{-1}\right)$.

Hence, $N^{-2\psi_f}T^{-1}\Phi'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\Phi_g$ is $O_p\left(\frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2}\Gamma_{N_g,T}}\right)$.

It is also easy to see from the proof presented above, that, depending on whether ψ_f is greater, equal to or less than ψ_g , $N^{-2\psi_f}T^{-1}\Phi'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\Phi_g$ is either $O_p\left(N^{-\psi_f/2}\Gamma_{N_f,T}^{-1}\right)$ (when $\psi_f > \psi_g$) or $O_p\left(\frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2}\Gamma_{N_g,T}}\right)$ (when $\psi_f < \psi_g$). When $\psi_f = \psi_g$, the two rates are equal.

Therefore, the matrix $N^{-2\psi_f}T^{-1}\Phi'\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\Phi$ is $O_p\left(N^{-\psi_f/2}\Gamma_{N_f,T}^{-1} \vee \frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2}\Gamma_{N_g,T}}\right)$.

Item 2 is given by

$$N^{-2\psi_f}T^{-2}\Phi'\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\mathbf{F} = \begin{bmatrix} N^{-2\psi_f}T^{-2}\Phi'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\mathbf{F} \\ N^{-2\psi_f}T^{-2}\Phi'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\mathbf{F} \end{bmatrix}.$$

We first show (below) that the $K_f \times K$ matrix $N^{-2\psi_f}T^{-2}\Phi'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\mathbf{F}$ is $O_p\left(T^{-1/2}N^{-\frac{3\psi_f}{2}+1}\delta_{N,T}^{-1}\right)$.

The matrix $N^{-2\psi_f}T^{-2}\Phi'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\mathbf{F}$ has generic (m_1, m_2) element,

$$\begin{aligned} & N^{-2\psi_f}T^{-2} \sum_{i,j,s,t} \phi_{jf}(m_1) \varepsilon_{jt} \varepsilon_{it} \varepsilon_{is} F_s(m_2) - N^{-2\psi_f}T^{-3} \sum_{i,j,s,t,u} \phi_{jf}(m_1) \varepsilon_{ju} \varepsilon_{iu} \varepsilon_{it} F_s(m_2) \\ & - N^{-2\psi_f-1}T^{-2} \sum_{i,j,k,s,t} \phi_{kf}(m_1) \varepsilon_{kt} \varepsilon_{jt} \varepsilon_{is} F_s(m_2) + N^{-2\psi_f-1}T^{-3} \sum_{i,j,k,s,t,u} \phi_{kf}(m_1) \varepsilon_{ku} \varepsilon_{ju} \varepsilon_{it} F_s(m_2) \\ & - N^{-2\psi_f}T^{-3} \sum_{i,j,s,t,u} \phi_{jf}(m_1) \varepsilon_{ju} \varepsilon_{it} \varepsilon_{is} F_s(m_2) + N^{-2\psi_f}T^{-4} \sum_{i,j,s,t,u,v} \phi_{jf}(m_1) \varepsilon_{jv} \varepsilon_{iu} \varepsilon_{it} F_s(m_2) \\ & + N^{-2\psi_f-1}T^{-3} \sum_{i,j,k,s,t,u} \phi_{kf}(m_1) \varepsilon_{ku} \varepsilon_{jt} \varepsilon_{is} F_s(m_2) - N^{-2\psi_f-1}T^{-4} \sum_{i,j,k,s,t,u,v} \phi_{kf}(m_1) \varepsilon_{kv} \varepsilon_{ju} \varepsilon_{it} F_s(m_2) \\ & - N^{-2\psi_f-1}T^{-2} \sum_{i,j,k,s,t} \phi_{kf}(m_1) \varepsilon_{jt} \varepsilon_{it} \varepsilon_{is} F_s(m_2) + N^{-2\psi_f-1}T^{-3} \sum_{i,j,k,s,t,u} \phi_{kf}(m_1) \varepsilon_{ju} \varepsilon_{iu} \varepsilon_{it} F_s(m_2) \\ & + N^{-2\psi_f-2}T^{-2} \sum_{i,j,k,l,s,t} \phi_{lf}(m_1) \varepsilon_{kt} \varepsilon_{jt} \varepsilon_{is} F_s(m_2) - N^{-2\psi_f-2}T^{-3} \sum_{i,j,k,l,s,t,u} \phi_{lf}(m_1) \varepsilon_{ku} \varepsilon_{ju} \varepsilon_{it} F_s(m_2) \\ & + N^{-2\psi_f-1}T^{-3} \sum_{i,j,k,s,t,u} \phi_{kf}(m_1) \varepsilon_{ju} \varepsilon_{it} \varepsilon_{is} F_s(m_2) - N^{-2\psi_f-1}T^{-4} \sum_{i,j,k,s,t,u,v} \phi_{kf}(m_1) \varepsilon_{jv} \varepsilon_{iu} \varepsilon_{it} F_s(m_2) \\ & - N^{-2\psi_f-2}T^{-3} \sum_{i,j,k,l,s,t,u} \phi_{lf}(m_1) \varepsilon_{ku} \varepsilon_{jt} \varepsilon_{is} F_s(m_2) + N^{-2\psi_f-2}T^{-4} \sum_{i,j,k,l,s,t,u,v} \phi_{lf}(m_1) \varepsilon_{kv} \varepsilon_{ju} \varepsilon_{it} F_s(m_2) \\ & = 2.I - \dots - 2.XVI. \end{aligned}$$

2.I: Using the Cauchy–Schwarz inequality, this term is bounded by

$$\begin{aligned}
& N^{-\frac{3\psi_f}{2}+1}T^{-1/2} \left(T^{-1} \sum_t [N^{-\psi_f/2} \sum_j \phi_{jf}(m_1) \varepsilon_{jt}]^2 \right)^{\frac{1}{2}} \left(T^{-1} \sum_t [N^{-1}T^{-1/2} \sum_{i,s} \varepsilon_{is} F_s(m_2) \varepsilon_{it}]^2 \right)^{\frac{1}{2}} \\
&= N^{-\frac{3\psi_f}{2}+1}T^{-1/2} O_p(\delta_{N,T}^{-1}) O_p(1) \\
&= O_p(T^{-1/2} N^{-\frac{3\psi_f}{2}+1} \delta_{N,T}^{-1})
\end{aligned}$$

by Lemma 1.12 and Assumption 3.7.

2.II: Using the Cauchy–Schwarz inequality, this term is bounded by,

$$\begin{aligned}
& N^{-\frac{3\psi_f}{2}+1}T^{-1/2} \left(T^{-1} \sum_u [N^{-\psi_f/2} \sum_j \phi_{jf}(m_1) \varepsilon_{ju}]^2 \right)^{\frac{1}{2}} \left(T^{-1} \sum_u [N^{-1}T^{-1/2} \sum_{i,t} \varepsilon_{it} \varepsilon_{iu}]^2 \right)^{\frac{1}{2}} \\
& \times \left(T^{-1} \sum_s F_s(m_2) \right).
\end{aligned}$$

This is $O_p(\delta_{N,T}^{-1} T^{-1/2} N^{-\frac{3\psi_f}{2}+1})$ by Lemma 1.10(a), Assumption 2.1 and 3.7.

2.III can be written as

$$\begin{aligned}
& N^{-\frac{3\psi_f}{2}+1}T^{-1/2} \left(N^{-\psi_f} \sum_k \phi_{kf}(m_1) \left[N^{-1/2}T^{-1} \sum_{j,t} \varepsilon_{kt} \varepsilon_{jt} \right] \right) \\
& \times \left(N^{-1} \sum_i T^{-1/2} \sum_s \varepsilon_{is} F_s(m_2) \right).
\end{aligned}$$

This is equal to $N^{\frac{1}{2}-\psi_f}T^{-1/2}O_p(\delta_{N,T}^{-1})O_p(1)$ which is $O_p(\delta_{N,T}^{-1}T^{-1/2}N^{\frac{1}{2}-\psi_f})$ by Lemma 1.10(b) and Assumption 2.2 and 3.6.

2.IV can be written as

$$\begin{aligned}
& N^{-\frac{3\psi_f+1}{2}}T^{-1/2} \left(N^{-1} \sum_j N^{-\psi_f/2}T^{-1} \sum_{k,u} \phi_{kf}(m_1) \varepsilon_{ku} \varepsilon_{ju} \right) \\
& \times \left(N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} \right) \left(T^{-1} \sum_s F_s(m_2) \right).
\end{aligned}$$

This is $O_p(\Gamma_{N_f,T}^{-1} N^{-\frac{3\psi_f+1}{2}}T^{-1/2})$ by Lemma 1.3 and 1.16, Assumption 2.1.

2.V can be written as

$$N^{-\frac{3\psi_f}{2}+1}T^{-1/2} \left(T^{-1} \sum_u N^{-\psi_f/2} \sum_{j=1} \phi_{jf}(m_1) \varepsilon_{ju} \right) \\ \times \left(T^{-1} \sum_t N^{-1}T^{-1/2} \sum_{i,s} F_s(m_2) \varepsilon_{is} \varepsilon_{it} \right).$$

This is $O_p(\delta_{N,T}^{-1}T^{-1/2}N^{-\frac{3\psi_f}{2}+1})$ by Assumption 3.7 and Lemma 1.12.

2.VI can be written as

$$N^{-\frac{3\psi_f}{2}+1}T^{-1/2} \left(T^{-1} \sum_v \left(N^{-\psi_f/2} \sum_j \phi_{jf}(m_1) \varepsilon_{jv} \right) \right) \\ \times \left(T^{-1} \sum_t N^{-1}T^{-1/2} \sum_{i,u} \varepsilon_{iu} \varepsilon_{it} \right) \left(T^{-1} \sum_s F_s(m_2) \right).$$

This is $O_p(\delta_{N,T}^{-1}N^{-\frac{3\psi_f}{2}+1}T^{-1/2})$ by Assumption 2.1, 3.7 and Lemma 1.10.

2.VII can be written as

$$N^{-\frac{3\psi_f+1}{2}}T^{-1} \left(T^{-1} \sum_u N^{-\psi_f/2} \sum_k \phi_{kf}(m_1) \varepsilon_{ku} \right) \\ \times \left(N^{-1/2}T^{-1/2} \sum_{j,t} \varepsilon_{jt} \right) \left(N^{-1} \sum_i T^{-1/2} \sum_s \varepsilon_{is} F_s(m_2) \right).$$

This is $O_p(N^{-\frac{3\psi_f+1}{2}}T^{-1})$ by Assumption 3.6, 3.7 and Lemma 1.3.

2.VIII can be written as

$$N^{-\frac{3\psi_f}{2}}T^{-1} \left(T^{-1} \sum_v N^{-\psi_f/2} \sum_k \phi_{kf}(m_1) \varepsilon_{kv} \right) \left(N^{-1/2}T^{-1/2} \sum_{j,u} \varepsilon_{ju} \right) \\ \times \left(N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} \right) \left(T^{-1} \sum_s F_s(m_2) \right).$$

This is $O_p(N^{-\frac{3\psi_f}{2}}T^{-1})$ by Assumption 3.7, Assumption 2.1 and Lemma 1.3.

2.IX: Using the Cauchy–Schwarz inequality, this term is bounded by

$$N^{-\psi_f + \frac{1}{2}} T^{-1/2} \left(N^{-\psi_f} \sum_k \phi_{kf}(m_1) \right) \left(T^{-1} \sum_t \left[N^{-\frac{1}{2}} \sum_j \varepsilon_{jt} \right]^2 \right)^{1/2} \\ \times \left(T^{-1} \sum_t \left[N^{-1} T^{-\frac{1}{2}} \sum_{i,s} F_s(m_2) \varepsilon_{is} \varepsilon_{it} \right]^2 \right)^{1/2},$$

which is $O_p(N^{-\psi_f + \frac{1}{2}} T^{-1/2} \delta_{N,T}^{-1})$ by Lemma 1.3, Lemma 1.12 and Assumption 2.2.

2.X: Using the Cauchy–Schwarz inequality, this term is bounded by

$$N^{-\psi_f + \frac{1}{2}} T^{-1/2} \left(N^{-\psi_f} \sum_k \phi_{kf}(m_1) \right) \left(T^{-1} \sum_u \left[N^{-\frac{1}{2}} \sum_j \varepsilon_{ju} \right]^2 \right)^{1/2} \\ \times \left(T^{-1} \sum_u \left[N^{-1} T^{-\frac{1}{2}} \sum_{i,t} \varepsilon_{iu} \varepsilon_{it} \right]^2 \right)^{1/2},$$

which is $O_p(N^{-\psi_f + \frac{1}{2}} T^{-1/2} \delta_{N,T}^{-1})$ by Assumptions 2.1, 2.2, Lemma 1.3 and Lemma 1.10.

2.XI can be written as

$$N^{-\psi_f + \frac{1}{2}} T^{-1/2} \left(N^{-\psi_f} \sum_l \phi_{lf}(m_1) \right) \left(N^{-1} \sum_j \left[N^{-\frac{1}{2}} T^{-1} \sum_{k,t} \varepsilon_{jt} \varepsilon_{kt} \right] \right) \\ \times \left(N^{-1} \sum_i T^{-1/2} \sum_s \varepsilon_{is} F_s(m_2) \right).$$

This is $O_p(N^{-\psi_f + \frac{1}{2}} T^{-1/2} \delta_{N,T}^{-1})$ by Assumptions 2.2, 3.6 and Lemma 1.10.

2.XII can be written as

$$N^{-\psi_f} T^{-1/2} \left(N^{-\psi_f} \sum_l \phi_{lf}(m_1) \right) \left(N^{-1} \sum_k \left[N^{-\frac{1}{2}} T^{-1} \sum_{j,u} \varepsilon_{ju} \varepsilon_{ku} \right] \right) \\ \times \left(N^{-\frac{1}{2}} T^{-\frac{1}{2}} \sum_{i,t} \varepsilon_{it} \right) \left(T^{-1} \sum_s F_s(m_2) \right).$$

This is $O_p(N^{-\psi_f} T^{-1/2} \delta_{N,T}^{-1})$ by Assumption 2.2 and Lemma 1.3 and 1.10.

2.XIII: Using the Cauchy–Schwarz inequality, this term is bounded by

$$N^{-\psi_f+1/2}T^{-1/2} \left(N^{-\psi_f} \sum_k \phi_{kf}(m_1) \right) \left(N^{-1/2}T^{-1/2} \sum_{j,u} \varepsilon_{ju} \right) \\ \times \left(N^{-1} \sum_i [T^{-1/2} \sum_t \varepsilon_{it}^2] \right)^{1/2} \left(N^{-1} \sum_i [T^{-1/2} \sum_s \varepsilon_{is} F_s(m_2)]^2 \right)^{1/2},$$

which is $O_p(N^{-\psi_f+1/2}T^{-1/2}\delta_{N,T}^{-1})$ by Assumption 2.2, 3.6 and Lemma 1.3.

2.XIV can be written as

$$N^{-\psi_f+1/2}T^{-1/2} \left(N^{-\psi_f} \sum_k \phi_{kf}(m_1) \right) \left(N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} \right) \\ \times \left(N^{-1} \sum_u \left[N^{-1/2}T^{-1} \sum_{i,t} \varepsilon_{it} \varepsilon_{iu} \right] \right) \left(T^{-1} \sum_s F_s(m_2) \right).$$

This is $O_p(N^{-\psi_f}T^{-1/2}\delta_{N,T}^{-1})$ by Assumption 2.2, Lemma 1.3 and 1.10.

2.XV can be written as

$$N^{-\psi_f}T^{-3/2} \left(N^{-\psi_f} \sum_l \phi_{lf}(m_1) \right) \left(N^{-1/2}T^{-1/2} \sum_{k,u} \varepsilon_{ku} \right) \\ \times \left(N^{-1/2}T^{-1/2} \sum_{j,t} \varepsilon_{jt} \right) \left(N^{-1} \sum_i T^{-1/2} \sum_s \varepsilon_{is} F_s(m_2) \right).$$

This is $O_p(N^{-\psi_f}T^{-3/2})$ by Assumption 2.2, 3.6 and Lemma 1.3.

2.XVI can be written as

$$N^{-\psi_f-\frac{1}{2}}T^{-3/2} \left(N^{-\psi_f} \sum_l \phi_{lf}(m_1) \right) \left(N^{-1/2}T^{-1/2} \sum_{k,v} \varepsilon_{kv} \right) \\ \times \left(N^{-1/2}T^{-1/2} \sum_{j,u} \varepsilon_{ju} \right) \left(N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} \right) \left(T^{-1} \sum_s F_s(m_2) \right).$$

This is $O_p(N^{-\psi_f-\frac{1}{2}}T^{-3/2})$ by Assumption 2.1, 2.2, Lemma 1.3.

Since $0 < \psi_f \leq 1$, the initial terms dominate the order. Thus, summing all these terms yields:

$$N^{-2\psi_f} T^{-2} \Phi'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = \mathcal{O}_p \left(T^{-\frac{1}{2}} N^{-\frac{3\psi_f}{2}+1} \delta_{N,T}^{-1} \right).$$

By a symmetric argument, we have:

$$N^{-2\psi_g} T^{-2} \Phi'_g \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = \mathcal{O}_p \left(T^{-\frac{1}{2}} N^{-\frac{3\psi_g}{2}+1} \delta_{N,T}^{-1} \right).$$

Hence, we can conclude that:

$$N^{-2\psi_f} T^{-2} \Phi'_g \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = \mathcal{O}_p \left(\frac{N^{2(\psi_g - \psi_f)}}{T^{1/2} N^{\frac{3\psi_g}{2}-1} \delta_{N,T}} \right).$$

Therefore, we have:

$$N^{-2\psi_f} T^{-2} \Phi' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} = \mathcal{O}_p \left(T^{-\frac{1}{2}} N^{-\frac{3\psi_f}{2}+1} \delta_{N,T}^{-1} \vee \frac{N^{2(\psi_g - \psi_f)}}{T^{1/2} N^{\frac{3\psi_g}{2}-1} \delta_{N,T}} \right).$$

Item 3 is a $K \times L$ matrix and the proof follows the same logic as Item 2, replacing $F_s(m_2)$ by $\omega_s(m_2)$.

A3 Proofs of Theorem 1(a), 2, 3(a) and 4(a)

We introduce two lemmas which are utilized in proofs of Stage 1 results of 3PRF LASSO for the cases when $\gamma = 0$, and $\zeta = 0$. The proofs are presented subsequent to these lemmas.

Lemma 4. Recall that $\Xi_{N,T}^{-1} \equiv T^{-1/2} \vee N^{-\psi_f/2} \vee \left(\frac{N^{\psi_g - \psi_f}}{\Gamma_{N_g,T}} \right)$ which is equivalent to $T^{-1/2} \vee N^{-\psi_f/2} \vee N^{-\psi_f + \psi_g/2} \vee \left(\frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \right)$.

Under Assumptions 1-6, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, if $\gamma = 0$, and $\zeta = 0$, we have:

$$1. \hat{\mathbf{F}}_A = T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} = \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f + \boldsymbol{\Delta}_\omega + \mathcal{O}_p(T^{-1/2}).$$

$$2. \hat{\mathbf{F}}_B = N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f + \mathcal{O}_p \left(\Xi_{N,T}^{-1} \right).$$

$$3. \hat{\mathbf{F}}_{C,t} = N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t = N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}_t + \mathcal{O}_p \left(\Xi_{N,T}^{-1} \right).$$

Furthermore, the probability limit of $\hat{\mathbf{\Phi}}'$ and an asymptotic representation of $\hat{\mathbf{F}}_t$ are

$$\hat{\mathbf{\Phi}}' \xrightarrow[T \rightarrow \infty]{p} (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Lambda}_f' + \mathbf{\Delta}_\omega)^{-1} \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Phi}_f',$$

and

$$\hat{\mathbf{F}}_t = (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Lambda}_f' + \mathbf{\Delta}_\omega) (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathbf{\Lambda}_f')^{-1} \left(N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}_t \right) + o_p(1).$$

Proof:

First we note that

$$\begin{aligned} N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} &= \begin{bmatrix} N^{-\psi_f} \mathbf{\Phi}_f' \mathbf{J}_N \mathbf{\Phi}_f & N^{-\psi_f} \mathbf{\Phi}_f' \mathbf{J}_N \mathbf{\Phi}_g \\ N^{-\psi_f} \mathbf{\Phi}_g' \mathbf{J}_N \mathbf{\Phi}_f & N^{\psi_g - \psi_f} (N^{-\psi_g} \mathbf{\Phi}_g' \mathbf{J}_N \mathbf{\Phi}_g) \end{bmatrix} \\ &= \mathcal{O}_p \left(1 \vee N^{\psi_g - \psi_f} \right). \end{aligned} \tag{A3.1}$$

The final equality follows from Assumption 2.2 and 4.1.

Item 1 :

$$\begin{aligned} \hat{\mathbf{F}}_A &= T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} \\ &= \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega}) + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\omega} \\ &= \mathbf{\Lambda} \mathbf{\Delta}_F \mathbf{\Lambda}' + \mathbf{\Delta}_\omega + \mathcal{O}_p \left(T^{-1/2} \right) \\ &= \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Lambda}_f' + \mathbf{\Delta}_\omega + \mathcal{O}_p \left(T^{-1/2} \right). \end{aligned}$$

The first equality follows from Assumptions 2.1, 2.4 and Lemma 2.1 and the final equality follows from the fact that $\mathbf{\Delta}_F$ is block diagonal (Assumption 4) and $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$ by Assumption 5.

Item 2:

$$\begin{aligned}
\hat{\mathbf{F}}_B &= N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \\
&= \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega}) \\
&\quad + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} \right) \mathbf{\Lambda}' + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} \right) \\
&\quad + \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega}) \\
&\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \mathbf{\Lambda} \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} \right) \mathbf{\Lambda}' + \frac{N^{1-\psi_f}}{\sqrt{T}} \mathbf{\Lambda} \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} \right) \\
&\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega}) \\
&\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} \right) \mathbf{\Lambda}' + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} \right) \\
&\quad + \left(N^{-\psi_f} T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \left(N^{-\psi_f} T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega}) \\
&\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F} \right) \mathbf{\Lambda}' + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega} \right) \\
&= \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \mathcal{O}_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

The final equality follows from Assumptions 2.1 and 4, Lemma 2 and equation A3.1.

Now, using the fact that $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$ (Assumption 5), we have

$$\begin{aligned}
\mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' &= \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \mathbf{\Lambda}'_f \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \mathbf{f}) \mathbf{\Lambda}'_f \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \mathbf{\Lambda}'_f \\
&\quad + \frac{N^{\psi_g - \psi_f}}{T} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g} \right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g \right) \\
&\quad \times \left(T^{-1/2} \mathbf{g}' \mathbf{J}_T \mathbf{f} \right) \mathbf{\Lambda}'_f \\
&= \mathbf{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathbf{\Lambda}'_f + \mathcal{O}_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

The final equality follows from Assumptions 2.1 and 4.

Hence, $\hat{\mathbf{F}}_B$, which is $N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z}$ is equal to $\mathbf{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathbf{\Lambda}'_f + \mathcal{O}_p \left(\Xi_{N,T}^{-1} \right)$.

Item 3:

$$\begin{aligned}
\hat{\mathbf{F}}_{C,t} &= N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t \\
&= \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \phi_0 \right) + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) \mathbf{F}_t \\
&\quad + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \varepsilon_t \right) + \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon \mathbf{J}_N \phi_0 \right) \\
&\quad + \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon \mathbf{J}_N \mathbf{\Phi} \right) \mathbf{F}_t + \frac{N^{1-\psi_f}}{\sqrt{T}} \mathbf{\Lambda} \left(N^{-1} T^{-1/2} \mathbf{F}' \mathbf{J}_T \varepsilon \mathbf{J}_N \varepsilon_t \right) \\
&\quad + (T^{-1} \mathbf{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \phi_0 \right) + (T^{-1} \mathbf{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) \mathbf{F}_t \\
&\quad + (T^{-1} \mathbf{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \varepsilon_t \right) + \left(N^{-\psi_f} T^{-1} \mathbf{\omega}' \mathbf{J}_T \varepsilon \mathbf{J}_N \phi_0 \right) \\
&\quad + \left(N^{-\psi_f} T^{-1} \mathbf{\omega}' \mathbf{J}_T \varepsilon \mathbf{J}_N \mathbf{\Phi} \right) \mathbf{F}_t + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-1/2} \mathbf{\omega}' \mathbf{J}_T \varepsilon \mathbf{J}_N \varepsilon_t \right) \\
&= N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) \mathbf{F}_t + O_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

Assumptions 2.1 and 4, Lemma 2 and equation A3.1 give the final equality.

Again using Assumption 5, $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$, we have,

$$\begin{aligned}
\mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) \mathbf{F}_t &= \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_f \right) \mathbf{f}_t \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g \right) \mathbf{g}_t \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f \right) \mathbf{f}_t \\
&\quad + \frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g} \right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g \right) \mathbf{g}_t \\
&= \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}_t + O_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

The final equality again follows from Assumption 2.1, 4.

Hence, $\hat{\mathbf{F}}_{C,t} = N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t = N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}_t + O_p \left(\Xi_{N,T}^{-1} \right)$.

Combining the results for Items 1-3 and using Assumption 6, we have:

$$\begin{aligned}
\hat{\mathbf{F}}_t &= T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} \left(N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t \\
&= \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \hat{\mathbf{F}}_{C,t} \\
&= (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Lambda}'_f + \mathbf{\Delta}_\omega) (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathbf{\Lambda}'_f)^{-1} \left(N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}_t \right) + o_p(1).
\end{aligned}$$

Similarly expanding $T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{X}$ gives $T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{X} = \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathbf{\Phi}'_f + \mathbf{O}_p(T^{-1/2})$ and hence,

$$\hat{\mathbf{\Phi}}' = (T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{Z})^{-1} (T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{X}) \xrightarrow[T \rightarrow \infty]{p} (\mathbf{\Lambda}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f + \mathbf{\Delta}_\omega)^{-1} \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathbf{\Phi}'_f.$$

Lemma 5. Under Assumptions 1- 6, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, if $\gamma = 0$, and $\zeta = 0$, we have:

1. $\hat{\beta}_1 = T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{Z} = \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f + \mathbf{\Delta}_\omega + \mathbf{O}_p(T^{-1/2})$.
2. $\hat{\beta}_2 = N^{-\psi_f}T^{-2}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{Z} = \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f + \mathbf{O}_p\left(\Xi_{N,T}^{-1}\right)$.
3. $\hat{\beta}_3 = N^{-2\psi_f}T^{-3}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{Z} = \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f + \mathbf{O}_p\left(\Xi_{N,T}^{-1}\right)$.
4. $\hat{\beta}_4 = N^{-\psi_f}T^{-2}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{y} = \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathbf{\beta}_f + \mathbf{O}_p\left(\Xi_{N,T}^{-1}\right)$.

Therefore,

$$\begin{aligned} \hat{\beta} &= (T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{Z})^{-1} N^{-\psi_f}T^{-2}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{Z} \\ &\quad \times \left(N^{-2\psi_f}T^{-3}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{Z} \right)^{-1} N^{-\psi_f}T^{-2}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{y} \\ &= \hat{\beta}_1^{-1} \hat{\beta}_2 \hat{\beta}_3^{-1} \hat{\beta}_4 \end{aligned}$$

satisfies

$$\hat{\beta} \xrightarrow[T, N \rightarrow \infty]{p} (\mathbf{\Lambda}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f + \mathbf{\Delta}_\omega)^{-1} \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f (\mathbf{\Lambda}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathbf{\Lambda}'_f)^{-1} \mathbf{\Lambda}_f\mathbf{\Delta}_f\mathcal{P}_f\mathbf{\Delta}_f\mathbf{\beta}_f.$$

Proof:

Note that $\hat{\beta}_1 = \hat{\mathbf{F}}_A$ and $\hat{\beta}_2 = \hat{\mathbf{F}}_B$ and their probability limits are established in Lemma 4. The expressions for $\hat{\beta}_3$ and $\hat{\beta}_4$ are handled below.

[illegible]

$$\begin{aligned}
& + \omega' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T F \Lambda' + \omega' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T \omega \\
& + \omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T F \Lambda' + \omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T \omega \\
& + \omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F \Lambda' + \omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega \\
& + \omega' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T F \Lambda' + \omega' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T \omega \\
& + \omega' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T F \Lambda' + \omega' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T \omega \\
& + \omega' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T F \Lambda' + \omega' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T \omega \\
& + \omega' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T F \Lambda' + \omega' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T \omega \\
& = (N^{2\psi_f} T^3)^{-1} \Lambda F' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T F \Lambda' + O_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

The final equality follows from Assumption 2.1, Lemmas 2 and 3 and equation A3.1. Further, note that

$$\begin{aligned}
& (N^{-2\psi_f} T^{-3}) \Lambda F' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T F \Lambda' \\
& = \frac{1}{T} \left((N^{-\psi_f} T^{-1}) \Lambda F' J_T F \Phi' J_N \Phi F' \right) J_T \left((N^{-\psi_f} T^{-1}) \Lambda F' J_T F \Phi' J_N \Phi F' \right)' \\
& = \frac{1}{T} \left(\Lambda_f \Delta_f \mathcal{P}_f f' + O_p \left(\Xi_{N,T}^{-1} \right) \right) J_T \left(f \mathcal{P}_f \Delta_f \Lambda_f' + O_p \left(\Xi_{N,T}^{-1} \right) \right) \\
& = \Lambda_f \Delta_f \mathcal{P}_f \frac{f' J_T f}{T} \mathcal{P}_f \Delta_f \Lambda_f' + O_p \left(\Xi_{N,T}^{-1} \right) \\
& = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p \left(1/\sqrt{T} \right) + O_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

We have used the result from the proof of Lemma 4.3 in the third equality. Standard arguments, then, give us the fourth equality and the final equality follows from Assumption 2.1. Hence, $\hat{\beta}_3 = N^{-2\psi_f} T^{-3} \mathbf{Z}' J_T \mathbf{X} J_N \mathbf{X}' J_T \mathbf{X} J_N \mathbf{X}' J_T \mathbf{Z} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p \left(\Xi_{N,T}^{-1} \right)$.

$$\begin{aligned}
\hat{\beta}_4 = & N^{-\psi_f} T^{-2} (\Lambda F' J_T F \Phi' J_N \Phi F' J_T F \beta + \Lambda F' J_T F \Phi' J_N \Phi F' J_T \eta + \Lambda F' J_T F \Phi' J_N \varepsilon' J_T F \beta \\
& + \Lambda F' J_T F \Phi' J_N \varepsilon' J_T \eta + \omega' J_T F \Phi' J_N \Phi F' J_T F \beta + \omega' J_T F \Phi' J_N \Phi F' J_T \eta \\
& + \omega' J_T F \Phi' J_N \varepsilon' J_T F \beta + \omega' J_T F \Phi' J_N \varepsilon' J_T \eta + \Lambda F' J_T \varepsilon J_N \Phi F' J_T F \beta \\
& + \Lambda F' J_T \varepsilon J_N \Phi F' J_T \eta + \Lambda F' J_T \varepsilon J_N \varepsilon' J_T F \beta + \Lambda F' J_T \varepsilon J_N \varepsilon' J_T \eta \\
& + \omega' J_T \varepsilon J_N \Phi F J_T F \beta + \omega' J_T \varepsilon J_N \Phi F J_T \eta + \omega' J_T \varepsilon J_N \varepsilon' J_T F \beta \\
& + \omega' J_T \varepsilon J_N \varepsilon' J_T \eta) \\
= & N^{-\psi_f} T^{-2} (\Lambda F' J_T F \Phi' J_N \Phi F' J_T F \beta) + O_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

The final equality follows from Assumption 2.1, Lemma 2 and equation A3.1.

Further, note that, since $\Lambda = \begin{bmatrix} \Lambda_f & \mathbf{0} \end{bmatrix}$ (Assumption 5) and $\beta = (\beta'_f, 0')'$ (Assumption 1), we have

$$\begin{aligned}
\Lambda (T^{-1} F' J_T F) (N^{-\psi_f} \Phi' J_N \Phi) (T^{-1} F' J_T F) \beta = & \Lambda_f (T^{-1} f' J_T f) (N^{-\psi_f} \Phi'_f J_N \Phi_f) (T^{-1} f' J_T f) \beta_f \\
& + \Lambda_f (T^{-1} f' J_T f) (N^{-\psi_f} \Phi'_f J_N \Phi_g) (T^{-1} g' J_T f) \beta_f \\
& + \Lambda_f (T^{-1} f' J_T g) (N^{-\psi_f} \Phi'_g J_N \Phi_f) (T^{-1} f' J_T f) \beta_f \\
& + \frac{N^{\psi_g - \psi_f}}{T} \Lambda_f (T^{-1/2} f' J_T g) (N^{-\psi_g} \Phi'_g J_N \Phi_g) \\
& \times (T^{-1/2} g' J_T f) \beta_f \\
= & \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p \left(\Xi_{N,T}^{-1} \right).
\end{aligned}$$

The final equality follows from Assumption 2.1, 4 and Lemma 2 and equation A3.1.

Hence, $\hat{\beta}_4 = N^{-\psi_f} T^{-2} Z' J_T X J_N X' J_T y = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p \left(\Xi_{N,T}^{-1} \right)$. Combining these results gives the probability limit of $\hat{\beta}$ stated in Lemma 5.

Using Lemmas 1-5, we now prove Theorems 1(a), 2, 3(a) and 4(a).

Theorem 1(a) Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, then we have:

$$\hat{y}_{t+h,f} - \mathbb{E}_t y_{t+h} = O_p(\Xi_{N,T}^{-1})$$

Proof: Let $\bar{\mathbf{f}} = \frac{\sum_{s=1}^T \mathbf{f}_s}{T}$. We have,

$$\begin{aligned}
\hat{y}_{t+h,f} &= \bar{y} + \left(N^{-\psi_f} T^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) \mathbf{W}_{XZ} \right) \left(N^{-2\psi_f} T^{-3} \mathbf{W}'_{XZ} \mathbf{S}_{XX} \mathbf{W}_{XZ} \right)^{-1} \left(N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{Xy} \right) \\
&= \beta_0 + \bar{\mathbf{f}}' \beta_f + O_p(T^{-1/2}) + \left((\mathbf{f}_t - \bar{\mathbf{f}})' \mathcal{P}_f \Delta_f \Lambda'_f + O_p(\Xi_{N,T}^{-1}) \right) \\
&\quad \times \left[\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f + O_p(\Xi_{N,T}^{-1}) \right]^{-1} \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p(\Xi_{N,T}^{-1}) \right) \\
&= \beta_0 + \bar{\mathbf{f}}' \beta_f + O_p(T^{-1/2}) + (\mathbf{f}_t - \bar{\mathbf{f}})' \mathcal{P}_f \Delta_f \Lambda'_f \left[\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right]^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p(\Xi_{N,T}^{-1}) \\
&= \beta_0 + \mathbf{f}_t' \mathcal{P}_f \Delta_f \Lambda'_f \left[\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right]^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p(\Xi_{N,T}^{-1}) \\
&= \beta_0 + \mathbf{f}_t' \beta_f + O_p(\Xi_{N,T}^{-1}) \\
&= \mathbb{E}_t y_{t+h} + O_p(\Xi_{N,T}^{-1}).
\end{aligned}$$

The third equality follows since, for any invertible matrices $\mathbf{A}$ and $\mathbf{A} + \mathbf{B}$, we have:

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{A} + \mathbf{B})^{-1}, \quad (\text{A3.2})$$

which implies that:

$$\begin{aligned}
\left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f + O_p(\Xi_{N,T}^{-1}) \right)^{-1} &= \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right)^{-1} \\
&\quad - \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right)^{-1} O_p(\Xi_{N,T}^{-1}) \\
&\quad \times \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f + O_p(\Xi_{N,T}^{-1}) \right)^{-1} \\
&= \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right)^{-1} - O_p(1) O_p(\Xi_{N,T}^{-1}) O_p(1) \\
&= \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right)^{-1} - O_p(\Xi_{N,T}^{-1}).
\end{aligned}$$

The final equality comes from Assumptions 4 and 5, which require Λ_f , $\mathcal{P}_f$, and Δ_f to be nonsingular. The stochastic orders in the expression are obtained using Lemmas 4 and 5 and noting that $\frac{\sum_{s=1}^T \eta_{h+s}}{T} = O_p(T^{-1/2})$.

Theorem 2 Let $\hat{\alpha}_i$ denote the i^{th} element of $\hat{\alpha}$. Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$

and $\mathcal{P}_f = \mathbb{I}_{K_f}$, then for any i ,

$$N^{\psi_f} \hat{\alpha}_i - \left(\phi_{if} - N^{\psi_f-1} \bar{\phi}_f \right)' \beta_f \xrightarrow[T, N \rightarrow \infty]{p} 0.$$

Proof: $\hat{\alpha}_i = \mathbf{S}_i \hat{\alpha}$, where $\mathbf{S}_i$ is the $(1 \times N)$ selector vector with i^{th} element equal to one and the remaining elements equal to zero. Using the expression for $\hat{\alpha}$ we have,

$$\begin{aligned} \hat{\alpha}_i &= N^{-\psi_f} T^{-1} \mathbf{S}_i \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \left(N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} \\ &\quad \times N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y}. \end{aligned}$$

From Lemma 4 and 5, we have

$$\hat{\alpha}_i = N^{-\psi_f} \left[\mathbf{S}_i \mathbf{J}_N \mathbf{\Phi}_f \Delta_f \Lambda_f' \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' \right)^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + o_p(1) \right].$$

The expression $\mathbf{S}_i \mathbf{J}_N \mathbf{\Phi}_f$ can be written as $\left(\phi_{if} - N^{\psi_f-1} \bar{\phi}_f \right)' + o_p(1)$ as $N, T \rightarrow \infty$. Therefore, we have that

$$N^{\psi_f} \hat{\alpha}_i \xrightarrow[T, N \rightarrow \infty]{p} \left(\phi_{if} - N^{\psi_f-1} \bar{\phi}_f \right)' \Delta_f \Lambda_f' \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' \right)^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f.$$

Using the fact that $\mathcal{P}_f = \mathbb{I}$ this reduces to $\left(\phi_{if} - N^{\psi_f-1} \bar{\phi}_f \right)' \beta_f$.

Define $\mathbf{G}_\beta \equiv \hat{\beta}_1^{-1} \hat{\beta}_2 \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' \right)^{-1} \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \right)$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are defined in Lemma 5.

Theorem 3(a) Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, then we have:

$$\hat{\beta} - \mathbf{G}_\beta \beta_f = O_p(\Xi_{N,T}^{-1}).$$

Proof:

$$\begin{aligned}
\hat{\beta} &= (T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z})^{-1} N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \\
&\quad \times \left(N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} \\
&= \hat{\beta}_1^{-1} \hat{\beta}_2 \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p(\Xi_{N,T}^{-1}) \right)^{-1} \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p(\Xi_{N,T}^{-1}) \right) \\
&= \hat{\beta}_1^{-1} \hat{\beta}_2 (\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f')^{-1} (\Lambda_f \Delta_f \mathcal{P}_f \Delta_f) \beta_f + O_p(\Xi_{N,T}^{-1}) \\
&= \mathbf{G}_\beta \beta_f + O_p(\Xi_{N,T}^{-1}).
\end{aligned}$$

The stochastic orders in the expression are obtained using Lemmas 4 and 5. The second equality follows by employing the identity for the inverse of a sum of two matrices as in the proof of Theorem 1(a).

Define $\mathbf{H}_f \equiv \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \Lambda_f \Delta_f \mathcal{P}_f$ and $\mathbf{H}_0 \equiv \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} [N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0]$.

Theorem 4(a) Let Assumptions 1-6 hold and $\gamma = 0$ and $\zeta = 0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, we have:

$$\hat{\mathbf{F}}_t - (\mathbf{H}_0 + \mathbf{H}_f \mathbf{f}_t) = O_p(\Xi_{N,T}^{-1}).$$

Proof:

$$\begin{aligned}
\hat{\mathbf{F}}_t &= T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} \left(N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t \\
&= \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \left[N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \Lambda_f \Delta_f \mathcal{P}_f \mathbf{f}_t + O_p(\Xi_{N,T}^{-1}) \right] \\
&= \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \left[N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 \right] + \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \Lambda_f \Delta_f \mathcal{P}_f \mathbf{f}_t + O_p(\Xi_{N,T}^{-1}) \\
&= \mathbf{H}_0 + \mathbf{H}_f \mathbf{f}_t + O_p(\Xi_{N,T}^{-1}).
\end{aligned}$$

The second equality follows from employing the expression for $\hat{\mathbf{F}}_{C,t}$ in Lemma 4. $\mathbf{H}_f' \mathbf{G}_\beta = \mathbf{I}$ can be verified easily given Assumptions 4 and 5.

Remark 1. The proofs of Theorem 1, 2, and 4 can be approached in an alternative manner. We can demonstrate that, for the matrix $\mathbf{H}_2 = \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} [N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \Phi_f]$, $\hat{\mathbf{F}}_t$ converges to $\mathbf{H}_2 \mathbf{f}_t$ at the rate $\Xi_{N,T}$, while $\hat{\beta}$ converges to $\mathbf{H}_2^{-1} \beta$ at $\min(\sqrt{N^{\psi_f}}, \sqrt{T})$ rate, under the assumptions of

our model. Therefore, by specifying a different limit, we can establish faster convergence of $\hat{\beta}$ to that limit. Essentially, we require that rotations in $\hat{\mathbf{F}}_t$ and $\hat{\beta}$ be nullified upon multiplication, which occurs with this newly specified limit.

However, we specify the matrix $\mathbf{H}_f$ such that $\hat{\beta}$ converges to $\mathbf{H}_f'^{-1}\beta$ at the slower $\Xi_{N,T}$ rate. We do this for simplicity of exposition, noting that the convergence rate of the target depends on the convergence rates of both $\hat{\mathbf{F}}_t$ and $\hat{\beta}$. Consequently, any improvement in the convergence result for $\hat{\beta}$ is not useful unless the rate for $\hat{\mathbf{F}}_t$ improves as well.

In [Kelly & Pruitt \[2015\]](#), they specify the convergence of $\hat{\mathbf{F}}_t$ to $\mathbf{H}\mathbf{F}_t$, where $\mathbf{H} = \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} [N^{-1}T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\Phi]$ at a $\sqrt{N}$ rate. In our weak factor context, this would be a $\sqrt{N^{\psi_f}}$ rate. However, this matrix $\mathbf{H}$ is not square unless there are no irrelevant factors. Elements of $\mathbf{H}\mathbf{F}_t$ are linear combinations of relevant and irrelevant factors. Therefore, their convergence result cannot be employed in our context to obtain faster convergence for the target estimator. We must establish convergence of our factor estimates to some rotation of relevant factors.

A4 Proofs of Theorems 1(b), 3(b) and 4(b)

We introduce three lemmas which are employed in the subsequent proofs of Theorems 1(b), 3(b), and 4(b), which deal with the general setting of $\zeta \neq 0$ and $\gamma \neq 0$.

Lemma 6. *Let Assumptions [1-3](#), [6](#), [8](#) and [9](#) hold. Additionally, let $\frac{T}{N} = O(1)$. Then:*

1. $N^{-1}T^{-1/2}\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}_t = O_p\left(\delta_{N,T}^{-1}\right).$
2. $N^{-1}T^{-3/2}\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p\left(\delta_{N,T}^{-1}\right).$
3. $N^{-1}T^{-3/2}\mathbf{F}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p\left(\delta_{N,T}^{-1}\right).$
4. $N^{-\psi_f/2}T^{-1}\Phi'\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p\left(\Xi_{N,T}^{-1}\right).$
5. $N^{-1}T^{-3/2}\boldsymbol{\omega}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p\left(\delta_{N,T}^{-1}\right).$
6. $N^{-1}T^{-3/2}\boldsymbol{\eta}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p\left(\delta_{N,T}^{-1}\right).$
7. (a) $T^{-1/2}\mathbf{F}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p(1)$ and (b) $T^{-1/2}\boldsymbol{\omega}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p(1).$

To prove this lemma, we need to show the following.

Let Assumptions 1-3 and 6-8 hold. Then, for all t, m ,

$$N^{-1}T^{-1/2} \sum_{i,s} \varepsilon_s(m) \varepsilon_{is} \varepsilon_{it} = O_p \left(\delta_{N,T}^{-1} \right).$$

Proof: Adding and subtracting terms, we can write the above as

$$\begin{aligned} & N^{-1/2} \left(N^{-1/2} T^{-1/2} \sum_{i,s} \varepsilon_s(m) [\varepsilon_{is} \varepsilon_{it} - \sigma_{ii,st}] \right) + T^{-1/2} \left(N^{-1} \sum_{i,s} \varepsilon_s(m) \sigma_{ii,st} \right) \\ & = I + II. \end{aligned}$$

$$\mathbb{E} \left| N^{-1} \sum_{i,s} \varepsilon_s(m) \sigma_{ii,st} \right| \leq N^{-1} \max_s \mathbb{E} |\varepsilon_s(m)| \sum_{i,s} |\sigma_{ii,st}| = O_p(1) \text{ by Assumptions 3.1 and 2.3.}$$

Hence, the second term is $O_p(T^{-1/2})$.

For the first term to be $O_p(N^{-1/2})$, it is sufficient to show that

$$\mathbb{E} \left| (NT)^{-1/2} \sum_{i,s} \varepsilon_s(m) [\varepsilon_{is} \varepsilon_{it} - \sigma_{ii,st}] \right|^2 \leq M.$$

Proof: Using the Cauchy–Schwarz inequality twice,

$$\begin{aligned} & \mathbb{E} \left| (NT)^{-1/2} \sum_{i,s} \varepsilon_s(m) [\varepsilon_{is} \varepsilon_{it} - \sigma_{ii,st}] \right|^2 \\ & = \mathbb{E} \left[(NT)^{-1} \sum_{i,j,s,u} \varepsilon_s(m) \varepsilon_u(m) [\varepsilon_{is} \varepsilon_{it} - \sigma_{ii,st}] [\varepsilon_{ju} \varepsilon_{jt} - \sigma_{jj,ut}] \right] \\ & \leq \max_{s,u} \left(\mathbb{E} |\varepsilon_s(m) \varepsilon_u(m)|^2 \right)^{1/2} \\ & \quad \times \left(\mathbb{E} \left[(NT)^{-1} \sum_{i,j,s,u} [\varepsilon_{is} \varepsilon_{it} - \sigma_{ii,st}] [\varepsilon_{ju} \varepsilon_{jt} - \sigma_{jj,ut}] \right]^2 \right)^{1/2} \\ & \leq \max_{s,u} \left(\mathbb{E} |\varepsilon_s(m)|^4 \right)^{1/4} \left(\mathbb{E} |\varepsilon_u(m)|^4 \right)^{1/4} \left(\mathbb{E} \left| (NT)^{-1/2} \sum_{i,s} [\varepsilon_{is} \varepsilon_{it} - \sigma_{ii,st}] \right|^4 \right)^{1/2} < \infty \end{aligned}$$

by Assumptions 2.3 and 3.2. Therefore, we have

$$N^{-1}T^{-1/2} \sum_{i,s} \varepsilon_s(m) \varepsilon_{is} \varepsilon_{it} = O_p \left(\delta_{N,T}^{-1} \right). \quad (\text{A4.1})$$

Now we can prove Lemma 6.

Item 1 = $N^{-1}T^{-1/2} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}_t$ has generic m^{th} element given by

$$\begin{aligned} & N^{-1}T^{-1/2} \sum_{i,s} \varepsilon_s(m) \varepsilon_{is} \varepsilon_{it} - N^{-2}T^{-1/2} \sum_{i,j,s} \varepsilon_s(m) \varepsilon_{is} \varepsilon_{jt} \\ & - N^{-1}T^{-3/2} \sum_{i,s,u} \varepsilon_s(m) \varepsilon_{iu} \varepsilon_{it} + N^{-2}T^{-3/2} \sum_{i,j,s,u} \varepsilon_s(m) \varepsilon_{iu} \varepsilon_{jt} \\ & = 1.\text{I} - 1.\text{II} - 1.\text{III} + 1.\text{IV}. \end{aligned}$$

1.I is $O_p \left(\delta_{N,T}^{-1} \right)$ by equation A4.1.

$$\begin{aligned} 1.\text{II} &= N^{-1/2} \left(N^{-1} \sum_{i \in \Delta_{m,\varepsilon}} \left(T^{-1/2} \sum_s \varepsilon_s(m) \varepsilon_{is} \right) + \frac{T^{1/2}}{N} \sum_{i \in \Delta_{m,\varepsilon}^c} \left(T^{-1} \sum_s \varepsilon_s(m) \varepsilon_{is} \right) \right) \left(N^{-1/2} \sum_j \varepsilon_{jt} \right) \\ &= N^{-1/2} O_p(1) \\ &= O_p \left(N^{-1/2} \right) \end{aligned}$$

by Assumption 8 and Lemma 1.3.

1.III is $O_p \left(\delta_{N,T}^{-1} T^{-1/2} \right)$ by Lemma 1.3 and 1.10.

1.IV is $O_p \left(N^{-1} T^{-1/2} \right)$ by Lemma 1.3. Summing these terms, Item 1 is $O_p \left(\delta_{N,T}^{-1} \right)$.

Item 2: $N^{-1}T^{-3/2}\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}$ is an $N \times N$ matrix with generic (m_1, m_2) element

$$\begin{aligned}
& N^{-1}T^{-3/2} \sum_{i,s,t} \varepsilon_s(m_1) \varepsilon_{is} \varepsilon_{it} \varepsilon_t(m_2) - N^{-1}T^{-5/2} \sum_{i,s,t,u} \varepsilon_s(m_1) \varepsilon_{is} \varepsilon_{it} \varepsilon_u(m_2) \\
& - N^{-1}T^{-5/2} \sum_{i,s,t,u} \varepsilon_s(m_1) \varepsilon_{it} \varepsilon_{iu} \varepsilon_u(m_2) + N^{-1}T^{-7/2} \sum_{i,s,t,u,v} \varepsilon_s(m_1) \varepsilon_{it} \varepsilon_{iu} \varepsilon_v(m_2) \\
& - N^{-2}T^{-3/2} \sum_{i,j,s,t} \varepsilon_s(m_1) \varepsilon_{is} \varepsilon_{jt} \varepsilon_t(m_2) + N^{-2}T^{-5/2} \sum_{i,j,s,t,u} \varepsilon_s(m_1) \varepsilon_{is} \varepsilon_{jt} \varepsilon_u(m_2) \\
& + N^{-2}T^{-5/2} \sum_{i,j,s,t,u} \varepsilon_s(m_1) \varepsilon_{it} \varepsilon_{ju} \varepsilon_u(m_2) - N^{-2}T^{-7/2} \sum_{i,j,s,t,u,v} \varepsilon_s(m_1) \varepsilon_{it} \varepsilon_{ju} \varepsilon_v(m_2) \\
& = 2.I - \dots - 2.VIII.
\end{aligned}$$

2.I can be written as

$$\begin{aligned}
& T^{-1/2} \left(N^{-1} \left[\sum_{i \in (\Delta_{m_1, \varepsilon} \cup \Delta_{m_2, \varepsilon})} \left(T^{-1/2} \sum_s \varepsilon_s(m_1) \varepsilon_{is} \right) \left(T^{-1/2} \sum_t \varepsilon_t(m_2) \varepsilon_{it} \right) \right] \right. \\
& \quad \left. + \frac{T}{N} \left[\sum_{i \in (\Delta_{m_1, \varepsilon} \cup \Delta_{m_2, \varepsilon})^c} \left(T^{-1} \sum_s \varepsilon_s(m_1) \varepsilon_{is} \right) \left(T^{-1} \sum_t \varepsilon_t(m_2) \varepsilon_{it} \right) \right] \right) \\
& = O_p(T^{-1/2}),
\end{aligned}$$

by Assumption 8, given that $\frac{T}{N} = O(1)$.

2.II is $O_p(\delta_{N,T}^{-1}T^{-1/2})$ by Lemma 1.3 and equation A4.1. Item 2.III is identical.

2.IV is $O_p(\delta_{N,T}^{-1}T^{-1})$ by Lemma 1.3 and 1.10.

2.V can be written as

$$\begin{aligned}
& T^{-1/2} \left(\left[N^{-1} \sum_{i \in \Delta_{m_1, \varepsilon}} \left(T^{-1/2} \sum_s \varepsilon_s(m_1) \varepsilon_{is} \right) + \frac{T^{1/2}}{N} \sum_{i \in \Delta_{m_1, \varepsilon}^c} \left(T^{-1} \sum_s \varepsilon_s(m_1) \varepsilon_{is} \right) \right] \right. \\
& \quad \left. \times \left[N^{-1} \sum_{j \in \Delta_{m_2, \varepsilon}} \left(T^{-1/2} \sum_t \varepsilon_t(m_2) \varepsilon_{jt} \right) + \frac{T^{1/2}}{N} \sum_{j \in \Delta_{m_2, \varepsilon}^c} \left(T^{-1} \sum_t \varepsilon_t(m_2) \varepsilon_{jt} \right) \right] \right).
\end{aligned}$$

This is $O_p(T^{-1/2})$ by Assumption 8, given that $\frac{T^{1/2}}{N} = O(1)$.

2.VI is given by

$$\begin{aligned}
& N^{-1/2}T^{-1} \left(N^{-1} \sum_{i \in \Delta_{m_1, \varepsilon}} \left(T^{-1/2} \sum_s \varepsilon_s(m_1) \varepsilon_{is} \right) + \frac{T^{1/2}}{N} \sum_{i \in \Delta_{m_1, \varepsilon}^c} \left(T^{-1} \sum_s \varepsilon_s(m_1) \varepsilon_{is} \right) \right) \\
& \times \left(N^{-1/2}T^{-1/2} \sum_{j, t} \varepsilon_{jt} \right) \left(T^{-1/2} \sum_v \varepsilon_v(m_2) \right) \\
& = O_p \left(N^{-1/2}T^{-1} \right)
\end{aligned}$$

by Assumption 8 and Lemma 1.3. Item 2.VII is identical.

2.VIII is $O_p \left(N^{-1}T^{-3/2} \right)$ by Lemma 1.3.

Summing these terms, Item 2 is $O_p \left(\delta_{N, T}^{-1} \right)$.

Item 3 = $N^{-1}T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} = O_p \left(\delta_{N, T}^{-1} \right)$ is a $K \times N$ matrix with generic (m_1, m_2) element

$$\begin{aligned}
& N^{-1}T^{-3/2} \sum_{i, s, t} F_s(m_1) \varepsilon_{is} \varepsilon_{it} \varepsilon_t(m_2) - N^{-1}T^{-5/2} \sum_{i, s, t, u} F_s(m_1) \varepsilon_{is} \varepsilon_{it} \varepsilon_u(m_2) \\
& - N^{-1}T^{-5/2} \sum_{i, s, t, u} F_s(m_1) \varepsilon_{it} \varepsilon_{iu} \varepsilon_u(m_2) + N^{-1}T^{-7/2} \sum_{i, s, t, u, v} F_s(m_1) \varepsilon_{it} \varepsilon_{iu} \varepsilon_v(m_2) \\
& - N^{-2}T^{-3/2} \sum_{i, j, s, t} F_s(m_1) \varepsilon_{is} \varepsilon_{jt} \varepsilon_t(m_2) + N^{-2}T^{-5/2} \sum_{i, j, s, t, u} F_s(m_1) \varepsilon_{is} \varepsilon_{jt} \varepsilon_u(m_2) \\
& + N^{-2}T^{-5/2} \sum_{i, j, s, t, u} F_s(m_1) \varepsilon_{it} \varepsilon_{ju} \varepsilon_u(m_2) - N^{-2}T^{-7/2} \sum_{i, j, s, t, u, v} F_s(m_1) \varepsilon_{it} \varepsilon_{ju} \varepsilon_v(m_2) \\
& = 3.I - \dots - 3. VIII.
\end{aligned}$$

3.I can be written as

$$\begin{aligned}
& T^{-1/2} \left(N^{-1} \left[\sum_{i \in \Delta_{m_2, \varepsilon}} \left(T^{-1/2} \sum_s F_s(m_1) \varepsilon_{is} \right) \left(T^{-1/2} \sum_t \varepsilon_t(m_2) \varepsilon_{it} \right) \right] \right. \\
& \left. + \frac{T^{1/2}}{N} \left[\sum_{i \in \Delta_{m_2, \varepsilon}^c} \left(T^{-1/2} \sum_s F_s(m_1) \varepsilon_{is} \right) \left(T^{-1} \sum_t \varepsilon_t(m_2) \varepsilon_{it} \right) \right] \right) \\
& = O_p(T^{-1/2})
\end{aligned}$$

by Assumptions 3.6, 8 and the fact that $\frac{T^{1/2}}{N} = O(1)$.

3.II = $O_p\left(\delta_{N,T}^{-1}T^{-1/2}\right)$ by Lemma 1.3 and 1.12.

3.III = $O_p\left(\delta_{N,T}^{-1}\right)$ by Assumption 2.1 and equation A4.1.

3.IV = $O_p\left(\delta_{N,T}^{-1}T^{-1/2}\right)$ by Assumption 2.1, Lemma 1.3 and 1.10.

3.V can be written as

$$\begin{aligned} & T^{-1/2} \left(\left[N^{-1} \sum_i \left(T^{-1/2} \sum_s F_s(m_1) \varepsilon_{is} \right) \right] \right. \\ & \quad \times \left. \left[N^{-1} \sum_{j \in \Delta_{m_2, \varepsilon}} \left(T^{-1/2} \sum_t \varepsilon_t(m_2) \varepsilon_{jt} \right) + \frac{T^{1/2}}{N} \sum_{j \in \Delta_{m_2, \varepsilon}} \left(T^{-1} \sum_t \varepsilon_t(m_2) \varepsilon_{jt} \right) \right] \right) \\ & = O_p\left(T^{-1/2}\right) \end{aligned}$$

by Assumption 3.6, given that $\frac{T^{1/2}}{N} = O(1)$.

3.VI is given by

$$\begin{aligned} & N^{-1/2}T^{-1} \left(\left[N^{-1} \sum_i \left(T^{-1/2} \sum_s F_s(m_1) \varepsilon_{is} \right) \right] \right. \\ & \quad \times \left. \left(N^{-1/2}T^{-1/2} \sum_{j,t} \varepsilon_{jt} \right) \left(T^{-1/2} \sum_u \varepsilon_u(m_2) \right) \right) \\ & = O_p\left(N^{-1/2}T^{-1}\right) \end{aligned}$$

by Assumption 3.6 and Lemma 1.3.

3.VII is given by

$$\begin{aligned} & N^{-1/2}T^{-1} \left(\left(T^{-1} \sum_s F_s(m_1) \right) \left(N^{-1/2}T^{-1/2} \sum_{i,t} \varepsilon_{it} \right) \right. \\ & \quad \times \left. \left[N^{-1} \sum_{j \in \Delta_{m_2, \varepsilon}} \left(T^{-1/2} \sum_u \varepsilon_u(m_2) \varepsilon_{ju} \right) + \frac{T^{1/2}}{N} \sum_{j \in \Delta_{m_2, \varepsilon}} \left(T^{-1} \sum_u \varepsilon_u(m_2) \varepsilon_{ju} \right) \right] \right) \\ & = O_p\left(N^{-1/2}T^{-1}\right) \end{aligned}$$

by Assumption 8 and 2.1.

3.VIII is $O_p\left(N^{-1}T^{-1}\right)$ by Assumption 2.1 and Lemma 1.3. Summing these terms, Item 3 is

$$\mathcal{O}_p\left(\delta_{N,T}^{-1}\right).$$

Items 5 and 6 follow similar steps as Item 3.

Item 4 is a $K \times N$ matrix which can be partitioned as

$$N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = \begin{bmatrix} N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} \\ N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} \end{bmatrix}.$$

We show that $N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}$ is $\mathcal{O}_p(\Gamma_{N_f,T}^{-1})$. Similar arguments establish $N^{-\psi_g/2}T^{-1}\mathbf{\Phi}'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}$ is $\mathcal{O}_p(\Gamma_{N_g,T}^{-1})$, which implies

$$N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'_g\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = \mathcal{O}_p\left(\frac{N^{(\psi_g-\psi_f)/2}}{\Gamma_{N_g,T}}\right).$$

Hence,

$$N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = \mathcal{O}_p\left(\Gamma_{N_f,T}^{-1} \vee \frac{N^{(\psi_g-\psi_f)/2}}{\Gamma_{N_g,T}}\right) = \mathcal{O}_p\left(\Xi_{N,T}^{-1}\right).$$

Below, we show that $N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}$ is $\mathcal{O}_p(\Gamma_{N_f,T}^{-1})$.

$N^{-\psi_f/2}T^{-1}\mathbf{\Phi}'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}$ is a $K_f \times N$ matrix with generic (m_1, m_2) element

$$\begin{aligned} & N^{-\psi_f/2}T^{-1} \sum_{i,t} \phi_{if}(m_1) \varepsilon_t(m_2) \varepsilon_{it} - N^{-\psi_f/2-1}T^{-1} \sum_{i,j,t} \phi_{if}(m_1) \varepsilon_t(m_2) \varepsilon_{jt} \\ & - N^{-\psi_f/2}T^{-2} \sum_{j,s,t} \varepsilon_s(m_2) \phi_{jf}(m_1) \varepsilon_{jt} + N^{-\psi_f/2-1}T^{-2} \sum_{i,j,s,t} \varepsilon_s(m_2) \phi_{if}(m_1) \varepsilon_{jt} \\ & = 4.I - 4.II - 4.III + 4.IV. \end{aligned}$$

4.I is $\mathcal{O}_p\left(\Gamma_{N_f,T}^{-1}\right)$ by Lemma 1.16.

4.II can be written as

$$N^{\frac{\psi_f-1}{2}} \left(N^{-\psi_f} \sum_i \phi_{if}(m_1) \right) \left(N^{-1/2}T^{-1} \sum_{j,t} \varepsilon_t(m_2) \varepsilon_{jt} \right),$$

which is $O_p\left(N^{\frac{\psi_f-1}{2}}\delta_{N,T}^{-1}\right)$ by Assumption 2.2 and Lemma 1.10(b). Since $0 < \psi_f \leq 1$,

$$N^{\frac{\psi_f-1}{2}}\delta_{N,T}^{-1} = O\left(N^{-\psi_f/2} \vee T^{-1/2}\right) = O\left(\Gamma_{N_f,T}^{-1}\right).$$

4.III is $T^{-1/2}\left(T^{-1}\sum_t\left(\sum_j N^{-\psi_f/2}\phi_{jf}(m_1)\varepsilon_{jt}\right)\right)\left(T^{-1/2}\sum_s\varepsilon_s(m_2)\right)$, which is $O_p(T^{-1/2})$ by Assumption 3.7 and Lemma 1.3. Hence, it is $O_p\left(\Gamma_{N_f,T}^{-1}\right)$.

4.IV is equal to

$$N^{\frac{\psi_f-1}{2}}T^{-1}\left(N^{-\psi_f}\sum_i\phi_{if}(m_1)\right)\left(N^{-1/2}T^{-1/2}\sum_{j,t}\varepsilon_{jt}\right)\left(T^{-1/2}\sum_s\varepsilon_s(m_2)\right),$$

which is $O_p\left(N^{\frac{\psi_f-1}{2}}T^{-1}\right)$ by Assumption 2.2 and Lemma 1.3, and therefore is $O_p\left(\Gamma_{N_f,T}^{-1}\right)$.

Summing these terms, $N^{-\psi_f/2}T^{-1}\Phi'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}$ is $O_p\left(\Gamma_{N_f,T}^{-1}\right)$ and hence

$$N^{-\psi_f/2}T^{-1}\Phi'_f\mathbf{J}_N\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon} = O_p\left(\Gamma_{N_f,T}^{-1} \vee \frac{N^{(\psi_g-\psi_f)/2}}{\Gamma_{N_g,T}}\right) = O_p\left(\Xi_{N,T}^{-1}\right).$$

Item 7:

(a)

$$\begin{aligned} T^{-1/2}\mathbf{F}'\mathbf{J}_T\boldsymbol{\varepsilon} &= T^{-1/2}\sum_t\mathbf{F}_t\boldsymbol{\varepsilon}'_t - \left(T^{-1}\sum_t\mathbf{F}_t\right)\left(T^{-1/2}\sum_t\boldsymbol{\varepsilon}'_t\right) \\ &= O_p(1). \end{aligned}$$

This follows from Assumptions 2.1, 3.6 and Lemma 1.3. Item 7(b) follows using the same argument and employing Assumptions 2.4, 3.5, and Lemma 1.3.

Lemma 7. Under Assumptions 1-6, 8 and 9, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$, we have:

1. $\hat{\mathbf{F}}_A = T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{Z} = \Lambda_f\Delta_f\Lambda'_f + \zeta\left(T^{-1}\boldsymbol{\varepsilon}'\mathbf{J}_T\boldsymbol{\varepsilon}\right)\zeta' + \Delta_\omega + O_p(T^{-1/2})$.
2. $\hat{\mathbf{F}}_B = N^{-\psi_f}T^{-2}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{X}'\mathbf{J}_T\mathbf{Z} = \Lambda_f\Delta_f\mathcal{P}_f\Delta_f\Lambda'_f + O_p\left(\Xi_{N,T}^{-1}\right)$.
3. $\hat{\mathbf{F}}_{C,t} = N^{-\psi_f}T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\mathbf{x}_t = N^{-\psi_f}T^{-1}\mathbf{Z}'\mathbf{J}_T\mathbf{X}\mathbf{J}_N\phi_0 + \Lambda_f\Delta_f\mathcal{P}_f\mathbf{f}_t + O_p\left(\Xi_{N,T}^{-1}\right)$.

Consequently, an asymptotic representation of $\hat{\mathbf{F}}_t$ is

$$\begin{aligned}\hat{\mathbf{F}}_t &= (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Lambda}'_f + \boldsymbol{\zeta} (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + \mathbf{\Delta}_\omega) (\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathbf{\Lambda}'_f)^{-1} \\ &\quad \times \left(N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \boldsymbol{\phi}_0 + \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}_t \right) + \mathbf{o}_p(1).\end{aligned}$$

Proof:

$$\begin{aligned}\hat{\mathbf{F}}_t &= T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} \left(N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t \\ &= \hat{\mathbf{F}}_A \hat{\mathbf{F}}_B^{-1} \hat{\mathbf{F}}_{C,t}.\end{aligned}$$

We look at all these terms separately.

$$\begin{aligned}\hat{\mathbf{F}}_A &= T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} \\ &= \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega}) + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\omega} \\ &\quad + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + \boldsymbol{\zeta} (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F}) \mathbf{\Lambda}' + \boldsymbol{\zeta} (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\omega}) \\ &\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + \boldsymbol{\zeta} (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' \\ &= \mathbf{\Lambda} \mathbf{\Delta}_F \mathbf{\Lambda}' + \boldsymbol{\zeta} (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + \mathbf{\Delta}_\omega + \mathbf{O}_p(T^{-1/2}) \\ &= \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathbf{\Lambda}'_f + \boldsymbol{\zeta} (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + \mathbf{\Delta}_\omega + \mathbf{O}_p(T^{-1/2}).\end{aligned}$$

The final equality follows from Assumptions 2.1, 5, Lemmas 4 and 6, and the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries by Assumption 9.

$$\begin{aligned}\hat{\mathbf{F}}_B &= N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \\ &= N^{-\psi_f} T^{-2} (\boldsymbol{\nu}_T \boldsymbol{\lambda}'_0 + \mathbf{F} \mathbf{\Lambda}' + \boldsymbol{\omega})' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T (\boldsymbol{\nu}_T \boldsymbol{\lambda}'_0 + \mathbf{F} \mathbf{\Lambda}' + \boldsymbol{\omega}) \\ &\quad + N^{-\psi_f} T^{-2} \boldsymbol{\zeta} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T (\boldsymbol{\nu}_T \boldsymbol{\lambda}'_0 + \mathbf{F} \mathbf{\Lambda}' + \boldsymbol{\omega} + \boldsymbol{\varepsilon} \boldsymbol{\zeta}') \\ &\quad + N^{-\psi_f} T^{-2} (\boldsymbol{\nu}_T \boldsymbol{\lambda}'_0 + \mathbf{F} \mathbf{\Lambda}' + \boldsymbol{\omega} + \boldsymbol{\varepsilon} \boldsymbol{\zeta}')' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \boldsymbol{\varepsilon} \boldsymbol{\zeta}' \\ &= \text{I} + \text{II} + \text{III}.\end{aligned}$$

The asymptotic representation of I was established in Lemma 4. Also note that II is the transpose

of III. We establish the stochastic order of III below.

$$\begin{aligned}
\text{III} &= N^{-\psi_f} T^{-2} (\iota_T \lambda'_0 + \mathbf{F} \Lambda' + \omega + \varepsilon \zeta')' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \varepsilon \zeta' \\
&= \Lambda (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \Phi' \mathbf{J}_N \Phi \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon) \zeta' + \Lambda (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \Phi' \mathbf{J}_N \varepsilon' \mathbf{J}_T \varepsilon \right) \zeta' \\
&\quad + \Lambda \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon \mathbf{J}_N \Phi \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon) \zeta' + (T^{-1} \omega' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \Phi' \mathbf{J}_N \Phi \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon) \zeta' \\
&\quad + (T^{-1} \omega' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \Phi' \mathbf{J}_N \varepsilon' \mathbf{J}_T \varepsilon \right) \zeta' + \left(N^{-\psi_f} T^{-1} \omega' \mathbf{J}_T \varepsilon \mathbf{J}_N \Phi \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon) \zeta' \\
&\quad + \zeta (T^{-1} \varepsilon' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \Phi' \mathbf{J}_N \Phi \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon) \zeta' + \zeta (T^{-1} \varepsilon' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \Phi' \mathbf{J}_N \varepsilon' \mathbf{J}_T \varepsilon \right) \zeta' \\
&\quad + \zeta \left(N^{-\psi_f} T^{-1} \varepsilon' \mathbf{J}_T \varepsilon \mathbf{J}_N \Phi \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \varepsilon) \zeta' + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \omega' \mathbf{J}_T \varepsilon \mathbf{J}_N \varepsilon' \mathbf{J}_T \varepsilon \right) \zeta' \\
&\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \Lambda \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \varepsilon \mathbf{J}_N \varepsilon' \mathbf{J}_T \varepsilon \right) \zeta' + \frac{N^{1-\psi_f}}{\sqrt{T}} \zeta \left(N^{-1} T^{-3/2} \varepsilon' \mathbf{J}_T \varepsilon \mathbf{J}_N \varepsilon' \mathbf{J}_T \varepsilon \right) \zeta' \\
&= O_p(\Xi_{N,T}^{-1}).
\end{aligned}$$

The final equality follows from Lemmas 2 and 6, equation A3.1, Assumptions 2.1 and 4, and the fact that ζ has finitely many non-zero entries given Assumption 9. Hence, $\hat{\mathbf{F}}_B = N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p(\Xi_{N,T}^{-1})$.

$$\begin{aligned}
\hat{\mathbf{F}}_{C,t} &= N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{x}_t \\
&= N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + N^{-\psi_f} T^{-1} (\iota_T \lambda'_0 + \mathbf{F} \Lambda' + \omega)' \mathbf{J}_T \mathbf{X} \mathbf{J}_N (\Phi \mathbf{F}_t + \varepsilon_t) \\
&\quad + \zeta (T^{-1} \varepsilon' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \Phi' \mathbf{J}_N \Phi \right) \mathbf{F}_t + \zeta (T^{-1} \varepsilon' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \Phi' \mathbf{J}_N \varepsilon_t \right) \\
&\quad + \zeta \left(N^{-\psi_f} T^{-1} \varepsilon' \mathbf{J}_T \varepsilon \mathbf{J}_N \Phi \right) \mathbf{F}_t + \frac{N^{1-\psi_f}}{\sqrt{T}} \zeta \left(N^{-1} T^{-1/2} \varepsilon' \mathbf{J}_T \varepsilon \mathbf{J}_N \varepsilon_t \right) \\
&= N^{-\psi_f} T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \phi_0 + \Lambda_f \Delta_f \mathcal{P}_f \mathbf{f}_t + O_p(\Xi_{N,T}^{-1}).
\end{aligned}$$

The stochastic order of the second term follows from Lemma 4, while all the remaining terms are $O_p(\Xi_{N,T}^{-1})$ by Lemma 6 and the fact that ζ has finitely many non-zero entries by Assumption 9.

Combining Items 1-3 gives the asymptotic representation of $\hat{\mathbf{F}}_t$ stated in Lemma 7.

Lemma 8. Under Assumptions 1-6, 8 and 9, if $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$,

$$1. \hat{\beta}_1 = T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z} = \Lambda_f \Delta_f \Lambda_f' + \zeta (T^{-1} \varepsilon' \mathbf{J}_T \varepsilon) \zeta' + \Delta_\omega + O_p(T^{-1/2}).$$

$$2. \hat{\beta}_2 = N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p \left(\Xi_{N,T}^{-1} \right).$$

$$3. \hat{\beta}_3 = N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p \left(\Xi_{N,T}^{-1} \right).$$

$$4. \hat{\beta}_4 = N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p \left(\Xi_{N,T}^{-1} \right).$$

Therefore,

$$\begin{aligned} \hat{\beta} &= (T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z})^{-1} N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \\ &\quad \times \left(N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} \\ &= \hat{\beta}_1^{-1} \hat{\beta}_2 \hat{\beta}_3^{-1} \hat{\beta}_4 \end{aligned}$$

has the following asymptotic representation:

$$\begin{aligned} \hat{\beta} &= (\Lambda_f \Delta_f \Lambda_f' + \zeta (T^{-1} \varepsilon' \mathbf{J}_T \varepsilon) \zeta' + \Delta_\omega)^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' (\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f')^{-1} \\ &\quad \times \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + o_p(1). \end{aligned}$$

Proof:

$$\begin{aligned} \hat{\beta} &= (T^{-1} \mathbf{Z}' \mathbf{J}_T \mathbf{Z})^{-1} N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \\ &\quad \times \left(N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} \right)^{-1} N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} \\ &= \hat{\beta}_1^{-1} \hat{\beta}_2 \hat{\beta}_3^{-1} \hat{\beta}_4. \end{aligned}$$

Note that $\hat{\beta}_1 = \hat{\mathbf{F}}_A$ and $\hat{\beta}_2 = \hat{\mathbf{F}}_B$ and their asymptotic representations are established in Lemma

7. The expressions for $\hat{\beta}_3$ and $\hat{\beta}_4$ are handled below. Note that,

$\hat{\beta}_3 = N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z}$ is equal to the product $\frac{\hat{\mathbf{F}}_C \mathbf{J}_T \hat{\mathbf{F}}_C'}{T}$ where $\hat{\mathbf{F}}_C$ is obtained by stacking $\hat{\mathbf{F}}_{C,t}$ horizontally. Using the asymptotic representation of $\hat{\mathbf{F}}_{C,t}$ obtained in Lemma 7, standard arguments imply that

$$\frac{\hat{\mathbf{F}}_C \mathbf{J}_T \hat{\mathbf{F}}_C'}{T} = \Lambda_f \Delta_f \mathcal{P}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \mathcal{P}_f \Delta_f \Lambda_f' + O_p(\Xi_{N,T}^{-1}),$$

which is equal to $\mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathbf{\Lambda}_f' + \mathbf{O}_p(\Xi_{N,T}^{-1})$ given Assumption 2.1. Hence,

$$N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \mathbf{\Lambda}_f' + \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

$$\begin{aligned} \hat{\beta}_4 &= N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} \\ &= N^{-\psi_f} T^{-2} (\iota_T \boldsymbol{\lambda}'_0 + \mathbf{F} \boldsymbol{\Lambda}' + \boldsymbol{\omega})' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T (\iota_T \beta_0 + \mathbf{F} \boldsymbol{\beta} + \boldsymbol{\eta}) \\ &\quad + N^{-\psi_f} T^{-2} \boldsymbol{\zeta} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T (\iota_T \beta_0 + \mathbf{F} \boldsymbol{\beta} + \boldsymbol{\eta}) \\ &\quad + N^{-\psi_f} T^{-2} (\iota_T \boldsymbol{\lambda}'_0 + \mathbf{F} \boldsymbol{\Lambda}' + \boldsymbol{\omega} + \boldsymbol{\varepsilon} \boldsymbol{\zeta}')' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \boldsymbol{\varepsilon} \boldsymbol{\gamma} \\ &= \text{I} + \text{II} + \text{III}. \end{aligned}$$

The stochastic order of I was established in Lemma 5. III is $\mathbf{O}_p(\Xi_{N,T}^{-1})$, which follows by replacing $\boldsymbol{\zeta}'$ by $\boldsymbol{\gamma}$ in the III term of the expression for $\hat{\mathbf{F}}_B$ in Lemma 7 and noting that $\gamma_i = 0$ implies $\zeta'_i = 0$ by Assumption 9. We look at the transpose of II below.

$$\begin{aligned} \text{II}' &= N^{-\psi_f} T^{-2} (\iota_T \beta_0 + \mathbf{F} \boldsymbol{\beta} + \boldsymbol{\eta})' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \boldsymbol{\varepsilon} \boldsymbol{\zeta}' \\ &= \boldsymbol{\beta}' (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + \boldsymbol{\beta}' (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \right) \boldsymbol{\zeta}' \\ &\quad + \boldsymbol{\beta}' \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' + (T^{-1} \boldsymbol{\eta}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' \\ &\quad + (T^{-1} \boldsymbol{\eta}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \right) \boldsymbol{\zeta}' + \left(N^{-\psi_f} T^{-1} \boldsymbol{\eta}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}) \boldsymbol{\zeta}' \\ &\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \boldsymbol{\eta}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \right) \boldsymbol{\zeta}' \\ &\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \boldsymbol{\beta}' \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \right) \boldsymbol{\zeta}' \\ &= \mathbf{O}_p(\Xi_{N,T}^{-1}). \end{aligned}$$

The final equality follows from Lemmas 2 and 6 and the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries given Assumption 9. Therefore,

$$\hat{\beta}_4 = \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \boldsymbol{\beta}_f + \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

Combining the asymptotic representations of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ gives the asymptotic representation of $\hat{\beta}$ stated in Lemma 8.

Theorem 1(b) Let Assumptions 1-6, 8-9 hold, $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$, then

$$\hat{y}_{t+h,f} - \mathbb{E}(y_{t+h} | \mathbf{F}_t) = O_p(\Xi_{N,T}^{-1}).$$

Proof: From Lemmas 4, 5, 7 and 8, we have established that

- $N^{-\psi_f} T^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \left(\hat{\mathbf{F}}_{C,t} - \bar{\mathbf{F}}_C \right)' = (\mathbf{f}_t - \bar{\mathbf{f}})' \mathcal{P}_f \Delta_f \Lambda_f' + O_p(\Xi_{N,T}^{-1}).$
- $N^{-2\psi_f} T^{-3} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{Z} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda_f' + O_p(\Xi_{N,T}^{-1}).$
- $N^{-\psi_f} T^{-2} \mathbf{Z}' \mathbf{J}_T \mathbf{X} \mathbf{J}_N \mathbf{X}' \mathbf{J}_T \mathbf{y} = \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \beta_f + O_p(\Xi_{N,T}^{-1}).$

Given these results and the fact that for all i , $T^{-1/2} \sum_t \varepsilon_{it} = O_p(1)$ by Lemma 1.3, we obtain, using the same steps as in the proof of Theorem 1(a),

$$\hat{y}_{t+h,f} = \beta_0 + \mathbf{F}_t' \boldsymbol{\beta} + O_p(\Xi_{N,T}^{-1}) = \mathbb{E}(y_{t+h} | \mathbf{F}_t) + O_p(\Xi_{N,T}^{-1}).$$

The proofs of Theorems 3(b) and 4(b), respectively, follow similarly from the rates derived in Lemmas 7 and 8.

A5 Proof of Theorem 5

We now proceed to prove Theorem 5, which deals with Stage 2 of 3PRF LASSO. We introduce two lemmas which are used in the subsequent proof of Theorem 5.

Lemma 9. Define $\hat{x}_{it} \equiv \hat{\phi}_{0,i} + \hat{\mathbf{F}}_t' \hat{\boldsymbol{\phi}}_i$, where $\hat{\phi}_{0,i}$ and $\hat{\boldsymbol{\phi}}_i$ are obtained from the Stage 2 Pass 4 regression. Let Assumptions 1-6, 7.1, 8 and 9 hold, $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ and $\frac{T}{N} = O(1)$. Then,

$$\hat{\varepsilon}_{it} - (\varepsilon_{it} + \mathbf{g}_t' \boldsymbol{\phi}_{ig} - \bar{\mathbf{g}}' \boldsymbol{\phi}_{ig}) = O_p(\Xi_{N,T}^{-1}).$$

Proof: The proof proceeds in a similar manner to the proof of Theorem 1(b). The target $\mathbf{y}$ can be replaced by $\mathbf{x}_i$ and the proof follows similar steps.

First notice that, using the same steps as in the proof of Lemma 7 for $\hat{\mathbf{F}}_B$, we obtain

$$N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{X\mathbf{x}_i} = \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \phi_i + O_p \left(\Xi_{N,T}^{-1} \right).$$

Employing the fact that $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$ (Assumption 5), we have

$$\begin{aligned} & \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \phi_i \\ &= \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \phi_{if} \\ & \quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \mathbf{f}) \phi_{if} \\ & \quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \phi_{if} \\ & \quad + \frac{N^{\psi_g - \psi_f}}{T} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g} \right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g \right) \left(T^{-1/2} \mathbf{g}' \mathbf{J}_T \mathbf{f} \right) \phi_{if} \\ & \quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \phi_{ig} \\ & \quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \mathbf{g}) \phi_{ig} \\ & \quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \phi_{ig} \\ & \quad + \frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g} \right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \mathbf{g}) \phi_{ig} \\ &= \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \phi_{if} + O_p \left(\Xi_{N,T}^{-1} \right) + O_p \left(\Xi_{N,T}^{-1} \right) \phi_{ig} \\ &= \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{\Delta}_f \phi_{if} + O_p \left(\Xi_{N,T}^{-1} \right). \end{aligned}$$

The result follows from Assumptions 2.1, 4, and 7.1. Substituting this in the expression for $\hat{x}_{it}$ gives

$$\begin{aligned}
\hat{x}_{it} &= \bar{x}_i + \left(N^{-\psi_f} T^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) \mathbf{W}_{XZ} \right) \left(N^{-2\psi_f} T^{-3} \mathbf{W}'_{XZ} \mathbf{S}_{XX} \mathbf{W}_{XZ} \right)^{-1} \left(N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{X\mathbf{x}_i} \right) \\
&= \phi_{0,i} + \bar{\mathbf{F}}' \phi_i + O_p(T^{-1/2}) + \left((\mathbf{f}_t - \bar{\mathbf{f}})' \mathcal{P}_f \Delta_f \Lambda'_f + O_p(\Xi_{N,T}^{-1}) \right) \\
&\quad \times \left[\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f + O_p(\Xi_{N,T}^{-1}) \right]^{-1} \left(\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} + O_p(\Xi_{N,T}^{-1}) \phi_{ig} + O_p(\Xi_{N,T}^{-1}) \right) \\
&= \phi_{0,i} + \bar{\mathbf{f}}' \phi_{if} + \bar{\mathbf{g}}' \phi_{ig} + O_p(T^{-1/2}) + (\mathbf{f}_t - \bar{\mathbf{f}})' \mathcal{P}_f \Delta_f \Lambda'_f \\
&\quad \times \left[\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right]^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} + O_p(\Xi_{N,T}^{-1}) + O_p(\Xi_{N,T}^{-1}) \phi_{ig} \\
&= \phi_{0,i} + \bar{\mathbf{g}}' \phi_{ig} + O_p(T^{-1/2}) + \mathbf{f}_t' \mathcal{P}_f \Delta_f \Lambda'_f \\
&\quad \times \left[\Lambda_f \Delta_f \mathcal{P}_f \Delta_f \mathcal{P}_f \Delta_f \Lambda'_f \right]^{-1} \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} + O_p(\Xi_{N,T}^{-1}) + O_p(\Xi_{N,T}^{-1}) \phi_{ig} \\
&= \phi_{0,i} + \bar{\mathbf{g}}' \phi_{ig} + \mathbf{f}_t' \phi_{if} + O_p(\Xi_{N,T}^{-1}) + O_p(\Xi_{N,T}^{-1}) \phi_{ig}, \\
\hat{\varepsilon}_{it} &= x_{it} - \hat{x}_{it} \\
&= (\phi_{0,i} + \mathbf{g}_t' \phi_{ig} + \mathbf{f}_t' \phi_{if} + \varepsilon_{it}) - \left(\phi_{0,i} + \bar{\mathbf{g}}' \phi_{ig} + \mathbf{f}_t' \phi_{if} + O_p(\Xi_{N,T}^{-1}) + O_p(\Xi_{N,T}^{-1}) \phi_{ig} \right),
\end{aligned}$$

which implies

$$\hat{\varepsilon}_{it} - (\varepsilon_{it} + \mathbf{g}_t' \phi_{ig} - \bar{\mathbf{g}}' \phi_{ig}) = O_p(\Xi_{N,T}^{-1}).$$

The stochastic orders for the matrices $N^{-\psi_f} T^{-1} (\mathbf{x}_t - \bar{\mathbf{x}}) \mathbf{W}_{XZ}$ and $N^{-2\psi_f} T^{-3} \mathbf{W}'_{XZ} \mathbf{S}_{XX} \mathbf{W}_{XZ}$ were obtained in Lemmas 7 and 8, respectively. Noting that $\sum_{s=1}^T \varepsilon_{is}/T = O_p(T^{-1/2})$ by Lemma 1.3, we obtain the second equality.

Lemma 10. Define $\tilde{\boldsymbol{\eta}} \equiv \hat{\mathbf{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}$, where $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}_f = \mathbf{y} - \boldsymbol{\nu}_T \bar{\mathbf{y}} - \mathbf{J}_T \hat{\mathbf{F}} \hat{\boldsymbol{\beta}}$. Then, Under Assumptions 1-9 and 10.1,

$$\max_i \left(\frac{\Xi_{N,T}}{T} \right) |\hat{\boldsymbol{\varepsilon}}'_i \tilde{\boldsymbol{\eta}} - \boldsymbol{\varepsilon}'_i \boldsymbol{\eta}| = O_p(1).$$

Consequently,

$$\max_i \left(\frac{\Xi_{N,T}}{T} \right) |\hat{\boldsymbol{\varepsilon}}'_i \tilde{\boldsymbol{\eta}}| \leq \max_i \left(\frac{\Xi_{N,T}}{T} \right) |\boldsymbol{\varepsilon}'_i \boldsymbol{\eta}| + O_p(1).$$

Proof: Adding and subtracting terms and applying the triangle inequality,

$$\begin{aligned}
& \max_i \left(\frac{\Xi_{N,T}}{T} \right) |\hat{\varepsilon}'_i \tilde{\boldsymbol{\eta}} - \varepsilon'_i \boldsymbol{\eta}| \\
& \leq \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| (\mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig})' \boldsymbol{\eta} \right| \\
& \quad + \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| (\varepsilon_i + \mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig})' (\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \right| \\
& \quad + \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| [\hat{\varepsilon}_i - (\varepsilon_i + \mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig})]' \boldsymbol{\eta} \right| \\
& \quad + \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| [\hat{\varepsilon}_i - (\varepsilon_i + \mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig})]' (\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \right| \\
& \equiv \text{I} + \text{II} + \text{III} + \text{IV}.
\end{aligned}$$

We show below that Items I, II, III and IV are all $O_p(1)$.

Item I is

$$\text{I} = \left(\frac{\Xi_{N,T}}{T} \right) \max_i |\boldsymbol{\phi}'_{ig} \mathbf{g}' \mathbf{J}_T \boldsymbol{\eta}|.$$

Since $\Xi_{N,T}/T \leq T^{-1/2}$, $\max_i \|\boldsymbol{\phi}_{ig}\| = O_p(1)$ by Assumption 7.1, and $T^{-1/2} \mathbf{g}' \mathbf{J}_T \boldsymbol{\eta} = O_p(1)$ by Lemma 2.2, we obtain

$$\text{I} = O_p(1).$$

Item II is

$$\text{II} = \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| (\varepsilon_i + \mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig})' (\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \right|.$$

Note that

$$\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}} = (\hat{\mathbf{y}}_f - \mathbf{y}_f) + (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}) \boldsymbol{\gamma}.$$

Therefore, applying the triangle inequality,

$$\text{II} \leq \max_i |\mathcal{A}_i| + \max_i |\mathcal{B}_i| + \max_i |\mathcal{C}_i| + \max_i |\mathcal{D}_i|,$$

where

$$\begin{aligned}\mathcal{A}_i &= \left(\frac{\Xi_{N,T}}{T} \right) (\hat{\mathbf{y}}_f - \mathbf{y}_f)' \boldsymbol{\varepsilon}_i, \\ \mathcal{B}_i &= \left(\frac{\Xi_{N,T}}{T} \right) (\hat{\mathbf{y}}_f - \mathbf{y}_f)' \mathbf{J}_T \mathbf{g} \phi_{ig}, \\ \mathcal{C}_i &= \left(\frac{\Xi_{N,T}}{T} \right) \boldsymbol{\gamma}' (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})' \boldsymbol{\varepsilon}_i, \\ \mathcal{D}_i &= \left(\frac{\Xi_{N,T}}{T} \right) \boldsymbol{\gamma}' (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})' \mathbf{J}_T \mathbf{g} \phi_{ig}.\end{aligned}$$

We first consider $\mathcal{A}_i$. From Lemma 2 and the expressions used in the proofs of Theorems 1(a) and 1(b), we have

$$(\hat{\mathbf{y}}_f - \mathbf{y}_f)' = \mathbf{O}_p(\Xi_{N,T}^{-1}) \mathbf{F}' \mathbf{J}_T + \mathbf{O}_p(1) \left(\hat{\mathbf{F}}_C - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \mathbf{f}' \right) \mathbf{J}_T + \boldsymbol{\gamma}'_S (\boldsymbol{\nu}_T \bar{\boldsymbol{\varepsilon}}_S)' + (\boldsymbol{\nu}_T \bar{\boldsymbol{\eta}})'$$

Therefore,

$$\begin{aligned}\mathcal{A}_i &= \Xi_{N,T} \mathbf{O}_p(\Xi_{N,T}^{-1}) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\ &\quad + \mathbf{O}_p(1) \left(\frac{\Xi_{N,T}}{T} \right) \left(\hat{\mathbf{F}}_C - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \mathbf{f}' \right) \mathbf{J}_T \boldsymbol{\varepsilon}_i \\ &\quad + \left(\frac{\Xi_{N,T}}{T} \right) \boldsymbol{\gamma}'_S (\boldsymbol{\nu}_T \bar{\boldsymbol{\varepsilon}}_S)' \boldsymbol{\varepsilon}_i + \left(\frac{\Xi_{N,T}}{T} \right) (\boldsymbol{\nu}_T \bar{\boldsymbol{\eta}})' \boldsymbol{\varepsilon}_i \\ &\equiv \mathcal{A}_{1i} + \mathcal{A}_{2i} + \mathcal{A}_{3i} + \mathcal{A}_{4i}.\end{aligned}$$

The first term $\mathcal{A}_{1i}$ can be expanded as

$$\begin{aligned}\mathcal{A}_{1i} &= \mathbf{O}_p(1) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\ &= \mathbf{O}_p(1) (T^{-1} \mathbf{F}' \boldsymbol{\varepsilon}_i) - \mathbf{O}_p(1) (T^{-1} \mathbf{F}' \boldsymbol{\nu}_T \bar{\boldsymbol{\varepsilon}}_i).\end{aligned}$$

Therefore, by the triangle inequality,

$$\begin{aligned}
\max_i |\mathcal{A}_{1i}| &\leq \|\mathbf{O}_p(1)\|_\infty \max_i \|T^{-1} \mathbf{F}' \boldsymbol{\varepsilon}_i\|_\infty \\
&\quad + \|T^{-1} \mathbf{O}_p(1) \mathbf{F}'\|_\infty \max_i |\bar{\boldsymbol{\varepsilon}}_i| \\
&= \mathbf{O}_p\left(\frac{(\log N)^{r_3}}{\sqrt{T}}\right) + \mathbf{O}_p\left(\frac{(\log N)^{r_2}}{\sqrt{T}}\right) = O_p(1),
\end{aligned}$$

where the penultimate equality follows from Assumptions 7.3 and 7.4, and the final equality follows from Assumption 10.1.

We next consider $\mathcal{A}_{2i}$. Using Lemma 7, we can expand it as

$$\begin{aligned}
\mathcal{A}_{2i} &= O_p(1) \left(\frac{\Xi_{N,T}}{T} \right) \left(\hat{\mathbf{F}}_C - \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}' \right) \mathbf{J}_T \boldsymbol{\varepsilon}_i \\
&= O_p(1) \Xi_{N,T} \left\{ \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \right. \\
&\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\
&\quad + \left(N^{-\psi_f} T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\
&\quad + \zeta (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\
&\quad + \zeta \left(N^{-\psi_f} T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\
&\quad + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\
&\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\
&\quad + \frac{N^{1-\psi_f}}{\delta_{N,T} \sqrt{T}} \mathbf{\Lambda} \left(\delta_{N,T} N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\
&\quad + \frac{N^{1-\psi_f}}{\delta_{N,T} \sqrt{T}} \left(\delta_{N,T} N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\
&\quad + \frac{N^{1-\psi_f}}{\delta_{N,T} \sqrt{T}} \zeta \left(\delta_{N,T} N^{-1} T^{-3/2} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\
&\quad + \zeta (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \\
&\quad \left. + \frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g} \right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) \right\}.
\end{aligned}$$

Given that ζ has finitely many non-zero entries by Assumption 9, $\|\zeta\|_\infty$ and $\|\zeta\|_1$ are both $O_p(1)$. Also, given the orders derived in Lemmas 2 and 6, and the fact that $\Xi_{N,T} \leq \delta_{N,T}$, $\mathcal{A}_{2i}$ consists of the following three types of terms:

1. $\mathcal{A}_{2i}.1$: $O_p(1) (T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i)$, where the $O_p(1)$ term is a finite-dimensional matrix.
2. $\mathcal{A}_{2i}.2$: $\frac{N^{1-\psi_f}}{\sqrt{T}} \frac{\mathbf{D} \boldsymbol{\varepsilon}_i}{T}$, where $\mathbf{D}$ is an $L \times T$ matrix, invariant across i , whose normalized maximum order is $O_p(1)$.
3. $\mathcal{A}_{2i}.3$: $\Xi_{N,T} O_p(1) (N^{-\psi_f} T^{-1} \mathbf{\Phi}'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) + \Xi_{N,T} O_p(T^{-1/2}) (N^{-\psi_f} T^{-1} \mathbf{\Phi}'_g \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i)$, where

the $\mathbf{O}_p(1)$ and $\mathbf{O}_p(T^{-1/2})$ terms are finite-dimensional matrices invariant across i .¹

For $\mathcal{A}_{2i.1}$, we follow the same proof as for $\mathcal{A}_{1i}$ to show that its maximum over i is $\mathbf{O}_p(1)$ under Assumption 10.1.

For $\mathcal{A}_{2i.2}$,

$$\begin{aligned} \left\| \frac{N^{1-\psi_f} \mathbf{D} \boldsymbol{\varepsilon}_i}{\sqrt{T} T} \right\|_\infty &\leq \frac{N^{1-\psi_f}}{\sqrt{T}} \left\| \frac{\mathbf{D}}{T} \right\|_\infty \|\boldsymbol{\varepsilon}_i\|_\infty \\ &= \mathbf{O}_p(1) \frac{N^{1-\psi_f}}{\sqrt{T}} \max_t |\varepsilon_{it}|. \end{aligned}$$

Hence,

$$\max_i \left\| \frac{N^{1-\psi_f} \mathbf{D} \boldsymbol{\varepsilon}_i}{\sqrt{T} T} \right\|_\infty \leq \mathbf{O}_p(1) \frac{N^{1-\psi_f}}{\sqrt{T}} \max_{i,t} |\varepsilon_{it}| = \mathbf{O}_p(1)$$

by Assumption 7.2 and Assumption 10.1.

To show that the maximum over i of $\mathcal{A}_{2i.3}$ is $\mathbf{O}_p(1)$, first consider

$$\max_i \left[\Xi_{N,T} \left\| N^{-\psi_f} T^{-1} \boldsymbol{\Phi}'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right\|_1 \right].$$

Since K_f is finite, it suffices to derive the stochastic order of a generic component. A generic component is bounded by

$$\begin{aligned} &(\Xi_{N,T} N^{-\psi_f/2} T^{-1/2}) \max_i \left| N^{-\psi_f/2} T^{-1/2} \sum_{j,t} \phi_{jf}(m) \varepsilon_{jt} \varepsilon_{it} \right| \\ &+ (\Xi_{N,T} N^{-1/2} T^{-1/2}) \left| N^{-\psi_f} \sum_l \phi_{lf}(m) \right| \max_i \left| N^{-1/2} T^{-1/2} \sum_{j,t} \varepsilon_{it} \varepsilon_{jt} \right| \\ &+ (\Xi_{N,T} N^{-\psi_f/2} T^{-1/2}) \left| N^{-\psi_f/2} T^{-1} \sum_{j,s} \phi_{jf}(m) \varepsilon_{js} \right| \max_i \left| T^{-1/2} \sum_t \varepsilon_{it} \right| \\ &+ (\Xi_{N,T} N^{-1/2} T^{-1}) \left| N^{-\psi_f} \sum_l \phi_{lf}(m) \right| \left| N^{-1/2} T^{-1/2} \sum_{j,s} \varepsilon_{js} \right| \max_i \left| T^{-1/2} \sum_t \varepsilon_{it} \right| \\ &\equiv a + b + c + d. \end{aligned}$$

¹This follows from the observation that $\boldsymbol{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) (N^{-\psi_f} T^{-1} \boldsymbol{\Phi}'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i)$ equals $\boldsymbol{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) (N^{-\psi_f} T^{-1} \boldsymbol{\Phi}'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i) + T^{-1/2} \boldsymbol{\Lambda}_f (T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g}) (N^{-\psi_f} T^{-1} \boldsymbol{\Phi}'_g \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i)$, while the remaining terms premultiplying $N^{-\psi_f} T^{-1} \boldsymbol{\Phi}'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i$ are $\mathbf{O}_p(T^{-1/2})$.

By the definition of $\Xi_{N,T}$, $\Xi_{N,T}N^{-\psi_f/2} = O(1)$. Thus,

$$a = O_p\left(\frac{(\log N)^{r_4}}{\sqrt{T}}\right)$$

by Assumption 7.5,

$$b = O_p\left(\frac{(\log N)^{r_5}}{\sqrt{T}}\right)$$

by Assumptions 2.2 and 7.6,

$$c = O_p\left(\frac{(\log N)^{r_2}}{\sqrt{T}}\right)$$

by Assumptions 3.7 and 7.3, and

$$d = O_p\left(\frac{(\log N)^{r_2}}{T}\right)$$

by Lemma 1.3 and Assumptions 2.2 and 7.3. Therefore, by Assumption 10.1,

$$\max_i \left[\Xi_{N,T} \left\| N^{-\psi_f} T^{-1} \Phi'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right\|_1 \right] = O_p(1).$$

Given Assumption 6, similar reasoning establishes

$$\max_i \left[\Xi_{N,T} T^{-1/2} \left\| N^{-\psi_f} T^{-1} \Phi'_g \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right\|_1 \right] = O_p(1).$$

Hence,

$$\begin{aligned} \Xi_{N,T} \max_i \left\{ \left\| O_p(1) \left(N^{-\psi_f} T^{-1} \Phi'_f \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \right\|_1 \right. \\ \left. + \left\| O_p(T^{-1/2}) \left(N^{-\psi_f} T^{-1} \Phi'_g \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \right\|_1 \right\} = O_p(1), \end{aligned}$$

and therefore the maximum over i of $\mathcal{A}_{2i}$ is $O_p(1)$. Consequently,

$$\max_i |\mathcal{A}_{2i}| = O_p(1).$$

For $\mathcal{A}_{3i}$, since $\Xi_{N,T}/T \leq T^{-1/2}$,

$$\begin{aligned} \max_i \left| \left(\frac{\Xi_{N,T}}{T} \right) \gamma'_S (\iota_T \bar{\varepsilon}_S)' \varepsilon_i \right| &\leq \sum_{j \in S} |\gamma_j| |\bar{\varepsilon}_j| \max_i \left| T^{-1/2} \sum_t \varepsilon_{it} \right| \\ &= O_p \left(\frac{(\log N)^{r_2}}{\sqrt{T}} \right) = O_p(1), \end{aligned}$$

where the penultimate equality follows from Lemma 1.3, Assumption 7.3, and the fixed cardinality of S under Assumption 9; the final equality follows from Assumption 10.1.

Similarly,

$$\max_i \left| \left(\frac{\Xi_{N,T}}{T} \right) (\iota_T \bar{\eta})' \varepsilon_i \right| = O_p(1)$$

by Lemma 1.4, Assumption 7.3, and Assumption 10.1. Hence,

$$\max_i |\mathcal{A}_i| = O_p(1).$$

Next, we show that $\max_i |\mathcal{B}_i| = O_p(1)$. We have

$$\begin{aligned} \mathcal{B}_i &= \left(\frac{\Xi_{N,T}}{T} \right) (\hat{\mathbf{y}}_f - \mathbf{y}_f)' \mathbf{J}_T \mathbf{g} \phi_{ig} \\ &= \Xi_{N,T} \mathbf{O}_p(\Xi_{N,T}^{-1}) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\ &\quad + \mathbf{O}_p(1) \left(\frac{\Xi_{N,T}}{T} \right) (\hat{\mathbf{F}}_C - \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}') \mathbf{J}_T \mathbf{g} \phi_{ig} \\ &\quad + \left(\frac{\Xi_{N,T}}{T} \right) \gamma'_S (\iota_T \bar{\varepsilon}_S)' \mathbf{J}_T \mathbf{g} \phi_{ig} \\ &\quad + \left(\frac{\Xi_{N,T}}{T} \right) (\iota_T \bar{\eta})' \mathbf{J}_T \mathbf{g} \phi_{ig} \\ &\equiv \mathcal{B}_{1i} + \mathcal{B}_{2i} + \mathcal{B}_{3i} + \mathcal{B}_{4i}. \end{aligned}$$

These terms depend on i through ϕ_{ig} only. Since $\max_i |\phi_i(m)| = O_p(1)$ for every m by Assumption 7.1, it suffices to establish the stochastic order of the remaining factors. $\mathcal{B}_{1i} = O_p(1)$ by Assumption 2.1. Moreover, $\mathcal{B}_{3i} = \mathcal{B}_{4i} = 0$ because $\iota'_T \mathbf{J}_T = \mathbf{0}'$. We show $\mathcal{B}_{2i} = O_p(1)$ below.

Using Lemma 7, we can expand $\mathcal{B}_{2i}$ as

$$\begin{aligned}
\mathcal{B}_{2i} &= O_p(1) \left(\frac{\Xi_{N,T}}{T} \right) \left(\hat{\mathbf{F}}_C - \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}' \right) \mathbf{J}_T \mathbf{g} \phi_{ig} \\
&= O_p(1) \Xi_{N,T} \left\{ \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \right. \\
&\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\
&\quad + \left(N^{-\psi_f} T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\
&\quad + \zeta (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\
&\quad + \zeta \left(N^{-\psi_f} T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{\varepsilon} \mathbf{J}_N \mathbf{\Phi} \right) (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\
&\quad + \mathbf{\Lambda} (T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{g} \phi_{ig} \right) \\
&\quad + (T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{g} \phi_{ig} \right) \\
&\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \mathbf{\Lambda} \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \mathbf{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{g} \phi_{ig} \right) \\
&\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{g} \phi_{ig} \right) \\
&\quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \zeta \left(N^{-1} T^{-3/2} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{g} \phi_{ig} \right) \\
&\quad + \zeta (T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F}) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{g} \phi_{ig} \right) \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\
&\quad + \mathbf{\Lambda}_f (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f \right) (T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \\
&\quad + \frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g} \right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g \right) (T^{-1} \mathbf{g}' \mathbf{J}_T \mathbf{g} \phi_{ig}) \left. \right\} \\
&= O_p(1) \Xi_{N,T} O_p(\Xi_{N,T}^{-1}) = O_p(1).
\end{aligned}$$

The final equality follows from Lemmas 2 and 6, and the fact that ζ has finitely many non-zero entries by Assumption 9. Hence,

$$\max_i |\mathcal{B}_i| = O_p(1).$$

Next,

$$\max_i |\mathcal{C}_i| = \max_i \left| \left(\frac{\Xi_{N,T}}{T} \right) \boldsymbol{\gamma}' (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})' \boldsymbol{\varepsilon}_i \right|.$$

Since

$$\boldsymbol{\gamma}'(\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})' = \boldsymbol{\gamma}'_S(\hat{\boldsymbol{\varepsilon}}_S - \boldsymbol{\varepsilon}_S)'$$

and, for every $j \in S$, $\boldsymbol{\phi}_{jg} = \mathbf{0}$ by Assumption 1, Lemma 9 gives

$$\hat{\boldsymbol{\varepsilon}}_j - \boldsymbol{\varepsilon}_j = \mathbf{O}_p(\Xi_{N,T}^{-1}).$$

The proof therefore follows the same steps as for $\mathcal{A}_i$, and hence

$$\max_i |\mathcal{C}_i| = O_p(1).$$

Similarly,

$$\max_i |\mathcal{D}_i| = \max_i \left| \left(\frac{\Xi_{N,T}}{T} \right) \boldsymbol{\gamma}'(\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})' \mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig} \right| = O_p(1)$$

by the same argument as for $\mathcal{B}_i$, again using Lemma 9, Assumption 1, and the fixed cardinality of S under Assumption 9. Therefore,

$$\text{II} = O_p(1).$$

Item III is

$$\text{III} = \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| [\hat{\boldsymbol{\varepsilon}}_i - (\boldsymbol{\varepsilon}_i + \mathbf{J}_T \mathbf{g} \boldsymbol{\phi}_{ig})]' \boldsymbol{\eta} \right|.$$

From the discussion leading to Lemma 9, this can be expressed as the sum of three terms:

$$\begin{aligned} \text{III} &\leq \max_i \left| \left(N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{X\mathbf{x}_i} - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if} \right)' \right. \\ &\quad \times \left. \mathbf{O}_p(1) \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \frac{\mathbf{f}' \mathbf{J}_T \boldsymbol{\eta}}{T} \Xi_{N,T} \right| \\ &\quad + \max_i \left| (\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if})' \mathbf{O}_p(1) \left(\hat{\mathbf{F}}_C - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \mathbf{f}' \right) \mathbf{J}_T \boldsymbol{\eta} \left(\frac{\Xi_{N,T}}{T} \right) \right| \\ &\quad + \left| \mathbf{O}_p(\Xi_{N,T}^{-1}) \left(\frac{\Xi_{N,T}}{T} \sum_t \eta_{t+h} \right) \right| \\ &\equiv |\mathcal{P}| + |\mathcal{Q}| + |\mathcal{R}|. \end{aligned}$$

Since $\Xi_{N,T}/T \leq T^{-1/2}$,

$$\Lambda_f \Delta_f \mathcal{P}_f \frac{\mathbf{f}' \mathbf{J}_T \boldsymbol{\eta}}{T} \Xi_{N,T} = O_p(1)$$

by Lemma 2.2. Therefore, it suffices for $\mathcal{P}$ to establish that

$$\max_i \left\| N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{X\mathbf{x}_i} - \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} \right\| = O_p(1).$$

Expanding the expression gives

$$\begin{aligned} & N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{X\mathbf{x}_i} - \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} \\ &= \Lambda \left(T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F} \right) \left(N^{-\psi_f} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\Phi} \right) \left(T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \Lambda \left(T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F} \right) \left(N^{-\psi_f} T^{-1} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \Lambda \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\Phi} \right) \left(T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \Lambda \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \left(T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F} \right) \left(N^{-\psi_f} T^{-1} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \left(N^{-\psi_f} T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\Phi} \right) \left(T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + \left(T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\omega} \right) \left(N^{-\psi_f} \boldsymbol{\Phi}' \mathbf{J}_N \boldsymbol{\Phi} \right) \left(T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon}_i \right) \\ & \quad + O_p(\Xi_{N,T}^{-1}) \phi_i. \end{aligned}$$

The last term captures terms in the expansion whose maximum order is $\Xi_{N,T}^{-1}$ and which depend on i only through ϕ_i . Hence, by Assumption 7.1 and the same uniform boundedness arguments used for $\mathcal{A}_{2i}$,

$$\max_i \left\| N^{-\psi_f} T^{-2} \mathbf{W}'_{XZ} \mathbf{S}_{X\mathbf{x}_i} - \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} \right\| = O_p(1),$$

and therefore $\mathcal{P} = O_p(1)$.

For $\mathcal{Q}$, note that

$$\max_i \left\| \Lambda_f \Delta_f \mathcal{P}_f \Delta_f \phi_{if} \right\| = O_p(1)$$

by Assumption 7.1. Hence, it suffices to show

$$\left(\frac{\Xi_{N,T}}{T}\right) \left(\hat{\mathbf{F}}_C - \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}'\right) \mathbf{J}_T \boldsymbol{\eta} = O_p(1).$$

Expanding as in Lemma 7,

$$\begin{aligned} & \frac{\Xi_{N,T}}{T} \left(\hat{\mathbf{F}}_C - \mathbf{\Lambda}_f \mathbf{\Delta}_f \mathcal{P}_f \mathbf{f}'\right) \mathbf{J}_T \boldsymbol{\eta} \\ &= \mathbf{\Lambda} \left(T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}\right) \left(N^{-\psi_f} \frac{\Xi_{N,T}}{T} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \mathbf{\Lambda} \left(N^{-\psi_f} T^{-1} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi}\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{F}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \left(T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}\right) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi}\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{F}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \left(N^{-\psi_f} T^{-1} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi}\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{F}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \boldsymbol{\zeta} \left(T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \mathbf{F}\right) \left(N^{-\psi_f} \mathbf{\Phi}' \mathbf{J}_N \mathbf{\Phi}\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{F}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \boldsymbol{\zeta} \left(N^{-\psi_f} T^{-1} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \mathbf{\Phi}\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{F}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \Xi_{N,T} \mathbf{\Lambda} \left(T^{-1} \mathbf{F}' \mathbf{J}_T \mathbf{F}\right) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \left(\frac{\Xi_{N,T}}{T} \boldsymbol{\omega}' \mathbf{J}_T \mathbf{F}\right) \left(N^{-\psi_f} T^{-1} \mathbf{\Phi}' \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \Xi_{N,T} \frac{N^{1-\psi_f}}{\sqrt{T}} \mathbf{\Lambda} \left(N^{-1} T^{-3/2} \mathbf{F}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \Xi_{N,T} \frac{N^{1-\psi_f}}{\sqrt{T}} \left(N^{-1} T^{-3/2} \boldsymbol{\omega}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \Xi_{N,T} \frac{N^{1-\psi_f}}{\sqrt{T}} \boldsymbol{\zeta} \left(N^{-1} T^{-3/2} \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\varepsilon} \mathbf{J}_N \boldsymbol{\varepsilon}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \mathbf{\Lambda}_f \left(T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{f}\right) \left(N^{-\psi_f} \mathbf{\Phi}'_f \mathbf{J}_N \mathbf{\Phi}_g\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{g}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \mathbf{\Lambda}_f \left(T^{-1} \mathbf{f}' \mathbf{J}_T \mathbf{g}\right) \left(N^{-\psi_f} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_f\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{f}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &+ \frac{N^{\psi_g - \psi_f}}{\sqrt{T}} \mathbf{\Lambda}_f \left(T^{-1/2} \mathbf{f}' \mathbf{J}_T \mathbf{g}\right) \left(N^{-\psi_g} \mathbf{\Phi}'_g \mathbf{J}_N \mathbf{\Phi}_g\right) \left(\frac{\Xi_{N,T}}{T} \mathbf{g}' \mathbf{J}_T \boldsymbol{\eta}\right) \\ &= O_p(1). \end{aligned}$$

Since $\Xi_{N,T}/T \leq T^{-1/2}$ and $N^{1-\psi_f}/\sqrt{T} = O(1)$, the final equality follows from the rates derived in Lemmas 2 and 6, together with the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries by Assumption

9. Hence $\mathcal{Q} = O_p(1)$.

Finally,

$$\mathcal{R} = O_p(\Xi_{N,T}^{-1}) \left(\frac{\Xi_{N,T}}{T} \sum_t \eta_{t+h} \right) = O_p(1)$$

because $\Xi_{N,T}/T \leq T^{-1/2}$ and $T^{-1/2} \sum_t \eta_{t+h} = O_p(1)$ by Lemma 1.4. Therefore,

$$\text{III} = O_p(1).$$

Item IV is

$$\text{IV} = \left(\frac{\Xi_{N,T}}{T} \right) \max_i \left| [\hat{\epsilon}_i - (\epsilon_i + \mathbf{J}_T \mathbf{g} \phi_{ig})]' (\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \right|.$$

Using

$$\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta} = -(\hat{\mathbf{y}}_f - \mathbf{y}_f) - (\hat{\epsilon} - \epsilon) \boldsymbol{\gamma},$$

Item IV is the sum of the products of the generated-residual estimation error appearing in Item III with the two components of $\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}$ treated in Item II. Applying the same uniform bounds established in Items II and III, together with Lemma 9, Assumption 7, and the fixed cardinality of S under Assumption 9, gives

$$\text{IV} = O_p(1).$$

Consequently,

$$\max_i \left(\frac{\Xi_{N,T}}{T} \right) |\hat{\epsilon}'_i \tilde{\boldsymbol{\eta}} - \epsilon'_i \boldsymbol{\eta}| = O_p(1),$$

which completes the proof.

Theorem 5 Let the regularization parameter in the Stage 2 Pass 5 regression be given by $\lambda := 8 \frac{\sqrt{c + \kappa \log N}}{\Xi_{N,T}}$, where $c > 0$ and κ is defined in Assumption 10.2. Then, if Assumptions 1-10 hold and $\frac{T}{N} = O(1)$, with probability at least $1 - \exp\left(-\frac{c}{\kappa}\right) + o(1)$, we have

$$\frac{1}{\sqrt{T}} \|\hat{\epsilon} \hat{\boldsymbol{\gamma}} - \epsilon \boldsymbol{\gamma}\|_2 = O_p\left(\frac{\sqrt{\log N}}{\Xi_{N,T}}\right).$$

Proof: The first-stage regression gives the initial forecast $\hat{\mathbf{y}}_f = \boldsymbol{\iota}_T \bar{y} + \mathbf{J}_T \hat{\mathbf{F}} \hat{\boldsymbol{\beta}}$. Let $\hat{\mathbf{u}} = (\hat{u}_{1+h}, \dots, \hat{u}_{T+h})'$

denote the vector of stacked residuals from the first-stage regression, so that $\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}}_f$. The second stage involves the LASSO regression of $\hat{\mathbf{u}}$ on $\hat{\boldsymbol{\varepsilon}}$, where both are generated quantities:

$$\hat{\boldsymbol{\gamma}} = \arg \min_{\boldsymbol{\gamma}} \left\{ \|\hat{\mathbf{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}\|_2^2/T + \lambda \|\boldsymbol{\gamma}\|_1 \right\}.$$

The LASSO solution satisfies

$$\|\hat{\mathbf{u}} - \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\gamma}}\|_2^2/T + \lambda \|\hat{\boldsymbol{\gamma}}\|_1 \leq \|\hat{\mathbf{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}\|_2^2/T + \lambda \|\boldsymbol{\gamma}\|_1. \quad (\text{A5.1})$$

Since $\tilde{\boldsymbol{\eta}} = \hat{\mathbf{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}$ is defined in Lemma 10, equation A5.1 gives

$$\frac{\|\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\|_2^2}{T} + \lambda \|\hat{\boldsymbol{\gamma}}\|_1 \leq 2(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\boldsymbol{\varepsilon}}' \tilde{\boldsymbol{\eta}}/T + \lambda \|\boldsymbol{\gamma}\|_1.$$

Moreover,

$$2 |(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\boldsymbol{\varepsilon}}' \tilde{\boldsymbol{\eta}}| \leq \left(\max_{1 \leq j \leq N} 2 |\hat{\boldsymbol{\varepsilon}}'_j \tilde{\boldsymbol{\eta}}| \right) \|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_1.$$

Let $\lambda_0 := 4 \frac{\sqrt{c + \kappa \log N}}{\Xi_{N,T}}$, so that $\lambda = 2\lambda_0$, and define

$$\mathcal{T} := \left\{ \max_{1 \leq j \leq N} \frac{2 |\hat{\boldsymbol{\varepsilon}}'_j \tilde{\boldsymbol{\eta}}|}{T} \leq \lambda_0 \right\}.$$

By Lemma 10,

$$\max_j \frac{2\Xi_{N,T}}{T} |\hat{\boldsymbol{\varepsilon}}'_j \tilde{\boldsymbol{\eta}}| \leq 2 \max_j \frac{|\boldsymbol{\varepsilon}'_j \boldsymbol{\eta}|}{\sqrt{T}} + O_p(1),$$

where we have used $\Xi_{N,T}/T \leq T^{-1/2}$. Since $\sqrt{c + \kappa \log N} \rightarrow \infty$, the $O_p(1)$ term is smaller than $2\sqrt{c + \kappa \log N}$ with probability approaching one. Hence,

$$\begin{aligned} \mathbb{P}(\mathcal{T}^c) &\leq \mathbb{P} \left(\max_{1 \leq j \leq N} \frac{|\boldsymbol{\varepsilon}'_j \boldsymbol{\eta}|}{\sqrt{T}} > \sqrt{c + \kappa \log N} \right) + o(1) \\ &\leq N \exp \left(-\frac{c + \kappa \log N}{\kappa} \right) + o(1) \\ &= \exp \left(-\frac{c}{\kappa} \right) + o(1), \end{aligned}$$

where the second inequality follows from Assumption 10.2. Therefore,

$$\mathbb{P}(\mathcal{T}) \geq 1 - \exp\left(-\frac{c}{\kappa}\right) + o(1).$$

On $\mathcal{T}$, with $\lambda = 2\lambda_0$,

$$2\frac{\|\hat{\varepsilon}(\hat{\gamma} - \gamma)\|_2^2}{T} + \lambda\|\hat{\gamma}_{S^c}\|_1 \leq 3\lambda\|\hat{\gamma}_S - \gamma_S\|_1. \quad (\text{A5.2})$$

This follows from the standard LASSO basic-inequality argument; see Lemma 6.3 of [Bühlmann & Van De Geer \[2011\]](#). In particular,

$$\|\hat{\gamma}_{S^c}\|_1 \leq 3\|\hat{\gamma}_S - \gamma_S\|_1.$$

Define

$$\Delta_{\hat{\varepsilon}} := \frac{\hat{\varepsilon}'\hat{\varepsilon}}{T}.$$

Using Lemma 9 and the same uniform arguments employed in Lemma 10, we can obtain

$$\max_{i,j} \left| \Delta_{\hat{\varepsilon}}^{ij} - \Delta_{\varepsilon,g}^{ij} \right| = o_p(1). \quad (\text{A5.3})$$

By Assumption 10.3, with probability approaching one, for every Θ satisfying

$$\|\Theta_{S^c}\|_1 \leq 3\|\Theta_S\|_1,$$

we have

$$\|\Theta_S\|_1^2 \leq (\Theta' \Delta_{\varepsilon,g} \Theta) \frac{|S|}{\nu_0^2}.$$

For such vectors,

$$\|\Theta\|_1 \leq 4\|\Theta_S\|_1.$$

Therefore,

$$\begin{aligned} |\boldsymbol{\Theta}'(\Delta_{\varepsilon} - \Delta_{\varepsilon,g})\boldsymbol{\Theta}| &\leq \max_{i,j} |\Delta_{\varepsilon}^{ij} - \Delta_{\varepsilon,g}^{ij}| \|\boldsymbol{\Theta}\|_1^2 \\ &\leq 16o_p(1) \|\boldsymbol{\Theta}_S\|_1^2. \end{aligned}$$

Since $|S|$ is fixed, equation A5.3 implies that, with probability approaching one,

$$\boldsymbol{\Theta}'\Delta_{\varepsilon}\boldsymbol{\Theta} \geq \frac{\nu_0^2}{2|S|} \|\boldsymbol{\Theta}_S\|_1^2.$$

Thus the compatibility condition holds, with probability approaching one, for set S with respect to Δ_{ε} and compatibility constant $\nu_0/\sqrt{2}$.

Let $\boldsymbol{\delta} = \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}$. Using equation A5.2 and the compatibility condition just established,

$$\begin{aligned} 2\frac{\|\hat{\boldsymbol{\varepsilon}}\boldsymbol{\delta}\|_2^2}{T} + \lambda\|\boldsymbol{\delta}\|_1 &\leq 4\lambda\|\boldsymbol{\delta}_S\|_1 \\ &\leq \frac{4\sqrt{2}\lambda\sqrt{|S|}}{\nu_0} \frac{\|\hat{\boldsymbol{\varepsilon}}\boldsymbol{\delta}\|_2}{\sqrt{T}} \\ &\leq \frac{\|\hat{\boldsymbol{\varepsilon}}\boldsymbol{\delta}\|_2^2}{T} + \frac{8\lambda^2|S|}{\nu_0^2}. \end{aligned}$$

Hence,

$$\frac{\|\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\|_2^2}{T} + \lambda\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_1 \leq \frac{8\lambda^2|S|}{\nu_0^2}. \quad (\text{A5.4})$$

Since $|S|$ is finite,

$$\frac{\|\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\|_2}{\sqrt{T}} = O_p(\lambda) = O_p\left(\frac{\sqrt{\log N}}{\Xi_{N,T}}\right). \quad (\text{A5.5})$$

Finally, using the triangle inequality,

$$\begin{aligned} \frac{1}{\sqrt{T}} \|\hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\gamma}} - \boldsymbol{\varepsilon}\boldsymbol{\gamma}\|_2 &\leq \frac{1}{\sqrt{T}} \|\hat{\boldsymbol{\varepsilon}}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\|_2 + \frac{1}{\sqrt{T}} \|(\hat{\boldsymbol{\varepsilon}}_S - \boldsymbol{\varepsilon}_S)\boldsymbol{\gamma}_S\|_2 \\ &= O_p\left(\frac{\sqrt{\log N}}{\Xi_{N,T}}\right) + O_p(\Xi_{N,T}^{-1}) \\ &= O_p\left(\frac{\sqrt{\log N}}{\Xi_{N,T}}\right). \end{aligned}$$

The second term follows from Lemma 9, the fixed cardinality of S , and the fact that $\boldsymbol{\phi}_{jg} = \mathbf{0}$ for

every $j \in S$ by Assumption 1. This completes the proof.

Corollary 5.1 follows directly from the triangle inequality by combining Theorems 5 and 1(b).

Appendix B

B1 Additional Simulation Results

This section presents additional out-of-sample R^2 results that supplement those reported in the main paper. For $N = 100$ and $T = 100$, the main paper reports results for all four combinations of $\psi_f, \psi_g \in \{1, 0.8\}$. For $N = 200$ and $T = 100$, it reports only the benchmark case $(\psi_f, \psi_g) = (1, 1)$. We therefore report here the remaining three factor-strength configurations for $N = 200$ and $T = 100$: $(\psi_f, \psi_g) = (1, 0.8)$, $(0.8, 1)$, and $(0.8, 0.8)$.

Table B1: Out-of-sample R^2 values: $N = 200$, $T = 100$, $\psi_f = 1$, $\psi_g = 0.8$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.580	0.272	0.303	0.209	0.464 ^[1]	0.448 ^[2]	0.441 ^[3]
0.3	0.9	0.3	0	0.576	0.236	0.294	0.149	0.446 ^[1]	0.419 ^[2]	0.405 ^[3]
0.3	0.9	0.3	1	0.580	0.239	0.308	0.337	0.508 ^[1]	0.504 ^[2]	0.496 ^[3]
0.3	0.9	0.9	0	0.567	0.213	0.278 ^[1]	-1.389	0.239 ^[3]	0.216	0.275 ^[2]
0.3	0.9	0.9	1	0.588	0.167	0.262	-0.342	0.442 ^[1]	0.439 ^[2]	0.433 ^[3]
0.9	0.3	0.3	0	0.600	0.292	0.331	0.182	0.498 ^[1]	0.476 ^[2]	0.465 ^[3]
0.9	0.3	0.3	1	0.596	0.285	0.329	0.326	0.531 ^[1]	0.525 ^[3]	0.528 ^[2]
0.9	0.3	0.9	0	0.574	0.201	0.244	-1.027	0.260 ^[3]	0.272 ^[1]	0.271 ^[2]
0.9	0.3	0.9	1	0.594	0.214	0.268	-0.603	0.430 ^[1]	0.341 ^[3]	0.408 ^[2]

Notes: Entries report mean out-of-sample R^2 values across Monte Carlo replications. T denotes the training-sample size and is fixed at 100. The testing sample has the same size as the training sample. Superscripts ^[1], ^[2], and ^[3] denote, respectively, the highest, second-highest, and third-highest out-of-sample R^2 values among PCR, 3PRF, PCR LASSO, 3PRF LASSO, LASSO, and Elastic Net. The Oracle is excluded from these rankings.

Table B2: Out-of-sample R^2 values: $N = 200$, $T = 100$, $\psi_f = 0.8$, $\psi_g = 1$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.573	0.245	0.131	0.165	0.455 ^[3]	0.481 ^[2]	0.483 ^[1]
0.3	0.9	0.3	0	0.575	0.217	0.096	0.097	0.414 ^[3]	0.439 ^[2]	0.446 ^[1]
0.3	0.9	0.3	1	0.566	0.196	0.102	0.328	0.460 ^[3]	0.491 ^[2]	0.496 ^[1]
0.3	0.9	0.9	0	0.574	0.088	-0.028	-1.108	0.259 ^[3]	0.357 ^[2]	0.372 ^[1]
0.3	0.9	0.9	1	0.573	0.164	0.076	-0.306	0.389 ^[3]	0.426 ^[2]	0.449 ^[1]
0.9	0.3	0.3	0	0.612	0.291	0.143	0.253	0.466 ^[3]	0.481 ^[2]	0.491 ^[1]
0.9	0.3	0.3	1	0.590	0.274	0.158	0.335	0.511 ^[3]	0.529 ^[2]	0.531 ^[1]
0.9	0.3	0.9	0	0.575	0.205	0.111	-1.248	0.322 ^[3]	0.382 ^[2]	0.407 ^[1]
0.9	0.3	0.9	1	0.577	0.192	0.128	-0.077	0.453 ^[3]	0.481 ^[1]	0.480 ^[2]

Notes: See the notes to Table B1.

Table B3: Out-of-sample R^2 values: $N = 200$, $T = 100$, $\psi_f = 0.8$, $\psi_g = 0.8$

ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	PCR LASSO	3PRF LASSO	LASSO	Elastic Net
0	0	0	0	0.561	0.257	0.268	0.137	0.411 ^[3]	0.422 ^[1]	0.421 ^[2]
0.3	0.9	0.3	0	0.581	0.235	0.260	0.096	0.408 ^[3]	0.423 ^[1]	0.423 ^[2]
0.3	0.9	0.3	1	0.572	0.221	0.282	0.326	0.468 ^[3]	0.495 ^[1]	0.491 ^[2]
0.3	0.9	0.9	0	0.595	0.186	0.198 ^[3]	-1.444	0.193	0.229 ^[2]	0.263 ^[1]
0.3	0.9	0.9	1	0.595	0.199	0.224	-0.461	0.311 ^[3]	0.423 ^[2]	0.440 ^[1]
0.9	0.3	0.3	0	0.603	0.275	0.287	0.154	0.436 ^[3]	0.444 ^[1]	0.443 ^[2]
0.9	0.3	0.3	1	0.586	0.271	0.303	0.301	0.492 ^[3]	0.508 ^[1]	0.508 ^[2]
0.9	0.3	0.9	0	0.602	0.219 ^[3]	0.252 ^[1]	-1.250	0.026	0.228 ^[2]	0.163
0.9	0.3	0.9	1	0.557	0.192	0.247	-0.325	0.370 ^[3]	0.416 ^[2]	0.425 ^[1]

Notes: See the notes to Table B1.

B2 Variable Selection in 3PRF LASSO and PCR LASSO

We next compare the variable-selection performance of the second-stage LASSO regression in 3PRF LASSO and PCR LASSO. Tables B4 and B5 report the true-positive rate (TPR) and false-positive rate (FPR) for all four factor-strength configurations considered in the main paper. The first table considers $N = 100$ and $T = 100$, while the second considers $N = 200$ and $T = 100$.

The distinction between the two procedures is important for the estimated regressors entering the second-stage LASSO. 3PRF estimates the target-relevant factor and removes its contribution from the predictors. As a result, the components generated by the target-irrelevant factors remain in the residuals of predictors that load on these factors. This induces substantial dependence among many of the irrelevant regressors entering the second-stage LASSO. PCR, in contrast, estimates and removes both the relevant and irrelevant common factors. Consequently, much of this common dependence is removed from the residualized predictors used in PCR LASSO.

The results in Tables B4 and B5 show that 3PRF LASSO generally has a considerably lower FPR than PCR LASSO, while its TPR is often similar and in many configurations is higher. The difference in FPR is particularly pronounced when $N = 200$. This suggests that the stronger dependence among irrelevant regressors generated by the 3PRF residualization may actually facilitate sparse selection by reducing spurious inclusion of irrelevant predictors. We examine this mechanism directly in the following section, where we isolate the correlation among irrelevant predictors in a separate simulation exercise.

Table B4: True- and false-positive rates: $N = 100$, $T = 100$

Parameters				$\psi_f = 1, \psi_g = 1$		$\psi_f = 1, \psi_g = 0.8$		$\psi_f = 0.8, \psi_g = 1$		$\psi_f = 0.8, \psi_g = 0.8$	
ρ_f	ρ_g	a	d	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
<i>3PRF LASSO</i>											
0	0	0	0	0.995	0.130	0.980	0.112	0.988	0.139	0.980	0.128
0.3	0.9	0.3	0	1.000	0.114	0.990	0.125	0.985	0.151	0.935	0.142
0.3	0.9	0.3	1	0.845	0.074	0.792	0.072	0.748	0.098	0.690	0.078
0.3	0.9	0.9	0	0.868	0.115	0.797	0.106	0.792	0.124	0.640	0.098
0.3	0.9	0.9	1	0.685	0.070	0.660	0.070	0.632	0.087	0.545	0.076
0.9	0.3	0.3	0	0.990	0.110	0.975	0.105	0.978	0.153	0.940	0.146
0.9	0.3	0.3	1	0.833	0.079	0.815	0.062	0.705	0.110	0.703	0.076
0.9	0.3	0.9	0	0.892	0.105	0.780	0.103	0.812	0.120	0.733	0.112
0.9	0.3	0.9	1	0.738	0.069	0.647	0.055	0.655	0.092	0.605	0.070
<i>PCR LASSO</i>											
0	0	0	0	0.985	0.210	0.978	0.189	0.980	0.226	0.993	0.214
0.3	0.9	0.3	0	0.978	0.191	0.965	0.207	0.973	0.188	0.953	0.204
0.3	0.9	0.3	1	0.797	0.100	0.785	0.099	0.790	0.094	0.770	0.112
0.3	0.9	0.9	0	0.688	0.184	0.705	0.153	0.713	0.156	0.652	0.136
0.3	0.9	0.9	1	0.623	0.091	0.608	0.090	0.647	0.084	0.595	0.096
0.9	0.3	0.3	0	0.985	0.207	0.978	0.182	0.978	0.206	0.958	0.200
0.9	0.3	0.3	1	0.802	0.103	0.802	0.101	0.775	0.107	0.795	0.110
0.9	0.3	0.9	0	0.740	0.189	0.760	0.150	0.735	0.176	0.740	0.166
0.9	0.3	0.9	1	0.645	0.088	0.632	0.086	0.613	0.099	0.635	0.107

Notes: TPR is the proportion of the truly relevant idiosyncratic predictors that are selected, while FPR is the proportion of the remaining predictors that are selected. Entries are averages across Monte Carlo replications.

Table B5: True- and false-positive rates: $N = 200$, $T = 100$

Parameters				$\psi_f = 1, \psi_g = 1$		$\psi_f = 1, \psi_g = 0.8$		$\psi_f = 0.8, \psi_g = 1$		$\psi_f = 0.8, \psi_g = 0.8$	
ρ_f	ρ_g	a	d	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
<i>3PRF LASSO</i>											
0	0	0	0	0.990	0.068	0.980	0.095	0.988	0.105	0.975	0.112
0.3	0.9	0.3	0	0.985	0.081	0.955	0.080	0.978	0.115	0.950	0.090
0.3	0.9	0.3	1	0.828	0.047	0.787	0.052	0.743	0.066	0.718	0.054
0.3	0.9	0.9	0	0.812	0.060	0.690	0.065	0.757	0.078	0.510	0.068
0.3	0.9	0.9	1	0.670	0.044	0.630	0.041	0.590	0.059	0.547	0.059
0.9	0.3	0.3	0	0.983	0.073	0.963	0.073	0.980	0.112	0.958	0.113
0.9	0.3	0.3	1	0.840	0.049	0.780	0.048	0.743	0.068	0.720	0.046
0.9	0.3	0.9	0	0.885	0.083	0.672	0.060	0.812	0.070	0.588	0.088
0.9	0.3	0.9	1	0.743	0.037	0.657	0.044	0.738	0.055	0.552	0.043
<i>PCR LASSO</i>											
0	0	0	0	0.968	0.316	0.983	0.300	0.958	0.313	0.965	0.318
0.3	0.9	0.3	0	0.960	0.279	0.948	0.266	0.950	0.283	0.943	0.286
0.3	0.9	0.3	1	0.777	0.127	0.760	0.122	0.765	0.123	0.765	0.120
0.3	0.9	0.9	0	0.647	0.206	0.630	0.213	0.605	0.178	0.632	0.221
0.3	0.9	0.9	1	0.560	0.100	0.562	0.099	0.573	0.110	0.595	0.117
0.9	0.3	0.3	0	0.968	0.297	0.970	0.293	0.953	0.253	0.950	0.302
0.9	0.3	0.3	1	0.743	0.126	0.777	0.151	0.785	0.139	0.752	0.156
0.9	0.3	0.9	0	0.647	0.201	0.642	0.210	0.642	0.216	0.650	0.242
0.9	0.3	0.9	1	0.635	0.094	0.610	0.112	0.625	0.091	0.545	0.101

Notes: See the notes to Table B4.

We conduct an additional simulation exercise to examine how correlation among irrelevant predictors affects the variable-selection performance of LASSO. The design has sample size $n = 100$ and contains five relevant predictors and 100 irrelevant predictors. The observations are generated independently from a multivariate normal distribution with mean zero and covariance matrix

$$\Sigma = \begin{pmatrix} \mathbf{I}_5 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{100} + \rho_1 \boldsymbol{\iota} \boldsymbol{\iota}' \end{pmatrix},$$

where $\boldsymbol{\iota}$ is a 100×1 vector of ones. Thus, the five relevant predictors are mutually uncorrelated and are also uncorrelated with the irrelevant predictors. The diagonal elements of the covariance matrix for the irrelevant predictors are $1 + \rho_1$, while its off-diagonal elements are ρ_1 . Consequently, the pairwise correlation between two distinct irrelevant predictors is

$$\rho = \frac{\rho_1}{1 + \rho_1}.$$

We consider

$$\rho_1 \in \{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.5, 1.8\},$$

which corresponds to

$$\rho \in \{0, 0.167, 0.286, 0.375, 0.444, 0.5, 0.545, 0.6, 0.643\}.$$

After generation, each predictor is demeaned and standardized. The target is generated as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + 2\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_n),$$

where

$$\boldsymbol{\theta} = (\beta, \beta, \beta, \beta, \beta, 0, \dots, 0)', \quad \beta \in \{0.1, 0.2, \dots, 1\}.$$

Therefore, the first five predictors are relevant and the remaining 100 predictors are irrelevant. LASSO is estimated using 10-fold cross-validation. The true-positive rate is the proportion of the five relevant predictors selected by LASSO, while the false-positive rate is the proportion of the 100 irrelevant predictors selected. The reported results are averages across 100 Monte Carlo

replications.

Table B6 shows that stronger correlation among irrelevant predictors generally reduces the false-positive rate. This pattern is especially clear at moderate and strong signal levels. At the same time, the true-positive rate is generally maintained and frequently increases as the correlation among irrelevant predictors becomes stronger. Hence, in this design, stronger dependence within the irrelevant predictor block does not impair the selection of relevant predictors and tends to discourage LASSO from selecting irrelevant predictors. The results are reported on the following page.

Table B6: Effect of correlation among irrelevant predictors on LASSO variable selection

β	$\rho = 0$		$\rho = 0.167$		$\rho = 0.286$		$\rho = 0.375$		$\rho = 0.444$		$\rho = 0.5$		$\rho = 0.545$		$\rho = 0.6$		$\rho = 0.643$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
0.1	0.044	0.038	0.074	0.030	0.084	0.031	0.082	0.031	0.084	0.033	0.088	0.028	0.094	0.029	0.128	0.033	0.126	0.029
0.2	0.102	0.042	0.106	0.028	0.136	0.028	0.176	0.035	0.178	0.036	0.180	0.033	0.228	0.034	0.234	0.035	0.268	0.036
0.3	0.210	0.058	0.320	0.076	0.346	0.075	0.344	0.071	0.340	0.067	0.370	0.061	0.386	0.059	0.410	0.055	0.448	0.064
0.4	0.330	0.060	0.420	0.079	0.438	0.075	0.452	0.074	0.492	0.074	0.530	0.072	0.534	0.061	0.560	0.058	0.598	0.057
0.5	0.636	0.108	0.704	0.119	0.746	0.118	0.762	0.110	0.802	0.104	0.820	0.102	0.830	0.101	0.850	0.100	0.854	0.091
0.6	0.784	0.131	0.852	0.133	0.872	0.132	0.892	0.122	0.910	0.122	0.914	0.116	0.928	0.103	0.922	0.088	0.912	0.089
0.7	0.906	0.136	0.950	0.153	0.962	0.146	0.960	0.140	0.966	0.130	0.976	0.123	0.982	0.112	0.976	0.102	0.978	0.093
0.8	0.964	0.147	0.972	0.164	0.976	0.153	0.990	0.147	0.992	0.149	0.992	0.130	0.998	0.131	0.998	0.119	1.000	0.105
0.9	0.988	0.152	0.998	0.155	1.000	0.138	1.000	0.137	1.000	0.129	0.998	0.117	0.998	0.114	0.998	0.112	0.998	0.106
1	0.996	0.159	1.000	0.146	1.000	0.137	1.000	0.128	1.000	0.120	1.000	0.109	1.000	0.109	1.000	0.104	1.000	0.098

Notes: Entries report average LASSO true-positive rates and false-positive rates across 100 Monte Carlo replications. The sample size is $n = 100$. There are five relevant and 100 irrelevant predictors. The signal coefficient β is common to the five relevant predictors. The correlation ρ is the common pairwise correlation between distinct irrelevant predictors.

References

- Bühlmann, Peter, & Van De Geer, Sara. 2011. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.
- Kelly, Bryan, & Pruitt, Seth. 2015. The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, **186**(2), 294–316.