Forecasting Using Supervised Factors and Idiosyncratic Elements

Daanish Padha* and Tae-Hwy Lee†

January 20, 2025

Abstract

We extend the Three-Pass Regression Filter (3PRF) in two key dimensions: (1) accommodating weak factors and, (2) allowing for a correlation between the target variable and the predictors, even after adjusting for common factors, driven by correlations in the idiosyncratic components of the covariates and the prediction target. Our theoretical contribution is to establish the consistency of 3PRF under these flexible assumptions, showing that relevant factors can be consistently estimated even when they are weak, albeit at slower rates. Stronger relevant factors improve 3PRF convergence to the infeasible best forecast, while weaker relevant factors dampen it. Conversely, stronger irrelevant factors hinder the rate of convergence, whereas weaker irrelevant factors enhance it. We compare 3PRF with Principal Component Regression (PCR), highlighting scenarios where 3PRF performs better. Methodologically, we extend 3PRF by integrating a LASSO step to develop the 3PRF LASSO estimator, which effectively captures the target's dependency on the predictors' idiosyncratic components. We derive the rate at which the average prediction error from this step converges to zero, accounting for generated regressor effects. Simulation results confirm that 3PRF performs well under these broad assumptions, with the LASSO step delivering a substantial gain. In an empirical application using the FRED-QD dataset, 3PRF LASSO delivers reliable forecasts of key macroeconomic variables across multiple horizons.

Keywords: Weak Factors, Forecasting, high dimension, supervision, three pass regression filter, LASSO.

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

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.

*Department of Economics, University of California, Riverside, CA 92521. E-mail: dpadh002@ucr.edu

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

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, $\boldsymbol{x}_t$, 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:

$$y_{t+h} = \beta_0 + \boldsymbol{\beta}' \boldsymbol{F}_t + u_{t+h}, \tag{1.1}$$

$$\boldsymbol{x}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi} \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t. \tag{1.2}$$

The N-dimensional vector of covariates $\boldsymbol{x}_t$ is decomposed into three latent components: an intercept term $\boldsymbol{\phi}_0$, a low-rank component $\boldsymbol{\Phi F}_t$, 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 factor(s) $\boldsymbol{F}_t$, 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 $\boldsymbol{X}$. This method does not incorporate information about the target variable $\boldsymbol{y}$ during factor estimation, which limits its predictive utility. To address this limitation, Kelly & Pruitt [2015] introduced the Three-Pass Regression Filter (3PRF), a supervised framework that leverages additional information from 'proxy' variables, denoted as $\boldsymbol{Z}$. These proxies, either pre-specified or constructed based on $\boldsymbol{y}$, enable the estimation of factor(s) relevant to $\boldsymbol{y}$, which often constitute a strict subset of the factors driving $\boldsymbol{X}$. 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 factor(s) 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 factor(s), a concept commonly referred to as the "strength" of the factor(s). 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 $\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}$ 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 factor(s) assumption often fails in practice. Factors may instead be weak, i.e., $\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N^{\psi}}$, 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 factor(s) can primarily be attributed to two reasons, or a combination thereof: (i) when factor(s) 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 factor(s) using the principal components method. In contrast, our focus is on evaluating the effect of weak factor(s) 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 Principal Component Regression (PCR) when the relevant factor(s) 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 factor(s) 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 factor(s) 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 $\beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t$ is optimal only if $\mathbb{E}(u_{t+h} \mid \boldsymbol{\varepsilon}_t) = 0$, as assumed by Kelly & Pruitt [2015]. However, in high-dimensional settings, it is unlikely that all idiosyncratic components $(\boldsymbol{\varepsilon}_i \mid i \in \{1, \dots, N\})$ 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 $\boldsymbol{X}$.

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.* [2023b] 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. They note, "This suggests the presence of sparsity — alongside a dense component — in widely studied economic applications." To further motivate this, consider the empirical application in Kelly & Pruitt [2015], wherein they examine the forecastability of key macroeconomic aggregates using a comprehensive set of predictors - specifically, 108 macroeconomic variables compiled by Stock & Watson [2012]. The predictors are assumed to follow an approximate factor structure, as introduced by Chamberlain & Rothschild [1983], accommodating weak cross-sectional and temporal dependence in the idiosyncratic components. Although Kelly & Pruitt [2015] do not explicitly assert that the chosen target variables are fundamentally different from other predictors, their methodology implicitly relies on the assumption that the idiosyncratic components of the target variables are martingale difference sequences and are uncorrelated with the idiosyncratic components of other predictors.[1] This implies a strict factor structure for the target variables rather than an approximate one, which is restrictive.

Few studies in the literature have sought to address this limitation, namely, the assumption that idiosyncratic components are orthogonal to the target variable—by leveraging the predictive content of these components. Notable examples include Kneip & Sarda [2011], Fan *et al.* [2020] and Fan *et al.* [2023b]. These works augment the principal components-based factor forecasting model by incorporating

[1] They assume that $\boldsymbol{y}_{t+1} - \mathbb{E}(\boldsymbol{y}_{t+1}|\boldsymbol{F}_t)$ is serially uncorrelated and independent of all idiosyncratic components.

a small subset of idiosyncratic elements relevant to the target as additional predictors. This is accomplished using regularized M-estimation methods, which capture residual dependencies between the target and predictors that are not explained by common factors and are thus attributed to the idiosyncratic components of the predictors. In doing so, these studies aim to bridge two distinct methodologies for high-dimensional estimation: sparse modeling and dense, factor-based modeling.

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 the Least Absolute Shrinkage and Selection Operator (LASSO; Tibshirani [1996]) to select relevant idiosyncratic components as additional predictors. We refer to this augmented estimator as 3PRF LASSO. Our empirical analysis demonstrates that idiosyncratic components can exhibit substantial predictive power for certain macroeconomic variables. This is evidenced by the enhanced performance of the 3PRF LASSO method compared to the original 3PRF approach, underscoring the value of integrating sparse and dense components in economic forecasting models.

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 $\boldsymbol{y}$. However, in the 3PRF framework, allowing idiosyncratic elements to possess predictive content for the target $\boldsymbol{y}$, and thus for its proxies $\boldsymbol{Z}$, introduces a form of 'corruption' in the supervision process, as clarified in Section 2.[2] 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

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

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 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 underscores the effectiveness of the 3PRF LASSO approach. We forecast four key U.S. macroeconomic variables—GDP, Exports, the GDP Deflator, and Housing Starts—using a comprehensive set of macroeconomic variables from the FRED-QD dataset by Clark & McCracken [2023]. The 3PRF LASSO method demonstrates competitive performance compared to established methods, highlighting its reliability in macroeconomic forecasting.

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 key findings and offers concluding remarks.

1.1 Definitions and notations

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

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

Given an index set $S \subset \{1, \ldots, N\}$ and a vector $\mathscr{X}$ with i-th component $\mathscr{X}_i$, define $\mathscr{X}_{i,S} = \mathscr{X}_i \mathbb{1}\{i \in S\}$, where $\mathbb{1}$ is the indicator function. For a set A, $|A|$ denotes its cardinality. For a vector $\boldsymbol{v}$, $v(m)$

denotes its m-th component. The norms we use in the paper are:

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

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

$$\|A\|_\infty = \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|, \quad \|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|.$$

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, $\boldsymbol{O}_p$ and $\boldsymbol{o}_p$ denote element-wise stochastic orders. A matrix $\boldsymbol{A}$ is $\boldsymbol{O}_p(1)$ if all elements are $O_p(1)$, and $\boldsymbol{o}_p(1)$ 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 $O_p(\min(a, b))$. $\max_i\{\boldsymbol{A}_i\}_{i\in\{1,\dots,N\}}$ denotes the element-wise maximum of matrices $\{\boldsymbol{A}_i\}_{i\in\{1,\dots,N\}}$. The abbreviation 'w.p.' stands for 'with probability' and 'w.r.t.' stands for with respect to.

2 The Estimator

We predict the target $\boldsymbol{y}$ 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 factor(s). The procedure relies on a set of 'proxies' $\boldsymbol{Z}$, which, as in Kelly & Pruitt [2015], are required to be driven by target-relevant factor(s) while remaining unaffected by target-irrelevant factor(s). This requirement, explained in greater detail in the next section, is a crucial element in identifying the target-relevant factor(s). Once we obtain the target relevant factor(s) from Stage 1, we regress each predictor $\mathbf{x}_i$ on them and estimate the residual(s). Thereafter, we perform a LASSO regression to extract any predictive content in these residuals for our target $\boldsymbol{y}$. 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 $\boldsymbol{Z}$ for $i = 1, \ldots, N$,
	$x_{i,t} = \tilde{\phi}_{0,i} + \tilde{\boldsymbol{\phi}}_i' \boldsymbol{z}_t + \hat{v}_{it}$, retain slope estimate $\tilde{\boldsymbol{\phi}}_i$.
2.	Run cross-section regression of $\boldsymbol{x}_t$ on $\tilde{\boldsymbol{\phi}}_i$ for $t = 1, \ldots, T$,
	$x_{i,t} = \tilde{\phi}_{0,t} + \tilde{\boldsymbol{\phi}}_i' \hat{\boldsymbol{F}}_t + \tilde{\varepsilon}_{it}$, retain slope estimate $\hat{\boldsymbol{F}}_t$.
3.	Run time series regression of y_{t+h} on predictive factors $\hat{\boldsymbol{F}}_t$,
	$y_{t+h} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{F}}_t + \hat{u}_{t+h}$, delivers initial forecast $\hat{y}_{t+h,f} = \hat{\beta}_0 + \hat{\boldsymbol{F}}_t' \hat{\boldsymbol{\beta}}$.
	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 $\hat{\boldsymbol{F}}$ for $i = 1, \ldots, N$,
	$x_{i,t} = \hat{\phi}_{0,i} + \hat{\boldsymbol{\phi}}_i' \hat{\boldsymbol{F}}_t + \hat{\varepsilon}_{it}$, 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 $\hat{\varepsilon}_{it}$,
	$\hat{u}_{t+h} = \hat{\boldsymbol{\gamma}}' \hat{\boldsymbol{\varepsilon}}_t + \hat{\eta}_{t+h}$.
	The final forecast is given by
6.	$\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}' \hat{\boldsymbol{F}}_t + \hat{\boldsymbol{\gamma}}' \hat{\boldsymbol{\varepsilon}}_t$

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

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

where the Stage 1 forecast is given by

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

The estimated factor(s) are given by

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

and the estimated coefficient(s) of the factor(s) are given by

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

Alternatively, we can rewrite the Stage 1 forecast as

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

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

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

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

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

$$\boldsymbol{Z} = \boldsymbol{\iota}_T\boldsymbol{\lambda}'_0 + \boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\varepsilon}\boldsymbol{\zeta}' + \boldsymbol{\omega}. \tag{2.1}$$

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 $\boldsymbol{\zeta} = \mathbf{0}$ and $\boldsymbol{\omega}$ is uncorrelated with $\boldsymbol{\varepsilon}$, then, through Pass 1 of Stage 1, we obtain $\tilde{\boldsymbol{\phi}}_i = c\boldsymbol{\phi}_i + O_p(T^{-1/2})$, 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 $\boldsymbol{\zeta} \neq 0$, since in that case, for all j such that $\zeta_j \neq 0$, we would have $\tilde{\boldsymbol{\phi}}_j = c\boldsymbol{\phi}_j + d_j + O_p(T^{-1/2})$. This d_j term is $O_p(1)$ and arises from the correlation between $\{\boldsymbol{\varepsilon}_j | \zeta_j \neq 0\}$ in the data generating process of $\boldsymbol{Z}$ and predictor(s) in the set $\Delta_j = \{\mathbf{x}_i | \boldsymbol{\varepsilon}_i \text{ is correlated with } \boldsymbol{\varepsilon}_j\}$. 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 criteria 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 original data generating process (DGP) of $\boldsymbol{y}$, 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 is generated as follows:*

$$\boldsymbol{x}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \qquad y_{t+h} = \beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t + \boldsymbol{\gamma}'\boldsymbol{\varepsilon}_t + \eta_{t+h}, \qquad \boldsymbol{z}_t = \boldsymbol{\lambda}_0 + \boldsymbol{\Lambda}\boldsymbol{F}_t + \boldsymbol{\zeta}\boldsymbol{\varepsilon}_t + \boldsymbol{\omega}_t,$$
$$\boldsymbol{X} = \boldsymbol{\iota}_T\boldsymbol{\phi}_0' + \boldsymbol{F}\boldsymbol{\Phi}' + \boldsymbol{\varepsilon}, \qquad \boldsymbol{y} = \boldsymbol{\iota}_T\beta_0 + \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\boldsymbol{\gamma} + \boldsymbol{\eta}, \qquad \boldsymbol{Z} = \boldsymbol{\iota}_T\boldsymbol{\lambda}_0' + \boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\varepsilon}\boldsymbol{\zeta}' + \boldsymbol{\omega},$$

where $\boldsymbol{F}_t = \left(\boldsymbol{f}_t', \boldsymbol{g}_t'\right)', \boldsymbol{\Phi} = \left(\boldsymbol{\Phi}_f, \boldsymbol{\Phi}_g\right), \boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_f, \boldsymbol{\Lambda}_g)$*, and* $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_f', \mathbf{0}'\right)'$ *with* $|\boldsymbol{\beta}_f| > \mathbf{0}$*.* $K_f > 0$ *is the dimension of vector* $\boldsymbol{f}_t, K_g \geq 0$ *is the dimension of vector* $\boldsymbol{g}_t, L > 0$ *is the dimension of vector* $\boldsymbol{z}_t$*, and* $K = K_f + K_g$*. Furthermore, for 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, factor(s) act as a convenient conduit relating $\boldsymbol{X}$ to $\boldsymbol{y}$. This involves an implicit assumption that $\boldsymbol{X}$ 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 factor(s) 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.* [2023a]. The latter two papers assume a DGP for the target, which takes the following form,

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\iota}_T\beta_0^* + \boldsymbol{F}\boldsymbol{\beta}^* + \boldsymbol{X}\boldsymbol{\gamma}^* + \boldsymbol{\eta} \\ &= \boldsymbol{\iota}_T\left(\beta_0^* + \boldsymbol{\phi}_0'\boldsymbol{\gamma}^*\right) + \boldsymbol{F}\left(\boldsymbol{\beta}^* + \boldsymbol{\phi}'\boldsymbol{\gamma}^*\right) + \boldsymbol{\varepsilon}\boldsymbol{\gamma}^* + \boldsymbol{\eta}. \end{aligned}$$

Comparing it with the DGP of $\boldsymbol{y}$ given in Assumption 1, one can clearly see that, $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$ and $\boldsymbol{\beta}(m) = \boldsymbol{\beta}^*(m) + \sum_i \boldsymbol{\phi}_i(m)\gamma_i^* = 0$. We assume that $\boldsymbol{\beta}^*(m) + \sum_i \boldsymbol{\phi}_i(m)\gamma_i^* = 0$ only if $\boldsymbol{\beta}^*(m) = 0$ and for all $j \in \{i|\boldsymbol{\gamma}_i^* \neq 0\}$ we have $\boldsymbol{\phi}_j(m) = 0$. We are ruling out the pathological cases where both these aforementioned quantities are not zero but the sum $\boldsymbol{\beta}^*(m) + \sum_i \boldsymbol{\phi}_i(m)\gamma_i^* = 0$. This assumption is succinctly expressed by stating that for all $j \in \{i|\boldsymbol{\gamma}_i \neq 0\}$, $\boldsymbol{\beta}(m) = 0$ 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. $\mathbb{E}\left\|\boldsymbol{F}_t\right\|^4 < M, T^{-1}\sum_{s=1}^T \boldsymbol{F}_s \xrightarrow[T\to\infty]{p} \boldsymbol{\mu}$ *and* $T^{1/2}\left(\dfrac{\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}}{T} - \boldsymbol{\Delta}_F\right) = \boldsymbol{O}_p(1)$*.*
2. $\mathbb{E}\left\|\boldsymbol{\phi}_i\right\|^4 \leq M$*. For* $v = f, g$*,* $N^{-\psi_v}\sum_{j=1}^N \boldsymbol{\phi}_{vj} \xrightarrow[N\to\infty]{p} \overline{\boldsymbol{\phi}}_v < \infty$*,* $N^{\psi_v/2}\left(\dfrac{\boldsymbol{\Phi}_v{}'\boldsymbol{J}_N\boldsymbol{\Phi}_v}{N^{\psi_v}} - \mathcal{P}_v\right) = \boldsymbol{O}_p(1)$

and for $\psi_s = \min(\psi_f, \psi_g)$, $N^{\psi_s/2}\left(\frac{\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\Phi}_g}{N^{\psi_s}} - \mathcal{P}_{fg}\right) = \boldsymbol{O}_p(1)$.

3. $\mathbb{E}(\varepsilon_{it}) = 0, \mathbb{E}\|\varepsilon_{it}\|^8 \leq M$.

4. $\mathbb{E}(\boldsymbol{\omega}_t) = \mathbf{0}, \mathbb{E}\|\boldsymbol{\omega}_t\|^4 \leq M, T^{-1/2}\sum_{s=1}^T \boldsymbol{\omega}_s = \boldsymbol{O}_p(1)$ *and* $T^{1/2}\left(\frac{\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\omega}}{T} - \boldsymbol{\Delta}_\omega\right) = \boldsymbol{O}_p(1)$.

5. $\mathbb{E}_t(\eta_{t+h}) = \mathbb{E}(\eta_{t+h} \mid y_t, F_t, y_{t-1}, F_{t-1}, \ldots) = 0, \mathbb{E}(\eta_{t+h}^2) = \delta_\eta < \infty$, *and* η_{t+h} *is independent of* $\phi_i(m)$ *and* $\varepsilon_{i,t}$ *for any* $h > 0$.

The vector $\boldsymbol{\mu}$ is non-stochastic. The vectors $\overline{\boldsymbol{\phi}}_v$ for $v = f, g$ are non-stochastic. The matrix $\boldsymbol{\Delta}_F$, 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 $\boldsymbol{\Delta}_\omega$ 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 $\beta_0 + \boldsymbol{\beta}_f'\boldsymbol{f}_t + \boldsymbol{\gamma}'\boldsymbol{\varepsilon}_t$ 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_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. $\mathbb{E}\left|N^{-\psi_v/2}T^{-1/2}\sum_{t=1}^T\sum_{i=1}^N \phi_{iv}(m)[\varepsilon_{it}\varepsilon_{jt} - \sigma_{ij,tt}]\right|^2 \leq M$.[3]

4. $\mathbb{E}\left|T^{-1/2}\sum_{t=1}^T F_t(m_1)\,\omega_t(m_2)\right|^2 \leq M$.

[3] If weakness in loadings is induced by sparsity, i.e., the factor(s) 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(s) may not be local; instead, all loadings are weak, and hence we introduce this assumption.

5. $\mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}\omega_t\left(m\right)\varepsilon_{it}\right|^2 \leq M$.

6. $\mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}F_t\left(m\right)\varepsilon_{it}\right|^2 \leq M$.

7. $\mathbb{E}\left|N^{-\psi_v/2}\sum_{i=1}^{N}\phi_{iv}\left(m_1\right)\varepsilon_{it}\right|^2 \leq M$.

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 $\mathbf{\Delta}_F$, featuring in Assumption 2, we require the following,*

1. *$\mathcal{P}_f$ is positive definite and $\mathcal{P}_{fg} = \mathbf{0}$.*

2. $\mathbf{\Delta}_F \equiv \begin{pmatrix} \mathbf{\Delta}_f & \mathbf{\Delta}_{fg} \\ \mathbf{\Delta}_{fg} & \mathbf{\Delta}_g \end{pmatrix}$ *is positive definite and $\mathbf{\Delta}_{fg} = \mathbf{0}$.*

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.[4]

Assumption 5. *(Relevant Proxies).*

1. $\mathbf{\Lambda} = \left[\begin{array}{cc} \mathbf{\Lambda}_f & \mathbf{0} \end{array}\right]$.

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

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

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

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 $r_1, \ldots, r_5$,*

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.* [2023] 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.[5] Similarly, Giglio *et al.* [2023] 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

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

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 $i \in \{1, \ldots, N\}$, if $\mathbb{E}(\mathcal{X}_i^k) < M$ for some finite k, then by Markov Inequality we have*

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

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 $\Delta_{i,\varepsilon} \equiv \left\{ j \left| \mathbb{E} \left| T^{-1/2} \sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{jt} \right|^2 \leq M < \infty \right. \right\}$. 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 $\boldsymbol{\zeta}$ introduces 'corruption' in the supervisor, which necessitates assuming weaker dependence in the idiosyncratic terms to enable consistent estimation.

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 $\boldsymbol{\zeta}_i' = 0$, where $\boldsymbol{\zeta}_i'$ denotes the i^{th} row of matrix $\boldsymbol{\zeta}'$.*

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 $\boldsymbol{y}$, the extent of corruption in the Stage 1 Pass 1 (supervising step) is greater. The assumption that $\boldsymbol{\gamma}_i = 0$ if and only if $\boldsymbol{\zeta}_i' = 0$ can be relaxed. We only require that $|\{i | \boldsymbol{\zeta}_i' \neq 0\}| < \infty$ to establish our theoretical results. However, since $\boldsymbol{Z}$ is used as a proxy for $\boldsymbol{y}$, 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. *(Stage 2 LASSO Regression). The following assumptions are necessary to establish the convergence result in Theorem 5 for the Stage 2 LASSO regression, which involves generated regressors. For $r_1, \ldots, r_5$ as defined in Assumption 7, we require:*

1. *(a)* $r \in \{r_1, \ldots r_5\}$, $\dfrac{(\log N)^r}{\sqrt{T}} = O(1)$.

   (b) $\dfrac{N^{1-\psi_f}}{\sqrt{T}} \left[(\log N)^{r_1} + (\log T)^{r_1}\right] = O(1)$.

2. *There exists a large enough constant $\kappa > 0$ s.t. $\forall i \in \{1, \ldots, N\}$, and $\forall\, T$, 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 $\Delta_{\varepsilon,g} := \left(\boldsymbol{\varepsilon} + \boldsymbol{J}_T \boldsymbol{g} \boldsymbol{\Phi}_g'\right)' \left(\boldsymbol{\varepsilon} + \boldsymbol{J}_T \boldsymbol{g} \boldsymbol{\Phi}_g'\right)/T$. For the $N \times N$ matrix $\Delta_{\varepsilon,g}$, we say that, w.r.t. $\Delta_{\varepsilon,g}$, the compatibility condition is met for some set $A \subset \{1, \ldots, N\}$, if for some compatibility constant $\nu > 0$, and for all $N \times 1$ vectors $\boldsymbol{\Theta}$ satisfying $\|\boldsymbol{\Theta}_{A^c}\|_1 \leq 3\,\|\boldsymbol{\Theta}_A\|_1$, it holds that*

$$\|\boldsymbol{\Theta}_A\|_1^2 < \left(\boldsymbol{\Theta}' \Delta_{\varepsilon,g} \boldsymbol{\Theta}\right) |A|/\nu^2.$$

   We assume that, w.p. approaching one, the compatibility condition is met for set S defined in Assumption 9, w.r.t. $\Delta_{\varepsilon,g}$ 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 factor(s), 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 Appendix A.

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

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

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

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, $\boldsymbol{\gamma} \neq 0$ *and* $\boldsymbol{\zeta} \neq 0$. *Furthermore,* $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ *and* $\frac{T}{N} = O(1)$, *then*

$$\hat{y}_{t+h,f} - \mathbb{E}\left(y_{t+h}|\boldsymbol{F}_t\right) = O_p(\Xi_{NT}^{-1}).$$

Theorem 1 (a) specifies the rate of convergence of the Stage 1 forecast when $\boldsymbol{\gamma} = 0$ and $\boldsymbol{\zeta} = 0$ but $0 < \psi_f \leq 1$ and $0 < \psi_g \leq 1$, hence generalizing Kelly & Pruitt [2015] by accommodating weak factor(s). 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 $\boldsymbol{\gamma}$ 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 $y_{t+h} - \mathbb{E}(y_{t+h}|\boldsymbol{F}_t)$ has a conditional expectation of zero with respect to information set in period t.

Remark 4. *If factor(s) are strong, Theorem 1(a) would imply that* $\hat{y}_{t+h,f} - \mathbb{E}_t y_{t+h} = O_p(\delta_{NT}^{-1})$. *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* $\sqrt{T}\tilde{y}_{t+h} \underset{T,N\to\infty}{\longrightarrow} \mathbb{E}_t y_{t+h}$. *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* $\sqrt{T}\left(\tilde{y}_{t+h} - \mathbb{E}_t y_{t+h}\right) \underset{T,N\to\infty}{\longrightarrow} 0$, *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* $O_p(N^{-1/2}) + O_p(T^{-1/2}) = O_p(\delta_{NT}^{-1})$.

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

$$N^{\psi_f}\hat{\alpha}_i \xrightarrow[T,N\to\infty]{p} \left(\boldsymbol{\phi}_{if} - N^{\psi_f - 1}\bar{\boldsymbol{\phi}}_f\right)'\boldsymbol{\beta}_f.$$

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 $\boldsymbol{f}$ 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 factor(s). When the factor(s) are weak, they may either be local or have weak (local to zero) loadings or an amalgam thereof. If the factor(s) are not pervasive, the predictive information contained within the vector $\boldsymbol{f}$ 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 factor(s) 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 $N\hat{\alpha}_i \xrightarrow[T,N\to\infty]{p} \left(\tilde{\phi}_{if} - \bar{\tilde{\boldsymbol{\phi}}}_f\right)'\boldsymbol{\beta}_f$. Consequently, when factors are pervasive but all loadings are weak (local to zero), the predictive information in $\boldsymbol{f}$ 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 factor(s).

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

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

$$\hat{\boldsymbol{\beta}} - \boldsymbol{G}_\beta\boldsymbol{\beta}_f = \boldsymbol{O}_p(\Xi_{NT}^{-1}).$$

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

$$\hat{\boldsymbol{\beta}} - \boldsymbol{G}_\beta\boldsymbol{\beta}_f = \boldsymbol{O}_p(\Xi_{NT}^{-1}).$$

Theorem 3 (a) specifies the convergence rate of the vector of predictive coefficient(s) of the factor(s), i.e., $\hat{\boldsymbol{\beta}}$ to a rotated version of the true coefficient vector $\boldsymbol{\beta}$. This generalizes Theorem 5 of Kelly & Pruitt [2015] by accommodating weak factor(s). 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 $\boldsymbol{G}_\beta$, 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 $\boldsymbol{H}_f \equiv \hat{\boldsymbol{F}}_A \hat{\boldsymbol{F}}_B^{-1} \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f$ and $\boldsymbol{H}_0 \equiv \hat{\boldsymbol{F}}_A \hat{\boldsymbol{F}}_B^{-1} \left[N^{-\psi_f} T^{-1} \boldsymbol{Z}' \boldsymbol{J}_T \boldsymbol{X} \boldsymbol{J}_N \boldsymbol{\phi}_0\right]$, where, $\hat{\boldsymbol{F}}_A = T^{-1} \boldsymbol{Z}' \boldsymbol{J}_T \boldsymbol{Z}$ and $\hat{\boldsymbol{F}}_B = N^{-\psi_f} T^{-2} \boldsymbol{Z}' \boldsymbol{J}_T \boldsymbol{X} \boldsymbol{J}_N \boldsymbol{X}' \boldsymbol{J}_T \boldsymbol{Z}$.

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

$$\hat{\boldsymbol{F}}_t - (\boldsymbol{H}_0 + \boldsymbol{H}_f \boldsymbol{f}_t) = \boldsymbol{O}_p(\Xi_{NT}^{-1}).$$

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

$$\hat{\boldsymbol{F}}_t - (\boldsymbol{H}_0 + \boldsymbol{H}_f \boldsymbol{f}_t) = \boldsymbol{O}_p(\Xi_{NT}^{-1}).$$

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 factor(s) 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 factor(s) to a rotation of the true relevant factor(s) and provide the corresponding rate. Our convergence result diverges from the one presented in Kelly & Pruitt [2015]. They demonstrate the convergence of $\hat{\boldsymbol{F}}_t$ to a vector $\boldsymbol{H}\boldsymbol{F}_t$ ($\boldsymbol{H} \neq \boldsymbol{H}_f$) at a $\sqrt{N}$ rate. However, the matrix $\boldsymbol{H}$, 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 $\boldsymbol{H}_f$ and $\boldsymbol{G}_\beta$ in Theorem 3 and Theorem 4 stems from our estimation of a vector space. These Theorems "pertain to the difference between $\left[\hat{\boldsymbol{F}}_t / \hat{\boldsymbol{\beta}}\right]$ and the space spanned by $[\boldsymbol{F}_t/\boldsymbol{\beta}]$". The product $\boldsymbol{H}_f' \boldsymbol{G}_\beta$ equals an identity matrix, thereby nullifying the rotations in the predictive coefficients and relevant factors and preserving the direction spanned by $\boldsymbol{\beta}' \boldsymbol{F}_t$. However, this characteristic is absent in Theorems 5 and 6 of Kelly & Pruitt [2015]. The matrices $\boldsymbol{H}$ and $\boldsymbol{G}_\beta$ 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 $\lambda := 2\frac{\sqrt{c + \kappa \log N}}{\Xi_{NT}}$, $c > 0$ and κ is defined in Assumption 10. Then, if Assumptions 1-10 hold, w.p. at least $1 - \left(\exp\left[-\frac{c}{\kappa}\right] + o(1)\right)$, we have,*

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

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

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

Theorem 5 establishes the rate at which the average prediction error of the Stage 2 LASSO regression converges to 0. The rate $\frac{\sqrt{\log N}}{\Xi_{NT}}$, 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 and Comparison with Principal Component regression

Given the theorems we have derived, we can now discuss the theoretical advantages that the 3PRF, and by extension Partial Least Squares (PLS), which is equivalent to auto-proxy 3PRF, offers over PCR. If $N^{\psi_f} = O(T)$, an advantage arises when relevant factors are stronger than irrelevant factors, i.e., $\psi_f > \psi_g$.[6]

To see this, consider a PCR model that uses the first K principal components, where K is equal to the number of factors, as regressors. From the discussion in Section 2.2 of Bai & Ng [2006], leading up to their Theorem 3, one can deduce the following:

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

where $\hat{y}_{t+h}^{PCR}$ is the predicted value of the target at time $t+h$, formed using predictors available at time t via PCR; $\mathbb{E}_t y_{t+h}$ denotes the conditional expectation of the target variable at $t+h$ given the information set at t; $\boldsymbol{H}$ is an invertible rotation matrix; and $\hat{\boldsymbol{F}}_t^{PC}$ represents the estimated factors obtained as the leading K principal components.

The convergence of PCR prediction to $\mathbb{E}_t y_{t+h}$ depends on two main components: the convergence of $\hat{\boldsymbol{\beta}}^{PCR}$ to a rotated version of the true coefficient vector $\boldsymbol{\beta}$, and the convergence of $\hat{\boldsymbol{F}}_t^{PC}$ to a rotated version of the true factors $\boldsymbol{F}_t$. Adapting the results of Lemma 4 from Bai & Ng [2023] to the case of heterogeneous factor strengths, if $\sqrt{T}/N^{\psi_g} = o(1)$, the first term in Equation 4.1 converges to zero at a rate of $\sqrt{T}$. For the second term, the convergence rate depends on the elements of $\hat{\boldsymbol{F}}_t^{PC} - \boldsymbol{H}\boldsymbol{F}_t$ and the structure of $\boldsymbol{\beta}'\boldsymbol{H}^{-1}$.

Even if $\boldsymbol{\beta}$ contains zero elements (for irrelevant factors), the rotation matrix $\boldsymbol{H}$ generally results in $\boldsymbol{\beta}'\boldsymbol{H}^{-1}$ having nonzero elements. Although the components of $\boldsymbol{\beta}'\boldsymbol{H}^{-1}$ corresponding to zero entries in $\boldsymbol{\beta}$ may converge to zero under specific settings, this convergence occurs more slowly than the convergence of the estimated factors to a rotated version of the true factors. To illustrate, consider a case with two factors, F_{1t} and F_{2t}, where only the first factor is relevant to the target variable (i.e., $\boldsymbol{\beta} = (\beta_1, 0)'$). If $\boldsymbol{H}$

[6] The condition $N_f^{\psi} = O(T)$ is assumed to show the impact of factor convergence rates; we avoid allowing T to determine the rates, as this would trivially render both regressions with an equivalent rate.

were the identity matrix, the convergence of the second term in (4.1) would depend on how fast $\hat{F}_{1t}^{PC}$, the first estimated factor, converges to F_{1t}.

It has been established that, under certain conditions—more restrictive than those considered in this paper— the rotation matrix $\boldsymbol{H}$ converges to the identity matrix.[7] However, this convergence occurs at a slower rate compared to the rate at which the estimated factors converge to the rotated versions of the true factors (see Lemma 3 in Freyaldenhoven [2022]). Consequently, the convergence of the second term in Equation (4.1) is determined by the rate at which the elements of the vector $\hat{\boldsymbol{F}}_t^{PC}$ converge to a rotated version of $\boldsymbol{F}_t$ and this rate is determined by the slowest converging element within $\hat{\boldsymbol{F}}_t^{PC}$. In our simplified two-factor setting, if the irrelevant factor is weaker, then from Proposition 7 of Bai & Ng [2023], we have $\hat{\boldsymbol{F}}_t^{PC} - \boldsymbol{H}\boldsymbol{F}_t = O_p(N^{-\psi_g/2})$, and hence the convergence rate of PCR, which is determined by the two terms in Equation 4.1, and would be equal to $\min(N^{\psi_g/2}, T^{1/2}) = N^{\psi_g/2}$ as $N^{\psi_g} = o(N^{\psi_f}) = O(T)$.

One may, however, wonder what if we included only the leading K_f principal components as regressors instead of all. Intuitively, this might seem advantageous, as the factors estimated using the leading K_f principal components of $\boldsymbol{X}$ are expected to converge to the strongest K_f factors. These strongest factors correspond to the relevant factors when $\psi_f > \psi_g$. Thus, it might seem that using only the first K_f principal components could result in faster convergence of PCR. However, this is not the case. The estimated factors obtained via the principal component method using the first K_f principal components converge to a rotated version of the true factors, not the true factors themselves. Consequently, unless the rotation matrix converges to the identity matrix, the leading K_f principal components will provide a noisy estimate of the relevant factors. Furthermore, the convergence of the rotation matrix to the identity matrix occurs at a slow rate, limiting the potential benefits of using only the leading K_f principal components in PCR.

To make things clear, once again, consider a case with two factors: one relevant and one irrelevant, and let $\psi_f > \psi_g$. Then, according to Bai & Ng [2023], the PC estimate of the leading factor, i.e., $\hat{F}_{1t}^{PC}$, converges to hF_t where $h = [h_1, h_2]$. For $\hat{F}_{1t}$ to serve as a reliable predictor, we need h_2 to converge to zero; otherwise, $\hat{F}_{1t}^{PC}$ would represent a linear combination of relevant and irrelevant factors, not the relevant factor alone, leading to the inconsistency of PCR (which uses only the leading principal component). Under Assumption 1 of Freyaldenhoven [2022], which is more restrictive than the assumptions in our paper, along with additional assumptions that are comparable to those we impose, we can apply Theorem

[7]See Assumptions 1(b) and 1(c) in Freyaldenhoven [2022], which are substantially more restrictive than Assumption 4 imposed in our paper regarding the covariance structure of factors. Our assumptions are placed on the population covariance matrices, whereas Freyaldenhoven [2022] requires certain assumptions to hold for sample covariance matrices. Additionally, Freyaldenhoven [2022] imposes the restrictive condition that $\boldsymbol{\Phi}'\boldsymbol{\Phi}$ (without any normalization) is diagonal, which is a much stronger requirement. Despite these more restrictive assumptions, we compare our results to their setting and demonstrate better properties.

2 of Freyaldenhoven [2022] to deduce:

$$\hat{F}_{1t}^{PC} - F_{1t} = O_p\left(N^{-\frac{\psi_f}{4}}\right) + O_p\left(N^{\frac{1}{2}-\psi_f}\right) + O_p\left(N^{1-2\psi_f}\right).$$

This follows since, one can derive the following using Lemma 3 of Freyaldenhoven [2022]:

$$h - [1,0] = O_p\left(N^{-\frac{\psi_f}{4}}\right) + O_p\left(N^{\frac{1}{2}-\psi_f}\right),$$

and $\hat{F}_{1t}^{PC} - hF_{1t} = O_p(1-2\psi_f) + O_p(-\psi_f/2)$. Therefore, if we use the leading principal component in this simplified two-factor case (with a strong relevant factor and a weak irrelevant factor), our convergence rate cannot exceed $N^{\psi_f/4}$, which is much slower than $N^{\psi_f/2}$. Hence, when comparing PCR and 3PRF in a case with strong relevant factors, we restrict our attention to cases where PCR includes all estimated factors rather than just the leading principal components corresponding to the first K_f factor(s).

3PRF performs sub-optimally when irrelevant factors are stronger than relevant factors. In such cases, as can be inferred from Proposition 7 of Bai & Ng [2023], PCR's convergence rate to the infeasible best forecast will be $\min(N^{\psi_f/2}, T^{1/2})$, whereas for 3PRF it is strictly slower, as can be seen in Theorem 1. However, this scenario is improbable, as the predictors are typically driven by multiple irrelevant factors, with at least some of these factors being probably weaker than the relevant factors. This is discussed further in Remark 6. Figure 1 illustrates the performance of 3PRF and PCR when all relevant factors have the same strength, all irrelevant factors have the same strength, and $N \asymp T$. **D** at $(1,1)$ represents the case where both relevant and irrelevant factors are strong, as described in Kelly & Pruitt [2015]. The figure highlights three distinct regions. In **Region A (Green)**, relevant factors are relatively stronger, determining the convergence rates, and 3PRF converges faster than PCR. Along the line $\psi_f = \psi_g$, the convergence rates of PCR and 3PRF are identical. In **Region B (Yellow)**, weak factor strength slows convergence rates, but consistency is preserved, and PCR outperforms 3PRF. Finally, in **Region C (Red)**, 3PRF becomes inconsistent.

Remark 6. *Introducing weaker irrelevant factors alongside stronger ones does not alter the theoretical properties of 3PRF; the strength of the strongest irrelevant factor determines its convergence rate. In contrast, the theoretical properties of PCR are affected by variations in the strength of irrelevant factors. PCR requires the inclusion of all estimated factors as regressors, as omitting even a subset introduces issues due to rotational indeterminacy. This indeterminacy arises because PCR estimates factors as linear combinations of the true underlying factors, making it difficult to distinguish relevant factors from irrelevant ones at a sufficiently fast rate. As a result, the leading principal components (equal to the number of factors driving* $\boldsymbol{X}$*) must be included as regressors in PCR.*

If weaker estimated factors are excluded, the noise in the rotation matrix can prevent PCR from con-

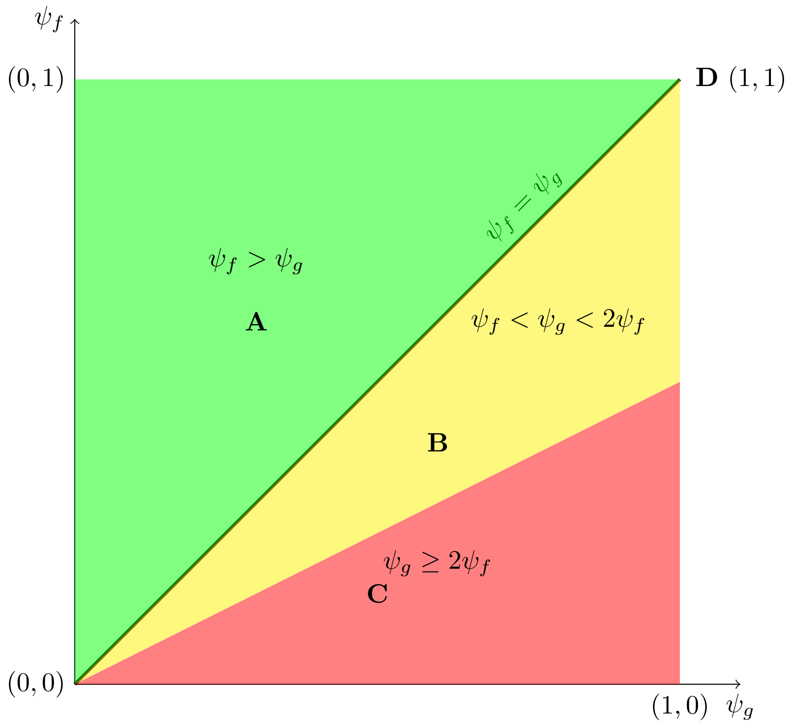

Figure 1: Convergence rate based on Factor strength. (Drawn for the case when $N \asymp T$).

verging at a fast enough rate. Including all PC estimated factors resolves this issue but results in slower convergence compared to 3PRF, as the weakest irrelevant factor then determines the rate of convergence in PCR. Therefore, while 3PRF's convergence rate is robust to variations in the strengths across irrelevant factors—being determined by the strongest irrelevant factor—PCR's performance deteriorates when there are weaker irrelevant factors, as all such factors must be estimated and included in the predictive regression. When irrelevant factors vary in strength, with some being stronger and others weaker than the relevant factors, the comparison between 3PRF and PCR becomes less clear; One must evaluate the trade-off between excluding weaker estimated factors, which can lead to issues arising from rotational indeterminacy, and including them, which may introduce problems due to their estimates being very noisy.

5 Simulation Analysis

To evaluate the performance of our estimator in finite samples, we undertake Monte Carlo experiments. The data is generated based on Assumption 1. We explore scenarios where $K_f = 1$ and $K_g = k$, with k taking values of either 4 or 5. The relevant and irrelevant factors are generated as follows: we begin by drawing the first observation from the $N(0, 1)$ distribution, and then draw subsequent observations as $\tilde{f}_t = \rho_f \tilde{f}_{t-1} + u_{f,t}$ and $\tilde{\boldsymbol{g}}_t = \rho_g \tilde{\mathbf{g}}_{t-1} + \boldsymbol{u}_{g,t}$ with $u_{f,t} \sim \text{IIN}(0, 1)$ and $\boldsymbol{u}_{g,t} \sim \text{IIN}(\mathbf{0}, \boldsymbol{\Sigma}_g)$, $u_{f,t}$ and $\boldsymbol{u}_{g,t}$ are uncorrelated and $\boldsymbol{\Sigma}_g$ is an identity matrix of order k. We divide each factor by its standard deviation to obtain $\boldsymbol{f}$ and $\boldsymbol{g}$. The parameters ρ_f and ρ_g dictate the serial correlation among factors, and they take values of 0, 0.3, or 0.9 in our setup, similar to Kelly & Pruitt [2015]. The idiosyncratic elements are generated as, $\varepsilon_{i,t} = a\varepsilon_{i,t-1} + \tilde{\varepsilon}_{i,t}$. $\tilde{\varepsilon}_{i,t} = \left(1 + d^2\right) v_{i,t} + dv_{i-1,t} + dv_{i+1,t}$ where $v_{i,t}$ is standard normal. The parameter a controls their serial correlation and takes values 0, 0.3 and 0.9 whereas d, which governs the strength of cross-sectional correlation takes values of 0 or 1. For each predictor $\mathbf{x}_i$, the loading on the relevant factor is independently drawn from a standard normal distribution and scaled by $N^{-(1-\psi_f)}$. Similarly, for each predictor, the loadings on the irrelevant factors are drawn independently from a multivariate normal distribution with zero mean and an identity covariance matrix, where the dimension of the covariance matrix is either 4 or 5. These loadings are then scaled by $N^{-(1-\psi_g)}$. Here, ψ_f represents the strength of the relevant factor, while ψ_g denotes the strength of all irrelevant factors. ψ_f and ψ_g take the values 0.7 or 1. For the set $\{i|\gamma_i \neq 0\}$, we make $\boldsymbol{\phi}_{ig} = \mathbf{0}$, in line with Assumption 1. The target variable is generated as $y_{t+1} = f_t + \alpha\boldsymbol{\gamma}'\boldsymbol{\varepsilon}_t + \eta_{t+1}$, where $\eta_{t+1} \sim \text{IIN}(0,1)$, and $\gamma = (0,1,1,1,1,\mathbf{0}_{N-5})'$. We set $\alpha = 0.3$ when $d = 1$ and $\alpha = 0.375$ when $d = 0$, ensuring that the explained variation by the factors and idiosyncratic elements remains within a narrow band. For our simulations, we use the target $\boldsymbol{y}$ as the auto-proxy.

We compare the out-of-sample performance of five methods: PCR (as described in Stock & Watson [2002]), 3PRF by Kelly & Pruitt [2015], LASSO by Tibshirani [1996], PCR LASSO, and 3PRF LASSO (our method). The PCR LASSO method is a two-stage procedure: initially, a regression of the target is performed on the leading principal components (similar to Stock & Watson [2002]), followed by regressing the residual from the initial regression on the idiosyncratic components. This process resembles our method, with the key distinction being that the factors and idiosyncratic components are estimated using an unsupervised technique, i.e., the principal components method. The hyper-parameter tuning for the LASSO regressions in our simulations is performed using 10-fold cross-validation, following the approach in Fan *et al.* [2020]. The column labeled 'Oracle' displays the average in-sample R-square value (across repeated samples) derived from the infeasible regression of $\boldsymbol{y}$ on $\boldsymbol{f}$ and the 'relevant' idiosyncratic elements, i.e., $\{\boldsymbol{\varepsilon}_i|\gamma_i \neq 0\}$. The following five columns report the average (across repeated samples) out-of-sample R-square values for the 5 aforementioned methods. We consider 100 repeated samples. To compute the out-of-sample R-squared values, we partition the sample into two halves: a training window and a testing window, each comprising 100 observations. We use a fixed estimation window as described in West & McCracken [1998].

The simulation results in Tables 1-4 show that, for a given strength of irrelevant factors, the performance of all methods improves as the strength of the relevant factors increases. Conversely, all methods, perform poorly when the strength of irrelevant factors increases for a fixed strength of relevant factors. This outcome is anticipated, as higher factor strength enhances the signal-to-noise ratio in all predictors, positively impacting these methods. However, when the strength of the relevant factor diminishes relative to irrelevant factors, the 3PRF and 3PRF LASSO methods perform the poorest. This result aligns with theoretical expectations, as the 3PRF, being a supervised method, is inherently more sensitive to the variance in predictors attributable to relevant factors compared to alternative approaches. In contrast, when the strength of relevant factors increases relative to irrelevant factors—the 3PRF and 3PRF LASSO methods exhibit the best performance.

When comparing PCR and 3PRF, we observe that 3PRF outperforms PCR in most cases, except when the irrelevant factors are stronger than the relevant ones. Therefore, even when the idiosyncratic terms possess predictive power for the target variable, 3PRF demonstrates strong performance. This indicates that the 'corruption' in the 3PRF procedure, as discussed in Section 2, has a minimal effect on its efficacy, and the advantages conferred by its supervised nature persist when compared to PCR. Consequently, the 3PRF procedure is robust to minor perturbations in the DGP of the auto-proxy.

The simulation results reveal that, in most cases, 3PRF and PCR are outperformed by either LASSO, 3PRF LASSO, or PCR LASSO. This outcome is expected, as 3PRF and PCR do not leverage the predictive power of idiosyncratic elements. When comparing 3PRF LASSO and PCR LASSO, we find that 3PRF LASSO outperforms PCR LASSO in the majority of cases. Even when 3PRF LASSO does not outperform PCR LASSO, its performance closely trails that of PCR LASSO. Conversely, when 3PRF LASSO outperforms PCR LASSO, the margin of superiority is often substantial. This suggests that augmenting the factor-based prediction model with an additional LASSO step is more effective when factors are estimated using 3PRF rather than the principal component method. This superiority arises from the supervised nature of 3PRF, which leads to better estimation of the relevant factors in Stage 1 of 3PRF LASSO relative to PCR. This improvement in Stage 1 percolates over to Stage 2, where we deal with generated regressors. Additionally, we observe that the false positive rates in Stage 2 of PCR LASSO are substantially higher than those in 3PRF LASSO, while the true positive rates remain comparable, as shown in Online Appendix Tables B11-B14. The potential reasons for this are discussed in Section B2 of the Online Appendix, where we also provide an additional experiment to corroborate these findings.

The performance differential between 3PRF LASSO and PCR LASSO is more pronounced when $K_g = 5$ compared to the case where $K_g = 4$, as evident from Tables 1 and 5.[8] Furthermore, when the training sample size $T = 100$ is half of the cross-sectional size $N = 200$, 3PRF LASSO exhibits much better performance than PCR LASSO, as shown by comparing Tables 1 and 6.[9]

When comparing LASSO and 3PRF LASSO, we observe that 3PRF LASSO outperforms LASSO when the relevant factors are relatively stronger than the irrelevant ones. Conversely, LASSO demon-

[8]The results reported in this paper examine the effect of increasing irrelevant factors with $T = 100$, $N = 100$, and $\psi_g = \psi_f = 1$. Additional simulations in the Online Appendix confirm that these findings are consistent across different sample sizes and factor strengths.

[9]This finding is robust beyond the limited comparison provided in Tables 1 and 6. Variations in the number of irrelevant factors and factor strengths have little effect on this conclusion.

strates superior performance when the relevant factors are weaker than the irrelevant factors. This result aligns with intuition: stronger relevant factors improve the convergence rate of the Stage 1 estimates in 3PRF LASSO, which in turn enhances the accuracy of Stage 2 in 3PRF LASSO. However, when the relevant factors are relatively weaker, the convergence rate of 3PRF is significantly impaired, leading to suboptimal performance in Stage 2. The benefits of explicitly modeling the factor structure become evident when the factors are strong, as this enhances the effectiveness of Stage 1 in 3PRF LASSO. The improved estimation in Stage 1 carries over to Stage 2, leading 3PRF LASSO to outperform LASSO. However, in cases where the factor structure is weak, estimation errors in Stage 1 can offset these advantages, reducing the overall effectiveness of 3PRF LASSO relative to LASSO.

Overall, 3PRF LASSO demonstrates robust performance, often outperforming its competitors, particularly when the relevant factors are relatively stronger than the irrelevant ones. Even when the strengths of relevant and irrelevant factors are similar—whether all factors are strong or all are weak—3PRF LASSO performs comparably to, and frequently better than, alternative methods. However, when the irrelevant factors are relatively stronger than the relevant ones, the performance of 3PRF LASSO declines and falls short of some of its competitors.

The tables in this paper present a subset of the simulation results discussed in Section 5. Additional simulation results, evaluating the effect of sample sizes across varying numbers of factors and factor strength combinations, are provided in the Online Appendix. True and false positive rates and the simulation results for an additional example to corroborate some of our findings are also reported in Online Appendix.

6 Empirical Application

In our empirical analysis, we assess the forecastability of four key U.S. macroeconomic aggregates: Gross Domestic Product, Exports, the GDP Deflator, and Housing Starts. The first two variables are production-related, while the GDP Deflator reflects price movements. Housing Starts is included due to its role in previous empirical work by Fan *et al.* [2023a], which found that combining factor and sparse regression methods outperformed both PCR and LASSO in predicting Housing Starts in Northeast United States. For our study, we analyze overall Housing Starts across the United States.

These variables, along with their predictors, are obtained from the FRED-QD dataset, published by Clark & McCracken [2023]. The target variables—Gross Domestic Product, Exports, the GDP Deflator, and Housing Starts—correspond respectively to the dataset codes 'GDPC1', 'EXPGSC1', 'GDPCTPI', and 'HOUST'. Each series is transformed following the method by Hamilton & Xi [2024] to address non-stationarity, a common challenge in macroeconomic data analysis as noted by Beveridge & Nelson [1981] and Nelson & Plosser [1982]. All variables are standardized to account for the sensitivity of the 3PRF method to differences in scale, similar to the scaling sensitivity in PCR and LASSO. Standardization ensures that no variable disproportionately influences the results due to its scale.

Prior to forecasting, both target and predictor data undergo partial transformations with respect to a constant and four lags of the target variable, following the approach by Kelly & Pruitt [2015], Bai & Ng [2008], Stock & Watson [2012]. This generates the following variables:

$$\ddot{y}_{t+h} = y_{t+h} - \hat{\mathbb{E}}\left(y_{t+h} \mid y_t, y_{t-1}, y_{t-2}, y_{t-3}\right),$$
$$\ddot{\boldsymbol{x}}_t = \boldsymbol{x}_t - \hat{\mathbb{E}}\left(\boldsymbol{x}_t \mid y_t, y_{t-1}, y_{t-2}, y_{t-3}\right),$$

where $\hat{\mathbb{E}}(\cdot \mid \Omega)$ denotes linear projection on Ω and a constant. As a consequence of these transformations, some observations are lost, leading to a dataset spanning from 1963:Q3 to 2019:Q3. The COVID-19 period is deliberately excluded due to its outlier nature, rendering it unforecastable. Our forecasting horizon spans one quarter ($h = 1$) to one year ($h = 4$).

The determination of the number of factors in PCR is based on the eigenvalue ratio method introduced by Ahn & Horenstein [2013], yielding one factor for all training samples. This aligns with the findings of Kelly & Pruitt [2015], who also observed one factor using the information criteria by Bai & Ng [2002] in their dataset. Since the number of target-relevant factors is equal to or fewer than those that drive the set of predictors, we use one factor for our 3PRF forecasts. Also, as argued in Remark 1, choosing a single factor may be a prudent choice under many circumstances.

We employ a recursive window approach to construct out-of-sample forecasts of the aforementioned series similar to Kelly & Pruitt [2015] and present OOS R-squared values for different methods. The initial training sample (which expands recursively) spans the following time periods, 1963:Q3 - 1997:Q3, encompassing 60 percent of the total observations.

For LASSO regression, whether implemented as a standalone method or as an auxiliary regression in 3PRF LASSO and PCR LASSO, we use a 10-fold cross-validation technique to estimate parameters within each training sample, following the methodology outlined in Fan *et al.* [2020]. The results are presented in Table 7. We forecast four variables across four horizons, yielding a total of 16 comparisons (i.e., across four variables and four horizons).

In each of these comparisons, 3PRF LASSO and PCR LASSO consistently outperform the standard 3PRF and PCR methods. Additionally, the standalone LASSO method emerges as the best performer in only two out of the sixteen instances, and even then, by a narrow margin. This underscores the utility of combining sparse and dense regression techniques for forecasting macroeconomic variables.

Out of the sixteen combinations, 3PRF LASSO emerges as the best-performing method in nine cases and the second-best in five, with only two instances where it does not feature among the top two methods. Furthermore, it closely trails the second-best method in these two instances, underscoring its reliability.

Overall, 3PRF LASSO proves to be a robust forecasting technique for high-dimensional datasets, especially where both factor structure and sparsity are likely to be present.

7 Conclusion

Our paper extends the framework developed in Kelly & Pruitt [2015] by addressing two critical aspects of factor-based forecasting models: accommodating weak factors and leveraging potentially informative idiosyncratic components by incorporating them as additional regressors (in addition to factors) through a penalized regression (LASSO). From a theoretical standpoint, our contribution lies in demonstrating how the weakness of factors impacts the convergence rate of the estimator, leading to slower rates as relevant factors become less pervasive. Importantly, we establish that when relevant factors are stronger than irrelevant ones, 3PRF achieves a faster convergence rate to the infeasible best forecast compared to PCR. This advantage arises because 3PRF can isolate and utilize relevant factors more effectively in the presence of irrelevant ones.

Allowing idiosyncratic components to have predictive power for the target and, by extension, the proxies (which have a similar DGP as the target) necessitates expanding the underlying model assumptions. We show that, under mild assumptions, this extension does not impose any penalty on convergence rates. If factors are strong, as in Kelly & Pruitt [2015], the convergence rate of 3PRF within a framework where idiosyncratic dependence is allowed would be identical to the case where idiosyncratic dependence is absent, i.e., the setting in the paper by Kelly & Pruitt [2015]. This result underscores the robustness of 3PRF in adapting to more general DGPs without sacrificing theoretical efficiency.

On the methodological front, we enhance 3PRF by incorporating a Stage 2 LASSO regression to capture predictive content from idiosyncratic components. This integration allows the model to effectively utilize residual variation that is not explained by common factors. Our empirical analysis, using macroeconomic data, highlights the practical significance of this extension, demonstrating substantial improvements in the predictability of key macroeconomic variables when idiosyncratic information is included.

Table 1: OOS R-squared across competing methods

$K_f = 1$, $K_g = 4$				$N = 100$, $T = 100$			$\psi_f = 1$, $\psi_g = 1$		
ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	LASSO	PCR+L	3PRF+L
0	0	0	0	0.61963	0.35711	0.35713	0.50183	0.483	**0.50297**
0.3	0.9	0.3	0	0.61963	0.34567	0.36254	0.48815	0.41808	**0.4941**
0.3	0.9	0.3	1	0.65218	0.2887	0.38393	0.52197	0.50523	**0.53013**
0.3	0.9	0.9	0	0.62498	0.32803	0.35824	**0.49378**	0.41771	0.49123
0.3	0.9	0.9	1	0.64927	0.27884	0.37397	0.5074	**0.51562**	0.5142
0.9	0.3	0.3	0	0.61978	0.36505	0.34815	0.49502	0.38609	**0.50948**
0.9	0.3	0.3	1	0.6419	0.31822	0.39491	0.52705	**0.53859**	0.52448
0.9	0.3	0.9	0	0.617	0.39879	0.38352	0.52131	0.52027	**0.53489**
0.9	0.3	0.9	1	0.64385	0.33854	0.41707	0.53639	**0.54622**	0.54224

Notes: K_f, K_g , ρ_f ρ_g, a , d ψ_f, ψ_g are defined in Section 5. Oracle denotes the infeasible regression, as described in Section 5. PCR denotes the regression of $\boldsymbol{y}$ on first 'K' principal components, where $K = K_f + K_g$. 3PRF denotes the auto-proxy 3PRF with K_f auto-proxies. LASSO denotes the the LASSO regression of $\boldsymbol{y}$ on $\boldsymbol{X}$. 3PRF+L is 3PRF LASSO procedure where Stage 1 (3PRF) uses K_f proxies. PCR+L is analogously a 2 Stage regression where Stage 1 is PCR involving leading $K = K_f + K_g$ PCs as predictors, and Stage 2 is a LASSO regression involving the idiosyncratic components estimated using principal component method. The highest R^2 value across competing methods in in bold.

Table 2: OOS R-squared across competing methods

$K_f = 1$, $K_g = 4$				$N = 100$, $T = 100$			$\psi_f = 1$, $\psi_g = 0.7$		
ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	LASSO	PCR+L	3PRF+L
0	0	0	0	0.61586	0.35254	0.3763	0.44874	0.43334	**0.45638**
0.3	0.9	0.3	0	0.61272	0.34334	0.37507	0.4475	0.32952	**0.46164**
0.3	0.9	0.3	1	0.64204	0.34798	0.40109	0.51072	0.50934	**0.5275**
0.3	0.9	0.9	0	0.61562	0.34182	0.37569	0.43835	0.42262	**0.47136**
0.3	0.9	0.9	1	0.65034	0.35345	0.41414	0.53489	0.53643	**0.55218**
0.9	0.3	0.3	0	0.61676	0.40086	0.43177	0.5093	0.4897	**0.52916**
0.9	0.3	0.3	1	0.6445	0.38991	0.45023	0.54269	0.55545	**0.56651**
0.9	0.3	0.9	0	0.6191	0.40294	0.42346	0.49675	0.41185	**0.49669**
0.9	0.3	0.9	1	0.64742	0.38952	0.45598	0.56103	0.56505	**0.56816**

Notes: See Table 1

Table 3: OOS R-squared across competing methods

$K_f = 1,\ K_g = 4$				$N = 100,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 1$		
ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	LASSO	PCR+L	3PRF+L
0	0	0	0	0.61984	0.25428	0.042604	0.3525	**0.3717**	0.36652
0.3	0.9	0.3	0	0.6142	0.21833	-0.0001	**0.3204**	0.30434	0.31485
0.3	0.9	0.3	1	0.6526	-0.039526	0.042867	**0.25753**	0.18481	0.14079
0.3	0.9	0.9	0	0.60923	0.19586	-0.020731	0.31079	**0.3338**	0.30475
0.3	0.9	0.9	1	0.65025	-0.016048	0.050238	**0.2339**	0.053706	0.14786
0.9	0.3	0.3	0	0.62011	0.24858	0.014442	0.35448	0.34412	**0.36225**
0.9	0.3	0.3	1	0.64539	0.012379	0.067817	**0.24279**	0.17838	0.16922
0.9	0.3	0.9	0	0.60969	0.17364	0.024531	0.34554	**0.35578**	0.34891
0.9	0.3	0.9	1	0.65026	0.010274	0.078063	**0.25798**	0.13377	0.18915

Notes: See Table 1

Table 4: OOS R-squared across competing methods

$K_f = 1,\ K_g = 4$				$N = 100,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 0.7$		
ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	LASSO	PCR+L	3PRF+L
0	0	0	0	0.61494	0.27246	0.31847	0.31304	0.34248	**0.35345**
0.3	0.9	0.3	0	0.61829	0.2715	0.32075	0.31823	0.34769	**0.36768**
0.3	0.9	0.3	1	0.64903	0.039788	0.23694	**0.2993**	0.25044	0.24138
0.3	0.9	0.9	0	0.6153	0.28658	0.3367	0.33782	0.36746	**0.37318**
0.3	0.9	0.9	1	0.65476	0.064438	0.25242	**0.28539**	0.20769	0.2527
0.9	0.3	0.3	0	0.61824	0.321	0.3625	0.36587	0.35551	**0.3989**
0.9	0.3	0.3	1	0.65205	0.066997	0.23593	**0.2994**	0.23966	0.23904
0.9	0.3	0.9	0	0.61817	0.27051	0.35604	0.34206	0.32838	**0.38636**
0.9	0.3	0.9	1	0.64375	0.051624	0.21217	**0.25925**	0.18316	0.21405

Notes: See Table 1

Table 5: OOS R-squared across competing methods

$K_f = 1,\ K_g = 5$				$N = 100,\ T = 100$			$\psi_f = 1,\ \psi_g = 1$		
ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	LASSO	PCR+L	3PRF+L
0	0	0	0	0.61952	0.35065	0.36033	0.49697	0.45746	**0.50415**
0.3	0.9	0.3	0	0.61747	0.32665	0.35071	0.47532	0.3951	**0.48162**
0.3	0.9	0.3	1	0.64156	0.24943	0.35685	**0.49404**	0.46676	0.48814
0.3	0.9	0.9	0	0.61054	0.31013	0.33968	0.46534	0.30602	**0.46885**
0.3	0.9	0.9	1	0.64935	0.25122	0.3529	0.50154	0.48852	**0.50764**
0.9	0.3	0.3	0	0.61706	0.39123	0.38815	0.52875	0.38402	**0.5384**
0.9	0.3	0.3	1	0.65268	0.31479	0.40096	**0.53045**	0.54751	0.52677
0.9	0.3	0.9	0	0.62003	0.39536	0.38733	0.52423	0.40492	**0.53273**
0.9	0.3	0.9	1	0.64097	0.31841	0.38816	0.52378	**0.52466**	0.51459

Notes: See Table 1

Table 6: OOS R-squared across competing methods

$K_f = 1,\ K_g = 4$				$N = 200,\ T = 100$			$\psi_f = 1,\ \psi_g = 1$		
ρ_f	ρ_g	a	d	Oracle	PCR	3PRF	LASSO	PCR+L	3PRF+L
0	0	0	0	0.62649	0.37007	0.36536	0.49511	0.251	**0.5001**
0.3	0.9	0.3	0	0.61362	0.31505	0.33425	**0.44706**	0.23926	0.43754
0.3	0.9	0.3	1	0.64683	0.29327	0.3462	**0.51792**	0.3642	0.5174
0.3	0.9	0.9	0	0.61788	0.32355	0.33426	0.44939	0.25414	**0.4643**
0.3	0.9	0.9	1	0.64526	0.28446	0.34314	0.48878	0.35585	**0.50658**
0.9	0.3	0.3	0	0.61543	0.40231	0.38629	0.51894	0.33766	**0.52398**
0.9	0.3	0.3	1	0.64284	0.3531	0.38736	0.52788	0.39192	**0.54455**
0.9	0.3	0.9	0	0.61923	0.39344	0.37964	0.51337	0.32802	**0.52474**
0.9	0.3	0.9	1	0.64907	0.34188	0.38216	0.53225	0.49887	**0.55153**

Notes: See Table 1

Table 7: Forecasting performance of different models

Target Variable	PCR	3PRF	LASSO	PCR-LASSO	3PRF-LASSO
One Quarter Ahead Forecast, OOS $R^2(\%)$					
GDP	3.27	4.56	**16.87****	12.79	**17.05***
Exports	−1.36	−2.24	−1.90	**4.96****	**6.66***
Housing Starts	−22.83	−27.61	**18.60***	**14.90****	14.73
GDP Deflator	0.04	0.84	5.58	**8.90****	**9.95***
Two Quarters Ahead Forecast, OOS $R^2(\%)$					
GDP	7.14	9.04	**41.99****	**42.88***	41.13
Exports	1.13	1.15	19.08	**31.97***	**24.23****
Housing Starts	−25.08	−34.74	**32.29****	15.02	**37.65***
GDP Deflator	−1.04	2.19	32.52	**40.38***	**37.75****
Three Quarters Ahead Forecast, OOS $R^2(\%)$					
GDP	9.85	13.28	**58.84****	57.03	**59.63***
Exports	3.55	3.86	46.23	**54.31***	**53.23****
Housing Starts	−20.32	−32.13	**16.40****	4.22	**25.78***
GDP Deflator	−5.02	−0.22	15.48	**31.33***	**27.24****
One Year Ahead Forecast, OOS $R^2(\%)$					
GDP	12.20	17.59	**69.99***	67.18	**68.45****
Exports	6.62	7.36	52.55	**57.14****	**58.97***
Housing Starts	−10.08	−14.13	**16.39****	11.68	**19.50***
GDP Deflator	−9.25	−0.45	20.45	**24.42****	**27.76***

*Notes: The highest two entries have been put in bold. In each row, entry marked with * is the highest entry and entry marked with ** is the second highest entry.*

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

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 Datzig 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, Lou, Zhipeng, & Yu, Mengxin. 2023a. Are latent factor regression and sparse regression adequate? *Journal of the American Statistical Association*, 1–13.

Fan, Jianqing, Masini, Ricardo P, & Medeiros, Marcelo C. 2023b. Bridging factor and sparse models. *The Annals of Statistics*, **51**(4), 1692–1717.

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. 2023. Prediction when factors are weak. *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

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, & Xi, Jin. 2024. *Principal Component Analysis for Nonstationary Series*. Working Paper 32068. National Bureau of Economic Research.

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.

West, Kenneth, & McCracken, Michael. 1998. Regression-Based Tests of Predictive Ability. *International Economic Review*, **39**(4), 817–40.

Appendix A

This Appendix A is organized as follows. Section A1 provides supplementary lemmas that are used throughout the proofs. Section A2 includes proofs of Theorems 1(a), 2, 3(a), and 4(a), covering stage 1 results of 3PRF-Lasso for the cases when $\boldsymbol{\gamma} = 0$, and $\boldsymbol{\zeta} = 0$. The lemmas used in these proofs are also contained in Section A2. Section A3 addresses Theorems 1(b), 3(b), and 4(b), which focus on stage 1 results of 3PRF-Lasso for $\boldsymbol{\gamma} \neq 0$ and $\boldsymbol{\zeta} \neq 0$. Section A4 presents the stage 2 results of 3PRF-Lasso, i.e., the proof of Theorem 5 and the supporting lemmas for this proof. Additional Simulation results are in the Online Appendix B. We refer to this as Appendix A and label the sections within it as A1–A4 to differentiate them from the Online Appendix B, which we refer to as Online Appendix B, with sections labeled B1–B2.

A1 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. $\mathbb{E}\left|(NT)^{-1/2}\sum_{i,s} F_s(m)\left[\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}\right]\right|^2 \leq M$.
2. $\mathbb{E}\left|(NT)^{-1/2}\sum_{i,s} \omega_s(m)\left[\varepsilon_{is}\varepsilon_{it} - \sigma_{ii,st}\right]\right|^2 \leq M$.
3. $N^{-1/2}T^{-1/2}\sum_{i,t}\varepsilon_{it} = O_p(1)$, $\quad N^{-1/2}\sum_i \varepsilon_{it} = O_p(1)$ *and* $T^{-1/2}\sum_t \varepsilon_{it} = O_p(1)$.
4. $T^{-1/2}\sum_t \eta_{t+h} = O_p(1)$.
5. $T^{-1/2}\sum_t F_t(m)\eta_{t+h} = O_p(1)$ *and* $T^{-1/2}\sum_t \omega_t(m)\eta_{t+h} = O_p(1)$.
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)* $N^{-1}T^{-1/2}\sum_{i,s}\varepsilon_{is}\varepsilon_{it} = O_p\left(\delta_{NT}^{-1}\right)$ *and* *(b)* $N^{-1/2}T^{-1}\sum_{i,t}\varepsilon_{it}\varepsilon_{jt} = O_p\left(\delta_{NT}^{-1}\right)$.
11. $N^{-1}T^{-3/2}\sum_{i,s,t}\varepsilon_{is}\varepsilon_{it}\eta_{t+h} = O_p\left(\delta_{NT}^{-1}\right)$.
12. $N^{-1}T^{-1/2}\sum_{i,s}F_s(m)\varepsilon_{is}\varepsilon_{it} = O_p\left(\delta_{NT}^{-1}\right)$.
13. $N^{-1}T^{-1/2}\sum_{i,s}\omega_s(m)\varepsilon_{is}\varepsilon_{it} = O_p\left(\delta_{NT}^{-1}\right)$.
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. $N^{-\psi_v/2}T^{-1}\sum_{i,t}\phi_{iv}(m)\varepsilon_{it}\varepsilon_{jt} = O_p\left(\Gamma_{N_vT}^{-1}\right)$.

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

Proof: Item 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

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

by Assumption 2.1 and 2.5. Therefore,

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

and similarly,

$$\sum_t \omega_t(m)\eta_{t+h} = O_p(1)$$

using Assumption 2.4, 2.5.

Item 7: For $v \in \{f, g\}$, Using the Cauchy Schwartz inequality,

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

by Assumptions 3.7 and 2.1.

Item 8: $v \in \{f, g\}$, Using the Cauchy Schwartz inequality,

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

by Assumptions 3.7 and 2.4.

Item 9 :Since $\mathbb{E}\left[\eta_{t+1}\eta_{s+1}\right] = 0$ for $t \neq s$ and η_{t+h} is independent of $\phi_i(m)$ and $\varepsilon_{i,t}$, $\forall i, t$ for any $h > 0$ by

Assumption 2.5, we have,

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

Last line follows from from Assumptions 3.7 and 2.5.

Item 10(b):

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

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

Item 16 : We have,

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

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

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

Lemma 2. *Under Assumption(s) 2-3, we have the following*

1. $T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\omega}=\boldsymbol{O}_p(1)$.
2. $T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\eta}=\boldsymbol{O}_p(1)$.
3. $T^{-1/2}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta}=\boldsymbol{O}_p(1)$.
4. $N^{-\psi_f}\boldsymbol{\varepsilon}_t'\boldsymbol{J}_N\boldsymbol{\Phi}=\boldsymbol{O}_p\left(N^{\frac{-\psi_f}{2}}\vee N^{-\psi_f+\frac{\psi_g}{2}}\right)$.
5. $N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}=\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\vee\dfrac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$.
6. $N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}=\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\vee\dfrac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$.

7. $N^{-\psi_f}T^{-1/2}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta} = \boldsymbol{O}_p\left(N^{\frac{-\psi_f}{2}} \vee N^{-\psi_f+\frac{\psi_g}{2}}\right)$.

8. $N^{-1}T^{-3/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

9. $N^{-1}T^{-3/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

10. $N^{-1}T^{-3/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

11. $N^{-1}T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}_t = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

12. $N^{-1}T^{-1/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon_t = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

13. $N^{-1}T^{-3/2}\boldsymbol{\eta}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

14. $N^{-1}T^{-3/2}\boldsymbol{\eta}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

Proof: Item 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,

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

m^{th} element of $N^{-\psi_f}\boldsymbol{\varepsilon}_t'\boldsymbol{J}_N\boldsymbol{\Phi}_f$ 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 $\boldsymbol{O}_p(1)$ by Assumption 3.7.

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

Therefore,

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

Similarly,

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

which implies

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

Hence, the whole matrix has the following order.

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

Item 5 can be expanded as

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

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

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

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

5.I is $O_p\left(N^{-\psi_f/2}\right)$ by Lemma 1.7.
5.II is $O_p\left(T^{-1/2}\right)$ since $N^{-\psi_f}\sum_i\phi_{if}\left(m_1\right) = O_p(1)$ by Assumption 2.2 and $N^{-1}\sum_j\left(T^{-1/2}\sum_t F_t\left(m_2\right)\varepsilon_{jt}\right) = O_p(1)$ by Assumption 3.6.
5.III is $O_p\left(N^{-\psi_f/2}\right)$ since $T^{-1}\sum_s F_s\left(m_2\right) = O_p(1)$ by Assumption 2.1 and $T^{-1}\sum_t\left(N^{-\psi_f/2}\sum_j\phi_{jf}\left(m_1\right)\varepsilon_{jt}\right) = O_p(1)$ by Assumption 3.7.
5.IV is $O_p\left(T^{-1/2}N^{-1/2}\right)$ by Assumptions 2.1, 2.2 and Lemma 1.3.
Summing these terms, $N^{-\psi_f}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}$ is $\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\right)$.

Item 6 can be expanded as

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

As in the case of Item 5, it suffices to show that $N^{-\psi_f}T^{-1}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}$ is $\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\right)$.

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

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

6.I is $O_p\left(N^{-\psi_f/2}\right)$ by Lemma 1.8.
6.II is $O_p\left(T^{-1/2}\right)$ since $N^{-\psi_f}\sum_i \phi_{if}(m_1) = O_p(1)$ by Assumption 2.2 and $N^{-1}\sum_j\left(T^{-1/2}\sum_t \omega_t(m_2)\,\varepsilon_{jt}\right) = O_p(1)$ by Assumption 3.5.
6.III is $O_p\left(N^{-\psi_f/2}T^{-1/2}\right)$ 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 $O_p\left(T^{-1}N^{-1/2}\right)$ by Assumption 2.2, 2.4 and Lemma 1.3.
Summing these terms ,we have that $N^{-\psi_f}T^{-1}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}$ is $\boldsymbol{O}_p\left(\Gamma^{-1}_{N_fT}\right)$.

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

m^{th} element of $N^{-\psi_f}T^{-1/2}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta}$ is given by,

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

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

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

This follows from Assumption 2.2 Lemma 1.3 and 1.4.

Summing these terms, $N^{-\psi_f}T^{-1/2}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta}$ is $\boldsymbol{O}_p(N^{-\psi_f/2})$.

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

1. $N^{-2\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi} = \boldsymbol{O}_p\left(N^{-\psi_f/2}\Gamma_{N_fT}^{-1} \vee \dfrac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2}\Gamma_{N_gT}}\right)$.
2. $N^{-2\psi_f}T^{-2}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(T^{-1/2}N^{-\frac{3\psi_f}{2}+1}\delta_{NT}^{-1} \vee \dfrac{N^{2(\psi_g-\psi_f)}}{T^{1/2}N^{\frac{3\psi_g}{2}-1}\delta_{NT}}\right)$.
3. $N^{-2\psi_f}T^{-2}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p\left(T^{-1/2}N^{-\frac{3\psi_f}{2}+1}\delta_{NT}^{-1} \vee \dfrac{N^{2(\psi_g-\psi_f)}}{T^{1/2}N^{\frac{3\psi_g}{2}-1}\delta_{NT}}\right)$.
4. $N^{-2\psi_f}T^{-3}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F} = \boldsymbol{O}_p\left(\left(\dfrac{N^{1-\psi_f}}{\sqrt{T}\delta_{NT}}\right)^2\right)$.
5. $N^{-2\psi_f}T^{-3}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p\left(\left(\dfrac{N^{1-\psi_f}}{\sqrt{T}\delta_{NT}}\right)^2\right)$.
6. $N^{-2\psi_f}T^{-3}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega} = \boldsymbol{O}_p\left(\left(\dfrac{N^{1-\psi_f}}{\sqrt{T}\delta_{NT}}\right)^2\right)$.

Proof: We prove 1-3 below .4-6 have been proved in Kelly & Pruitt [2015], Online Appendix, Lemma Web. They show that $N^{-2}T^{-3}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F}$, $N^{-2}T^{-3}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}$ and $N^{-2}T^{-3}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}$ are all $\boldsymbol{O}_p\left(T^{-1}\delta_{NT}^{-2}\right)$. 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 $K\times K$ matrix given as,

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

First, We show that $N^{-2\psi_f}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}_f$ is $\boldsymbol{O}_p\left(N^{-\psi_f/2}\Gamma_{N_fT}^{-1}\right)$.

$N^{-2\psi_f}T^{-1}\mathbf{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\mathbf{\Phi}_f$ has generic (m_1, m_2) element given as

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

1.I can be written as

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

by Assumption 3.7.

1.II can be written as

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

by lemma 1.16 and Assumption 2.2.

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

1.IV can be written as

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

by Assumption 2.2 and Lemma 1.10.

1.V is $O_p\left(N^{-\psi_f}\right)$. Identical to 1.I.

1.VI can be written as

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

by Assumption 2.2, 3.7 and Lemma 1.10.

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

1.VIII can be written as

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

by Assumptions 2.2 and Lemma 1.3.

Summing all these terms gives us $N^{-2\psi_f}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}_f$ is $\boldsymbol{O}_p\left(N^{-\psi_f/2}\Gamma_{N_fT}^{-1}\right)$.

By a symmetrical argument $N^{-2\psi_g}T^{-1}\boldsymbol{\Phi}_g'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}_g$ is $\boldsymbol{O}_p\left(N^{-\psi_g/2}\Gamma_{N_gT}^{-1}\right)$.

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

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, $N^{-2\psi_f}T^{-1}\boldsymbol{\Phi}_g'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}_f$ is either $\boldsymbol{O}_p\left(N^{-\psi_f/2}\Gamma_{N_fT}^{-1}\right)$ (when $\psi_f > \psi_g$) or $\boldsymbol{O}_p\left(\frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2}\Gamma_{N_gT}}\right)$ (when $\psi_f < \psi_g$). When $\psi_f = \psi_g$, the two rates are equal.

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

Item 2 is given by

$$
N^{-2\psi_f}T^{-2}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}=\begin{bmatrix}N^{-2\psi_f}T^{-2}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\\ N^{-2\psi_f}T^{-2}\boldsymbol{\Phi_g}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\end{bmatrix}.
$$

We first show (below) that $K_f\times K$ matrix $N^{-2\psi_f}T^{-2}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}$ is $\boldsymbol{O}_p\left(N^{-\psi_f/2}\Gamma_{N_fT}^{-1}\right)$.

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

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

2.I Using Cauchy Schwartz inequality, this term is bounded by

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

by Lemma 1.12 and Assumption 3.7.

2.II: Using Cauchy Schwartz inequality, this term is bounded by,

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

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

2.III can be written as

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

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

2.IV can be written as

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

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

2.V can be written as

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

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

2.VI can be written as

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

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

2.VII can be written as

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

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

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

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

2.IX: Using Cauchy Schwartz, this term is bounded by

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

which is $O_p(N^{-\psi_f+\frac{1}{2}}T^{-1/2}\delta_{NT}^{-1})$ by Lemma 1.3, Lemma 1.14 and Assumption 2.2.

2.X: Using Cauchy Schwartz, this term is bounded by

$$N^{-\psi_f+\frac{1}{2}}T^{-1/2}\left(N^{-\psi_f}\sum_{k}\phi_{k_f}\left(m_1\right)\right)\left(T^{-1}\sum_{u}\left[N^{-\frac{1}{2}}\sum_{j}\varepsilon_{ju}\right]^2\right)^{1/2}$$
$$\times\left(T^{-1}\sum_{u}\left[N^{-1}T^{-\frac{1}{2}}\sum_{i,t}\varepsilon_{iu}\varepsilon_{it}\right]\right)^{1/2},$$

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

2.XI can be written as

$$N^{-\psi_f+\frac{1}{2}}T^{-1/2}\left(N^{-\psi_f}\sum_{l}\phi_{l_f}\left(m_1\right)\right)\left(N^{-1}\sum_{j}\left[N^{-\frac{1}{2}}T^{-1}\sum_{k,t}\varepsilon_{jt}\varepsilon_{kt}\right]\right)$$
$$\times\left(N^{-1}\sum_{i}T^{-1/2}\sum_{s}\varepsilon_{is}F_s\left(m_2\right)\right).$$

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

2.XII can be written as

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

This is $O_p(N^{-\psi_f}T^{-1/2}\delta_{NT}^{-1})$ by Assumption 2.2, 2.3 and Lemma 1.10.

2.XIII Using Cauchy Schwartz, this term is bounded by

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

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

2.XIV can be written as

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

This is $O_p(N^{-\psi_f}T^{-1/2}\delta_{NT}^{-1})$ by Assumption 2.2, 2.3, Lemma 1.3 and 1.10.

2.XV can be written as

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

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

2.XVI can be wriiten as

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

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:

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

By a symmetrical argument, we have:

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

Hence, we can conclude that:

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

Therefore, we have:

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

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

A2 Proofs of Theorem 1(a), 2, 3(a) and 4(a)

We introduce 2 Lemmas which are utilized in proofs of stage 1 results of 3PRF-Lasso for the cases when $\boldsymbol{\gamma} = 0$, and $\boldsymbol{\zeta} = 0$. The proofs are presented subsequent to these Lemmas.

Lemma 4. *Recall,* $\Xi_{NT}^{-1} \equiv T^{-1/2} \vee N^{-\psi_f/2} \vee \left(\frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$ *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* $\boldsymbol{\gamma} = 0$*, and* $\boldsymbol{\zeta} = 0$*, we have,*

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

Furthermore, the probability limits of $\hat{\boldsymbol{\Phi}}'$ *and* $\hat{\boldsymbol{F}}_t$ *are*

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

and

$$\hat{\boldsymbol{F}}_t \xrightarrow[T,N\to\infty]{p} \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{\Delta}_\omega\right)\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1}\left(N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\phi_0 + \boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{f}_t\right).$$

Proof:

First we note that

$$\begin{aligned} N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi} &= \begin{bmatrix} N^{-\psi_f}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\Phi}_f & N^{-\psi_f}\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\Phi}_g \\ N^{-\psi_f}\boldsymbol{\Phi_g}'\boldsymbol{J}_N\boldsymbol{\Phi}_f & N^{\psi_g-\psi_f}\left(N^{-\psi_g}\boldsymbol{\Phi_g}'\boldsymbol{J}_N\boldsymbol{\Phi}_g\right) \end{bmatrix} \\ &= \boldsymbol{O}_p\left(1 \vee N^{\psi_g-\psi_f}\right). \end{aligned} \tag{A2.1}$$

The final equality follows from Assumption 2.2 and 4.1.

Item 1 :

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

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

Item 2:

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

The final equality follows from Assumption 2.1, 4 , Lemma 2 and equation A2.1.

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

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

The final equality follows from Assumption 2.1, 4.

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

Item 3:

$$\begin{aligned}
\hat{\boldsymbol{F}}_{C,t} =& \quad N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{x}_t \\
=& \quad \boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\phi}_{\mathbf{0}}\right) + \boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t \\
&+ \boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right) + \boldsymbol{\Lambda}\left(N^{-\psi_f}T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\phi_0\right) \\
&+ \boldsymbol{\Lambda}\left(N^{-\psi_f}T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t + \frac{N^{1-\psi_f}}{\sqrt{T}}\boldsymbol{\Lambda}\left(N^{-1}T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right) \\
&+ \left(T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\phi_{\mathbf{0}}\right) + \left(T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t \\
&+ \left(T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right) + \left(N^{-1}T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\phi_{\mathbf{0}}\right) \\
&+ \left(N^{-\psi_f}T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t + \frac{N^{1-\psi_f}}{\sqrt{T}}\left(N^{-1}T^{-1/2}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right) \\
=& N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\phi_0 + \boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right).
\end{aligned}$$

Assumption 2.1, 4 , Lemma 2 and equation A2.1 give the final equality.

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

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

The final equality again follows from Assumption 2.1, 4.

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

Combining results for Items 1-3, Using Assumption 6, we have,

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

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

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

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

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

Therefore,

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

satisfies

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

Proof:

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

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_3 =&(N^{2\psi_f}T^3)^{-1}(\boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\omega' J_T F \Phi' J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \Phi F' J_T \varepsilon J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\Lambda F' J_T \varepsilon J_N \varepsilon' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\omega' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\omega' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\omega' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\omega' J_T \varepsilon J_N \Phi F' J_T F \Phi' J_N \varepsilon' J_T \omega} \\
&+ \boldsymbol{\omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T F \Lambda'} + \boldsymbol{\omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \Phi F' J_T \omega} \\
&+ \boldsymbol{\omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T F \Lambda'} + \boldsymbol{\omega' J_T \varepsilon J_N \varepsilon' J_T \varepsilon J_N \varepsilon' J_T \omega}
\end{aligned}
$$

$$
\begin{aligned}
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\Lambda' + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega})\\
=&(N^{2\psi_f}T^3)^{-1}\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right).
\end{aligned}
$$

The final equality follows from Assumption 2.1, Lemmas 2 and 3 and equation A2.1. Further, note that

$$
\begin{aligned}
&(N^{-2\psi_f}T^{-3})\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Lambda}'\\
=&\frac{1}{T}\left((N^{-\psi_f}T^{-1})\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\right)\boldsymbol{J}_T\left((N^{-\psi_f}T^{-1})\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\right)'\\
=&\frac{1}{T}\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{f}' + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\right)\boldsymbol{J}_T\left(\boldsymbol{f}\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\right)\\
=&\boldsymbol{\Lambda_f}\boldsymbol{\Delta_f}\mathcal{P}_f\frac{\boldsymbol{f}'\boldsymbol{J}_T\boldsymbol{f}}{T}\mathcal{P}_f\boldsymbol{\Delta_f}\boldsymbol{\Lambda_f}' + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\\
=&\boldsymbol{\Lambda_f}\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\boldsymbol{\Lambda_f}' + \boldsymbol{O}_p\left(1/\sqrt{T}\right) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right).
\end{aligned}
$$

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

$$
\begin{aligned}
N^{\psi_f}T^2\hat{\boldsymbol{\beta}}_4 =&\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\Phi'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta}\\
&+\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta} + \omega'\boldsymbol{J}_T\boldsymbol{F}\Phi'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\eta}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\Phi'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \omega'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta}\\
&+\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\eta}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \omega'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\eta\\
=&N^{-\psi_f}T^{-2}\left(\boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta}\right) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right).
\end{aligned}
$$

The final equality follows from Assumption 2.1, Lemma 2 and equation A2.1.

Further, note that, since $\boldsymbol{\Lambda}=\left[\begin{array}{cc}\boldsymbol{\Lambda}_f & \mathbf{0}\end{array}\right]$ (Assumption 5) and $\boldsymbol{\beta}=(\boldsymbol{\beta}_f', 0')'$ (Assumption 1), we have

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

The final equality follows from Assumption 2.1, 4 and Lemma 2 and equation A2.1.

Hence, $\hat{\boldsymbol{\beta}}_4=\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{y}=\boldsymbol{\Lambda_f}\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\boldsymbol{\beta}_f+\boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)$. Combining these results give the probability limit of $\hat{\boldsymbol{\beta}}$ 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 $\boldsymbol{\gamma}=0$ and $\boldsymbol{\zeta}=0$. Additionally, if $\frac{N^{1-\psi_f}}{\sqrt{T}}=O(1)$, then we have,

$$\hat{y}_{t+h,f}-\mathbb{E}_t y_{t+h}=\boldsymbol{O}_p(\Xi_{NT}^{-1})$$

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

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

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

$$\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{A}+\boldsymbol{B}\right)^{-1}, \tag{A2.2}$$

which implies that:

$$
\begin{aligned}
\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right)^{-1} &= \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1} \\
&\quad - \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1}\boldsymbol{O}_p(\Xi_{NT}^{-1}) \\
&\quad\; \times \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right)^{-1} \\
&= \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1} - \boldsymbol{O}_p(1)\boldsymbol{O}_p(\Xi_{NT}^{-1})\boldsymbol{O}_p(1) \\
&= \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1} - \boldsymbol{O}_p(\Xi_{NT}^{-1}).
\end{aligned}
$$

The final equality comes from Assumptions 4 and 5, which require $\boldsymbol{\Lambda}_f$, $\mathcal{P}_f$, and $\boldsymbol{\Delta}_f$ to be non-singular. 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 $\hat{\boldsymbol{\alpha}}$. Let Assumptions 1-6 hold and $\boldsymbol{\gamma} = 0$ and $\boldsymbol{\zeta} = 0$ and $\mathcal{P}_f = \mathbb{I}$, Then for any i,

$$N^{\psi_f}\hat{\alpha}_i \xrightarrow[T,N\to\infty]{p} \left(\boldsymbol{\phi}_{if} - N^{\psi_f - 1}\bar{\boldsymbol{\phi}}_{\boldsymbol{f}}\right)'\boldsymbol{\beta}_f.$$

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

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

From Lemma 4 and 5, we have

$$\hat{\alpha}_i = N^{-\psi_f}\boldsymbol{S}_i\boldsymbol{J}_N\boldsymbol{\phi}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1}\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\beta}_f + o_p(1).$$

The expression $\boldsymbol{S}_i\boldsymbol{J}_N\boldsymbol{\Phi}_f$ has the probability limit $\phi_{if} - N^{\psi_f-1}\overline{\phi_f}$ as $N, T \to \infty$. Therefore, we have that

$$N^{\psi_f}\hat{\alpha}_i \xrightarrow[T,N\to\infty]{p} \left(\phi_{if} - N^{\psi_f-1}\overline{\boldsymbol{\phi}_f}\right)'\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1}\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\beta}_f.$$

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

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

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

$$\hat{\boldsymbol{\beta}} - \boldsymbol{G}_\beta\boldsymbol{\beta}_f = \boldsymbol{O}_p(\Xi_{NT}^{-1}).$$

Proof:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} = & \left(T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{Z}\right)^{-1} N^{-\psi_f}T^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z} \\
& \times \left(N^{-2\psi_f}T^{-3}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}\right)^{-1} N^{-\psi_f}T^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{y} \\
= & \hat{\boldsymbol{\beta}}_1^{-1}\hat{\boldsymbol{\beta}}_2\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right)^{-1}\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\beta}_f + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right) \\
= & \hat{\boldsymbol{\beta}}_1^{-1}\hat{\boldsymbol{\beta}}_2\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1}\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\right)\boldsymbol{\beta}_f + \boldsymbol{O}_p(\Xi_{NT}^{-1}) \\
= & \boldsymbol{G}_\beta\boldsymbol{\beta}_f + \boldsymbol{O}_p(\Xi_{NT}^{-1}).
\end{aligned}$$

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 $\boldsymbol{H}_f \equiv \hat{\boldsymbol{F}}_A\hat{\boldsymbol{F}}_B^{-1}\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f$ and $\boldsymbol{H}_0 \equiv \hat{\boldsymbol{F}}_A\hat{\boldsymbol{F}}_B^{-1}\left[N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\phi_0\right]$.

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

$$\hat{\boldsymbol{F}}_t - (\boldsymbol{H}_0 + \boldsymbol{H}_f\boldsymbol{f}_t) = \boldsymbol{O}_p(\Xi_{NT}^{-1}).$$

Proof:

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

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

Remark 7. *The proofs of Theorem 1, 2, and 4 can be approached in an alternative manner. We can demonstrate that, for the matrix* $\boldsymbol{H}_2 = \hat{\boldsymbol{F}}_A\hat{\boldsymbol{F}}_B^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{\Phi}_f$, $\hat{\boldsymbol{F}}_t$ *converges to* $\boldsymbol{H}_2\boldsymbol{f}_t$ *at the rate* Ξ_{NT}*, while* $\hat{\boldsymbol{\beta}}$ *converges to* ${\boldsymbol{H}_2'}^{-1}\boldsymbol{\beta}$ *at* $\min\left(\sqrt{N^{\psi_f}}, \sqrt{T}\right)$ *rate, under the assumptions of our model. Therefore, by specifying a different limit, we can establish faster convergence of* $\hat{\boldsymbol{\beta}}$ *to that limit. Essentially, we require that rotations in* $\hat{\boldsymbol{F}}_t$ *and* $\hat{\boldsymbol{\beta}}$ *be nullified upon multiplication, which occurs with this newly specified limit.*

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

In Kelly & Pruitt [2015], they specify the convergence of $\hat{\boldsymbol{F}}_t$ to $\boldsymbol{HF}_t$, where $\boldsymbol{H} = \boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{\Phi}$ at a $\sqrt{N}$ rate. In our weak factor context, this would be a $\sqrt{N^{\psi_f}}$ rate. However, this matrix $\boldsymbol{H}$ is not square unless there are zero irrelevant factors. Elements of $\boldsymbol{HF}_t$ are linear combinations of relevant and irrelevant factors. Therefore, their convergence result can not 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.

A3 Proofs of Theorems 1(b), 3(b) and 4(b)

We introduce 3 Lemmas which shall be employed for subsequent proofs (Theorems 1(b), 3(b), 4(b)) which deal with the general setting of $\boldsymbol{\zeta} \neq 0$ and $\boldsymbol{\gamma} \neq 0$.

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

1. $N^{-1}T^{-1/2}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon_t = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.
2. $N^{-1}T^{-3/2}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.
3. $N^{-1}T^{-3/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.
4. $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)$.
5. $N^{-1}T^{-3/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.
6. $N^{-1}T^{-3/2}\boldsymbol{\eta}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.
7. (a) $T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p(1)$ *and* $T^{-1/2}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon} = \boldsymbol{O}_p(1)$.

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

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

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

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

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

$$= I + II$$

$\mathbb{E}\left|N^{-1}\sum_{i,s}\varepsilon_s(m)\sigma_{ii,st}\right| \leq N^{-1}\max_s \mathbb{E}\left|\varepsilon_s(m)\right|\sum_{i,s}\left|\sigma_{ii,st}\right| = O_p(1)$ by Assumption 3.1 2.3. Hence, the second term is $O_p\left(T^{-1/2}\right)$.

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

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

Proof: Using Cauchy Schwartz inequality twice,

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

by Assumptions 2.3 and 3.2. Therefore we have that,

$$N^{-1}T^{-1/2}\sum_{i,s}\varepsilon_s(m)\varepsilon_{is}\varepsilon_{it} = O_p\left(\delta_{NT}^{-1}\right). \tag{A3.1}$$

Now we can prove Lemma 6

Item 1 $= N^{-1}T^{-1/2}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}_t$ 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 $O_p\left(\delta_{NT}^{-1}\right)$ by equation A3.1.

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

by Assumption 8 and Lemma 1.3.

1.III is $O_p\left(\delta_{NT}^{-1}T^{-1/2}\right)$ 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 $\boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

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

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

2.I can be wriiten as,

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

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

2.II is $O_p\left(\delta_{NT}^{-1}T^{-1/2}\right)$ by Lemma 1.3 and equation A3.1. Item 2.III is identical.

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

2.V can be written as,

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

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

2.VI is given by

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

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 $\boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

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

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

3.I can be written as,

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

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

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

3.III $= O_p\left(\delta_{NT}^{-1}\right)$ by 2.1 and equation A3.1.

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

3.V can be written as,

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

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

3.VI is given by

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

by Assumption 3.6 and Lemma 1.3.

3.VII is given by

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

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 $\boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$. Item 5 and Item 6 follow similar steps as 3.

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

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

We show that $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p(\Gamma_{N_fT}^{-1})$. Similar arguments would establish $N^{-\psi_g/2}T^{-1}\boldsymbol{\Phi}_g'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p(\Gamma_{N_gT}^{-1})$, which implies $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}_g'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p\left(\frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$.
Hence the matrix $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p\left(\Gamma_{N_fT} \vee \frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right) = \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)$.
Below, we show that $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p(\Gamma_{N_fT}^{-1})$.
$N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is a $K_g \times N$ matrix with generic (m_1, m_2) element

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

4.I is $O_p\left(\Gamma_{N_fT}^{-1}\right)$ by Lemma 1.16.
4.II is $N^{\psi_f/2}T^{-1/2}\delta_{NT}^{-1}\left(N^{-1}T^{-1/2}\sum_{j,t}\varepsilon_t\left(m_2\right)\varepsilon_{jt}\right)\left(N^{-\psi_f}\phi_{if}\left(m_1\right)\right)$ which is $O_p\left(N^{\psi_f/2}T^{-1/2}\delta_{NT}^{-1}\right)$ by Assumption 2.2 and Lemma 1.10. If $\frac{N^{\psi_f}}{T} = O(1)$ then this is $O_p\left(\Gamma_{N_fT}^{-1}\right)$.
4.III is $T^{-1/2}\left(T^{-1}\sum_t\left(\sum_j N^{-\psi_f/2}\phi_{jf}\left(m_1\right)\varepsilon_{jt}\right)\right)\left(T^{-1/2}\sum_s\varepsilon_s\left(m_2\right)\right)$, which is $O_p\left(T^{-1/2}\right)$ by Assumption 3.7 and Lemma 1.3.
4.IV is equal to $(N^{\frac{\psi_f-1}{2}}T^{-1}\left(N^{-\psi_f}\sum_i\phi_{if}\left(m_1\right)\right)\left(N^{-1/2}T^{-1/2}\sum_{j,t}\varepsilon_{jt}\right)\left(T^{-1/2}\sum_s\varepsilon_s\left(m_2\right)\right)$ which is $O_p\left(N^{\frac{\psi_f-1}{2}}T^{-1}\right)$ by Assumption 2.2 and Lemma 1.3.
Summing these terms, $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\right)$ and hence $N^{-\psi_f/2}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}$ is $\boldsymbol{O}_p\left(\Gamma_{N_fT} \vee \frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right) = \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)$.

Item 7:

(a)

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

by Assumption 2.1, 3.6 and Lemma 1.3. 7(b) follows using the same argument and employing Assumption 2.4, 3.5, and Lemma 1.3.

Lemma 7. *Under Assumptions 1-5,8 and 9, if* $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ *and* $\frac{T}{N} = O(1)$*, we have,*

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

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

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

Consequently, the probability limit of $\hat{\boldsymbol{F}}_t$ *is*

$$\hat{\boldsymbol{F}}_t \xrightarrow[T,N\to\infty]{p} \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{\zeta}\left(T^{-1}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\right)\boldsymbol{\zeta}' + \boldsymbol{\Delta}_\omega\right)\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'\right)^{-1}\left(N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\phi_0 + \boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{f}_t\right).$$

Proof:

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

We look at all these terms separately,

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

The limit follows using Assumption 2.1, 5, Lemma 4 and 6 and noting that $\boldsymbol{\zeta}$ has a finite number of non-zero entries by Assumption 9.

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

The probability limit of I was established in Lemma 4. Also note that II is transpose of III. We establish the probability limit of III below.

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

The final equality comes from Lemma 2 ,6, equation A2.1, Assumption 2.1, 4 and noting the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries given Assumption 9. Hence, $\hat{\boldsymbol{F}}_B = N^{-\psi_f}T^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}$ $= \boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)$.

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

The stochastic order of the second term was established in Lemma 4 which yields the order here since all terms except the first two are $\boldsymbol{O}_p(\Xi_{NT}^{-1})$ given Lemma 6 and the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries by Assumption 9.

Continuous mapping theorem yields the plm of $\hat{\boldsymbol{F}}_t$.

Lemma 8. *Under Assumptions 1-5,8 and 9, if* $\dfrac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$ *and* $\dfrac{T}{N} = O(1)$,

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

Therefore,

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

has the the probability limit under Assumption 6 given by,

$$\begin{aligned}\hat{\boldsymbol{\beta}} \xrightarrow[T,N\to\infty]{p} & \left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f' + \boldsymbol{\zeta}\left(T^{-1}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\right)\boldsymbol{\zeta}' + \boldsymbol{\Delta}_\omega\right)^{-1} \boldsymbol{\Lambda_f}\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\boldsymbol{\Lambda_f}' \left(\boldsymbol{\Lambda_f}\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\boldsymbol{\Lambda_f}'\right)^{-1} \\ &\times \boldsymbol{\Lambda_f}\boldsymbol{\Delta_f}\mathcal{P}_f\boldsymbol{\Delta_f}\boldsymbol{\beta}_f.\end{aligned}$$

Proof:

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

Note that $\hat{\boldsymbol{\beta}}_1 = \hat{\boldsymbol{F}}_A$ and $\hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{F}}_B$ and their probability limits are established in Lemma 7. The expressions for $\hat{\boldsymbol{\beta}}_3$ and $\hat{\boldsymbol{\beta}}_4$ are handled below. Note that, $\hat{\boldsymbol{\beta}}_3 = N^{-2\psi_f}T^{-3}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}$ is essentially the product $\frac{\hat{\boldsymbol{F}}_C\boldsymbol{J}_T\hat{\boldsymbol{F}}_C'}{T}$ where $\hat{\boldsymbol{F}}_C$ is obtained by stacking $\hat{\boldsymbol{F}}_{C,t}$ horizontally. Using the probability limit of $\hat{\boldsymbol{F}}_{C,t}$ obtained in Lemma 7 standard arguments would imply that $plim\left(\frac{\hat{\boldsymbol{F}}_C\boldsymbol{J}_T\hat{\boldsymbol{F}}_C'}{T}\right) = \boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\left(T^{-1}\boldsymbol{f}\boldsymbol{J}_T\boldsymbol{f}\right)\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda_f}'$ which is equal to $\boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda_f}'$ given Assumption 2.1. Using Lemma 6 and the expression for $\hat{\boldsymbol{F}}_{C,t}$ in Lemma 7 we can establish that $N^{-2\psi_f}T^{-3}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z} = \frac{\hat{\boldsymbol{F}}_C\boldsymbol{J}_T\hat{\boldsymbol{F}}_C'}{T} = \boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda_f}' + \boldsymbol{O}_p(\Xi_{NT}^{-1})$.

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

The stochastic order of I was established in Lemma 5. III is $\boldsymbol{O}_p(\Xi_{NT}^{-1})$, which can be seen simply by replacing ζ' by γ in the III term of expression of $\hat{\boldsymbol{F}}_B$ in Lemma 7 and noting that $\left(\gamma_i = 0 \implies \boldsymbol{\zeta}_i' = 0\right)$

by Assumption 9. We look at the transpose of II below.

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

The final equality comes from Lemma 2, 6 and noting the fact that $\boldsymbol{\zeta}$ and $\boldsymbol{\gamma}$ have finitely many non-zero entries given Assumption 9. Therefore, $\hat{\boldsymbol{\beta}}_4=\boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\beta}+\boldsymbol{O}_p(\Xi_{NT}^{-1})$.

Given the probability limits of $\hat{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_2$, $\hat{\boldsymbol{\beta}}_3$ and , $\hat{\boldsymbol{\beta}}_4$, Continuous mapping theorem yields the probability limit in the statement of Lemma 8, i.e.,

$$\hat{\boldsymbol{\beta}}\underset{T,N\to\infty}{\xrightarrow{p}}\left(\boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\boldsymbol{\Lambda_f}'+\boldsymbol{\zeta}\left(T^{-1}\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\right)\boldsymbol{\zeta}'+\boldsymbol{\Delta}_\omega\right)^{-1}\boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda_f}'\left(\boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda_f}'\right)^{-1}\boldsymbol{\Lambda_f}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\beta}.$$

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

$$\hat{y}_{t+h,f}-\mathbb{E}\left(y_{t+h}|\boldsymbol{F}_t\right)=O_p(\Xi_{NT}^{-1}).$$

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

- $N^{-\psi_f}T^{-1}\left(\boldsymbol{x}_t-\bar{\boldsymbol{x}}\right)\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}=\hat{\boldsymbol{F}}_{C,t}-\bar{\boldsymbol{F}}_C=\left(\boldsymbol{f}_t-\bar{\boldsymbol{f}}\right)'\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'+\boldsymbol{O}_p(\Xi_{NT}^{-1})$.
- $N^{-2\psi_f}T^{-3}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z}=\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_f'+\boldsymbol{O}_p(\Xi_{NT}^{-1})$.
- $N^{-\psi}T^{-2}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{y}=\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\beta}_f+\boldsymbol{O}_p(\Xi_{NT}^{-1})$.

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 get that $\hat{y}_{t+h,f}=\beta_0+\boldsymbol{F}_t'\boldsymbol{\beta}+O_p(\Xi_{NT}^{-1})$ using the same steps as in the proof of Theorem 1(a).

The Proofs for Theorem3(b) and Theorem4(b) respectively, follow similarly given the rates derived in Lemmas 7 and 8.

A4 Proof of Theorem 5

We now proceed to prove Theorem 5 which deals with stage 2 of 3PRF-Lasso. We introduce 2 Lemmas which shall be used in the subsequent proof of Theorem 5

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

$$\hat{\varepsilon}_{it} - \left(\varepsilon_{it} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} - \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig}\right) = O_p(\Xi_{NT}^{-1}).$$

.

Proof: The proof proceeds in a similar manner to Theorem1(b). The target $\boldsymbol{y}$ can be replaced by $\boldsymbol{x}_i$ and the proof follows similar steps.

First notice that, using the same steps as in proof of Lemma 4 for $\hat{\boldsymbol{F}}_B$, we can get

$$N^{-\psi_f}T^{-2}\boldsymbol{W}'_{XZ}\boldsymbol{S}_{X\boldsymbol{x}_i} = \boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\boldsymbol{\phi}_i + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right).$$

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

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

The result follows from Assumption 2.1 and 4. Substituting this in the expression of $\hat{x}_{it}$ we get,

$$
\begin{aligned}
\hat{x}_{it} &= \bar{\boldsymbol{x}}_i + \left(N^{-\psi_f}T^{-1}\left(\boldsymbol{x}_t - \bar{\boldsymbol{x}}\right)\boldsymbol{W}_{XZ}\right)\left(N^{-2\psi_f}T^{-3}\boldsymbol{W}'_{XZ}\boldsymbol{S}_{XX}\boldsymbol{W}_{XZ}\right)^{-1}\left(N^{-\psi_f}T^{-2}\boldsymbol{W}'_{XZ}\boldsymbol{S}_{X\boldsymbol{x}_i}\right) \\
&= \phi_{0,i} + \bar{\boldsymbol{F}}'\boldsymbol{\phi}_i + O_p(T^{-1/2}) + \left(\left(\boldsymbol{f}_t - \bar{\boldsymbol{f}}\right)'\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}'_f + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right) \\
&\quad \times \left[\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}'_f + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right]^{-1}\left(\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\phi}_{if} + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig} + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right) \\
&= \phi_{0,i} + \bar{\boldsymbol{f}}'\boldsymbol{\phi}_{if} + \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig} + O_p(T^{-1/2}) + \left(\boldsymbol{f}_t - \bar{\boldsymbol{f}}\right)'\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\boldsymbol{\Lambda}'_f \\
&\quad \times \left[\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_{f,1}\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\boldsymbol{\Lambda}_f\right]^{-1}\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_{f,1}\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\boldsymbol{\phi}_{if} + \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig} \\
&= \phi_{0,i} + \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig} + O_p(T^{-1/2}) + \boldsymbol{f}'_t\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\boldsymbol{\Lambda}'_f \\
&\quad \times \left[\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_{f,1}\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\boldsymbol{\Lambda}_f\right]^{-1}\boldsymbol{\Lambda}_f\boldsymbol{\Delta}_{f,1}\mathcal{P}_1\boldsymbol{\Delta}_{f,1}\boldsymbol{\phi}_{if} + \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig} \\
&= \phi_{0,i} + \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig} + \boldsymbol{f}'_t\boldsymbol{\phi}_{if} + \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig} \\
\therefore \hat{\varepsilon}_{it} &= x_{it} - \hat{x}_{it} = \left(\phi_{0,i} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} + \boldsymbol{f}'_t\boldsymbol{\phi}_{if} + \varepsilon_{it}\right) - \left(\phi_{0,i} + \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig} + \boldsymbol{f}'_t\boldsymbol{\phi}_{if} + \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig}\right) \\
\implies & \hat{\varepsilon}_{it} - \left(\varepsilon_{it} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} - \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig}\right) = \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig} \\
\implies & \hat{\varepsilon}_{it} - \left(\varepsilon_{it} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} - \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig}\right) = \boldsymbol{O}_p(\Xi_{NT}^{-1}).
\end{aligned}
$$

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

Lemma 10. *Define* $\tilde{\boldsymbol{\eta}} \equiv \hat{\boldsymbol{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}$*, where* $\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}}_f = \boldsymbol{y} - \iota_T\bar{y} - \boldsymbol{J}_T\hat{\boldsymbol{F}}\hat{\boldsymbol{\beta}}$.

$$
\max_i\left(\frac{\Xi_{NT}}{T}\right)\hat{\boldsymbol{\varepsilon}}'_i\tilde{\boldsymbol{\eta}} = \max_i\left(\frac{\Xi_{NT}}{T}\right)\boldsymbol{\varepsilon}'_i\boldsymbol{\eta} + O_p(1).
$$

Proof: Adding and subtracting terms,

$$
\begin{aligned}
\max_i\left(\frac{\Xi_{NT}}{T}\right)\hat{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\eta}} = &\left(\frac{\Xi_{NT}}{T}\right)\max_i\left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig}\right)'\boldsymbol{\eta} + \left(\frac{\Xi_{NT}}{T}\right)\max_i\left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig}\right)'\left(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}\right) \\
&+ \left(\frac{\Xi_{NT}}{T}\right)\max_i\left(\hat{\boldsymbol{\varepsilon}}_i - \left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig}\right)\right)'\boldsymbol{\eta} + \left(\frac{\Xi_{NT}}{T}\right)\max_i\left(\hat{\boldsymbol{\varepsilon}}_i - \left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig}\right)\right)'\left(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}\right) \\
= &\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}.
\end{aligned}
$$

We show that $\mathrm{I} = \max_i\left(\frac{\Xi_{NT}}{T}\right)\boldsymbol{\varepsilon}'_i\boldsymbol{\eta} + O_p(1)$ and II, III and IV are $O_p(1)$.

Item I:

$$
\begin{aligned}
\max_i\left(\frac{\Xi_{NT}}{T}\right)\left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig}\right)'\boldsymbol{\eta} &= \left(\frac{\Xi_{NT}}{T}\right)\max_i\boldsymbol{\varepsilon}'_i\boldsymbol{\eta} + \left(\frac{\Xi_{NT}}{T}\right)\left(\max_i\boldsymbol{\phi}_{ig}\right)\boldsymbol{g}'\boldsymbol{J}_T\boldsymbol{\eta} \\
&= \max_i\left(\frac{\Xi_{NT}}{T}\right)\boldsymbol{\varepsilon}'_i\boldsymbol{\eta} + O_p(1). \qquad \text{(A4.1)}
\end{aligned}
$$

Where the final line follows from the fact that $\dfrac{\Xi_{NT}}{T} \leq \dfrac{1}{\sqrt{T}}$, $\max_i \boldsymbol{\phi}_{ig}$ is $O_p(1)$ by Assumption 7.1 and $\dfrac{\boldsymbol{g}'\boldsymbol{J}_T\boldsymbol{\eta}}{T^{1/2}}$ is $\boldsymbol{O}_p(1)$ by Lemma 2.2.

Item II : Note that $\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}} = (\boldsymbol{y} - \boldsymbol{y}_f - \boldsymbol{\varepsilon}\gamma) - (\boldsymbol{y} - \hat{\boldsymbol{y}}_f - \hat{\boldsymbol{\varepsilon}}\gamma) = (\hat{\boldsymbol{y}}_f - \boldsymbol{y}_f) + (\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})\gamma$. Therefore,

$$\begin{aligned}
\left(\frac{\Xi_{NT}}{T}\right)(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig})'(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) &= \left(\frac{\Xi_{NT}}{T}\right)(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta})'(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig}) \\
&= \left(\frac{\Xi_{NT}}{T}\right)(\hat{\boldsymbol{y}}_f - \boldsymbol{y}_f)'\boldsymbol{\varepsilon}_i + \left(\frac{\Xi_{NT}}{T}\right)(\hat{\boldsymbol{y}}_f - \boldsymbol{y}_f)'\boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig} \\
&\quad + \left(\frac{\Xi_{NT}}{T}\right)\gamma'(\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})'\boldsymbol{\varepsilon}_i + \left(\frac{\Xi_{NT}}{T}\right)\gamma'(\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon})'\boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\phi}_{ig} \\
&= \mathcal{A}_i + \mathcal{B}_i + \mathcal{C}_i + \mathcal{D}_i.
\end{aligned}$$

Notice that terms $\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i$ and $\mathcal{D}_i$ are all scalars. Therefore $\mathcal{A}_i = \|\mathcal{A}_i\|_1 = \|\mathcal{A}_i\|_\infty$. Same holds for $\mathcal{B}_i, \mathcal{C}_i$ and $\mathcal{D}_i$. We use this fact throughout the proof. We look at all these terms separately. From Lemma 2 and the expression for $\boldsymbol{y}$ in the proof of Theorem 1(a) and Theorem 1(b), we have that $(\hat{\boldsymbol{y}}_f - \boldsymbol{y}_f)' = \boldsymbol{O}_p(\Xi_{NT}^{-1})\boldsymbol{F}'\boldsymbol{J}_T + \boldsymbol{O}_p(1)\left(\hat{\boldsymbol{F}}_C - \boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{f}'\right)\boldsymbol{J}_T \; + \gamma_s'(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)' + (\iota_T\bar{\boldsymbol{\eta}})'$. Therefore,

$$\begin{aligned}
\mathcal{A}_i &= \left(\frac{\Xi_{NT}}{T}\right)(\hat{\boldsymbol{y}}_f - \boldsymbol{y}_f)'\boldsymbol{\varepsilon}_i \\
&= \Xi_{NT}\boldsymbol{O}_p(\Xi_{NT}^{-1})\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) \\
&\quad + \boldsymbol{O}_p(1)\left(\frac{\Xi_{NT}}{T}\right)\left(\hat{\boldsymbol{F}}_C - \boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{f}'\right)\boldsymbol{J}_T\boldsymbol{\varepsilon}_i + \left(\frac{\Xi_{NT}}{T}\right)\gamma_s'(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)'\boldsymbol{\varepsilon}_i + \left(\frac{\Xi_{NT}}{T}\right)(\iota_T\bar{\boldsymbol{\eta}})'\boldsymbol{\varepsilon}_i \\
&= \mathcal{A}_{1i} + \mathcal{A}_{2i} + \mathcal{A}_{3i} + \mathcal{A}_{4i}.
\end{aligned}$$

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

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

Therefore, by Triangle inequality,

$$\|\mathcal{A}_{1i}\|_\infty \leq \|\boldsymbol{O}_p(1)\|_\infty\|\frac{1}{T}\boldsymbol{F}'\boldsymbol{\varepsilon}_i\|_\infty + \|\frac{1}{T}\boldsymbol{O}_p(1)\boldsymbol{F}'\|_\infty\|\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_i\|_\infty.$$

Since the $\boldsymbol{O}_p(1)$ matrix listed above is a finite dimensional matrix, it's L_∞ norm will have the same order as it's elements. Also, $\boldsymbol{O}_p(1)\boldsymbol{F}'$ is a $L \times T$,($L < \infty$) matrix with all elements having bounded second moments given Assumptions 2 and 3. Hence, its L_∞ norm will scale with at most T and therefore $\|\dfrac{1}{T}\boldsymbol{O}_p(1)\boldsymbol{F}'\|_\infty$ has a maximum order $O_p(1)$. $\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_i$ has all same elements, hence its L_∞ norm is equal to any element in this vector. Hence, we have, Therefore

$$\max_i \|\mathcal{A}_{1i}\|_\infty \leq \|\boldsymbol{O}_p(1)\|_\infty \max_i \|\frac{1}{T}\boldsymbol{F}'\boldsymbol{\varepsilon}_i\|_\infty + \|\frac{1}{T}\boldsymbol{O}_p(1)\boldsymbol{F}'\|_\infty \max_i |\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_i|$$
$$= \boldsymbol{O}_p\left(\frac{(\log N)^{r_3}}{\sqrt{T}}\right) + \boldsymbol{O}_p\left(\frac{(\log N)^{r_2}}{\sqrt{T}}\right).$$

The final line follows from Assumptions 7.3 and 7.4. Hence by Assumption 10, $\max_i \mathcal{A}_{1i}$ is $O_p(1)$. We can expand $\mathcal{A}_{2i}$ using Lemma 4 as

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

Given that $\boldsymbol{\zeta}$ has finitely many non-zero entries by Assumption 9, $\|\boldsymbol{\zeta}\|_\infty$ and $\|\boldsymbol{\zeta}\|_1$ are both $O_p(1)$. Also, given the orders derived in 2 and 6, and the fact that $\Xi_{NT} < \delta_{NT}$, we can say that $\mathcal{A}_{2i}$ is a sum of 3 type of terms.

1. $\mathcal{A}_{2i}.1$: $\boldsymbol{O}_p(1)\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right)$ where the $O_p(1)$ term is a finite dimensional matrix.
2. $\mathcal{A}_{2i}.2$: $\dfrac{N^{1-\psi_f}}{\sqrt{T}}\dfrac{\boldsymbol{D}\boldsymbol{\varepsilon}_i}{T}$ where $\boldsymbol{D}$ is a $L \times T$ $\boldsymbol{O}_p(1)$ matrix, invariant across t and i, with all terms having bounded second moments.
3. $\mathcal{A}_{2i}.3$: $\Xi_{NT}\boldsymbol{O}_p(1)\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) + \Xi_{NT}\boldsymbol{O}_p(T^{-1/2})\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}_g'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right)$, where the $\boldsymbol{O}_p(1)$ and $\boldsymbol{O}_p(T^{-1/2})$ terms are finite dimensional matrices invariant across t and i.[10]

For $\mathcal{A}_{2i}.1$, we follow the same proof as in $\mathcal{A}_{1i}$ to show that its maximum value over i is bounded under Assumption 10.

[10]This follows from the observation that $\boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) = \boldsymbol{\Lambda}_f\left(T^{-1}\boldsymbol{f}'\boldsymbol{J}_T\boldsymbol{f}\right)\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) + T^{-1/2}\boldsymbol{\Lambda}_f\left(T^{-1/2}\boldsymbol{f}'\boldsymbol{J}_T\boldsymbol{g}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}_g'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right)$ and noticing that the rest of the terms pre-multiplying $\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right)$ are $\boldsymbol{O}_p(T^{-1/2})$.

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

$$\left\|\frac{N^{1-\psi_f}}{\sqrt{T}}\frac{\boldsymbol{D}\boldsymbol{\varepsilon}_i}{T}\right\|_\infty \leq \frac{N^{1-\psi_f}}{\sqrt{T}}\left\|\frac{\boldsymbol{D}}{T}\right\|_\infty \|\boldsymbol{\varepsilon}_i\|_\infty$$
$$= O_p(1)\frac{N^{1-\psi_f}}{\sqrt{T}}\max_t |\varepsilon_{it}|.$$

We get the final equality from the fact that $\boldsymbol{D}$ is $L \times T$ matrix with all elements having bounded second moments given Assumptions 2 and 3. Hence, its L_∞ norm will scale with at most T and therefore $\left\|\frac{\boldsymbol{D}}{T}\right\|_\infty$ will have a maximum order $O_p(1)$. Hence we have,

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

Under Assumption 10, we have $\frac{N^{1-\psi_f}}{\sqrt{T}}\left[(\log N)^{r_1} + (\log T)^{r_1}\right]$ is $O(1)$. Hence,

$$\max_i \left\|\frac{N^{1-\psi_f}}{\sqrt{T}}\frac{\boldsymbol{D}\boldsymbol{\varepsilon}_i}{T}\right\|_\infty \quad \text{is } O_p(1).$$

To show that $\max_i \mathcal{A}_{2i}.3$ is $O_p(1)$, we first show that $\max_i \left[\Xi_{NT}\left\|N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_f\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right\|_1\right]$ is $\boldsymbol{O}_p(1)$ given assumptions of our model. to do so, we derive the stochastic order of a generic $(m,1)^{th}$ element of the $K_f \times 1$ matrix $(\max_i \Xi_{NT}N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_f\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i)$. Since K_f is finite and the stochastic order is invariant across $m \in \{1,\dots,K_f\}$, the stochastic order of a generic element will be equal to the stochastic order of the L_1 norm the matrix.

A generic $(m,1)^{th}$ element of the $K_f \times 1$ matrix $(\max_i \Xi_{NT}N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_f\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i)$ is bounded by

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

By the definition of Ξ_{NT}, $\Xi_{NT}N^{-\psi_f/2} = O(1)$, which implies $a = \boldsymbol{O}_p\left(\frac{(logN)^{r_4}}{\sqrt{T}}\right)$ by Assumption 7.5. $b = \boldsymbol{O}_p\left(\frac{(logN)^{r_5}}{\sqrt{T}}\right)$ by Assumptions 2.2 and 7.6. $c = \boldsymbol{O}_p\left(\frac{(logN)^{r_2}}{\sqrt{T}}\right)$ by Assumption 3.7 and

7.3. $d = \boldsymbol{O}_p\left(\frac{(logN)^{r_2}}{T}\right)$ by Lemma 1.3 and Assumptions 2.2 and 7.3. Therefore, by Assumption 10, $\max_i \left[\Xi_{NT}\left\|N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_f\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right\|_1\right]$ is $\boldsymbol{O}_p(1)$. Given Assumption 6, similar reasoning would establish, $\max_i \left[\Xi_{NT}T^{-1/2}\left\|N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_g\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right\|_1\right]$ is $\boldsymbol{O}_p(1)$. Hence, we have

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

Hence $\mathcal{A}_{2i}.3 = O_p(1)$.
For $\mathcal{A}_{3i}$, since $\frac{\Xi_{NT}}{T} \leq \frac{1}{\sqrt{T}}$, we have, $\left(\frac{\Xi_{NT}}{T}\right)\boldsymbol{\gamma}'_s(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)'\boldsymbol{\varepsilon}_i \leq \left(\frac{1}{\sqrt{T}}\right)\boldsymbol{\gamma}'_s(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)'\boldsymbol{\varepsilon}_i$ and we have

$$\left(\frac{1}{\sqrt{T}}\right)\boldsymbol{\gamma}'_s(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)'\boldsymbol{\varepsilon}_i = \sum_{j\in S}\gamma_j\bar{\boldsymbol{\varepsilon}}_j\left(\frac{1}{\sqrt{T}}\sum_t \varepsilon_{it}\right).$$

Therefore,

$$\begin{aligned}
\max_i\left(\frac{1}{\sqrt{T}}\right)\boldsymbol{\gamma}'_s(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)'\boldsymbol{\varepsilon}_i &\leq \sum_{j\in S}\gamma_j\bar{\boldsymbol{\varepsilon}}_j\left(\max_i\left|\frac{1}{\sqrt{T}}\sum_t\varepsilon_{it}\right|\right)\\
&=\sum_{j\in S}O_p\left(\frac{1}{\sqrt{T}}\right)O_p((\log N)^{r_2})\\
&=O_p\left(\frac{(\log N)^{r_2}}{\sqrt{T}}\right)\\
&=O_p(1),
\end{aligned}$$

where the second last line line follows from Assumption 7 and Lemma 1.3. The second last line follows from the fact that the cardinality of S is bounded by Assumption 9. The final line follows from Assumption 10.
Using a similar logic, $\max_i\left(\frac{\Xi_{NT}}{T}\right)(\iota_T\bar{\boldsymbol{\eta}})'\boldsymbol{\varepsilon}_i$ is $O_p(1)$. Hence $\max_i\mathcal{A}_i$ $=\max_i(\mathcal{A}_{1i}+\mathcal{A}_{2i}+\mathcal{A}_{3i}+\mathcal{A}_{4i})$ is $O_p(1)$.
Next, we show that $\max_i\mathcal{B}_i = O_p(1)$.

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

$\mathcal{B}_{1i} \ldots \mathcal{B}_{4i}$ depend on i through $\boldsymbol{\phi}_i$ only. Since $\forall m$, $\max_i \phi_i(m) = O_p(1)$ by Assumption 7.1, it suffices to show that $\mathcal{B}_{1i} + \mathcal{B}_{2i} + \mathcal{B}_{3i} + \mathcal{B}_{4i} = O_p(1)$ in order to prove $\max_i \mathcal{B}_i = O_p(1)$. $\mathcal{B}_{1i} = O_p(1)$ by assumption 2.1. $\mathcal{B}_{3i}$ and $\mathcal{B}_{4i}$ are $O_p(1)$ by Assumption 2.1 and lemmas 1.3 and 1.4. We show $\mathcal{B}_{2i} = O_p(1)$ below. Using Lemma 4, we can expand $\mathcal{B}_{2i}$ as,

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

The final line follows from Lemma 2 and 6 and the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries by Assumption 9.

$\max_i \mathcal{C}_i = O_p(1)$ follows from similar argument as for $\max_i \mathcal{A}_i = O_p(1)$. $\max_i \mathcal{C}_i = \max_i \boldsymbol{\gamma}' \left(\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}\right)' \boldsymbol{\varepsilon}_i = \max_i \boldsymbol{\gamma}_S' \left(\hat{\boldsymbol{\varepsilon}}_S - \boldsymbol{\varepsilon}_S\right)' \boldsymbol{\varepsilon}_i$. The expression for $\hat{\boldsymbol{\varepsilon}}_S - \boldsymbol{\varepsilon}_S$ can be obtained using lemma 9 as $\forall i \in S$, $\phi_{ig} = 0$. The proof then follows similar steps as for $\mathcal{A}_i$.

Similarly, $\max_i \mathcal{D}_i = O_p(1)$ follows from an analogous argument as for $\max_i \mathcal{B}_i = O_p(1)$. Therefore $\text{II} = \max_i(\mathcal{A}_i + \mathcal{B}_i + \mathcal{C}_i + +\mathcal{D}_i)$ is $O_p(1)$.

We now show item III : $\left(\frac{\Xi_{NT}}{T}\right) \max_i \left(\hat{\boldsymbol{\varepsilon}}_i - \left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T \boldsymbol{g} \boldsymbol{\phi}_{ig}\right)\right)' \boldsymbol{\eta}$ is $O_p(1)$.

From the discussion leading upto lemma 9, we can express $\left(\frac{\Xi_{NT}}{T}\right) \max_i \left(\hat{\boldsymbol{\varepsilon}}_i - \left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T \boldsymbol{g} \boldsymbol{\phi}_{ig}\right)\right)' \boldsymbol{\eta}$ as a sum of three terms.

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

Since $\frac{\Xi_{NT}}{T} \leq \frac{1}{\sqrt{T}}$, we have $\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \frac{\boldsymbol{f}' \boldsymbol{J}_T \boldsymbol{\eta}}{T} \Xi_{NT} \leq \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \frac{\boldsymbol{f}' \boldsymbol{J}_T \boldsymbol{\eta}}{\sqrt{T}} = O_p(1)$ by Lemma 1.2. Therefore we have $\mathcal{P} = \max_i (N^{-\psi_f} T^{-2} \boldsymbol{W'}_{XZ} \boldsymbol{S}_{X\boldsymbol{x}_i} - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if})' \boldsymbol{O}_p(1)$. We need to show that $\max_i (N^{-\psi_f} T^{-2} \boldsymbol{W'}_{XZ} \boldsymbol{S}_{X\boldsymbol{x}_i} - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if}) = O_p(1)$.

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

The last term $\boldsymbol{O}_p(\Xi_{NT}^{-1}) \boldsymbol{\phi}_i$ captures all the terms in the expansion which have a maximum order of Ξ_{NT}^{-1} and depend on i through ϕ_i only. Therefore $\max_i (N^{-\psi_f} T^{-2} \boldsymbol{W'}_{XZ} \boldsymbol{S}_{X\boldsymbol{x}_i} - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if}) = \boldsymbol{O}_p(1)$ since $\max_i \boldsymbol{\phi}_i$ is $\boldsymbol{O}_p(1)$ by Assumption 7.1. Stochastic order for the max of other terms follows similar arguments as in Item I for the term $\mathcal{A}$.

For $\mathcal{Q}$, note that since $\max_i \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if} = O_p(1)$ by Assumption 7.1, it suffices to show that

$\left(\hat{\boldsymbol{F}}_C - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{f}'\right) \boldsymbol{J}_T \boldsymbol{\eta} \left(\frac{\Xi_{NT}}{T}\right) = \boldsymbol{O}_p(1)$, which we show below.

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

Since $\frac{\Xi_{NT}}{T} \leq \frac{1}{\sqrt{T}}$ and $\frac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, the final equality comes from rates derived in Lemma 2, 6, employing the fact that $\boldsymbol{\zeta}$ has finitely many non-zero entries by Assumption 9. IV $= O_p(1)$ can be deduced similarly and this concludes the proof.

Theorem 5 Let the regularization parameter in Stage-2 Pass 1 regression be given by $\lambda := 2\frac{\sqrt{c+\kappa\log N}}{\Xi_{NT}}$, $c > 0$ and κ is defined in assumption 10 . Then, if Assumptions 1-10 hold, w.p at least $1-\exp\left[-\frac{c}{\kappa}\right]+o(1)$, we have,

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

Proof: First stage regression gives initial forecast $\hat{\boldsymbol{y}}_f = \iota_T\bar{y} + \boldsymbol{J}_T\hat{\boldsymbol{F}}\hat{\boldsymbol{\beta}}$. Let $\hat{\boldsymbol{u}} = (\hat{u}_{1+h}, \ldots, \hat{u}_{T+h})'$ denote the vector of stacked residuals from the first stage regression. $\hat{\boldsymbol{u}} = \boldsymbol{y} - \hat{\boldsymbol{y}}_f = \boldsymbol{y} - \iota_T\bar{y} - \boldsymbol{J}_T\hat{\boldsymbol{F}}\hat{\boldsymbol{\beta}}$. The second stage involves the Lasso regression of $\hat{\boldsymbol{u}}$ on $\hat{\boldsymbol{\varepsilon}}$, where both $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{\varepsilon}}$ are generated regressors ,i.e., $\boldsymbol{\gamma}$ is estimated by the following penalized regression,

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

The lasso solution must satisfy

$$\| \hat{\boldsymbol{u}} - \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\gamma}}\|_2^2/T + \lambda\|\hat{\boldsymbol{\gamma}}\|_1 \leq \| \hat{\boldsymbol{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}\|_2^2/T + \lambda\|\boldsymbol{\gamma}\|_1. \tag{A4.2}$$

Since $\tilde{\boldsymbol{\eta}} = \hat{\boldsymbol{u}} - \hat{\boldsymbol{\varepsilon}}\boldsymbol{\gamma}$(Defined in Lemma 9), Equation A4.2 can be written as,

$$\frac{(\hat{\boldsymbol{u}} - \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\eta}})' (\hat{\boldsymbol{u}} - \hat{\boldsymbol{\varepsilon}}\hat{\boldsymbol{\gamma}} + \tilde{\boldsymbol{\eta}})}{T} + \lambda\|\hat{\boldsymbol{\gamma}}\|_1 - \lambda\|\boldsymbol{\gamma}\|_1 \leq 0.$$

Substituting values of $\tilde{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{u}}$, this simplifies to

$$\frac{(-\hat{\boldsymbol{\varepsilon}}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}))' \,(-\hat{\boldsymbol{\varepsilon}}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + 2\tilde{\boldsymbol{\eta}})}{T} + \lambda\|\hat{\boldsymbol{\gamma}}\|_1 - \lambda\|\boldsymbol{\gamma}\|_1 \leq 0.$$

This gives the “Basic Inequality” for Lasso. See Bühlmann & Van De Geer [2011] (Page 103)

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

Note that,

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

Next, we show that for an appropriate choice λ_0 the set

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

has a high probability.

Let $\lambda_0 := \dfrac{\sqrt{c + \kappa\log N}}{\Xi_{NT}}$, From Assumption 10,

$$\begin{aligned}
\mathbb{P}(\mathscr{T}) &= \mathbb{P}\left(\max_{1\leq j\leq N} \frac{2\left|\hat{\boldsymbol{\varepsilon}}_j'\tilde{\boldsymbol{\eta}}\right|}{T} \leq \frac{\sqrt{c + \kappa\log N}}{\Xi_{NT}}\right)\\
&= 1 - \mathbb{P}\left(\max_{1\leq j\leq N} \frac{\Xi_{NT}}{T} 2\left|\hat{\boldsymbol{\varepsilon}}_j'\tilde{\boldsymbol{\eta}}\right| \geq \sqrt{c + \kappa\log N}\right)\\
&\geq 1 - \mathbb{P}\left(\max_{1\leq j\leq N} \frac{2\left|\hat{\boldsymbol{\varepsilon}}_j'\tilde{\boldsymbol{\eta}}\right|}{\sqrt{T}} \geq \sqrt{c + \kappa\log N}\right)\\
&\geq 1 - \mathbb{P}\left(\left(\max_{1\leq j\leq N} \frac{2\left|\boldsymbol{\varepsilon}_j{}'\boldsymbol{\eta}\right|}{\sqrt{T}}\right) + |O_p(1)| \geq \sqrt{c + \kappa\log N}\right)\\
&\geq 1 - N\exp\left[\frac{-(c + \kappa\log N)}{\kappa}\right] + o(1)\\
&= 1 - \left(\exp\left(\frac{-c}{\kappa}\right) + o(1)\right).
\end{aligned}$$

The third inequality follows from the fact that $\dfrac{\Xi_{NT}}{T} \leq \dfrac{1}{\sqrt{T}}$, the fourth inequality comes from Lemma

10 and the final inequality is by Assumption 10. By making c arbitrarily large, the probability can be made arbitrarily small.

We have on $\mathscr{T}$, with $\lambda = 2\lambda_0$,

$$2\frac{\|\hat{\boldsymbol{\varepsilon}}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)\|_2^2}{T} + \lambda \left\|\hat{\boldsymbol{\gamma}}_{s^c}\right\|_1 \leq 3\lambda \left\|\hat{\boldsymbol{\gamma}}_s - \boldsymbol{\gamma}_s\right\|_1. \tag{A4.3}$$

Proof: See Lemma 6.3, page 105 , Bühlmann & Van De Geer [2011].

Define $\Delta_{\hat{\varepsilon}} := \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/T$. Given that the set of 'relevant' idiosyncratic terms S has a fixed cardinality, it is easy to show that if the comparability condition for set S, holds w.r.t $\Delta_{\varepsilon,g}$ (defined in Assumption 10(c)) it implies that the compatibility condition also holds for set S w.r.t $\Delta_{\hat{\varepsilon}}$.

Proof: Through similar steps as in Lemma 10, we can show that

$$\max_{ij}\left(\frac{1}{T}\right)\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_j = \max_{ij}\left(\frac{\left(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\Phi}_{ig}\right)'\left(\boldsymbol{\varepsilon}_j + \boldsymbol{J}_T\boldsymbol{g}\boldsymbol{\Phi}_{jg}\right)}{T}\right) + o_p(1). \tag{A4.4}$$

Hence,

$$\max_{ij}\Delta_{\hat{\varepsilon}}^{ij} = \max_{ij}\Delta_{\varepsilon,g}^{ij} + o_p(1). \tag{A4.5}$$

Compatibility condition for set S w.r.t $\Delta_{\varepsilon,g}$ implies, for all $N\times 1$ vectors $\boldsymbol{\Theta}$ satisfying $\|\boldsymbol{\Theta}_{S^c}\|_1 < 3\|\boldsymbol{\Theta}_S\|_1$, we have,

$$\|\boldsymbol{\Theta}_S\|_1^2 < \left(\boldsymbol{\Theta}'\Delta_{\varepsilon,g}\boldsymbol{\Theta}\right)|S|/\nu_0^2,$$

which equivalently can be stated as,

$$1 < \frac{\left(\boldsymbol{\Theta}'\Delta_{\varepsilon,g}\boldsymbol{\Theta}\right)|S|}{\|\boldsymbol{\Theta}_S\|_1^2\,\nu_0^2}.$$

Therefore we have,

$$\begin{aligned}
1 &< \frac{\left(\boldsymbol{\Theta}'\Delta_{\varepsilon,g}\boldsymbol{\Theta}\right)|S|}{\|\boldsymbol{\Theta}_S\|_1^2\,\nu_0^2} \\
&< C_o\sum_{i,j}\frac{\Theta_i\Theta_j}{\|\boldsymbol{\Theta}_S\|_1^2}\left(\Delta_{\varepsilon,g}^{ij}\right) \\
&< C_o\sum_{i,j}\frac{\Theta_i\Theta_j}{\|\boldsymbol{\Theta}_S\|_1^2}\left(\Delta_{\hat{\varepsilon}}^{ij}\right) + \max_{ij} C_o\left(\Delta_{\varepsilon,g}^{ij} - \Delta_{\hat{\varepsilon}}^{ij}\right)\sum_{i,j}\frac{\Theta_i\Theta_j}{\|\boldsymbol{\Theta}_S\|_1^2} \\
&< C_o\sum_{i,j}\frac{\Theta_i\Theta_j}{\|\boldsymbol{\Theta}_S\|_1^2}\left(\Delta_{\hat{\varepsilon}}^{ij}\right) + o_p(1)O(1) \\
&< C_o\sum_{i,j}\frac{\Theta_i\Theta_j}{\|\boldsymbol{\Theta}_S\|_1^2}\left(\Delta_{\hat{\varepsilon}}^{ij}\right) + o_p(1).
\end{aligned} \tag{A4.6}$$

The second last line follows from A4.5. $C_o = \frac{|S|}{\nu_0^2}$ is a constant. $\sum_{i,j}\frac{\Theta_i\Theta_j}{\|\boldsymbol{\Theta}_S\|_1^2}$ is bounded since we need to look at only those vectors which satisfy $\|\boldsymbol{\Theta}_{S^c}\|_1 < 3\|\boldsymbol{\Theta}_S\|_1$. Therefore, compatibility condition for the S w.r.t $\Delta_{\varepsilon,g}$ implies that compatibility condition for the S holds w.r.t $\Delta_{\hat{\varepsilon}}$ with probability approaching one.

Using equation A4.3 and Assumption 10(b), A4.6, we have, on $\mathscr{T}$, with probability approaching one, for $\lambda = 2\lambda_0$,

$$\|\hat{\varepsilon}(\hat{\gamma}-\gamma)\|_2^2/T + \lambda\|\hat{\gamma}-\gamma\|_1 \le 4\lambda^2|S|/\nu_0^2. \tag{A4.7}$$

Proof:

$$\begin{aligned}2\|\hat{\varepsilon}(\hat{\gamma}-\gamma)\|_2^2/T + \lambda\|\hat{\gamma}-\gamma\|_1 &= 2\|\hat{\varepsilon}(\hat{\gamma}-\gamma)\|_2^2/T + \lambda\|\hat{\gamma}_s-\gamma_s\|_1 + \lambda\|\hat{\gamma}_{s^c}\|_1\\ &\le 4\lambda\|\hat{\gamma}_s-\gamma_s\|_1\\ &\le 4\lambda\sqrt{|S|}\,\|\hat{\varepsilon}(\hat{\gamma}-\gamma)\|_2/\left(\sqrt{T}\nu_0\right)\\ &\le \|\hat{\varepsilon}(\hat{\gamma}-\gamma)\|_2^2/T + 4\lambda^2|S|/\nu_0^2.\end{aligned}$$

where we have used equation A4.3 in the first inequality and the compatibility condition (Assumption 10) is used in the second inequality.[11] The Last inequality uses that fact that for any u, v, $4uv \le u^2+4v^2$. Concluding from the discussion above, we have that, Using the regularization parameter $\lambda = 2\lambda_0$, on the set $\mathscr{T}$, w.p approaching one, we have

$$\begin{aligned}\|\hat{\varepsilon}(\hat{\gamma}-\gamma)\|_2^2/T &\le 4\lambda^2|S|/\nu_0^2\\ &= O_p(\lambda^2)\\ &= O_p\left(\frac{logN}{{\Xi_{NT}}^2}\right).\end{aligned} \tag{A4.8}$$

The final equality uses the fact that $|S|$ is finite. Finally, using triangle inequality we have,

$$\begin{aligned}\frac{1}{T}\|\hat{\varepsilon}\hat{\gamma}-\varepsilon\gamma\|_2 &\le \frac{1}{T}\|\hat{\varepsilon}\hat{\gamma}-\hat{\varepsilon}\gamma\|_2 + \frac{1}{T}\|(\hat{\varepsilon_s}-\varepsilon_s)\gamma_s\|_2\\ &\le \frac{1}{T}\|\hat{\varepsilon}\hat{\gamma}-\hat{\varepsilon}\gamma\|_2 + O_p\left(\Xi_{NT}^{-1}\right)\\ &= O_p\left(\frac{\sqrt{\log N}}{\Xi_{NT}}\right) + O_p\left(\Xi_{NT}^{-1}\right)\\ &= O_p\left(\frac{\sqrt{\log N}}{\Xi_{NT}}\right).\end{aligned}$$

We have used triangle inequality in first step and Lemma 9 in the second step, noting that for $j \in \{i|\gamma_i \neq 0\}$, $\boldsymbol{\phi}_{jg} = 0$ by Assumption 1. The third step invokes equation A4.8.

Corollary 5.1 follows directly using triangle inequality combining Theorem 5 and 1 (b).

[11]In the compatibility condition we have used $\boldsymbol{\Theta} = \hat{\gamma}-\gamma$ since $\hat{\gamma}-\gamma$ satisfies the condition $\|\hat{\gamma}_{s^c}-\gamma_{s^c}\|_1 = \|\hat{\gamma}_{s^c}\|_1 \le 3\|\hat{\gamma}_s-\gamma_s\|_1$ by A4.3