Forecasting Economic Time Series Using Supervised Factors and Idiosyncratic Elements

Daanish Padha*

Tae-Hwy Lee†

Abstract

We extend the paper on Three-Pass Regression Filter (3PRF) by Kelly & Pruitt [2015] along two significant dimensions. First, we account for scenarios where the factor(s) may be weak. Second, we allow for a correlation between the target variable and the predictors, even after adjusting for common factor(s), which arises from the correlation between the idiosyncratic components of the covariates and the prediction target.

Our primary theoretical contribution is to establish the consistency of the Three-Pass Regression Filter in estimating target-relevant factors under these broader assumptions. We show that these factors can be consistently estimated even when weak, though at a slower rate than strong factors. The convergence rate for the target variable, coefficients, and factors improves with the strength of relevant factors but decreases with stronger irrelevant factors.

Our second contribution is methodological: we introduce a Lasso step, creating the 3PRF-Lasso estimator, to model the dependence of the target on the idiosyncratic components of the predictors. We derive the rate at which the average prediction from this step converges to zero, taking into consideration the impact of generated regressors.

Extensive simulations show that while 3PRF remains competitive under these modified assumptions, its advantage over other methods diminishes in some scenarios. Adding the Lasso step significantly improves performance, even when only a few idiosyncratic components are correlated with the target variable.

In our empirical application, we use the 3PRF-Lasso framework to forecast GDP, Employment, and private fixed investment from the FRED-QD dataset. The method demonstrates strong performance compared to alternatives across all three variables in both the short run and the longer horizon.

*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

Keywords: Forecasting, high dimension, supervision, three pass regression filter, factor model, 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 and applications thereof: high-dimensional covariance estimation and forecasting. This paper delves into the latter domain. The literature on forecasting with factor models is extensive. Some prominent papers include Ludvigson & Ng [2016], which highlights their effectiveness in financial forecasting; Stock & Watson [2002] and Stock & Watson [2003], which demonstrate their importance in macroeconomic prediction; Ciccarelli & Mojon [2010], which uses them to predict global inflation, and Andreou *et al.* [2013], which applies them to forecast U.S. real GDP growth, 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] corroborates these findings by showing that factor-augmented models generally exhibit superior predictive performance compared to a wide array of machine learning methods.

To introduce the framework for forecasting with factor models, consider a scenario where we have a large number of predictors organized in a vector, $\boldsymbol{x}_t$, and 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:

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

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

In developing the framework for forecasting with factor models, certain foundational assumptions are typically made. Specifically, the forecast error η_{t+h} is assumed to have a conditional expectation of zero with respect to the information available in period t, the idiosyncratic components, denoted by $\boldsymbol{\varepsilon}$, are allowed only weak serial and cross-sectional correlation and the latent factors are assumed to be strong; i.e., $\frac{\boldsymbol{\Phi}'\boldsymbol{\Phi}}{N}$ is assumed to have a non-zero limit.

This formulation has three notable drawbacks. First, some factors driving the predictors may be irrelevant for forecasting the target variable. Unsupervised factor estimation methods, such as Principal Components (Bai [2003], Stock & Watson [2002]), estimate all factors in $\boldsymbol{X}$ and use them to predict $\boldsymbol{y}$, which can introduce inefficiencies by including irrelevant factors. Second, in equation 1.1, $\beta_0 + \boldsymbol{\beta}'\boldsymbol{F}_t$ provides an optimal forecast only if $\mathbb{E}(\eta_{t+h} \mid \boldsymbol{\varepsilon}_t) = 0$. However, when $\boldsymbol{X}$ is high-dimensional, it is unlikely that none of the idiosyncratic components $(\boldsymbol{\varepsilon}_i \mid i \in \{1, \ldots, N\})$ are correlated with the target. If even a small subset of these components is correlated with the target, the forecast constructed using equation 1.2 is sub-optimal, as all the predictive information in $\boldsymbol{X}$ is not being utilized to forecast $\boldsymbol{y}$. Third, the assumption of strong factors may not always be valid, as evidenced by various studies, including Bailey *et al.* [2021] and Freyaldenhoven [2022]. When factors are weak—due to their local nature or weak loadings—the properties of estimation procedures, such as Principal Components, are adversely affected. For PC-factors, when all factors have the same weak strength, additional conditions are required to ensure consistent forecasts, as discussed in Bai & Ng [2023]. When factors have varying strengths, the implications on forecasting are unknown. For other estimators like Kelly & Pruitt [2015], the specific conditions needed to ensure consistency under weak factors are not known.

The first limitation was addressed by Kelly & Pruitt [2015], who introduced the Three-Pass Regression Filter (3PRF), a supervised method that uses 'proxy' variables, denoted by $\boldsymbol{Z}$, to estimate factors relevant to the target variable $\boldsymbol{y}$. Their approach stems from the recognition that the factors influencing the target may form a subset of the factors driving the predictors $\boldsymbol{X}$. This filter estimates only the factors that are relevant to the target, thereby yielding more efficient forecasts. Building on their work, our study presents a generalized framework that relaxes the aforementioned remaining assumptions and introduces an additional step involving Lasso to capture the predictive content of idiosyncratic components. We call this augmented estimator as 3PRF-Lasso.

Few papers in the literature have attempted to address the aforementioned second limitation—namely, the assumption that idiosyncratic components are orthogonal to the target variable—by trying to harness the predictive content of these idiosyncratic components. Notable examples include Fan *et al.* [2023a], Kneip & Sarda [2011], and Fan *et al.* [2023b]. In these works, the factor-based forecasting model (1.1) is augmented by incorporating a small subset of idiosyncratic elements as additional predictors. This is achieved through regularized

M-estimation techniques that select the relevant idiosyncratic components and estimate their coefficients, thereby combining two orthogonal approaches for high-dimensional estimation: sparse modeling and dense factor modeling.

Augmenting a PC-based factor model to address this limitation is relatively straightforward because the factor estimation process is not influenced by the data-generating process of $\boldsymbol{y}$ due to the unsupervised nature of principal component-based factor estimation. However, in the 3PRF framework, allowing idiosyncratic elements to possess predictive content for the target $\boldsymbol{y}$, and thus for $\boldsymbol{Z}$[1], introduces a form of 'corruption' in the supervision process, as clarified in Section 2. Essentially, estimating factors as principal components ensures that the factor estimation procedure remains unaffected by the correlation between idiosyncratic components and the target, as neither the target nor the proxies mimicking the target are used in the factor estimation. However, within the 3PRF framework, where proxies are used for supervision, correlation between idiosyncratic components and the target can potentially undermine the advantages of supervision. 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, the extraction process becomes less precise, as it retrieves information that is unrelated to the factors but derived from the correlation between the proxies and the idiosyncratic components. We specify the assumptions in Section 3 to ensure that this 'corruption' does not adversely affect the asymptotic rates of convergence.

Allowing idiosyncratic components to be correlated with the target variable is realistic in many economic forecasting scenarios, especially when the number of predictors is large. Fan *et al.* [2023b] provides numerous examples illustrating this. To further motivate this, consider the study by Kelly & Pruitt [2015], which explores the forecastability of macroeconomic aggregates using a large set of predictors -specifically, 108 macroeconomic variables compiled by Stock & Watson [2012]. They demonstrate the effectiveness of the Three-Pass Regression Filter by assessing its forecasting performance, using various variables from the dataset as targets and the remaining variables as predictors. The predictors are assumed to follow an approximate factor structure, as introduced by Chamberlain & Rothschild [1983], accommodating weak serial and temporal correlations in the idiosyncratic components. Although Kelly & Pruitt [2015] do not claim that target variables are inherently distinct from other predictors, their methodology im-

[1] $\boldsymbol{Z}$ 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.

plicitly assumes that the idiosyncratic components[2] of target variables are martingale difference sequences, independent of other idiosyncratic components. This assumption implies a strict factor structure for the target variables rather than an approximate one, which is restrictive. Our empirical analysis reveals that idiosyncratic components can indeed exhibit substantial predictive power for certain macroeconomic variables, as evidenced by the enhanced performance of our 3PRF-Lasso method compared to the traditional 3PRF approach.

The third limitation, i.e., the strong factor assumption, has also been addressed in recent literature on PCA factors; see Bai & Ng [2023] and Freyaldenhoven [2022]. The primary focus of their papers is to examine the implications of a weak factor structure on PC estimation of factors rather than on forecasting. Bai & Ng [2023] includes a discussion on the implications of using PC-estimated factors under a weak factor setting for prediction when all factors have the same strength.

We extend the theory of the Three-Pass Regression Filter (3PRF) to include settings where predictors follow a weak factor structure. Additionally, our framework allows target-relevant factors to have a different strength compared to target-irrelevant factors[3]. Our theoretical results provide bounds on how weak the target-relevant factors can be. The lack of sufficient factor strength necessitates a larger sample size T for consistent estimation compared to the case when all factors are strong. Furthermore, if irrelevant factors are too strong, the convergence rate is severely reduced, and beyond a certain limit $\psi_g > 2\psi_f$[4], we encounter inconsistency. For instance, if irrelevant factors have a strength of 1, this implies a bound of $\frac{1}{2}$[5] on the strength of relevant factors.

Idiosyncratic elements can be viewed as factors with zero strength; they may exhibit local correlations but lack a factor structure. Extremely weak factors with near-zero strength can be treated as part of the idiosyncratic components. Estimating such factors is costly, as their low strength makes accurate estimation challenging or infeasible. A practical approach is to use predictors that load heavily on these weak factors as a proxies[6]. This is akin to regression with

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

[3] Assuming all irrelevant factors to have the same strength is primarily for notational convenience. When irrelevant factors have varying strengths, the theoretical results can be reformulated in terms of the strength of the strongest irrelevant factor.

[4] ψ_g is the strength of irrelevant factors and ψ_f is the strength of relevant factors, as clarified in Assumption 2.

[5] Factors with $\psi < \frac{1}{2}$ are considered very weak in the literature and not useful from the standpoint of economic theory; see Freyaldenhoven [2022] for further discussion.

[6] The term 'proxy' here is used in its more conventional sense, referring to a measurable variable that is used in place of an unobservable predictor, such as an omitted weak but relevant factor. This differs from the use of term 'proxies' in the 3PRF framework, which are specifically used to estimate the relevant latent factors.

measurement error, where these predictors serve as imperfectly measured weak factors. The second-stage Lasso step can then select such predictors, thereby indirectly capturing the impact of very weak (near-zero strength) factors on the target variable by selecting variables that load heavily on the omitted weak factors.

Finally, the existing literature on Principal Component (PC)-based factors typically addresses the issues of weak factors and idiosyncratic dependence in isolation without assessing their combined impact. In the context of the Three-Pass Regression Filter (3PRF), we extend the theoretical framework to account for both phenomena simultaneously. Specifically, we derive the rate at which the regularization parameter in the second-stage Lasso should approach zero asymptotically, with this rate depending on the strength of the factors. The convergence rate of the Lasso step is influenced by the rate at which the Lasso regularization parameter approaches 0, and hence, we provide a more comprehensive understanding of how both weak factors and idiosyncratic dependence interact and affect the model's performance.

Our simulation results show that under these aforementioned general assumptions, the performance of the Three-Pass Regression Filter (3PRF) is close to that of its closest competitor, Principal Component Regression (PCR). Although the performance is somewhat subdued due to idiosyncratic dependence and weak factors, it remains comparable. The augmented regression method, 3PRF-Lasso, which incorporates both the 3PRF first-stage regression and the second-stage Lasso involving idiosyncratic components, outperforms both 3PRF and PCR. In many scenarios, it also surpasses PCR augmented with a Lasso step. When N is large relative to T, this out performance is almost universal, across various factor strengths and serial / cross-sectional correlations in factors and idiosyncratic components. When factor strength is uniform across relevant and irrelevant factors, weakness in factors diminishes the performance of both our estimator and competing methods compared to a scenario with strong factors. When relevant factors are stronger than irrelevant factors, 3PRF-Lasso performs remarkably well. However, when relevant factors are weak compared to irrelevant factors, both 3PRF and 3PRF-Lasso perrfom poorly. This highlights a cautionary aspect of using supervised factor models: while they offer many advantages, as discussed, they may perform worse than unsupervised methods when irrelevant factors are relatively stronger.

Our empirical application demonstrates the effectiveness of the 3PRF-Lasso procedure. We predict three macroeconomic variables—GDP, Employment, and Investment—using a large pool of macroeconomic indicators. The 3PRF-Lasso not only outperforms both 3PRF and

PCR, emphasizing the importance of idiosyncratic components, but also surpasses PCR-Lasso in the short run for all variables and in the long run for two of them: Investment and GDP.

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

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 $S \subset \{1, \dots, N\}$ 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. 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 $O_p(\min(a,b))$. $\max_i \boldsymbol{A}_i$ denotes the element-wise maximum of matrices $\{\boldsymbol{A}_i\}_{i\in\{1,\ldots,N\}}$.

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 three-pass filter by Kelly & Pruitt [2015]. The three-pass regression filter is essentially a sequence of linear regressions aimed at consolidating information from a large set of predictors in a small set of factor(s). Once we obtain the target relevant factor(s) from the first stage, we regress each $\mathbf{x}_i$ 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.

3PRF-Lasso Procedure

Stage-1	
Pass	Description
1.	Run time series regression of $\mathbf{x}_i$ on $\boldsymbol{Z}$ for $i = 1, \ldots, N$, $x_{i,t} = \tilde{\phi}_{0,i} + \boldsymbol{z}_t'\tilde{\boldsymbol{\phi}}_i + \hat{v}_{1it}$, retain slope estimate $\tilde{\boldsymbol{\phi}}_i$.
2.	Run cross-section regression of $\boldsymbol{x}_t$ on $\tilde{\boldsymbol{\phi}}_i$ for $t = 1, \ldots, T$, $x_{i,t} = \tilde{\phi}_{0,t} + \tilde{\boldsymbol{\phi}}_i'\hat{\boldsymbol{F}}_t + \hat{v}_{2it}$, retain slope estimate $\hat{\boldsymbol{F}}_t$.
3.	Run time series regression of y_{t+h} on predictive factors $\hat{\boldsymbol{F}}_t$, $y_{t+h} = \hat{\beta}_0 + \hat{\boldsymbol{F}}_t'\hat{\boldsymbol{\beta}} + \hat{u}_{t+h}$, 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 initial forecast initial forecast $\hat{y}_{t+h,f}$.
Stage-2	
Pass	Description
1.	Run time series regression of $\mathbf{x}_i$ on $\boldsymbol{F}$ for $i = 1, \ldots, N$, $x_{i,t} = \hat{\phi}_{0,i} + \hat{\boldsymbol{F}}_t'\hat{\boldsymbol{\phi}}_i + \hat{\varepsilon}_{it}$, retain the residual $\hat{\varepsilon}_{it}$
2.	Run Lasso regression of $\hat{u}_{t+h}$ obtained from Pass 3 in stage 1 on the estimated residuals $\hat{\epsilon}_{it}$, $\hat{u}_{t+h} = \hat{\boldsymbol{\varepsilon}}_t'\hat{\boldsymbol{\gamma}} + \hat{\eta}_{t+h}$.
	The final forecast is given by $\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\boldsymbol{F}}_t'\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\varepsilon}}_t'\hat{\boldsymbol{\gamma}}$

The final forecast is a sum of stage-1 and stage-2 forecasts. We can rewrite the forecast as

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

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.

Auto-Proxy Algorithm

0. Initialize $\boldsymbol{r}_0 = \boldsymbol{y}$. For $k = 1, \dots, L$ (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 the k3PRF 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] mention that "Fluctuations in the latent factors cause the cross section of predictors to fan out and compress over time. First-stage 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 $\boldsymbol{X}$. In the framework of Kelly & Pruitt [2015], $\boldsymbol{\zeta} = \mathbf{0}$, and it is assumed that $\boldsymbol{\omega}$ is uncorrelated with $\boldsymbol{\varepsilon}$, which is highly restrictive. The set of proxies typically is a small subset of the predictors or is created using the target variable (see Auto-Proxy Procedure). The assumption made by Kelly & Pruitt [2015] is akin to assuming a strict, rather than the usual 'approximate', factor structure among the proxies. Such a formulation lacks substantiation unless there is reason to believe that the target and the proxies are characteristically disparate from other predictors, which is typically not the case in empirical applications.

When this assumption fails, we encounter what we refer to as the 'corrupted supervisor' problem; the supervisor is imperfect in the sense that it cannot estimate all the loadings consistently in the first pass of the three-pass procedure. To illustrate this, consider a simplified case where there is 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 the three-pass procedure 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 we ensure that pass-1 functions properly, "Second-stage cross-section regressions use the estimated mapping in Pass-1 to back out estimates of the factors at each point in time" as in Kelly & Pruitt [2015], enabling consistent estimation in Stage-1 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 the rotation of underlying factors, with the first rotation 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. For any vector $\boldsymbol{x}$, let $x(m)$ denote the m^{th} element of $\boldsymbol{x}$*

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

$$\boldsymbol{X} = \boldsymbol{\iota}_T\boldsymbol{\phi}_0' + \boldsymbol{F}\boldsymbol{\Phi}' + \boldsymbol{\varepsilon} \qquad \boldsymbol{y} = \boldsymbol{\iota}_T\beta_0 + \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\boldsymbol{\gamma} + \boldsymbol{\eta} \qquad \boldsymbol{Z} = \boldsymbol{\iota}_T\boldsymbol{\lambda}_0' + \boldsymbol{F}\boldsymbol{\Lambda}' + \boldsymbol{\varepsilon}\boldsymbol{\zeta}' + \boldsymbol{\omega}$$

where $\boldsymbol{F}_t = \left(\boldsymbol{f}_t', \boldsymbol{g}_t'\right)', \boldsymbol{\Phi} = \left(\boldsymbol{\Phi}_f, \boldsymbol{\Phi}_g\right), \boldsymbol{\Lambda} = \left(\boldsymbol{\Lambda}_f, \boldsymbol{\Lambda}_g\right)$, and $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_f', \mathbf{0}'\right)'$ with $\left|\boldsymbol{\beta}_f\right| > \mathbf{0}$. $K_f > 0$ 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 $K = K_f + K_g$. $\forall j \in \{i | \gamma_i \neq 0\}$, $\beta(m) = 0 \implies \phi_j(m) = 0$

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, the target is allowed 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, like the target, are also driven by a subset of 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 $\forall j \in \{i|\boldsymbol{\gamma}_i^* \neq 0\}$ we have $\boldsymbol{\phi}_j(m) = 0$. We are ruling out the pathological cases where both these aforementioned quantities are not 0 but the sum $\boldsymbol{\beta}^*(m) + \sum_i \boldsymbol{\phi}_i(m)\gamma_i^* = 0$. This assumption is succinctly expressed as $\forall j \in \{i|\boldsymbol{\gamma}_i \neq 0\}$, $\boldsymbol{\beta}(m) = 0 \implies \phi_j(m) = 0$. This assumption allows us to consistently recover the true idiosyncratic terms in stage-2 Pass 1 for the relevant $\mathbf{x}_i$ (i.e., $\{\mathbf{x}_i|\gamma_i \neq 0\}$) and subsequently implement pass-2 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* $v = f, g$, $N^{-\psi_v}\sum_{j=1}^N \boldsymbol{\phi}_{vj} \xrightarrow[N\to\infty]{p} \overline{\boldsymbol{\phi}_{\boldsymbol{v}}} < \infty$, $N^{\psi_v/2}\left(\dfrac{\boldsymbol{\Phi_v}'\boldsymbol{J}_N\boldsymbol{\Phi}_v}{N^{\psi_v}} - \mathcal{P}_v\right) = \boldsymbol{O}_p(1)$ *and for* $\psi_s = \min\left(\psi_f, \psi_g\right)$, $N^{\psi_s/2}\left(\dfrac{\boldsymbol{\Phi_f}'\boldsymbol{J}_N\boldsymbol{\Phi}_g}{N^{\psi_s}} - \mathcal{P}_{fg}\right) = \boldsymbol{O}_p(1)$
3. $\mathbb{E}\left(\varepsilon_{it}\right) = 0, \mathbb{E}\left\|\varepsilon_{it}\right\|^8 \leq M$
4. $\mathbb{E}\left(\boldsymbol{\omega}_t\right) = \boldsymbol{0}, \mathbb{E}\left\|\boldsymbol{\omega}_t\right\|^4 \leq M, T^{-1/2}\sum_{s=1}^T \boldsymbol{\omega}_s = \boldsymbol{O}_p(1)$ *and* $T^{1/2}\left(\dfrac{\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\omega}}{T} - \boldsymbol{\Delta}_\omega\right) = \boldsymbol{O}_p(1)$

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* $h > 0$.

Assumptions 2.1-2.3 and 2.5, with $\psi_f = \psi_g = 1$, are standard in the literature, as seen in Stock & Watson [2002], Bai & Ng [2006], and Kelly & Pruitt [2015], and characterize a 'strong factor' model. However, the 'strong factor' assumption is often unrealistic in practice, as noted by Bailey *et al.* [2021], who provide a framework for estimating ψ for different factors. Their work shows that even the strongest factor in large macroeconomic datasets has a strength of $\psi < 1$. Recent research has extended the study of factor strength to scenarios where it is strictly less than one, particularly in the context of Principal Component (PC) factors, as explored in Freyaldenhoven [2022] and Bai & Ng [2023]. In our framework, we allow for $0 < \psi_f \leq 1$ and $0 < \psi_g \leq 1$, with ψ_f and ψ_g potentially differing. This generalizes the Three-Pass Regression Filter (3PRF) approach of Kelly & Pruitt [2015] by incorporating weak factors.

Assumption 2.4 was introduced in Kelly & Pruitt [2015] to ensure that the proxy noise is well-behaved. The fact that conditional expectation of η_{t+h} w.r.t information set in time t is 0 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).*

Let $x(m)$ *denote the* m^{th} *element of* $\boldsymbol{x}$. *For* $M < \infty$ *and any* i, j, t, s, m_1, m_2 *and* $v = f, g$

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

[7] If weakness in loadings is induced by sparsity, i.e., the factor(s) are local, then we can use assumption 3.3 to prove 3.4 by slightly modifying the argument in Kelly & Pruitt [2015] Lemma 1.1 and 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.

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

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

8. $\mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}F_t\left(m\right)\eta_{t+h}\right|^2 \leq M$

Assumption 3.1-2 allow weak cross-sectional and temporal dependence in the idiosyncratic components; these 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], Bai [2003] and Stock & Watson [2002] etc.

Assumption 4. *(Uncorrelated loadings and Factors).*

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

2. $\mathbf{\Delta}_F \equiv \begin{pmatrix} \Delta_f & \Delta_{fg} \\ \Delta_{fg} & \Delta_g \end{pmatrix}$ *is positive definite and* $\Delta_{fg} = 0$.

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.[8]

Assumption 5. *(Relevant Proxies).*

1. $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$

2. $\mathbf{\Lambda}_f$ *is non-singular.*

[8] Normalization Assumption 5 in Kelly & Pruitt [2015].

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

Assumption 6. *For $v = f, g$, define*

$$\Gamma_{N_v,T} \equiv \min(\sqrt{N^{\psi_v}}, \sqrt{T}), \quad \delta_{N,T} \equiv \min(\sqrt{N}, \sqrt{T}).$$

We need the following:

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

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

This assumption 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, $\psi_f = \psi_g = \psi$, (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 $\psi_f \geq \psi_g$, (2) is automatically satisfied. Conversely, when $\psi_f < \psi_g$, (2) reflects the cost of having higher noise (irrelevant factors) relative to the signal (relevant factors). Furthermore, (2) imposes a limit on the weakness of the relevant factors compared to the irrelevant factors, 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 first stage does not adversely affect the theoretical results in second stage Lasso regression. Such assumptions are prevalent in the literature. Specifically, 1 is 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[9]. Similarly, Giglio *et al.* [2023] incorporates 2 with $r_1 = 1/2$. The partial sums (after centering) in 4–6 (without taking the maximum over i) are bounded by Assumptions 3.2, 3.3, and 3.6. We assume that the maximum of these empirical processes scales as some power of the logarithm of N. This assumption is equivalent to imposing a uniform tail bound (see remark 2) on the distribution of these partial sums.

Remark 2. *The scaling properties outlined in 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 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 second-stage Lasso as stated in Theorem 5 to hold.

Assumption 8. *(Weak cross sectional dependence)*
For each i, let $\Delta_{i,\varepsilon} \equiv \{j \,|\, \mathbb{E}\left|T^{-1/2}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{jt}\right|^2 \leq M \leq \infty\}$. Then $N - |\Delta_{i,\varepsilon}| \leq M_2 \leq \infty$.

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

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

Assumption 9. *(Active set and mimicking proxies)*
Let $S \equiv \{i|\gamma_i \neq 0\}$. $|S| \leq M \leq \infty$.
$\gamma_i = 0$ *iff* $\boldsymbol{\zeta}_i' = 0$*, where* $\boldsymbol{\zeta}_i'$ *denotes the* i^{th} *row of matrix* ζ'.

We require the set of 'relevant idiosyncratic' terms to be finite. If this set were allowed to grow in size, we would need to adjust the rates in various ensuing theorems. The dependence of the target (and thereby proxies) on idiosyncratic terms leads to noisier estimation in stage 1 of the 3PRF-Lasso Procedure. If many idiosyncratic terms have explanatory power for $\boldsymbol{y}$, the extent of corruption in the supervising step is greater. The assumption that $\gamma_i = 0 \implies \boldsymbol{\zeta}_i' = 0$ can be relaxed. We only require that $|\{i|\boldsymbol{\zeta}_i' \neq 0\}| \leq \infty$ 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-0 small coefficients, see He [2023]. In such cases, as shown in their paper, ridge regression is asymptotically efficient in capturing both factor and idiosyncratic information among the entire class of spectral regularized estimators. Our model, unlike their paper, assumes that the dependence of the target on idiosyncratic terms is sparse, more akin to the setting in Fan* et al. *[2020] and Kneip & Sarda [2011].*

Assumption 10. *(For Stage- 2 Lasso Regression)*
*(**a**) With* $r_1, \dots r_5$ *defined in Assumption 7,*

$$\text{(i)}\quad \text{For } r \in \{r_2, \dots, r_5\},\ \frac{(\log N)^r}{\sqrt{T}} = O(1)$$

$$\text{(ii)}\quad \frac{N^{1-\psi_f}}{\sqrt{T}}\left[(\log N)^{r_1} + (\log T)^{r_1}\right] = O(1)$$

*(**b**)* $\exists$ *a large enough constant* $\kappa > 0$ *s.t* $\forall i \in \{1, \dots, N\}$*, and* $\forall\, T$*, we have,*

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

*(**c**)(Compatibility condition) Define* $\Delta_{\varepsilon,g} := \left(\boldsymbol{\varepsilon} + \boldsymbol{g}\boldsymbol{\Phi}_g'\right)' \left(\boldsymbol{\varepsilon} + \boldsymbol{g}\boldsymbol{\Phi}_g'\right)/T$*. For the* $N \times N$ *matrix* $\Delta_{\varepsilon,g}$*, we say that 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*

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

We assume that, w.p approaching one, the compatibility condition holds for set S *defined in Assumption 9 and the associated compatibility constant is* ν_0*.*

Assumption 10(a) 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 (b) is a common assumption in lasso literature, as seen in Bühlmann & Van De Geer [2011]. This assumption essentially states that the relevant idiosyncratic components are not highly correlated.

4 Theoretical Results

We present each of the Theorems 1-4 in two parts, labeled as parts a and b. The first part accommodates the idea of weak factor(s), while the subsequent 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 relevant 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.

$$\textbf{Define } \Xi_{NT}^{-1} \equiv \max\left(T^{-1/2}, N^{-\psi_f/2}, \frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right) = \max\left(T^{-1/2}, N^{-\psi_f/2}, N^{-\psi_f+\psi_g/2}, \frac{N^{\psi_g-\psi_f}}{\sqrt{T}}\right)$$

The equality follows by substituting the value of Γ_{N_gT} defined in 6.

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

$O(1)$, then we have,

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

(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

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

Theorem 1 (a) specifies the rate of convergence of first stage 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 $\delta_{N,T}^{-1}$ ($\delta_{N,T} \equiv \min(\sqrt{N}, \sqrt{T})$).

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 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 $\boldsymbol{y} - \mathbb{E}(y_{t+h}|\boldsymbol{F}_t)$ has a conditional expectation of 0 w.r.t information set in period t.

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 $\boldsymbol{\zeta} = 0$. If $\mathcal{P}_f = \mathbb{I}$, Then for any i,

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

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 uniformly weak 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).

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,

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

$\boldsymbol{G}_\beta$ is defined in the appendix

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.

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,*

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

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

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

$\boldsymbol{H}_f$ *and* $\boldsymbol{H}_0$ *are defined in the Appendix and* $\boldsymbol{H}_f{}'\boldsymbol{G}_\beta = \boldsymbol{I}$

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.

Theorem 4 (a) and (b) establish the convergence and the corresponding rate of the estimated factor(s) to a rotation of the true relevant factor(s). 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 Stage-2 Pass 1 regression be given by* $\lambda := 2\dfrac{\sqrt{c + \kappa \log N}}{\Xi_{NT}}$, $c > 0$ *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* [10] *that*

$$\frac{1}{T}\left\|\hat{y}_{t+h}-\mathbb{E}_t y_{t+h}\right\|_2 = O_p\left(\frac{\sqrt{\log N}}{\Xi_{NT}}\right)$$

Theorem 5 establishes the rate at which the average prediction error of 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.

Discussion

The rate of convergence for our estimators in **Theorems** 1, 3, and 4 merits closer attention. Below, we present special cases that illustrate the impact of each of our assumptions:

- **When** $\psi_f = \psi_g = 1$:
  - The rate is $\delta_{NT}^{-1} = \left\{\min\left(\sqrt{N}, \sqrt{T}\right)\right\}^{-1}$
  - Our first two passes involve time series and cross-section regressions, respectively. The rate of convergence in cross-sectional regressions (pass 2) depends on N, while in time series regressions (pass 1), it depends on T.
  - Thus, overall, the convergence rate is determined by which of these dimensions diverges to infinity more slowly.
- **When** $\psi_f = \psi_g = \psi \neq 1$:
  - The rate is $\left\{\min\left(\sqrt{N^{\psi}}, \sqrt{T}\right)\right\}^{-1}$.
  - Essentially, the rate in cross-sectional regression (pass 2) slows down from $\sqrt{N}$ to $\sqrt{N^{\psi}}$, leading to an overall reduction in rate.

[10] under conditions laid out in **Theorem** 5 about the regularization parameter λ

- If T is small, $\sqrt{N^{\psi}}$ would dominate T, and the weakness in factors would not be a significant issue since the impediment to faster convergence in such a case is the small sample size, not the weakness of the factors.
- However, when T is large, the weakness in factors becomes detrimental and worsens the estimator's performance compared to the strong factor case.

- **When** $\psi = \psi_f > \psi_g$:
  - In this scenario, the relevant factors are stronger than the irrelevant factors, and the rate is same as in the case of $\psi_f = \psi_g = \psi$.
  - The variance of our estimators is lower compared to the case above. Although we do not derive an explicit expression for the variance, our proofs indicate that all terms involving weak factor loadings approach zero asymptotically in this case. Thus, the variance is primarily determined by the interaction terms between factors and noise variables, excluding the interaction terms involving weak factor loadings.

- **When** $\psi_g > \psi_f$:
  - This is the worst-case scenario, and our convergence rate is $\min\left(N^{-\psi_f+\frac{\psi_g}{2}}, \frac{N^{\psi_g-\psi_f}}{\sqrt{T}}\right)$
  - The rate in pass 2 is hindered by two factors: weak relevant factors and irrelevant factors that are relatively stronger compared to the relevant ones.

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 a $N(0,1)$ 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 $\boldsymbol{u}_{g,t} \sim \text{IIN}(0, \boldsymbol{\Sigma}_g)$, $u_{f,t}$ and $\boldsymbol{u}_{g,t}$ are uncorrelated and $\boldsymbol{\Sigma}_g$ is an identity matrix of order k, which is either 4 or 5. 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 drawn from an independent and identically distributed (iid) standard normal distribution, divided by $N^{1-\psi_f}$. Similarly, for each predictor, loadings on the irrelevant factors are drawn independently from a standard normal distribution, divided 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 predictors in active set, i.e. $\{i|\gamma_i \neq 0\}$, we make $\boldsymbol{\phi}_g = \mathbf{0}$, 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 auto-proxy($\boldsymbol{y}$).

We compare the out-of-sample performance of four methods: Principal Component Regression (as described in Stock & Watson [2002]), Three-Pass Regression Filter by Kelly & Pruitt [2015], 3PRF-Lasso (our method), and PCR-Lasso. 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 residuals 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 generated using an unsupervised technique, i.e. Principal Components. 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 $R^2(1)$ displays the average in-sample R-square value (across repeated samples) derived from the infeasible regression of $\boldsymbol{y}$ on $\boldsymbol{f}$. Similarly, the column labeled $R^2(2)$ presents the average in-sample R-square value obtained from the infeasible regression of $\boldsymbol{y}$ on $\boldsymbol{f}$ and the idiosyncratic elements within the active set, i.e. $\{\boldsymbol{\varepsilon}_i|\gamma_i \neq 0\}$. The last four columns report the average (across repeated samples) out-of-sample R-square values for the 4 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. All regression parameters, factors, and loadings are estimated using the in-sample data. Since out-of-sample factors are unknown, for each OOS period t_o, we regress $\boldsymbol{x}_{t_o}$ on the in-sample estimated factor loadings to obtain estimated OOS factor(s) for period t_o. Therefore, no out-of-sample information is utilized in the estimation process.

The simulation results in table 2-5 show that, for a given strength of irrelevant factors, performance of all methods improve 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 is consistent with expectations, as the 3PRF is a supervised method and thus more sensitive to variations explained by relevant factors compared to alternative approaches. The opposite holds true when the strength of irrelevant factors decreases relative to relevant factors.

When comparing PCR and 3PRF, we note that PCR generally outperforms 3PRF, albeit slightly, in the majority of cases. This finding is at variance with the simulations presented in Kelly & Pruitt [2015]. This discrepancy arises because permitting idiosyncratic elements to have predictive content for the target hampers the supervision process as explained in section 2; the target, which serves as an auto-proxy in our simulations has a diminished signal from the factor(s) compared to cases where all predictive content resides within the factor(s) and idiosyncratic elements have no predictive power for the target, as in Kelly & Pruitt [2015].

The simulations also reveal that in the majority of cases, 3PRF and PCR are outperformed by either 3PRF-Lasso or PCR-Lasso. This outcome is expected, as the former two techniques do not leverage the predictive power of idiosyncratic elements. When comparing 3PRF-Lasso with PCR-Lasso, we observe that neither method consistently outperforms the other. When we set the size of the cross-section equal to the training sample size, i.e., $N = T = 100$, both 3PRF-Lasso and PCR-Lasso show similar performance, regardless of whether all factors are weak or strong, as long as there is no difference in their relative strength. 3PRF-Lasso performs slightly better when $k_g = 5$, while PCR-Lasso performs slightly better when $k_g = 4$ as can be seen by comparing tables 2 and 6[11]. Tables 3 and 4 show that when relevant factors are relatively weak, 3PRF-Lasso has poorer performance, but when relevant factors are relatively strong, 3PRF-Lasso performs better. In general, there is no clear winner between these two

[11] The results in Appendix A examine the effect of increasing irrelevant factors with $T = 100$, $N = 100$, and $\psi_g = \psi_f = 1$. Additional simulations in the online appendix show that this finding is consistent across different sample sizes and factor strengths.

methods when the T and N size are identical.

However, when the training sample size $T = 100$ is half of the cross-sectional size $N = 200$, 3PRF-Lasso has a much superior performance. It outperforms other methods in almost all cases as seen by comparing tables 2 and 7[12]. Its superiority relative to PCR and 3PRF is understandable, as mentioned earlier since the latter two do not leverage the predictive information in idiosyncratic elements. However, the superior performance compared to PCR-Lasso is interesting. As observed in online Appendix tables Tables 18-21, this can be attributed to the high false positive rates in the case of PCR-Lasso. We provide a possible explanation below.

It is noteworthy that, with regards to the estimation of idiosyncratic elements in the active set, 3PRF-Lasso and PCR-Lasso yield very similar results[13]. However, with regard to the idiosyncratic elements not in the active set, their estimated value from our procedure differs markedly from their estimates in the PCR setup. For the case of PCR, it is known that $\hat{\varepsilon}_{it}^{PC} - \varepsilon_{it} = o_p(1)$, as shown in Proposition 5 of Bai & Ng [2023]. For our procedure, Lemma 9 in the online appendix shows that $\hat{\varepsilon}_{it} - \left(\varepsilon_{it} + \boldsymbol{g}_t'\boldsymbol{\phi}_{ig} - \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig}\right) = O_p(\Xi_{NT}^{-1})$. According to Assumption 1, for variables in the active set, $\boldsymbol{\phi}_{ig} = 0$, leading to an outcome similar to PCR, i.e. $\forall i \in S$, $\hat{\varepsilon}_{it}^{3prf} - \varepsilon_{it} = O_p(\Xi_{NT}^{-1}) = o_p(1)$. However, for variables not in the active set, as observed from lemma 9 of online Appendix, the convergent limit of $\hat{\varepsilon}_{it}^{3prf}$ differs from $\hat{\varepsilon}_{it}^{PC}$. The estimator of idiosyncratic elements in our setup is not consistent for predictors outside the active set. This inconsistency, however, is not detrimental since these predictors have a coefficient of 0.

Our setup naturally leads us to a situation where the estimated idiosyncratic elements in the active set are either orthogonal or weakly correlated, while the estimated idiosyncratic elements outside the active set are highly correlated due to the presence of irrelevant factors. As noted in Zou & Hastie [2005] and Wang *et al.* [2011], "*When the model includes several highly correlated variables, all of which are related to some extent to the response variable, lasso tends to pick only one or a few of them and shrink the rest to 0.*". This suggests that high correlation dampens the cardinality of the set of non-zero coefficients in Lasso. When this high correlation is between

[12] In Appendix A, we report results for $N = 200$ and $T = 100$ for all four methods with $k_g = 4$ and $\psi_f = \psi_g = 1$. Additional simulations in the online appendix confirm that these results hold across various factor strengths and numbers of irrelevant factors.

[13] We do not report these results in our simulations, but this can be easily checked by regressing estimated idiosyncratic elements (in the active set) from PCA and stage-2 3PRF on each other. We found that it was very close to 1

the relevant predictors and the irrelevant predictors (as in Fan *et al.* [2020]) or amongst relevant predictors, this is an undesirable feature, but when the correlation is high amongst the irrelevant predictors only, this could be a desirable feature since it suppresses the false positive rate. This rationalizes the result in Tables 18-21 of the Online Appendix. We corroborate our finding with a small experiment[14] discussed in the following paragraph.

We generate a vector $\boldsymbol{x}_{105\times 1}$ from a multivariate normal distribution. The first five elements of $\boldsymbol{x}$ are uncorrelated with others and amongst themselves. The rest of the variables in $\boldsymbol{x}$ have a covariance matrix with diagonal elements being 1 and off-diagonal elements being ρ. We generate 100 independent samples of the target $\boldsymbol{y}$ as follows: $y_{t+1} = \sum_{i=1}^{5} \beta x_{it} + u_{t+1}$. u_{t+1} is generated from a standard normal distribution for each t and is serially uncorrelated. We vary the signal (β) and the correlation between the irrelevant predictors (ρ). This experiment is repeated 100 times and we note the average false and true positive rates. Results are reported in Table 15 of the Online appendix. The false positive rate considerably decreases when we increase ρ (for all signal levels β). The true positive rate is relatively high, across ρ, even for lower signal levels, but at the lowest signal level, i.e., $\beta = 0.1$, we observe a jump in true positive rates as we increase ρ.

Therefore, the correlation between irrelevant predictors appears to be beneficial as it suppresses spurious selections into the active set. This also explains the improved performance of our estimator 3PRF-Lasso compared to PCR-Lasso when we increase the number of irrelevant predictors from 4 to 5.

Section 9 presents 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, as well as the simulation results for the stylized example in the final paragraph of this section, are also reported in the online appendix.

6 Empirical Application

For our empirical analysis, we evaluate the forecastability of three key US macroeconomic aggregates: Gross Domestic Product , Private Fixed Investment, and Total Non-farm employment. These variables, along with their predictors, are sourced from the FRED-QD dataset published

[14]The results are given in Online appendix

in 2023. The target variables, denoted as 'GDPC1', 'FPIx', and 'PAYEMS', respectively, in the FRED-QD dataset are transformed using the method by Hamilton & Xi [2024] to address their 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 Principal Component Analysis (PCA). 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:

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

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 encompasses one quarter (h=1) and one year (h=4).

The determination of the number of (PC) factors 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 the second stage lasso regression in 3PRF-Lasso and PCR-Lasso, we utilize a 10-fold cross-validation technique to estimate parameters within each training sample as done in Fan *et al.* [2020].The results are given in Table 1,

Our results show that 3PRF-Lasso performs best in the short horizon across all variables and outperforms all alternatives in the long horizon for two variables. Additionally, incorporating a second-stage Lasso regression following PCR or 3PRF substantially boosts the R-squared values compared to either PCR or 3PRF. This suggests that harnessing idiosyncratic information yields significant benefits for the variables under consideration.

7 Conclusion

Our paper generalizes the framework of Kelly & Pruitt [2015]. Our theoretical contribution is twofold: accommodating weak factor(s) and potentially informative idiosyncratic elements within the data-generating processes of the Target and proxies. We demonstrate how the weakness of factors induces a slower convergence rate of our estimator. Incorporating idiosyncratic information necessitates expanding the assumptions within the model, however it does not incur any costs on the convergence rates, i.e. we delineate a set of sufficient assumptions that maintain the convergence rates derived in Kelly & Pruitt [2015] when factors are strong, but idiosyncratic components are present in the DGP of proxies.

Our methodological contribution involves incorporating a second-stage lasso regression into 3PRF, thereby integrating idiosyncratic predictive content. Our empirical analysis illustrates that integrating idiosyncratic information indeed leads to significant improvements in the predictability of certain macroeconomic variables.

8 Empirical results

One Quarter ahead Forecast, $R^2(\%)$				
Target Variable	PCR	3PRF	3PRF-Lasso	PCR-Lasso
GDP	3.27	1.47	**17.47**	12.79
Investment	4.53	3.93	**16.20**	12.55
Employment	5.05	2.05	**19.54**	13.76
One Year ahead Forecast, $R^2(\%)$				
GDP	12.20	10.30	**68.44**	67.18
Investment	18.43	23.12	**83.12**	78.30
Employment	14.90	10.17	57.97	**61.58**

Table 1: Forecasting performance of different models

9 Simulation Results

Table 2: 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	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.391	0.609	0.363	0.310	0.464	**0.476**
0.3	0.9	0.3	0	0.401	0.619	0.344	0.311	**0.450**	0.442
0.3	0.9	0.3	1	0.362	0.647	0.301	0.261	0.460	**0.505**
0.3	0.9	0.9	0	0.387	0.621	0.317	0.249	**0.445**	0.434
0.3	0.9	0.9	1	0.360	0.642	0.285	0.266	0.455	**0.516**
0.9	0.3	0.3	0	0.396	0.614	0.4	0.322	0.486	**0.507**
0.9	0.3	0.3	1	0.359	0.642	0.344	0.263	0.463	**0.545**
0.9	0.3	0.9	0	0.404	0.624	0.418	0.317	**0.461**	0.423
0.9	0.3	0.9	1	0.371	0.65	0.328	0.288	0.463	**0.544**

Notes: K_f, K_g , ρ_f ρ_g, a , d ψ_f, ψ_g, $R^2(1)$ and $R^2(2)$ are defined in section 5. PCR-X (X=5 here) denotes the regression of y on first 'X' Principal components. 3PRF-Y, (Y = 1 here) denotes the auto-proxy 3PRF with $L = Y$ auto-proxies. 3PRF+L is 3PRF-Lasso procedure where first stage 3PRF uses $L = Y$ proxies. PCR+L is analogously PCR with X PCs augmented with second stage lasso step.

Table 3: 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	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.395	0.622	0.373	0.385	**0.492**	0.471
0.3	0.9	0.3	0	0.386	0.617	0.349	0.364	**0.489**	0.454
0.3	0.9	0.3	1	0.360	0.643	0.343	0.333	**0.518**	0.505
0.3	0.9	0.9	0	0.402	0.629	0.370	0.384	**0.497**	0.385
0.3	0.9	0.9	1	0.365	0.647	0.353	0.340	**0.524**	0.521
0.9	0.3	0.3	0	0.396	0.616	0.404	0.420	**0.516**	0.494
0.9	0.3	0.3	1	0.365	0.647	0.391	0.383	0.540	**0.553**
0.9	0.3	0.9	0	0.393	0.618	0.409	0.425	**0.491**	0.408
0.9	0.3	0.9	1	0.355	0.643	0.377	0.366	**0.527**	**0.527**

Notes: See Table 2

Table 4: 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	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.406	0.630	0.276	-0.246	0.008	**0.366**
0.3	0.9	0.3	0	0.397	0.621	0.217	-0.266	-0.0024	**0.257**
0.3	0.9	0.3	1	0.358	0.64	-0.0034	-0.236	-0.1126	**0.166**
0.3	0.9	0.9	0	0.383	0.61	0.1	-0.201	0.024	**0.176**
0.3	0.9	0.9	1	0.354	0.648	0.058	-0.284	-0.152	**0.191**
0.9	0.3	0.3	0	0.389	0.625	0.262	-0.145	0.154	**0.482**
0.9	0.3	0.3	1	0.354	0.645	0.022	-0.216	-0.058	**0.225**
0.9	0.3	0.9	0	0.39	0.616	0.173	-0.208	0.106	**0.385**
0.9	0.3	0.9	1	0.359	0.646	-0.001	-0.226	-0.072	**0.17**

Notes: See Table 2

Table 5: 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	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.394	0.616	0.283	0.235	**0.298**	**0.298**
0.3	0.9	0.3	0	0.386	0.622	0.276	0.262	**0.343**	0.340
0.3	0.9	0.3	1	0.361	0.643	0.0461	0.157	0.17	**0.213**
0.3	0.9	0.9	0	0.388	0.618	0.265	0.241	0.318	**0.356**
0.3	0.9	0.9	1	0.365	0.648	0.052	0.140	0.149	**0.225**
0.9	0.3	0.3	0	0.393	0.623	0.326	0.275	0.331	**0.398**
0.9	0.3	0.3	1	0.359	0.65	0.053	0.135	0.140	**0.178**
0.9	0.3	0.9	0	0.392	0.614	0.257	0.257	0.324	**0.354**
0.9	0.3	0.9	1	0.352	0.644	0.058	0.162	**0.169**	0.125

Notes: See Table 2

Table 6: 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	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.397	0.619	0.349	0.297	**0.467**	0.386
0.3	0.9	0.3	0	0.381	0.61	0.294	0.252	**0.425**	0.322
0.3	0.9	0.3	1	0.358	0.646	0.269	0.248	0.459	**0.510**
0.3	0.9	0.9	0	0.397	0.625	0.291	0.277	**0.473**	0.333
0.3	0.9	0.9	1	0.351	0.642	0.249	0.229	0.444	**0.464**
0.9	0.3	0.3	0	0.383	0.611	0.374	0.286	**0.467**	0.432
0.9	0.3	0.3	1	0.365	0.643	0.382	0.272	**0.459**	0.438
0.9	0.3	0.9	0	0.394	0.625	0.379	0.273	**0.478**	0.452
0.9	0.3	0.9	1	0.361	0.651	0.327	0.233	0.437	**0.547**

Notes: See Table 2

Table 7: 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	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.381	0.613	0.342	0.301	**0.448**	0.221
0.3	0.9	0.3	0	0.389	0.618	0.302	0.285	**0.441**	0.245
0.3	0.9	0.3	1	0.356	0.649	0.273	0.256	**0.483**	0.405
0.3	0.9	0.9	0	0.393	0.62	0.334	0.307	**0.46**	0.27
0.3	0.9	0.9	1	0.36	0.648	0.297	0.267	**0.478**	0.373
0.9	0.3	0.3	0	0.386	0.615	0.395	0.309	**0.454**	0.278
0.9	0.3	0.3	1	0.361	0.648	0.377	0.306	**0.498**	0.492
0.9	0.3	0.9	0	0.397	0.610	0.422	0.341	**0.464**	0.342
0.9	0.3	0.9	1	0.354	0.641	0.357	0.263	0.475	**0.486**

Notes: See Table 2

Online Appendix to “Forecasting economic time series using supervised factors and idiosyncratic elements”

A Mathematical Proofs

The proofs in the appendix are divided into two parts. The first part accommodates the case of weak factors but no idiosyncratic effect, while the other part considers the full generalized model.

Results when $\gamma = \zeta = 0$ and $0 < \psi_f \leq 1$, $0 < \psi_g \leq 1$

Lemma 1. *Under Assumptions 1-3, 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. $T^{-1/2}\sum_t F_t(m)\eta_{t+h} = O_p(1)$ $\quad T^{-1/2}\sum_t \omega_t(m)\eta_{t+h} = O_p(1)$
6. $N^{-1/2}T^{-1/2}\sum_{i,t}\varepsilon_{it}\eta_{t+h} = O_p(1)$
7. $N^{-\psi_v/2}T^{-1}\sum_{i,t}\phi_{iv}(m_1)\,\varepsilon_{it}F_t(m_2) = O_p(1)$
8. $N^{-\psi_v/2}T^{-1}\sum_{i,t}\phi_{iv}(m_1)\,\varepsilon_{it}\omega_t(m_2) = O_p(1)$
9. $N^{-\psi_v/2}T^{-1/2}\sum_{i,t}\phi_{iv}(m)\varepsilon_{it}\eta_{t+h} = O_p(1)$
10. (a) $N^{-1}T^{-1/2}\sum_{i,s}\varepsilon_{is}\varepsilon_{it} = O_p\left(\delta_{NT}^{-1}\right)$ *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, $\mathbb{E}\left|T^{-1/2}\sum_t F_t(m)\eta_{t+h}\right|^2=T^{-1}\sum_t\mathbb{E}\left[\eta_{t+h}^2\right]\mathbb{E}\left[F_t(m)^2\right]\leq T^{-1}\sum_t\delta_\eta M=0(1)$ by Assumption 2.1, 2.5. Therefore, $\sum_t F_t(m)\eta_{t+h}=O_p(1)$.

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 η_{t+h} is independent of $\phi_{iv}(m)$ and $\varepsilon_{i,t}$ for any $h>0$ by Assumption 2.5, we have

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

. Since $\mathbb{E}\left[\eta_{t+1}\eta_{s+1}\right]=0$ for $t\neq s$, this is in turn equal to

$$T^{-1}\sum_{t}\mathbb{E}\left[\eta_{t+h}^{2}\right]\mathbb{E}\left[\left(N^{-\psi_v/2}\sum_{i}\phi_{iv}(m)\varepsilon_{it}\right)^{2}\right]$$

which is $O_p(1)$ from Assumptions 3.7 and 2.5.

Item 10(b): $N^{-1/2}T^{-1}\sum_{i,t}\left[\varepsilon_{it}\varepsilon_{jt}-\sigma_{ij,tt}\right]+N^{-1/2}T^{-1}\sum_{i,t}\sigma_{ij,tt}=O_p\left(T^{-1/2}\right)+O_p\left(N^{-1/2}\right)$ by Assumption 3.2(b) and 3.1(d).

Item 16 : We have,

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

16.I $=O_p(T^{-1/2})$ by Assumption 3.4.

16.II $=O_p(N^{-\psi_v/2})$ 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) 1-3, 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. $N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}=\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\vee\frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$
6. $N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}=\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\vee\frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$
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. $N^{-1}T^{-1/2}\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}_t = \boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$

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,

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

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

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

$\mathrm{I} = \boldsymbol{O}_p(1)$ by Assumption 3.7.

II : Since $N^{-1/2}\sum_i \boldsymbol{\varepsilon}_{it} = O_p(1)$ 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, $N^{-\psi_f}\sum_i \boldsymbol{\phi}_{if}(m) = O_p(1)$. Hence, $(\mathrm{I}+\mathrm{II})$ is $\boldsymbol{O}_p(1)$.

Therefore, $N^{-\psi_f}\boldsymbol{\varepsilon}_t'\boldsymbol{J}_N\boldsymbol{\Phi}_f = N^{-\psi_f/2}\boldsymbol{O}_p(1) = \boldsymbol{O}_p\left(N^{-\psi_f/2}\right)$

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

Hence, the whole matrix,

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

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 $\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1} \vee \frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)$. Hence 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

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

$5.\mathrm{I} = O_p\left(N^{-\psi_f/2}\right)$ by Lemma 1.7.

$5.\mathrm{II} = O_p\left(T^{-1/2}\right)$ since $N^{-\psi_f}\sum_i \phi_{if}(m_1) = O_p(1)$ by Assumption 2.2 and $N^{-1}\sum_j\left(T^{-1/2}\sum_t F_t(m_2)\varepsilon_{jt}\right) = O_p(1)$ by Assumption 3.6 .

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

$5.\mathrm{IV} = O_p\left(T^{-1/2}N^{-1/2}\right)$ 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 $\boldsymbol{O}_p\left(\Gamma^{-1}_{N_fT}\right)$.

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 $\boldsymbol{O}_p\left(\Gamma^{-1}_{N_fT}\right)$.

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

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

$6.\mathrm{I} = O_p\left(N^{-\psi_f/2}\right)$ by Lemma 1.8.

$6.\mathrm{II} = O_p\left(T^{-1/2}\right)$ 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 .

$6.\mathrm{III} = O_p\left(N^{-\psi_f/2}T^{-1/2}\right)$ 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

$6.\mathrm{IV} = O_p\left(T^{-1}N^{-1/2}\right)$ 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 $\boldsymbol{O}_p\left(\Gamma_{N_fT}^{-1}\right)$

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

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

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

7.II can be written as

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

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

7.III can be written as

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

Hence, this term is $O_p\left(N^{-1/2}\right)$. by Assumption 2.2 and Lemma 1.6.

7.IV can be written as

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

Hence, this term is $O_p\left(N^{-1/2}\times T^{-1/2}\right)$ by 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. *(This lemma ensures the consistency of our coefficient estimate $\hat{\boldsymbol{\beta}}$). Under Assumptions 1-3, we have,*

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

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.

Item 1 is $K \times K$ matrix given as,

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

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 $\boldsymbol{O}_p\left(N^{-\psi_f/2}\Gamma_{N_fT}^{-1}\right)$.

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

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

1.I can be written as

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

by Assumption 3.7.

1.II can be written as

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

by lemma 1.16 and Assumption 2.2

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

1.IV can be written as

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

by Assumption 2.2 and Lemma 1.10

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

1.VI can be written as

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

by Assumption 2.2, 3.7 and Lemma 1.10.

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

1.VIII can be written as

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

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 $\boldsymbol{O}_p\left(N^{-\psi_f/2} \Gamma_{N_f T}^{-1} \right)$.

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 $\boldsymbol{O}_p\left(N^{-\psi_g/2} \Gamma_{N_g T}^{-1} \right)$.

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 $\boldsymbol{O}_p\left(\frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2} \Gamma_{N_g T}} \right)$.

It is also easy to see from the proof presented above, that, depending on whether ψ_f is greater, $=$ 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 $\boldsymbol{O}_p\left(N^{-\psi_f/2} \Gamma_{N_f T}^{-1} \right)$ (when $\psi_f > \psi_g$) or $\boldsymbol{O}_p\left(\frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2} \Gamma_{N_g T}} \right)$ (when $\psi_f < \psi_g$). When $\psi_f = \psi_g$, the two rates are equal.

Hence concluding from the discussion above, the matrix $N^{-2\psi_f} T^{-1} \boldsymbol{\Phi}' \boldsymbol{J}_N \boldsymbol{\varepsilon}' \boldsymbol{J}_T \boldsymbol{\varepsilon} \boldsymbol{J}_N \boldsymbol{\Phi}$ is $\boldsymbol{O}_p\left(N^{-\psi_f/2} \Gamma_{N_f T}^{-1} \vee \frac{N^{2(\psi_g-\psi_f)}}{N^{\psi_g/2} \Gamma_{N_g T}} \right)$

Item 2 is given by

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

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 $\boldsymbol{O}_p\left(N^{-\psi_f/2} \Gamma_{N_f T}^{-1} \right)$

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

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

2.I $- \cdots -$ 2.XVI.

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

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

by Lemma 1.12 and Assumption 3.7.

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

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

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

2.III can be written as

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

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

2.IV can be written as

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

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

2.V can be written as

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

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

2.VI can be written as

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

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

2.VII can be written as

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

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

2.VIII can be written as

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

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

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

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

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

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

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

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

2.XI can be written as

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

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

2.XII can be written as

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

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

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

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

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

2.XIV can be written as

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

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

2.XV

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

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

2.XVI

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

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

Since $0<\psi_f\leq 1$, the initial terms dominate the order. Hence, summing all these terms gives us that $N^{-2\psi_f}T^{-2}\boldsymbol{\Phi}_f'\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F}=\boldsymbol{O}_p(T^{-\frac{1}{2}}N^{-\frac{3\psi_f}{2}+1}\delta_{NT}^{-1})$.

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

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

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

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 $\frac{N^{1-\psi_f}}{\sqrt{T}}=O(1)$*,*

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

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

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

and

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

Proof:

First we note that

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

The final equality follows from Assumption 2.2 and 4.1.

Item 1 :

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

where 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 $\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_f & \mathbf{0} \end{bmatrix}$ by Assumption 5.

Item 2:

$$
\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} \\
&= \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 A.1.

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

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

Where the final equality again follows from Assumption 2.1, 4.

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

Item 3:

$$\begin{aligned}
\hat{\boldsymbol{F}}_{C,t} &= N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{x}_t \\
&= \boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\phi_{\boldsymbol{0}}\right)+\boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t+
\end{aligned}$$

$$\boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right)+\boldsymbol{\Lambda}\left(N^{-\psi_f}T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\phi_0\right)+\boldsymbol{\Lambda}\left(N^{-\psi_f}T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t+$$

$$\frac{N^{1-\psi_f}}{\sqrt{T}}\boldsymbol{\Lambda}\left(N^{-1}T^{-1/2}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right)+\left(T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\phi_{\boldsymbol{0}}\right)+\left(T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t$$

$$+\left(T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right)+\left(N^{-1}T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\phi_{\boldsymbol{0}}\right)+\left(N^{-\psi_f}T^{-1}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t$$

$$+\frac{N^{1-\psi_f}}{\sqrt{T}}\left(N^{-1}T^{-1/2}\boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{\varepsilon}\boldsymbol{J}_N\boldsymbol{\varepsilon}_t\right)$$

$$=N^{-\psi_f}T^{-1}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\phi_0+\boldsymbol{\Lambda}\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\right)\left(N^{-\psi_f}\boldsymbol{\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi}\right)\boldsymbol{F}_t+\boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)$$

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

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

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

Where 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,

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

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

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,

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

satisfies

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

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 \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F\Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\varepsilon' J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F} \Lambda' + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F} \Lambda' + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \omega' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \omega' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F} \Lambda' + \boldsymbol{\omega' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\varepsilon' J}_T \boldsymbol{\omega} \\
&+ \Lambda \boldsymbol{F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{F} \Lambda' + \boldsymbol{\Lambda F' J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Phi' J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\omega} \\
&+ \boldsymbol{\Lambda F' J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \boldsymbol{\varepsilon J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{F \Lambda'} + \boldsymbol{\Lambda F' J}_T \varepsilon \boldsymbol{J}_N \varepsilon' \boldsymbol{J}_T \varepsilon \boldsymbol{J}_N \boldsymbol{\Phi F' J}_T \boldsymbol{\omega}
\end{aligned}
$$

$$
\begin{aligned}
&+\boldsymbol{\Lambda F}'\boldsymbol{J}_T\boldsymbol{\varepsilon J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\Lambda F}'\boldsymbol{J}_T\boldsymbol{\varepsilon J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\varepsilon J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\Lambda F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\omega\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\omega'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F}\Lambda' + \omega'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F}\Lambda' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F}\Lambda' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{\omega}\\
&+\boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Lambda}' + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\omega})\\
&= (N^{2\psi_f}T^3)^{-1}\boldsymbol{\Lambda F}'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\Phi}'\boldsymbol{J}_N\boldsymbol{\Phi F}'\boldsymbol{J}_T\boldsymbol{F\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 A.1. Further, note that

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

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

Similarly expanding $\hat{\boldsymbol{\beta}}_4$, we get

$$
\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\varepsilon'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} \\
&+ \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\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\varepsilon'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\omega}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\Phi}'\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} \\
&+ \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{F}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{\eta} \\
&+ \omega'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} + \omega'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\boldsymbol{\Phi}\boldsymbol{F}\boldsymbol{J}_T\boldsymbol{\eta} + \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\varepsilon'\boldsymbol{J}_T\boldsymbol{F}\boldsymbol{\beta} \\
&+ \boldsymbol{\omega}'\boldsymbol{J}_T\varepsilon\boldsymbol{J}_N\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 A.1.

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

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

Where the final equality follows from Assumption 2.1, 4 and Lemma 2 and equation A.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.

Theorems when $\gamma = \zeta = 0$

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

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

The third equality follows since, for any invertible matrices $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{B}$, $(\boldsymbol{A} + \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1} \boldsymbol{B} (\boldsymbol{A} + \boldsymbol{B})^{-1}$ which implies that, $\left(\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f + \boldsymbol{O}_p(\Xi_{NT}^{-1})\right)^{-1} =$ $\left(\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f\right)^{-1} - \left(\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f\right)^{-1} \boldsymbol{O}_p(\Xi_{NT}^{-1}) \left(\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f + \boldsymbol{O}_p(\Xi_{NT}^{-1}))\right)^{-1}$ $= \left(\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f\right)^{-1} - \boldsymbol{O}_p(\Xi_{NT}^{-1}) \boldsymbol{O}_p(1) = \left(\boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\Lambda}'_f\right)^{-1} - \boldsymbol{O}_p(\Xi_{NT}^{-1})$.
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

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

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

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

Under Assumption 4 and 5 and 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 $\dfrac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$, then we have,

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

Proof:

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

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:

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

We use the 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 6. *The proofs of **Theorem** 1, 2, and 4 can be approached in an alternative manner. We can demonstrate that, for the matrix $\boldsymbol{H}_2 = \boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{\Phi}_f$, $\hat{\boldsymbol{F}}_t$ 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 $\boldsymbol{H}\boldsymbol{F}_t$, 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. The columns of $\boldsymbol{H}\boldsymbol{F}_t$ 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.

When $\delta \neq 0$ and $\gamma \neq 0$

. We need to introduce an additional Lemma which shall be employed for subsequent proofs under this general setting of $\delta \neq 0$ and $\gamma \neq 0$.

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

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

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

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

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

by Assumptions 2.3 and 3.2. Therefore we have that,

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

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

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

$1.\text{I}=O_p\left(\delta_{NT}^{-1}\right)$ by equation A.2

$1.\text{II}=N^{-1/2}\left(N^{-1}\sum_{i\in\Delta_{m,\varepsilon}}\left(T^{-1/2}\sum_s\varepsilon_s(m)\varepsilon_{is}\right)+\dfrac{T^{1/2}}{N}\sum_{i\in\Delta_{m,\varepsilon}^c}\left(T^{-1}\sum_s\varepsilon_s(m)\varepsilon_{is}\right)\right)\left(N^{-1/2}\sum_j\varepsilon_{jt}\right)$
$=N^{-1/2}O_p(1)=O_p\left(N^{-1/2}\right)$ by Assumption 8 and Lemma 1.3.

$1.\text{III}=O_p\left(\delta_{NT}^{-1}T^{-1/2}\right)$ by Lemma 1.3 and 1.10.

$1.\text{IV}=O_p\left(N^{-1}T^{-1/2}\right)$ by Lemma 1.3.

Summing these terms, Item 1 is $\boldsymbol{O}_p\left(\delta_{NT}^{-1}\right)$.

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

element

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

2.I is given by

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

by Assumption 8 and assuming that $\dfrac{T}{N}=O(1)$

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

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

2.V is given by

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

$=O_p\left(T^{-1/2}\right)$ by Assumption 8 and assuming that $\dfrac{T^{1/2}}{N}=O(1)$

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 is given by

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

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

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 A.2.

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

3.V is given by

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

$= O_p\left(T^{-1/2}\right)$ by Assumption 3.6 and assuming that $\dfrac{T^{1/2}}{N} = O(1)$

3.VI is given by

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

by Assumption 3.6 and Lemma 1.3.

3.VII is given by

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

$=O_p\left(N^{-1/2}T^{-1}\right)$ 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 $\boldsymbol{O}_p\left(\Gamma_{N_fT}\vee\frac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right)=\boldsymbol{O}_p\left(\Xi_{NT}\right)$.

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

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

4.I $= O_p\left(\Gamma_{N_fT}^{-1}\right)$ by Lemma 1.16.

4.II $= N^{\psi_f/2}T^{-1/2}\delta_{NT}^{-1}\left(N^{-1}T^{-1/2}\sum_{j,t}\varepsilon_t\left(m_2\right)\varepsilon_{jt}\right)\left(N^{-\psi_f}\phi_{if}\left(m_1\right)\right)$

$= O_p\left(N^{\psi_f/2}T^{-1/2}\delta_{NT}^{-1}\right)$ if $\dfrac{N^{\psi_f}}{T} = O(1)$ then this is $= O_p\left(\Gamma_{N_fT}^{-1}\right)$ by Assumption 2.2 and Lemma 1.10.

4.III $= T^{-1/2}\left(T^{-1}\sum_t\left(\sum_j N^{-\psi_f/2}\phi_{jf}\left(m_1\right)\varepsilon_{jt}\right)\right)\left(T^{-1/2}\sum_s\varepsilon_s\left(m_2\right)\right)$

$= O_p\left(T^{-1/2}\right)$ by Assumption 3.7 and Lemma 1.3.

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

$= O_p\left(N^{\frac{\psi_f-1}{2}}T^{-1}\right)$ 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 $\boldsymbol{O}_p\left(\Gamma_{N_fT}\vee\dfrac{N^{\psi_g-\psi_f}}{\Gamma_{N_gT}}\right) = \boldsymbol{O}_p\left(\Xi_{NT}\right)$.

Item 7:

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

We now establish the probability of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{F}}_t$ under the general setting of $\delta\neq 0$ and $\gamma\neq 0$

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

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 $\hat{\boldsymbol{F}}_t$, under Assumption 6 is

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

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.

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

Proof: 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}
&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 A.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* $\dfrac{N^{1-\psi_f}}{\sqrt{T}} = O(1)$

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,

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

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 $\dfrac{\hat{\boldsymbol{F}}_C\boldsymbol{J}_T\hat{\boldsymbol{F}}_C'}{T}$ 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 $plim\left(\dfrac{\hat{\boldsymbol{F}}_C\boldsymbol{J}_T\hat{\boldsymbol{F}}_C'}{T}\right) = \boldsymbol{\Lambda}_{\boldsymbol{f}}\boldsymbol{\Delta}_f\mathcal{P}_f\left(T^{-1}\boldsymbol{f}\boldsymbol{J}_T\boldsymbol{f}\right)\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_{\boldsymbol{f}}'$ which is equal to $\boldsymbol{\Lambda}_{\boldsymbol{f}}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_{\boldsymbol{f}}'$ given Assumption 2.1. Using Lemma 6 and the expression for $\hat{\boldsymbol{F}}_{C,t}$ in Lemma 7 we can establish that $N^{-2\psi_f}T^{-3}\boldsymbol{Z}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{X}\boldsymbol{J}_N\boldsymbol{X}'\boldsymbol{J}_T\boldsymbol{Z} = \dfrac{\hat{\boldsymbol{F}}_C\boldsymbol{J}_T\hat{\boldsymbol{F}}_C'}{T} = \boldsymbol{\Lambda}_{\boldsymbol{f}}\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{\Delta}_f\boldsymbol{\Lambda}_{\boldsymbol{f}}' + \boldsymbol{O}_p(\Xi_{NT}^{-1})$.

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

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}
&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.,

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

Theorem 1(b) Let Assumptions 1-9 hold, $\dfrac{N^{1-\psi}}{\sqrt{T}} = O(1)$ and $\dfrac{T}{N} = O(1)$, 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, $T^{-1/2}\sum_t \varepsilon_{it} = O_p(1)$, 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 **Theorem**3(b) and **Theorem**4(b) respectively, follow similarly given the rates derived in Lemma 4, 5, 7 and 8.

We now proceed to prove Theorem 5. We introduce 2 Lemmas which shall be used in the proof that follows

Second stage Lasso Results

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 **Theorem**1(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}
\end{aligned}
$$

$$
\begin{aligned}
&\therefore \hat{\varepsilon}_{it} = x_{it} - \hat{x}_{it} = \left(\phi_{0,i} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} + \boldsymbol{f}'_t\boldsymbol{\phi}_{if} + \varepsilon_{it}\right) - \left(\phi_{0,i} + \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig} + \boldsymbol{f}'_t\boldsymbol{\phi}_{if} + \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig}\right) \\
&\Longrightarrow \hat{\varepsilon}_{it} - \left(\varepsilon_{it} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} - \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig}\right) = \boldsymbol{O}_p(\Xi_{NT}^{-1}) + \boldsymbol{O}_p\left(\Xi_{NT}^{-1}\right)\boldsymbol{\phi}_{ig} \\
&\Longrightarrow \hat{\varepsilon}_{it} - \left(\varepsilon_{it} + \boldsymbol{g}'_t\boldsymbol{\phi}_{ig} - \bar{\boldsymbol{g}}'\boldsymbol{\phi}_{ig}\right) = \boldsymbol{O}_p(\Xi_{NT}^{-1})
\end{aligned}
$$

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 $\dfrac{\sum_{s=1}^{T}\varepsilon_{is}}{T} = O_p(T^{-1/2})$ 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 $\mathrm{I} = \max_i\left(\dfrac{\Xi_{NT}}{T}\right)\boldsymbol{\varepsilon}'_i\boldsymbol{\eta} + O_p(1)$ and II, III and IV are $O_p(1)$.

Item I:

$$\max_i \left(\frac{\Xi_{NT}}{T}\right)(\boldsymbol{\varepsilon}_i + \boldsymbol{J}_T \boldsymbol{g}\boldsymbol{\phi}_{ig})' \boldsymbol{\eta} = \left(\frac{\Xi_{NT}}{T}\right)\max_i \boldsymbol{\varepsilon}_i'\boldsymbol{\eta} + \left(\frac{\Xi_{NT}}{T}\right)\left(\max_i \boldsymbol{\phi}_{ig}\right)\boldsymbol{g}'\boldsymbol{J}_T\boldsymbol{\eta}$$
$$= \max_i \left(\frac{\Xi_{NT}}{T}\right)\boldsymbol{\varepsilon}_i'\boldsymbol{\eta} + O_p(1) \tag{A.3}$$

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

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

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

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 $(\hat{\boldsymbol{y}}_f - \boldsymbol{y}_f)' = \boldsymbol{O}_p(\Xi_{NT}^{-1})\boldsymbol{F}'\boldsymbol{J}_T + \boldsymbol{O}_p(1)\left(\hat{\boldsymbol{F}}_C - \boldsymbol{\Lambda}_f\boldsymbol{\Delta}_f\mathcal{P}_f\boldsymbol{f}'\right)\boldsymbol{J}_T \quad + \boldsymbol{\gamma}_s'(\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_s)' + (\iota_T\bar{\boldsymbol{\eta}})'$. Therefore,

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

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

$$\Xi_{NT}\boldsymbol{O}_p(\Xi_{NT}^{-1})\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) = \boldsymbol{O}_p(1)\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) \quad = \boldsymbol{O}_p(1)\left(T^{-1}\boldsymbol{F}'\boldsymbol{\varepsilon}_i\right) - \boldsymbol{O}_p(1)\left(T^{-1}\boldsymbol{F}'\boldsymbol{\iota}_T\bar{\boldsymbol{\varepsilon}}_i\right)$$

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 $\|\frac{1}{T}\boldsymbol{O}_p(1)\boldsymbol{F}'\|_\infty$ 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

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

Where 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}&\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. $\boldsymbol{O}_p(1)\left(T^{-1}\boldsymbol{F}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right)$ where the $O_p(1)$ term is a finite dimensional matrix.
2. $\dfrac{N^{1-\psi_f}}{\sqrt{T}}\dfrac{\boldsymbol{D}\boldsymbol{\varepsilon}_i}{T}$ 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. $\Xi_{NT}\boldsymbol{O}_p(1)\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_f\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right) + \Xi_{NT}\boldsymbol{O}_p(T^{-1/2})\left(N^{-\psi_f}T^{-1}\boldsymbol{\Phi}'_g\boldsymbol{J}_N\boldsymbol{\varepsilon}'\boldsymbol{J}_T\boldsymbol{\varepsilon}_i\right)$[15], where the $\boldsymbol{O}_p(1)$ and $\boldsymbol{O}_p(T^{-1/2})$ terms are finite dimensional matrices invariant across t and i.

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.

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 $\|\frac{\boldsymbol{D}}{T}\|_\infty$ 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 $\dfrac{N^{1-\psi_f}}{\sqrt{T}}\left[(\log N)^{r_1} + (\log T)^{r_1}\right]$ 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 $m\in\{1,\ldots,K_f\}$, 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

[15]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 $\boldsymbol{O}_p(T^{-1/2})$.

by

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

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. $c=\boldsymbol{O}_p\left(\frac{(logN)^{r_2}}{T}\right)$ 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

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

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

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

$\mathcal{B}_{1i} \ldots \mathcal{B}_{4i}$ 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,

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

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 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.

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

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 $\max_i (N^{-\psi_f} T^{-2} \boldsymbol{W'}_{XZ} \boldsymbol{S}_{X\boldsymbol{x}_i} - \boldsymbol{\Lambda}_f \boldsymbol{\Delta}_f \mathcal{P}_f \boldsymbol{\Delta}_f \boldsymbol{\phi}_{if}) = O_p(1)$.

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

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($\mathcal{A}$)

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}
&\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{\eta} = \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)\\
&\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)\\
&\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)\\
&+\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)\\
&+\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)\\
&+\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)\\
&+\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) +\\
&\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)\\
&+\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)\\
&= 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 6 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,

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

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,

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

The lasso solution must satisfy

$$\| \hat{\boldsymbol{u}} - \hat{\varepsilon}\hat{\gamma}\|_2^2/T + \lambda\|\hat{\gamma}\|_1 \leq \| \hat{\boldsymbol{u}} - \hat{\varepsilon}\gamma\|_2^2/T + \lambda\|\gamma\|_1 \tag{A.4}$$

Since $\tilde{\boldsymbol{\eta}} = \hat{\boldsymbol{u}} - \hat{\varepsilon}\gamma$(Defined in Lemma 9), Equation A.4 can be written as,

$$(\hat{\boldsymbol{u}} - \hat{\varepsilon}\hat{\gamma} - \tilde{\boldsymbol{\eta}})' (\hat{\boldsymbol{u}} - \hat{\varepsilon}\hat{\gamma} + \tilde{\boldsymbol{\eta}}) + \lambda\|\hat{\gamma}\|_1 - \lambda\|\gamma\|_1 \leq 0$$

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

$$(-\hat{\varepsilon}\,(\hat{\gamma} - \gamma))' \,(-\hat{\varepsilon}\,(\hat{\gamma} - \gamma) + 2\tilde{\boldsymbol{\eta}}) + \lambda\|\hat{\gamma}\|_1 - \lambda\|\gamma\|_1 \leq 0$$

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

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

Note that,

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

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

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

has a high probability.

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

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

The third inequality follows from the fact that $\dfrac{\Xi_{NT}}{T} \le \dfrac{1}{\sqrt{T}}$, 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 $\lambda = 2\lambda_0$,

$$2\frac{\left\|\hat{\varepsilon}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)\right\|_2^2}{T}+\lambda\left\|\hat{\boldsymbol{\gamma}}_{s^c}\right\|_1 \le 3\lambda\left\|\hat{\boldsymbol{\gamma}}_s-\boldsymbol{\gamma}_s\right\|_1. \tag{A.5}$$

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

Note that the event $\mathscr{C} = \left\{\left\|\boldsymbol{\Theta}_s\right\|_1^2 \le \left(\boldsymbol{\Theta}'\Delta_{\varepsilon,g}\boldsymbol{\Theta}\right)|S|/\nu_0^2 \implies \left\|\boldsymbol{\Theta}_s\right\|_1^2 \le \left(\boldsymbol{\Theta}'\Delta_{\hat{\varepsilon},\hat{g}}\boldsymbol{\Theta}\right)|S|/\nu_0^2\right\}$ has probability 1 asymptotically. This can be shown as a direct consequence of by Lemma 9.

Using equation A.5 and Assumption 10(b), i.e, the compatibility condition for set S defined in Assumption 9, on $\mathscr{T}$, with $\lambda = 2\lambda_0$, we have

$$\left\|\hat{\varepsilon}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)\right\|_2^2/T+\lambda\left\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right\|_1 \le 4\lambda^2|S|/\nu_0^2. \tag{A.6}$$

Proof:

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

where we have used equation A.5 and the compatibility condition [16] (Assumption 10) in the

[16] In the compatibility condition we have used $\boldsymbol{\Theta} = \hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ since $\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}$ satisfies the condition $\left\|\hat{\boldsymbol{\gamma}}_{s^c}-\boldsymbol{\gamma}_{s^c}\right\|_1 =$

first inequality, noting that, $\|\hat{\gamma}_s - \gamma_s\|_1 \leq \sqrt{|S|}\,\|\hat{\gamma}_s - \gamma_s\|_2$ by Cauchy Schwartz inequality. The second inequality uses that fact that for any u, v, $4uv \leq u^2 + 4v^2$.

Concluding from the discussion above, we have that, Using the regularization parameter $\lambda = 2\lambda_0$, on the set $\mathscr{T}$, we have

$$\|\hat{\varepsilon}\,(\hat{\gamma} - \gamma)\|_2^2\,/T \leq 4\lambda^2|S|/\nu_0^2 = O_p(\lambda^2) = O_p\left(\frac{logN}{\Xi_{NT}{}^2}\right) \tag{A.7}$$

The final equality uses the fact that $|S|$ is finite. Finally, noting that,

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

where we used triangle inequality in first step and Lemma 9 in the second step, noting that for $j \in \{i|\gamma_i \neq 0\}$, $\phi_{jg} = 0$ by Assumption 1. The third step invokes equation A.7.

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

$\|\hat{\gamma}_{s^c}\|_1 \leq 3\,\|\hat{\gamma}_s - \gamma_s\|_1$ by A.5

A.1 Additional Simulation Results

Table 8: OOS R-squared across competing methods

$K_f = 1$, $K_g = 5$				$N = 100$, $T = 100$			$\psi_f = 1$, $\psi_g = 0.7$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.392	0.617	0.357	0.381	**0.486**	0.426
0.3	0.9	0.3	0	0.388	0.616	0.349	0.378	**0.482**	0.35
0.3	0.9	0.3	1	0.348	0.640	0.345	0.325	**0.512**	0.462
0.3	0.9	0.9	0	0.39	0.611	0.349	0.37	**0.466**	0.28
0.3	0.9	0.9	1	0.355	0.649	0.348	0.33	**0.521**	**0.521**
0.9	0.3	0.3	0	0.398	0.625	0.414	0.439	**0.543**	0.495
0.9	0.3	0.3	1	0.352	0.646	0.398	0.382	0.544	**0.552**
0.9	0.3	0.9	0	0.378	0.616	0.392	0.403	**0.511**	0.461
0.9	0.3	0.9	1	0.361	0.647	0.397	0.382	**0.526**	0.520

Notes: K_f, K_g , ρ_f ρ_g, a , d ψ_f, ψ_g, $R^2(1)$ and $R^2(2)$ are defined in section 5 of the paper. PCR-X (X=5 here) denotes the regression of y on first 'X' Principal components. 3PRF-Y, ($Y = 1$ here) denotes the auto-proxy 3PRF with $L = Y$ auto-proxies. 3PRF+L is 3PRF-Lasso procedure where first stage 3PRF uses $L = Y$ proxies. PCR+L is analogously PCR with X PCs augmented with second stage lasso step.

Table 9: OOS R-squared across competing methods

$K_f = 1,\ K_g = 5$				$N = 100,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 1$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.386	0.611	0.203	-0.207	0.0482	**0.339**
0.3	0.9	0.3	0	0.382	0.605	0.138	-0.2020	0.021	**0.283**
0.3	0.9	0.3	1	0.365	0.644	-0.14	-0.23	-0.093	**0.154**
0.3	0.9	0.9	0	0.393	0.614	**0.079**	-0.184	0.076	0.033
0.3	0.9	0.9	1	0.35	0.642	-0.023	-0.248	-0.138	-0.025
0.9	0.3	0.3	0	0.393	0.619	0.266	-0.173	0.146	**0.358**
0.9	0.3	0.3	1	0.363	0.647	0.046	-0.186	-0.059	**0.104**
0.9	0.3	0.9	0	0.39	0.613	0.132	-0.175	0.1186	**0.209**
0.9	0.3	0.9	1	0.362	0.643	0.04	-0.199	-0.0399	**0.152**

Notes: See Table 8

Table 10: OOS R-squared across competing methods

$K_f = 1,\ K_g = 5$				$N = 100,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 0.7$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.395	0.617	0.29	0.251	**0.32**	0.285
0.3	0.9	0.3	0	0.387	0.617	0.256	0.209	0.284	**0.3**
0.3	0.9	0.3	1	0.366	0.65	0.074	0.145	**0.159**	0.147
0.3	0.9	0.9	0	0.387	0.611	0.272	0.235	**0.302**	0.291
0.3	0.9	0.9	1	0.365	0.649	0.057	0.144	0.15	**0.196**
0.9	0.3	0.3	0	0.392	0.618	0.318	0.272	**0.325**	0.286
0.9	0.3	0.3	1	0.358	0.646	0.071	0.155	0.164	**0.223**
0.9	0.3	0.9	0	0.394	0.624	0.283	0.248	0.326	**0.349**
0.9	0.3	0.9	1	0.361	0.653	0.063	0.134	**0.144**	0.14

Notes: See Table 8

Table 11: OOS R-squared across competing methods

$K_f = 1,\ K_g = 4$				$N = 200,\ T = 100$			$\psi_f = 1,\ \psi_g = 0.7$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.386	0.616	0.351	0.373	**0.437**	0.104
0.3	0.9	0.3	0	0.396	0.62	0.358	0.382	**0.438**	0.167
0.3	0.9	0.3	1	0.365	0.645	0.353	0.353	**0.538**	0.46
0.3	0.9	0.9	0	0.393	0.62	0.364	0.381	**0.442**	0.147
0.3	0.9	0.9	1	0.367	0.655	0.349	0.335	**0.53**	0.423
0.9	0.3	0.3	0	0.39	0.616	0.398	0.412	**0.48**	0.185
0.9	0.3	0.3	1	0.356	0.644	0.371	0.357	**0.54**	0.425
0.9	0.3	0.9	0	0.386	0.61	0.393	0.407	**0.472**	0.262
0.9	0.3	0.9	1	0.365	0.652	0.37	0.37	**0.555**	0.4

Notes: See Table 8

Table 12: OOS R-squared across competing methods

$K_f = 1,\ K_g = 4$				$N = 200,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 1$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.393	0.616	**0.26**	-0.188	0.175	0.096
0.3	0.9	0.3	0	0.393	0.618	**0.225**	-0.182	0.004	0.143
0.3	0.9	0.3	1	0.362	0.648	-0.0475	-0.319	-0.154	-0.102
0.3	0.9	0.9	0	0.385	0.623	0.0253	-0.198	-0.291	**0.048**
0.3	0.9	0.9	1	0.372	0.642	-0.034	-0.333	-0.21	-0.17
0.9	0.3	0.3	0	0.384	0.613	**0.251**	-0.14	0.122	0.191
0.9	0.3	0.3	1	0.353	0.640	-0.006	-0.297	-0.144	-0.016
0.9	0.3	0.9	0	0.391	0.621	0.037	-0.204	**0.076**	0.065
0.9	0.3	0.9	1	0.358	0.637	-0.012	-0.349	-0.183	-0.053

Notes: See Table 8

Table 13: OOS R-squared across competing methods

$K_f = 1,\ K_g = 4$				$N = 200,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 0.7$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-5	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.392	0.616	0.294	0.274	**0.325**	0.165
0.3	0.9	0.3	0	0.387	0.606	0.268	0.249	**0.298**	0.071
0.3	0.9	0.3	1	0.371	0.652	0.031	0.101	**0.102**	-0.166
0.3	0.9	0.9	0	0.393	0.619	0.275	0.273	**0.306**	0.094
0.3	0.9	0.9	1	0.367	0.653	0.029	0.128	**0.129**	0.025
0.9	0.3	0.3	0	0.386	0.619	0.316	0.287	**0.323**	0.197
0.9	0.3	0.3	1	0.357	0.640	0.027	0.095	**0.11**	-0.011
0.9	0.3	0.9	0	0.386	0.618	0.242	0.283	**0.329**	0.13
0.9	0.3	0.9	1	0.368	0.657	0.0527	0.136	**0.138**	-0.11

Notes: See Table 8

Table 14: OOS R-squared across competing methods

$K_f = 1,\ K_g = 5$				$N = 200,\ T = 100$			$\psi_f = 1,\ \psi_g = 1$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.386	0.612	0.335	0.287	**0.441**	0.153
0.3	0.9	0.3	0	0.395	0.618	0.326	0.293	**0.433**	0.166
0.3	0.9	0.3	1	0.37	0.657	0.286	0.256	**0.489**	0.301
0.3	0.9	0.9	0	0.403	0.619	0.329	0.29	**0.47**	0.133
0.3	0.9	0.9	1	0.358	0.65	0.261	0.232	**0.453**	0.101
0.9	0.3	0.3	0	0.394	0.616	0.387	0.282	**0.446**	0.123
0.9	0.3	0.3	1	0.372	0.654	0.341	0.269	**0.485**	0.268
0.9	0.3	0.9	0	0.399	0.623	0.390	0.301	**0.466**	0.235
0.9	0.3	0.9	1	0.347	0.639	0.339	0.26	**0.467**	0.249

Notes: See Table 8

Table 15: OOS R-squared across competing methods

$K_f = 1, K_g = 5$				$N = 200, T = 100$			$\psi_f = 1, \psi_g = 0.7$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.391	0.615	0.36	0.381	**0.452**	0.0546
0.3	0.9	0.3	0	0.396	0.624	0.357	0.386	**0.46**	0.083
0.3	0.9	0.3	1	0.358	0.643	0.356	0.351	**0.526**	0.203
0.3	0.9	0.9	0	0.397	0.625	0.357	0.386	**0.45**	0.076
0.3	0.9	0.9	1	0.359	0.65	0.347	0.34	**0.535**	0.269
0.9	0.3	0.3	0	0.392	0.619	0.411	0.428	**0.494**	0.144
0.9	0.3	0.3	1	0.356	0.648	0.379	0.377	**0.552**	0.377
0.9	0.3	0.9	0	0.397	0.615	0.41	0.43	**0.475**	0.171
0.9	0.3	0.9	1	0.352	0.644	0.387	0.385	**0.560**	0.327

Notes: See Table 8

Table 16: OOS R-squared across competing methods

$K_f = 1, K_g = 5$				$N = 200, T = 100$			$\psi_f = 0.7, \psi_g = 1$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.381	0.618	**0.232**	-0.137	0.0985	0.054
0.3	0.9	0.3	0	0.391	0.619	**0.182**	-0.184	-0.044	-0.054
0.3	0.9	0.3	1	0.358	0.647	-0.078	-0.333	-0.186	-0.375
0.3	0.9	0.9	0	0.377	0.621	0.027	-0.142	**0.03**	-0.122
0.3	0.9	0.9	1	0.366	0.647	-0.112	-0.378	-0.2	-0.403
0.9	0.3	0.3	0	0.402	0.62	**0.265**	-0.118	0.168	0.063
0.9	0.3	0.3	1	0.353	0.647	-0.016	-0.228	-0.0581	-0.245
0.9	0.3	0.9	0	0.396	0.623	0.071	-0.135	**0.154**	0.017
0.9	0.3	0.9	1	0.363	0.65	-0.013	-0.237	-0.064	-0.351

Notes: See Table 8

Table 17: OOS R-squared across competing methods

$K_f = 1,\ K_g = 5$				$N = 200,\ T = 100$			$\psi_f = 0.7,\ \psi_g = 0.7$		
ρ_f	ρ_g	a	d	$R^2(1)$	$R^2(2)$	PCR-6	3PRF-1	3PRF+L	PCR+L
0	0	0	0	0.394	0.619	0.289	0.277	**0.324**	-0.0282
0.3	0.9	0.3	0	0.374	0.606	0.25	0.219	**0.279**	-0.09
0.3	0.9	0.3	1	0.368	0.649	0.021	0.108	**0.113**	-0.227
0.3	0.9	0.9	0	0.396	0.615	0.256	0.239	**0.294**	-0.515
0.3	0.9	0.9	1	0.358	0.647	0.037	0.122	**0.126**	-0.164
0.9	0.3	0.3	0	0.387	0.623	0.29	0.236	**0.299**	0.013
0.9	0.3	0.3	1	0.357	0.645	0.029	0.092	**0.094**	-0.263
0.9	0.3	0.9	0	0.396	0.614	0.251	0.257	**0.297**	0.010
0.9	0.3	0.9	1	0.356	0.642	0.048	0.117	**0.119**	-0.222

Notes: See Table 8

In tables 18 -21, we list parameter configurations (ρ_f, ρ_g, a, d) as one entry to save space. These configurations are in the columns π, in exactly the same order as tables 8-17.

(π)	$\psi_f = 1$, $\psi_g = 1$		$\psi_f = 1$, $\psi_g = 0.7$		$\psi_f = 0.7$, $\psi_g = 1$		$\psi_f = 0.7$, $\psi_g = 0.7$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
True and False Positive rates 3PRF-Lasso, N=100, T=100, $K_f = 4$								
1	0.9625	0.12257	0.84167	0.13576	0.87083	0.2191	0.56667	0.06875
2	0.96667	0.14948	0.7875	0.12899	0.85417	0.2092	0.5875	0.062847
3	0.70833	0.075521	0.72083	0.051042	0.20417	0.12153	0.0125	0.01441
4	0.9375	0.12986	0.85	0.12604	0.8875	0.22726	0.575	0.088542
5	0.73333	0.10156	0.7125	0.056597	0.2	0.13906	0.025	0.025521
6	0.9375	0.14601	0.83333	0.11858	0.8625	0.20486	0.52083	0.072569
7	0.67917	0.089236	0.64583	0.069965	0.2125	0.12222	0.020833	0.023785
8	0.95	0.15156	0.7875	0.10174	0.86667	0.19913	0.50417	0.069618
9	0.69583	0.091493	0.62917	0.043056	0.2375	0.11684	0.025	0.017882
True and False Positive rates 3PRF-Lasso, N=100, T=100, $K_f = 5$								
1	0.92917	0.15451	0.9	0.15087	0.85833	0.21563	0.63333	0.090799
2	0.95833	0.12969	0.85	0.14201	0.88333	0.20399	0.6	0.12344
3	0.67917	0.099479	0.7125	0.062674	0.35	0.17587	0.0083333	0.0086806
4	0.97083	0.14792	0.84167	0.11476	0.8875	0.23368	0.55417	0.096181
5	0.72083	0.10642	0.6875	0.076389	0.27917	0.14792	0.025	0.019097
6	0.9625	0.16632	0.83333	0.12396	0.90417	0.21736	0.48333	0.08125
7	0.69167	0.12153	0.6375	0.042882	0.2625	0.13906	0.016667	0.012847
8	0.94167	0.1349	0.85833	0.12639	0.8625	0.23003	0.45417	0.087153
9	0.64167	0.11198	0.65	0.049653	0.24583	0.16406	0.020833	0.014583

Table 18: TPR/FPR in stage-2 Lasso regression

True and False Positive rates 3PRF, N=200, T=100, $K_f = 4$								
π	$\psi_f = 1$, $\psi_g = 1$		$\psi_f = 1$, $\psi_g = 0.7$		$\psi_f = 0.7$, $\psi_g = 1$		$\psi_f = 0.7$, $\psi_g = 0.7$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
1	0.8875	0.092517	0.7625	0.10417	0.9	0.18129	0.475	0.047109
2	0.9375	0.10085	0.79583	0.093367	0.9125	0.14983	0.52917	0.054337
3	0.7	0.060714	0.7	0.042517	0.3	0.10264	0.0083333	0.0056122
4	0.92917	0.075765	0.79583	0.10961	0.89167	0.16905	0.55	0.075085
5	0.71667	0.060459	0.68333	0.039966	0.35417	0.10935	0.0125	0.0037415
6	0.90833	0.10791	0.77083	0.10357	0.925	0.18206	0.50417	0.066497
7	0.7	0.057568	0.6875	0.038435	0.32083	0.11386	0.0083333	0.0060374
8	0.9625	0.084864	0.80417	0.11607	0.83333	0.16233	0.57083	0.075
9	0.74167	0.054762	0.72083	0.040561	0.31667	0.10068	0.029167	0.0059524
True and False Positive rates 3PRF-Lasso, N=200, T=100, $K_f = 5$								
1	0.91667	0.077976	0.84583	0.090306	0.92917	0.18104	0.55417	0.079507
2	0.95833	0.089116	0.72917	0.090901	0.86667	0.18827	0.62917	0.092177
3	0.74167	0.07449	0.72917	0.043197	0.37083	0.12151	0.016667	0.0053571
4	0.94583	0.11105	0.79167	0.11352	0.90833	0.1881	0.50833	0.059779
5	0.7375	0.069218	0.725	0.04966	0.39583	0.11327	0.020833	0.0048469
6	0.94583	0.096429	0.80417	0.094813	0.90833	0.19524	0.41667	0.051871
7	0.75	0.071173	0.675	0.027466	0.39583	0.11786	0.0125	0.0066327
8	0.9375	0.095493	0.79167	0.11412	0.89583	0.19116	0.50417	0.07568
9	0.75	0.077636	0.72083	0.042687	0.38333	0.12151	0	0.005017

Table 19: TPR/FPR in stage-2 Lasso regression

True and False Positive rates PCR-Lasso, N=100, T=100, $K_f = 4$								
(π)	$\psi_f = 1$, $\psi_g = 1$		$\psi_f = 1$, $\psi_g = 0.7$		$\psi_f = 0.7$, $\psi_g = 1$		$\psi_f = 0.7$, $\psi_g = 0.7$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
1	0.975	0.14809	0.89583	0.16719	0.9	0.13542	0.87917	0.19826
2	0.97083	0.16788	0.86667	0.17951	0.91667	0.20503	0.92083	0.18177
3	0.8125	0.1217	0.775	0.11181	0.63333	0.24583	0.66667	0.30608
4	0.9625	0.15781	0.87917	0.18299	0.88333	0.17778	0.8875	0.21059
5	0.775	0.090451	0.75833	0.13056	0.63333	0.28715	0.67917	0.26892
6	0.95833	0.15295	0.89583	0.16545	0.89583	0.13733	0.94583	0.18646
7	0.80417	0.10712	0.7625	0.1283	0.65	0.20781	0.63333	0.2717
8	0.97917	0.16285	0.84583	0.17118	0.87083	0.16858	0.89167	0.2349
9	0.80833	0.10712	0.75	0.13316	0.67083	0.26111	0.675	0.24861
True and False Positive rates PCR-Lasso, N=100, T=100, $K_f = 5$								
1	0.97083	0.14983	0.925	0.19948	0.825	0.12431	0.925	0.19062
2	0.97083	0.15069	0.90833	0.20538	0.8625	0.12812	0.9125	0.21215
3	0.79167	0.10451	0.76667	0.15278	0.64167	0.28819	0.64583	0.28924
4	0.9625	0.14375	0.88333	0.16007	0.86667	0.18628	0.90833	0.19861
5	0.7625	0.11892	0.70417	0.12448	0.66667	0.29045	0.6625	0.3026
6	0.9625	0.15208	0.8875	0.19705	0.84583	0.1283	0.87917	0.1691
7	0.79167	0.12118	0.7125	0.11858	0.6375	0.30295	0.70833	0.31372
8	0.97917	0.15486	0.9125	0.20521	0.8875	0.1724	0.9125	0.20243
9	0.8125	0.11771	0.7375	0.15104	0.6125	0.27135	0.65833	0.28299

Table 20: TPR/FPR in PCR-Lasso (Regressing $\boldsymbol{y} - \hat{\boldsymbol{y}}_{pcr}$ on idiosyncratic components generated after extracting PC-factors from $\boldsymbol{X}$)

True and False Positive rates PCR-Lasso, N=200, T=100, $K_f = 4$								
(π)	$\psi_f = 1$, $\psi_g = 1$		$\psi_f = 1$, $\psi_g = 0.7$		$\psi_f = 0.7$, $\psi_g = 1$		$\psi_f = 0.7$, $\psi_g = 0.7$	
	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
1	0.92917	0.21862	0.80417	0.23801	0.94583	0.271	0.89583	0.29694
2	0.9625	0.23163	0.925	0.30901	0.95417	0.24311	0.90417	0.29702
3	0.7875	0.1142	0.73333	0.14566	0.67917	0.20417	0.6625	0.25043
4	0.9625	0.23571	0.87917	0.30578	0.96667	0.304	0.91667	0.25969
5	0.79583	0.13104	0.77917	0.16335	0.66667	0.24252	0.64583	0.22619
6	0.975	0.26811	0.88333	0.29388	0.90417	0.26429	0.87917	0.32823
7	0.80833	0.15264	0.71667	0.12917	0.725	0.29702	0.63333	0.2284
8	0.95	0.23308	0.86667	0.25527	0.94583	0.32585	0.83333	0.26905
9	0.775	0.11003	0.775	0.1693	0.70833	0.23963	0.69167	0.28376
True and False Positive rates PCR-Lasso, N=200, T=100, $K_f = 5$								
1	0.95833	0.25859	0.93333	0.2932	0.9125	0.2676	0.9	0.30986
2	0.975	0.26862	0.88333	0.35655	0.88333	0.26497	0.91667	0.37041
3	0.77917	0.13146	0.73333	0.1591	0.65833	0.26973	0.62083	0.29957
4	0.975	0.22942	0.92083	0.31318	0.9375	0.3068	0.89583	0.30893
5	0.825	0.15366	0.8	0.20187	0.65833	0.2602	0.7125	0.38044
6	0.96667	0.24481	0.91667	0.29753	0.925	0.24107	0.9	0.29923
7	0.8125	0.10527	0.74583	0.22202	0.675	0.21973	0.65417	0.36811
8	0.96667	0.21403	0.9	0.31964	0.94583	0.29303	0.89167	0.30383
9	0.7875	0.12092	0.75417	0.26003	0.7	0.2824	0.6875	0.33801

Table 21: TPR/FPR in PCR-Lasso (Regressing $\boldsymbol{y}-\hat{\boldsymbol{y}}_{pcr}$ on idiosyncratic components generated after extracting PC-factors from $\boldsymbol{X}$)

Average TPR / Average FPR, 100 Samples									
(β)	ρ								
	0	0.17	0.29	0.37	0.44	0.5	0.55	0.6	0.65
0.1	0.278 / 0.068	0.312 /0.056	0.364/0.06	0.388 /0.059	0.396/0.047	0.478/0.054	0.474/0.058	0.528/0.057	0.518/0.066
0.2	0.916 / 0.159	0.92 /0.143	0.96/0.114	0.956 /0.112	0.954/0.114	0.974/0.129	0.958/0.116	0.964/0.095	0.964/0.096
0.3	1 / 0.150	1 /0.149	1/0.133	1 /0.128	1/0.123	1/0.128	1/0.109	1/0.112	1/0.086
0.4	1 / 0.157	1 /0.142	1/0.126	1 /0.131	1/0.116	1/0.124	1/0.098	1/0.113	1/0.099
0.5	1 / 0.154	1 /0.13	1/0.135	1 /0.127	1/0.124	1/0.115	1/0.104	1/0.106	1/0.103
0.6	1 / 0.152	1 /0.154	1/0.138	1 /0.14	1/0.117	1/0.117	1/0.135	1/0.116	1/0.093
0.7	1 / 0.163	1 /0.165	1/0.132	1 /0.139	1/0.113	1/0.096	1/0.102	1/0.108	1/0.102
0.8	1 / 0.178	1 /0.148	1/0.138	1 /0.12	1/0.126	1/0.112	1/0.108	1/0.115	1/0.098
0.9	1 / 0.163	1 /0.141	1/0.15	1 /0.126	1/0.133	1/0.126	1/0.114	1/0.109	1/0.094

Table 22: Additional Example

Note: *The values of ρ may appear arbitrarily chosen, but they result from a 2-step process. We generate the covariance matrix of the 105 predictors by combining two block diagonal matrices. The cov-matrix of the relevant predictors is an identity matrix $\mathbb{I}_5$, while the covariance matrix of irrelevant predictors is obtained as $\mathcal{I}_1 00+\rho_1 \boldsymbol{\iota}_{100}\boldsymbol{\iota}'_{100}$. We allow ρ_1 to take values from the set* $0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.5, 1.8$. *Finally, we combine both of these blocks to generate the covariance matrix of all 105 variables. To eliminate any scale effects, we standardize our variables. Standardization ensures that all variables have unit var, and for the irrelevant predictors, the covariance* (ρ_1) *on the off-diagonal is replaced by correlation* $\rho = \dfrac{\rho_1}{1+\rho_1}$.

References

Ahn, Seung C, & Horenstein, Alex R. 2013. Eigenvalue ratio test for the number of factors. *Econometrica*, **81**(3), 1203–1227.

Andreou, Elena, Ghysels, Eric, & Kourtellos, Andros. 2013. Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, **31**(2), 240–251.

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.

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.

Ciccarelli, Matteo, & Mojon, Benoit. 2010. Global inflation. *The Review of Economics and Statistics*, **92**(3), 524–535.

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. 2023. Ridge regression under dense factor augmented models. *Journal of the American Statistical Association*, 1–13.

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.

Wang, Sijian, Nan, Bin, Rosset, Saharon, & Zhu, Ji. 2011. Random lasso. *The annals of applied statistics*, **5**(1), 468.

Zou, Hui, & Hastie, Trevor. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **67**(2), 301–320.