Interactive Effects Panel Data Models with General Factors and Regressors

BIN PENG†, LIANGJUN SU†, JOAKIM WESTERLUND§,* AND YANRONG YANG‡

†Monash University, ‡Singapore Management University, §Deakin University, ¶Lund University, ∪Australian National University

August 4, 2020

Abstract

The presence of unobserved heterogeneity and its detrimental effect on inference has recently motivated the use of interactive effects panel data models. One of the workhorses of this literature is based on iterating between principal components estimation of the unknown factors and OLS estimation of the model parameters. We refer to this as the “PC” approach, the existing asymptotic theory for which is based on the requirement that all the factors and regressors are either stationary or unit root non-stationary. Deterministic factors are typically treated as known and are projected out prior to the application of PC. This is a problem in practice where there is typically great uncertainty over both the order of integration of the stochastic component of the data and the terms needed to capture the deterministic component. This paper relaxes the above mentioned assumptions by considering a model wherein both the factors and regressors are essentially unrestricted, up to mild regulatory conditions. An estimator based on iterating PC is proposed, which is shown to be not only asymptotically normal and oracle efficient, but under certain conditions also free of the otherwise so common asymptotic incidental parameters bias. Interestingly, the conditions required to achieve unbiasedness become weaker the stronger the trends in the factors, and if the trending is strong enough unbiasedness comes at no cost at all. Equally important as these theoretical properties is the ease with which the new approach can be applied. In particular, the approach does not require any knowledge of how many factors there are, or whether they are deterministic/stochastic. The order of integration of the factors is also treated as unknown, as is the order of integration of the regressors, which means that there is no need to pre-test for unit roots, or to decide on which deterministic terms to include in the model. The new estimation approach should therefore be attractive for practitioners.

Keywords: Panel data, non-stationarity, principal components, interactive effects.

Westerlund thanks the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship.

*Corresponding author: Department of Economics, Lund University, Box 7082, 220 07 Lund, Sweden. Telephone: +46 46 222 8997. Fax: +46 46 222 4613. E-mail address: joakim.westerlund@nek.lu.se.
1 Introduction

The use of panel data in regression analysis has attracted considerable attention in the empirical literature in economic and elsewhere. A major reason for this is the ability to deal with the presence of unobserved heterogeneity, and the problem that this causes when said heterogeneity is correlated with the regressors. The simplest approach is to assume that the unobserved heterogeneity is made up of additive individual- and time-specific constants, or “fixed effects”. Such effects can add hugely to the fit of the model, and are typically found to be much more important than the regressors predicted by economic theory. As an example, Lemmon et al. (2008) study the determinants of firm leverage, a problem that is the subject of a separate empirical literature. According to their results, firm and year fixed effects account for over 90% of the explained variation in leverage. Of course, as pointed out by DeAngelo and Roll (2015) in the same context, the type of heterogeneity that can be permitted using fixed effects is very limited, and is unlikely to be enough in general. As a partial solution, they suggest including not only year and firm fixed effects, but also their interaction. However, because a full set of interaction dummies is not feasible, the authors only include firm-decade interactions, which are found to account for over 40% of the explained variation. They therefore conclude that “[u]se of a purely additive specification, that is, one that excludes interaction effects, is not a mandate of the data” (page 382). In other words, the issue of unobserved heterogeneity cannot be ignored and is unlikely to be settled using fixed effects only.

Observations like the one reported in the previous paragraph have recently led to much interest in so-called “interactive effects” models, in which the individual and time effects enter in a multiplicative way. In the terminology of the classical factor literature in statistics, the unobserved heterogeneity is assumed to have a common factor structure with the time effects as factors and the individual effects as factor loadings. This common factor specification contains the conventional fixed effects model as a special case, but is much more flexible since it allows the factors to affect each cross-section unit differently. The two most common approaches to such interactive effects models are the common correlated effects (CCE) approach of Pesaran (2006), and the principal components (PC) approach of Bai (2009), which are both based on first estimating the unknown factors and then running ordinary least squares (OLS) conditional on the first-step factor estimates.\footnote{The PC approach is sometimes referred to as “concentrated least squares” or just “least squares” (see, for example, Bai, 2009, and Moon and Weidner, 2015). We choose “PC” to distinguish it from CCE, which is also based on OLS.} The CCE approach takes the cross-sectional averages of the observables as factor estimates, and is therefore very easy to implement. The PC approach, on the other hand, requires iteration. An initial factor estimate is obtained by applying the PC method to the residuals of the naive OLS estimator that ignores the interactive effects. The
model is then reestimated using OLS conditional on the first-step factor estimates, which leads to new residuals from which new factors are estimated. This is repeated until convergence. But while computationally more attractive, CCE suffers from (at least) two major limitations that PC does not. The first limitation is that the regressors must have a factor structure that features the same set of factors as the dependent variable, which is restrictive in itself but also because it rules out models involving, for example, lags, powers or products of the regressors. The second limitation is that the number of factors that can be permitted is bounded from above by the number of observables, which is typically a very small number. The number of factors can be larger but then the loadings have to be distributed independently of the regressors, which means that one can just as well use fixed effects OLS (see Westerlund, 2019).

Because of the shortcomings of CCE, many researchers have chosen to work with PC, and there is by now separate literature devoted to it (see Ando and Bai, 2017, Bai and Li, 2014, Fan et al., 2013, Fan et al., 2016, Li et al., 2016, Moon and Weidner, 2015, Moon and Weidner, 2017, Moon et al., 2018, to mention a few). The present paper aims to contribute to this strand of the literature, and it does so in at least three ways.

The first contribution of the paper is to consider a very general data generating process in which both the factors and regressors are essentially unrestricted, and that include most of the specifications considered previously in the literature as special cases. Examples of permissible factors include polynomial time trends of any finite degree, seasonal and structural break dummies, weak factors and factors of an unknown but finite order of integration. The regressors cannot be weak but their specification is otherwise just as general as that of the factors. In fact, the only requirement is that suitably normalized sample second moment matrices of the factors and regressors have positive definite limits. This is noteworthy because the existing literature is almost exclusively based on the assumption that both the factors and regressors are stationary. The only exceptions known to us are Bai et al. (2009), Dong et al. (2020) and Huang et al. (2020), but they assume instead that either the regressors, or the factors and regressors are pure unit root processes, which is equally unrealistic. Indeed, regressors and factors of different order of magnitude is likely to be the rule rather than the exception, especially in economic and financial data, due to differences in persistence over time.

The unrestricted data generating process is important in itself but also because it can be accommodated without requiring any knowledge thereof. Hence, not only do we treat the factors and their number as unknown, but we also do not require any knowledge of the order of magnitude of both factors and regressors. An important implication of this is that there is no need to distinguish between deterministic and stochastic factors, or stationary and non-stationary factors. In the existing literature, deterministic factors are often treated as known, and are projected out prior to the application of PC (see, for example, Moon and Weidner,
The problem here is that there is typically great uncertainty over which deterministic terms to include, which raises the issue of model misspecification. The fact that in the present paper deterministic terms are treated as additional factors means that the problem of deciding on which terms to include does not arise. Similarly, while the regressors can be tested for unit roots, and the estimation can be made conditional on the test outcome, this raises the issue of pre-testing bias. In the present paper we do not require any knowledge about the order of integration of the regressors, which means that there is no need for any pre-testing.

Equally as important as the general model formulation and its empirical appeal is the extension of the existing econometric theory, which has not yet ventured much outside the stationary or pure unit root environment. This is our second major contribution. The main difficulty here is not the unrestricted specification of the factors and regressors per se, but rather that the order of magnitude of the factors may differ. In particular, the problem is that the nonlinearity of the PC estimator distorts the signal coming from the factors, just as it does in estimation of nonlinear regression models with mixtures of integrated regressors (see, for example, Park and Phillips, 2000). This is true if both the number and order of the factors are known, and the problem does not become any simpler when these quantities are treated as unknown, as they are here. An additional problem that then arises is that existing studies on the selection of the number of factors all require that the data are stationary (see, for example, Bai and Ng, 2002, and Ahn and Horenstein, 2013), and it is not obvious how one should go about this when the order of magnitude of the factor is unknown.

Intuitively, the factors whose order is largest should dominate the PC estimator. This motivates the use of an iterative estimation procedure in which the factors and their number are estimated in order according to their magnitude with relatively larger factors being estimated first. We begin by prescribing a large but fixed number of factors, and estimate the resulting model by PC. The estimated factors only capture the most dominating factors whose order of magnitude is largest. In spite of this, we can show that the estimator is consistent, albeit at a relatively low rate of convergence. The rate is, however, high enough to ensure that the number of dominating factors can be consistently estimated using a version of the eigenvalue ratio statistics of Lam and Yao (2012), and Ahn and Horenstein (2013). We then apply PC conditional on the first-step factor estimates, and estimate the second most dominating factors. This procedure continues until we cannot identify any more factors, and we can show that both the estimated factors and their number are consistent. Because of the iterative fashion in which the factors are estimated, we refer to the new estimation procedure as “iterative PC” (IPC). The IPC estimator of the model parameters takes the form of a one-step Newton–Raphson update, and is shown to be asymptotically normal and “oracle efficient”, which means that it has the same asymptotic variance as the oracle estimator that takes the factors as known.
The fact that asymptotic variance is the same as when the factors are known greatly simplifies inference, which can then be based on standard covariance estimation procedures.

Our third contribution is to point to a “blessing” of trending factors. The blessing occurs if the magnitude of the factors is sufficiently large, in which case the otherwise so common asymptotic bias of the PC approach can be completely eliminated without imposing any additional restrictions on the cross-sectional and time series dependencies of the regression errors. This is noteworthy, because the conclusion made in the previous literature suggests that in order to eliminate the asymptotic bias, the errors have to be independent.

The reminder of the paper is organized as follows. We begin by describing the model that we will be considering and the proposed IPC approach that we will use to estimate it. This is done in Section 2. Section 3 presents the formal assumptions and our main asymptotic results, whose small-sample accuracy is evaluated using Monte Carlo simulations in Section 4. Section 5 reports the results obtained from two empirical applications to the returns to scale in the US banking industry and the long-run relationship between US house prices and income. Section 6 concludes. All proofs and results of secondary nature are relegated to the online appendix.

2 Model and estimation procedure

Consider the panel data variable \( y_{i,t} \), observable for \( i = 1, \ldots, N \) cross-sectional units and \( t = 1, \ldots, T \) time periods. The model of this variable that we will be considering is the same as in Bai (2009), and is given by

\[
y_{i,t} = x_{i,t}^{'} \beta + \gamma_{0}^{'} f_{0,t} + \varepsilon_{i,t},
\]

(2.1)

where \( x_{i,t} = (x_{1,i,t}, \ldots, x_{d_{x},i,t})^{'} \) is a \( d_{x} \times 1 \) vector of regressors, \( f_{0,t} = (f_{0,1,t}, \ldots, f_{0,d_{f},t})^{'} \) is a \( d_{f} \times 1 \) vector of unobservable common factors with \( \gamma_{0}^{'} = (\gamma_{0,1,i}, \ldots, \gamma_{0,d_{f},i})^{'} \) being a conformable vector of factor loadings, and \( \varepsilon_{i,t} \) is a largely idiosyncratic error term. The interactive effects are here given by \( \gamma_{0}^{'} f_{0,t} \). The factors are divided into groups according to their order of magnitude. There are \( G \) groups of size \( d_{1}, \ldots, d_{G} \), which means that \( d_{1} + \cdots + d_{G} = d_{f} \). Because the grouping is unknown, we may without loss of generality assume that the factors are ordered, such that the first \( d_{1} \) factors have the highest order of magnitude, the next \( d_{2} \) factors have the second highest order, and so on. Hence, if we denote by \( f_{g,t}^{0} \) and \( \gamma_{g,i}^{0} \) the \( d_{g} \times 1 \) vectors of factors and loadings associated with group \( g \), respectively, then \( \gamma_{0}^{'} f_{0,t} = \sum_{g=1}^{G} \gamma_{g,i}^{0} f_{g,t} \), where \( f_{t}^{0} = (f_{1,t}^{0}, \ldots, f_{G,t}^{0})^{'} \) and \( \gamma_{0}^{'} = (\gamma_{0,1,i}, \ldots, \gamma_{0,G,i})^{'} \). If \( d_{f} = 0 \), then we define \( G = 0 \) and \( \gamma_{0}^{'} f_{0,t} = 0 \).

Remark 2.1. It is important to note that \( d_{g} \) is not restricted in any way. We may therefore have \( d_{1} = \cdots = d_{G} = 1 \), such that \( (f_{1,t}^{0}, \ldots, f_{G,t}^{0})^{'} = (f_{1,t}^{0}, \ldots, f_{d_{f},t}^{0})^{'} \).
It is useful to write (2.1) on stacked vector form. Let us therefore introduce the $T \times 1$ matrices $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T})'$ and $\mathbf{\varepsilon}_i = (\varepsilon_{i,1}, \ldots, \varepsilon_{i,T})'$, the $T \times d_f$ matrix $\mathbf{F}^0 = (f^0_1, \ldots, f^0_T)'$, and the $T \times d_x$ matrix $\mathbf{X}_i = (x_{i,1}, \ldots, x_{i,T})'$. Analogous to $f^0_t$, $\mathbf{F}^0$ is further partitioned as $\mathbf{F}^0 = (\mathbf{F}^0_1, \ldots, \mathbf{F}^0_g)'$, where the $T \times d_g$ matrix $\mathbf{F}^0_g = (f^0_{g,1}, \ldots, f^0_{g,T})'$ contains the stacked factor observations for group $g$. In this notation, (2.1) can be written as

$$ y_i = \mathbf{X}_i \beta^0 + \mathbf{F}^0 \gamma^0_i + \mathbf{\varepsilon}_i. \quad (2.2) $$

The goal of this paper is to infer $\beta^0$. As mentioned in Section 1, however, because of the generality of the model being considered, the main difficulty in the estimation process is how to control for $\mathbf{F}^0$. Our proposed estimation procedure consists of three steps. We first initialize the estimation procedure by applying the PC estimator of Bai (2009). However, because the first group of factors dominates all the other groups in terms of order of magnitude, the first-step PC factor estimator will only be estimating (the space spanned by) $\mathbf{F}^0_1$. The second step of the procedure therefore involves iteratively applying PC conditional on previous factor estimates to estimate all subsequent groups of factors; hence, the “I” in IPC. In the third and final step, we estimate $\beta^0$ conditional on the second-step IPC estimator of $\mathbf{F}^0$ and the first-step PC estimator of $\beta^0$.

**Step 1 (Initial estimation).** The (concentrated) OLS objective function that we consider is the same as in Bai (2009). It is given by

$$ \text{SSR} (\beta, \mathbf{F}) = \sum_{i=1}^{N} (y_i - \mathbf{X}_i \beta)' \mathbf{M}_F (y_i - \mathbf{X}_i \beta), \quad (2.3) $$

where $\mathbf{F}$ is a $T \times d_{\text{max}}$ matrix satisfying $T^{-\delta} \mathbf{F}' \mathbf{F} = \mathbf{I}_{d_{\text{max}}}$, $d_{\text{max}} \geq d_f$ and $\delta$ are user-specified non-negative integers, and $\mathbf{M}_A = \mathbf{I}_T - \mathbf{A}(\mathbf{A}' \mathbf{A})^{-1} \mathbf{A}' = \mathbf{I}_T - \mathbf{P}_A$ for any $T$-rowed full column rank matrix $\mathbf{A}$. The choice of $\delta$ is arbitrary and does not affect the results. Hence, in practice one may simply let the software decide, which is of course very convenient. The initial estimator that we consider is the minimizer of $\text{SSR} (\beta, \mathbf{F})$:

$$ (\hat{\beta}_0, \hat{\mathbf{F}}_0) = \arg \min_{(\beta, \mathbf{F}) \in \mathcal{D}} \text{SSR} (\beta, \mathbf{F}), \quad (2.4) $$

where $\mathcal{D} = \mathbb{R}^{d_x} \times \mathcal{D}_F$, $\mathcal{D}_F = \{ \mathbf{F} : T^{-\delta} \mathbf{F}' \mathbf{F} = \mathbf{I}_{d_{\text{max}}} \}$. It is useful to note that $\hat{\beta}_0$ satisfies
\(\hat{\beta}_0 = \hat{\beta}(\hat{F}_0)\), where

\[
\hat{\beta}(F) = \left( \sum_{i=1}^{N} \mathbf{X}'_i \mathbf{M}_F \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}'_i \mathbf{M}_F \mathbf{y}_i
\]  

(2.5)

(see Bai, 2009).

**Step 2** (Iterative estimation of factors). As already pointed out, the factors are estimated in order according to magnitude. Therefore, \(\hat{F}_0\) is estimating \(F^0_1\). Of course, since \(d_1 \leq d_F \leq d_{max}\), in general the dimension of \(\hat{F}_0\) will be larger than that of \(F^0_1\). We therefore begin this step of the estimation procedure by estimating \(d_1\), and for this purpose we employ a version of the ratio of eigenvalue-based estimator considered by, for example, Lam et al. (2011), Lam and Yao (2012), and Ahn and Horenstein (2013), which is given by

\[
\hat{d}_1 = \text{argmin}_{0 \leq d \leq d_{max}} \left\{ \frac{\widehat{\lambda}_{1,d+1}}{\widehat{\lambda}_{1,d}} \cdot \mathbb{I} \left( \frac{\widehat{\lambda}_{1,d}}{\widehat{\lambda}_{1,0}} \geq \tau_N \right) \right\} + \mathbb{I} \left( \frac{\widehat{\lambda}_{1,d}}{\widehat{\lambda}_{1,0}} < \tau_N \right)
\]

(2.6)

where \(\mathbb{I}(A)\) is the indicator function for the event \(A\) taking the value one if \(A\) is true and zero otherwise, \(\tau_N = 1/\ln(\max\{\widehat{\lambda}_{1,0}, N\})\), \(\widehat{\lambda}_{1,0} = N^{-1} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0)' \mathbf{M}_{\hat{F}_0} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0)\) and \(\widehat{\lambda}_{1,1} \geq \cdots \geq \widehat{\lambda}_{1,d_{max}}\) are the \(d_{max}\) largest eigenvalues of the following \(T \times T\) matrix:

\[
\hat{\Sigma}_1 = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0)' (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0).
\]

(2.7)

The threshold \(\tau_N\), the “mock” eigenvalue \(\widehat{\lambda}_{1,0}\) and the indicator function are there to ensure that the estimator is consistent. The need for these will be explained later. Given \(\hat{d}_1\), we now update the estimate of \(F^0_1\) by setting \(\hat{F}_1\) equal to the eigenvectors associated with \(\widehat{\lambda}_{1,1}, \ldots, \widehat{\lambda}_{1,\hat{d}_1}\), and estimate \(\gamma^0_{g,1}\) by \(\widehat{\gamma}_{g,1} = T^{-\delta} \hat{F}_1 (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0)\).

The estimation of \(F^0_{2}, \ldots, F^0_{G}\) is analogous to that of \(F^0_1\). The main difference is that we have to condition on all previous estimates. Let us therefore use \(\hat{F}_{-g} = (\hat{F}_1, \ldots, \hat{F}_{g-1})\) and \(\hat{\gamma}_{-g,i} = (\gamma'_{1,i}, \ldots, \gamma'_{g-1,i})'\) to denote the matrices containing the previously estimated factors and loadings, respectively, when estimating group \(g > 1\). The estimator of \(d_g\) is given quite naturally by

\[
\hat{d}_g = \text{argmin}_{0 \leq d \leq d_{max}} \left\{ \frac{\widehat{\lambda}_{g,d+1}}{\widehat{\lambda}_{g,d}} \cdot \mathbb{I} \left( \frac{\widehat{\lambda}_{g,d}}{\widehat{\lambda}_{g,0}} \geq \tau_N \right) \right\} + \mathbb{I} \left( \frac{\widehat{\lambda}_{g,d}}{\widehat{\lambda}_{g,0}} < \tau_N \right)
\]

(2.8)

where \(\widehat{\lambda}_{g,0} = N^{-1} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0)' \mathbf{M}_{\hat{F}_{-g}} (\mathbf{y}_i - \mathbf{X}_i \hat{\beta}_0)\) and \(\widehat{\lambda}_{g,1} \geq \cdots \geq \widehat{\lambda}_{g,d_{max}}\) are the \(d_{max}\)
largest eigenvalues of

\[ \hat{\Sigma}_g = \frac{1}{N} \sum_{i=1}^{N} (y_i - X_i \hat{\beta}_0 - \hat{F}_{-g} \hat{\gamma}_{-g,i}) (y_i - X_i \hat{\beta}_0 - \hat{F}_{-g} \hat{\gamma}_{-g,i})' \].  

(2.9)

The resulting estimator \( \hat{F}_g \) of \( F_g^0 \) is given by the eigenvectors associated with \( \hat{\lambda}_{g,1}, \ldots, \hat{\lambda}_{g,\hat{d}_g} \) and \( \hat{\gamma}_{g,i} = T^{-\delta} \hat{F}_g' (y_i - X_i \hat{\beta}_0 - \hat{F}_{-g} \hat{\gamma}_{-g,i}) \). New groups of factors are estimated until \( \hat{d}_g = 0 \). At this point, we set \( \hat{G} = g - 1 \) and define \( \hat{F} = (\hat{F}_1, \ldots, \hat{F}_{\hat{G}}) \). This is the IPC estimator of \( F^0 \).

**Step 3** (Estimation of \( \beta^0 \)). The IPC-based estimator of \( \beta^0 \), which can be seen as a one-step Newton-Raphson update of \( \hat{\beta}_0 \), is given by

\[ \hat{\beta} = \hat{\beta}_0 + \left( \sum_{i=1}^{N} \hat{Z}_i \hat{Z}_i' \right)^{-1} \sum_{i=1}^{N} X_i' M_{\hat{F}} X_i (\hat{\beta}_1 - \hat{\beta}_0) \].  

(2.10)

where \( \hat{Z}_i = M_{\hat{F}} X_i - \sum_{j=1}^{\hat{N}} M_{\hat{F}} X_j a_{ij} \) with \( a_{ij} = \gamma_i' (\hat{F}' \hat{F})^{-1} \hat{\gamma}_j \), \( \hat{\gamma}_i = (\hat{\gamma}_{1,i}', \ldots, \hat{\gamma}_{G,i}')' \) and \( \hat{F} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_N)' \), and \( \hat{\beta}_1 = \hat{\beta} (\hat{F}) \).

**Remark 2.2.** The fact that in Step 1 \( \delta \) can be set arbitrarily is nice not only from an applied point of view, but also from a theoretical point of view. In the bulk of the previous literature, the appropriate value of \( \delta \) to use depends on whether \( F^0 \) is stationary or unit root non-stationary (see, for example, Bai, 2004). The assumed knowledge of \( \delta \) is therefore tantamount to assuming the order of integration of \( F^0 \) is known, which is again not a requirement here.

**Remark 2.3.** The basic idea behind the ratio of eigenvalue-based estimator considered in Step 2 is very simple, and is similar to the one used in ocular inspection of “scree plots”; one looks for the point at which the ordered eigenvalues drop substantially, and set the number of estimated factors accordingly. The most natural way to mimic this decision rule is to follow, for example, Lam et al. (2011), Lam and Yao (2012), and Ahn and Horenstein (2013), and to minimize \( \hat{\lambda}_{g,d+1}/\hat{\lambda}_{g,d} \) over \( 1 \leq d \leq d_{\text{max}} \). However, this raises two issues. One is that since \( \hat{\lambda}_{g,d+1}/\hat{\lambda}_{g,d} \) is not defined for \( d = 0 \), we cannot have \( d_f = 0 \), and we want to be able to entertain the possibility that there are no factors. The use of the mock eigenvalue \( \hat{\lambda}_{g,0} \) allows us to do just that. The other problem is that under the conditions of this paper (laid out in Section 3) the limiting behaviour of \( \hat{\lambda}_{g,d+1}/\hat{\lambda}_{g,d} \) when \( d > d_g \) is unknown. Lam and Yao (2012) discuss this issue at length. They conjecture that \( \hat{\lambda}_{g,d+1}/\hat{\lambda}_{g,d} \approx 1 \), where \( a \sim b \) means that \( a = O_P(b) \) and \( b = O_P(a) \), but there is no proof. The use of the indicator function allows us to circumvent this problem. The idea is to look at \( \hat{\lambda}_{g,d} \). If this eigenvalue is “small” we take it as a sign of \( d > d_g \) and set \( \hat{\lambda}_{g,d+1}/\hat{\lambda}_{g,d} \) to one. However, because the order of \( \mathbf{x}_{i,t} \) and \( \mathbf{f}_t^0 \) are both assumed to be
unknown, we cannot look at $\hat{\lambda}_{g,d}$ directly but rather we look at $\hat{\lambda}_{g,d}/\hat{\lambda}_{g,0}$, which in contrast to $\hat{\lambda}_{g,d}$ is self-normalizing.

Remark 2.4. Intuition suggests to take $\hat{\beta}_1$, the OLS estimator conditional on $\hat{F}$, as the final estimator of $\beta_0$ in Step 3. Interestingly, while consistent, because of the stepwise estimation of the factors, the asymptotic distribution of $\hat{\beta}_1$ is generally unknown. The proposed IPC estimator is not only asymptotically normal, but as we demonstrate in Section 3 also oracle efficient.

3 Assumptions and asymptotic results

Before we state our assumptions and asymptotic results, we introduce some notation. Specifically, if $A$ is a matrix, then $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ signify its smallest and largest eigenvalues, respectively, $\text{tr} A$ signifies its trace, and $\|A\| = \sqrt{\text{tr} A'A}$ and $\|A\|_2 = \sqrt{\lambda_{\max}(A)}$ to signify its Frobenius and spectral norms, respectively. We write $A > 0$ to signify that $A$ is positive definite. If $B$ is also a matrix, then $\text{diag}(A, B)$ denotes the block-diagonal matrix that takes $A$ ($B$) as the upper left (lower right) block. The symbols $\rightarrow_D$, $\rightarrow_P$ and $MN(\cdot, \cdot)$ signify convergence in distribution, convergence in probability and a mixed normal distribution, respectively. We use $N, T \rightarrow \infty$ to indicate that the limit has been taken while passing both $N$ and $T$ to infinity. Any restrictions on the relative expansion rate of $N$ and $T$ will be specified separately.

As usual, $a_T = O_P(T^r)$ will be used to signify that $a_T$ is at most order $T^r$ in probability, while $a_T = o_P(T^r)$ will be used in case $a_T$ is of smaller order in probability than $T^r$. We use w.p.a.1 to denote with probability approaching one.

Assumption 1 is concerned with the order of magnitude of $f^0_t$ and $x_{i,t}$. It is therefore key. The assumption is stated in terms of the required moment conditions rather than primitive assumptions on $f^0_t$ and $x_{i,t}$. This is convenient because it is this result that drives the asymptotic results, and we are not specifically interested here in the various sets of conditions under which they hold.

Assumption 1 (Moments).

(a) $E\|((NT)^{-1} \sum_{i=1}^N D_T X_i M_{F,0} X_i D_T - \Sigma_X)\|^2 = o(1)$ and $T^{-2}E\|D_T X_i\|^4 = O(1)$, where $D_T = \text{diag}(T^{-\kappa_1/2}, \ldots, T^{-\kappa_d/2})$ with $\kappa_j \geq 0$ for all $j = 1, \ldots, d_x$, $E\|\Sigma_X\|^2 < \infty$ and $0 < \lambda_{\min}(\Sigma_X) \leq \lambda_{\max}(\Sigma_X) < \infty$ w.p.a.1.

(b) $\|((NT)^{-1} \sum_{i=1}^N D_T X_i \varepsilon_i\| = O_P(1/ \min\{\sqrt{N}, \sqrt{T}\})$ and $\|\varepsilon\|_2 = O_P(\max\{\sqrt{N}, \sqrt{T}\})$ with $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$. 8
(c) $E\|C_T F^0 F^0 C_T - \Sigma_{F^0}\|^2 = o(1)$, where $C_T = \text{diag}(T^{-\nu_1/2}I_{d_1}, \ldots, T^{-\nu_G/2}I_{d_G})$ with $\nu_1 > \ldots > \nu_G > 1/2$, $E\|\Sigma_{F^0}\|^2 < \infty$ and $0 < \lambda_{\min}(\Sigma_{F^0}) \leq \lambda_{\max}(\Sigma_{F^0}) < \infty$ w.p.a.1.

(d) $\|N^{-1} \Gamma^0 - \Sigma_{\Gamma^0}\| = o_P(1)$ and $\max_{i \geq 1} E\|\gamma_i^0\|^4 < \infty$, where $\Gamma^0 = (\gamma_1^0, \ldots, \gamma_N^0)'$ is $N \times d_f$ and $0 < \Sigma_{\Gamma^0} < \infty$.

Assumption 1 (a) is very general in that it imposes almost no restrictions on the type of trending behaviour that $x_{i,t}$ may have. The trending can be deterministic but it can also be stochastic, as in the presence of unit roots. Either way, the degree of the trending is not restricted in any way, provided of course that it is finite. The only requirement is that $0 < \lambda_{\min}(\Sigma_X) \leq \lambda_{\max}(\Sigma_X) < \infty$ w.p.a.1, which means that the elements of $x_{i,t}$ cannot be asymptotically collinear when suitably normalized by $T$.\footnote{Note that $\Sigma_X$ is not required to be a constant matrix, as this would rule out regressors that are integrated (see Bai et al., 2009).} This is very different when compared to the bulk of the previous PC-based interactive effects literature in which $x_{i,t}$ is assumed to be either stationary (see, for example, Bai, 2009, and Moon and Weidner, 2015), such that $\kappa_1 = \ldots = \kappa_{d_x} = 0$, or unit-root non-stationary (see Bai et al., 2009, and Dong et al., 2020), such that $\kappa_1 = \ldots = \kappa_{d_x} = 1$, which of course need not be the case in practice.

The first requirement of Assumption 1 (b) is quite mild and holds if a central limit theorem in only one of the two panel dimensions applies to the normalized sum of $D_T X_i' \varepsilon_i$. The second requirement is quite common in the literature, and is expected to hold as long as $\varepsilon_{i,t}$ has zero mean, bounded fourth moment, and weak serial and cross-sectional correlation (see Moon and Weidner, 2015, for a discussion).

Assumption 1 (c) is similar to Assumption 1 (a) in that it leaves the trending behaviour of the factors essentially unrestricted. In fact, the factors are not even required to be strong ($\nu_G \geq 1$), but can also be weak ($\nu_G < 1$) as in Chudik et al. (2011), which means that the cross-sectional dependence can be spatial in nature. This is again very different from the previous PC-based literature, which typically assume that $T^{-1} F^0 F^0$ converges to positive definite matrix (see Bai, 2009, and Moon and Weidner, 2015). Notable exceptions include Bai (2004), Bai et al. (2009), and Choi (2017); however, these assume that $F^0$ is a pure unit root process, which is again not very realistic. Bai and Ng (2004) allow for a mix of stationary and unit root factors, which is dealt with by taking first differences. Because the factors of the first-differenced data are stationary, estimation can proceed in the usual fashion, and the resulting estimated first-differenced factors can then be accumulated up to levels. The problem with this approach is that it supposes that the order of integration of the data is at most one, and that any deterministic constant and trend terms are known to the researcher. The only study that comes close to ours in terms of the generality of the factors is that of Westerlund (2018). However, he uses
CCE, which means that his approach suffers from the same shortcomings as those laid out in Section 1. In particular, \( x_{i,t} \) is assumed to load linearly on the same set of factors as \( y_{i,t} \), which is restrictive. Indeed, even if \( x_{i,t} \) does in fact have a linear factor structure, the factors need not be the same as those affecting \( y_{i,t} \). The model considered in the present paper does not require \( x_{i,t} \) to have a factor structure, and is therefore more general in this regard.

Assumption 1 (d) is standard and ensures that each factor has a nontrivial contribution to the variance of \( y_{i,t} \).

**Assumption 2** (Identification). \( \inf_{F \in D} \lambda_{\min}(B(F)) \geq c_0 > 0 \) for all \( N \) and \( T \), where

\[
B(F) = \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i(F)' Z_i(F) D_T
\]

with \( Z_i(F) = M_F X_i - \sum_{j=1}^{N} M_F X_j a_{ij} \) and \( a_{ij} = \gamma_i^0 (\Gamma_0^0)^{-1} \gamma_j^0 \).

The requirement that the smallest eigenvalue of \( B(F) \) is positive is equivalent to requiring that \( B(F) \) is positive definite, which is a noncollinearity condition. It demands that the regressors in \( x_{i,t} \) have enough variation after projecting out all variation that can be explained by \( F^0 \) and \( \gamma_i^0 \). As in Bai (2009), had \( F^0 \) be observed, the identifying condition for \( \beta^0 \) would be that \( (NT)^{-1} \sum_{i=1}^{N} D_T X_i M_{F^0} X_i D_T \) is positive definite.

Together Assumptions 1 and 2 are enough to ensure that the initial estimator \( \hat{\beta}_0 \) of \( \beta^0 \) is consistent.

**Lemma 3.1** (Consistency of \( \hat{\beta}_0 \)). Under Assumptions 1 and 2,

\[
\min\{\sqrt{N}, \sqrt{T}\} D_T^{-1} (\hat{\beta}_0 - \beta^0) = O_P(1).
\]

Lemma 3.1 establishes that \( \hat{\beta}_0 \) is consistent for \( \beta^0 \) and that the rate of convergence is \( \|D_T\|/\min\{\sqrt{N}, \sqrt{T}\} = \max\{T^{-\kappa_1/2}, \ldots, T^{-\kappa_d/2}\}/\min\{\sqrt{N}, \sqrt{T}\} \). To put this into perspective, suppose that \( x_{i,t} \) is stationary such that \( \kappa_1 = \cdots = \kappa_d = 0 \). In this case, \( D_T = I_d \) and the rate of convergence is given by \( 1/\min\{\sqrt{N}, \sqrt{T}\} \), which is the slowest of the regular rates in pure time series and cross-section regressions. Still, the rate is fast enough for the estimation of the number of factors. This brings us to Step 2 of the estimation procedure. In order to be able to show that \( \hat{d}_1, \ldots, \hat{d}_{G+1} \) and \( \hat{F} \) are consistent, however, we need more structure.

**Assumption 3** (Errors).

(a) \( E \varepsilon_{i,t} = 0 \) and \( E \varepsilon_i \varepsilon_i' = \Sigma_{\varepsilon,i} \).
(b) Let \( \varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{N,t})' \) in this assumption only. \( \{ \varepsilon_t : t \geq 1 \} \) is strictly stationary and \( \alpha \)-mixing such that \( \max_{i \geq 1} E|\varepsilon_{i,1}|^{4+\mu} < \infty \) for some \( \mu > 0 \) and the mixing coefficient

\[
\alpha(t) = \sup_{A \in \mathcal{F}_{-\infty}, B \in \mathcal{F}_t} |P(A)P(B) - P(AB)| \]

satisfies \( \sum_{t=1}^{\infty} \alpha(t)^{(2+\mu)/2} < \infty \), where \( \mathcal{F}_{-\infty} \) and \( \mathcal{F}_t \) are the sigma-algebras generated by \( \{ \varepsilon_s : s \leq 0 \} \) and \( \{ \varepsilon_s : s \geq t \} \), respectively.

(c) Suppose that \( \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E\varepsilon_{i,t}\varepsilon_{j,s}| = O(NT) \) and \( \sum_{i=1}^{N} \sum_{j \neq i} |\sigma_{\varepsilon,ij}| = O(N) \), where \( \sigma_{\varepsilon,ij} = E\varepsilon_{i,t}\varepsilon_{j,t} \).

(d) \( \varepsilon_{i,t} \) is independent of \( \gamma^0_j, f^0_s \) and \( x_{j,s} \) for all \( i, j, t \) and \( s \).

Assumption 3 is similar to Assumptions C and D in Bai (2009). Assumption 3 (b) and (c) ensure that the serial and cross-sectional dependencies of \( \varepsilon_{i,t} \) are at most weak. Assumption 3 (d) requires that the regressors are exogenous, which rules out the presence of lagged dependent variables in \( x_{i,t} \). Assumption 3 (d) can be relaxed by instead requiring that \( \varepsilon_{i,t} \) satisfies a martingale type condition similar to Assumption 3 in Dong et al. (2020). However, since this will add to the already lengthy derivations and heavy notation, in the present paper we maintain Assumption 3. Note, however, that the composite error term \( \gamma^0_i f^0_t + \varepsilon_{i,t} \) may still be correlated with \( x_{i,t} \), which invalidates estimation by OLS. The estimation problem is therefore nontrivial.

If \( \nu_G \geq 1 \), so that all the factors in \( f^0_t \) are strong, Assumptions 1–3 are enough to ensure that \( \hat{d}_1 \) is consistent. If, however, \( \nu_G < 1 \), so that some of the factors are weak, then we also need Assumption 4.

**Assumption 4 (Weak factors).** If \( \nu_G < 1 \), then \( T/N^2 \to c_1 < \infty \).

As already pointed out, \( F^0_1 \) dominates all other factors and is thus easiest to estimate. It is therefore not surprising to find that \( \hat{d}_1 \) is consistent for \( d_1 \). Lemma 3.2 states this last result formally.

**Lemma 3.2 (Consistency of \( \hat{d}_1 \)).** Under Assumptions 1–4, as \( N, T \to \infty \),

\[
P(\hat{d}_1 = d_1) \to 1. \tag{3.3}
\]

**Remark 3.1.** In Lemma A.4 of the appendix, we show that \( \hat{F}_1 \) is rotationally consistent for \( F^0_1 \). This means that it is not possible to apply the original PC approach of Bai (2009) and to estimate all the factors in one go. The only exception is if \( G = 1 \), so that all the factors are of the same order of magnitude.

Hence, \( \hat{d}_1 \) is consistent. In order to ensure that also \( \hat{d}_2, \ldots, \hat{d}_{G+1} \) are consistent, the groups have to be “distinct”. The next assumption formalizes this requirement. The assumption is stated in terms of \( F^0_g \) and \( \Gamma^0_g \), where \( \Gamma^0_g = (\gamma^0_{g,1}, \ldots, \gamma^0_{g,N})' \) is the \( N \times d_g \) matrix of stacked factor loadings for group \( g \).
Assumption 5 (Orthogonal groups). $\max_{g \neq h} \| \Gamma_g^0 \Gamma_h^0 \| = O_P(N^p)$ and $\max_{g \neq h} \| F_g^0 F_h^0 \| = O_P(T^q)$, where $g, h = 1, \ldots, G$, $G > 1$, $p < 1$, $q < (\nu_G + \nu_{G-1})/2$ and $\nu_{G-1} \geq 1$.

The condition that $\nu_{G-1} \geq 1$ means that we only allow for one group of weak factors. This can be seen as a form of normalization and is not particularly restrictive. The value of $p$ measures the degree of orthogonality between $\Gamma_g^0$ and $\Gamma_h^0$. In the extreme case when $p = -\infty$, then $\| \Gamma_g^0 \Gamma_h^0 \| = 0$, and therefore $\Gamma_g^0$ and $\Gamma_h^0$ are exactly orthogonal. The condition that $p < 1$ rules out the case when $\Gamma_1^0, \ldots, \Gamma_G^0$ are constant matrices not depending on $N$ or $T$, in which case $\| \Gamma_g^0 \Gamma_h^0 \| = O_P(N)$. This condition is needed to separately identify the group-specific loadings and is similar to Assumption A of Ando and Bai (2017). The orthogonality condition on $F_0^0$ is similarly needed to ensure identification of the group-specific factors.

Lemma 3.3 (Consistency of $\hat{d}_2, \ldots, \hat{d}_{G+1}$ and $\hat{F}$). Suppose that Assumptions 1–5 are satisfied. Then, the following results hold as $N, T \to \infty$:

(a) $P(\hat{d}_g = d_g) \to 1$ for $g = 2, \ldots, G + 1$, where $G > 1$ and $d_{G+1} = 0$;

(b) $\| P_{\hat{F}} - P_{F_0^0} \| = o_P(1)$.

Together with Lemma 3.2, Lemma 3.3 (a) implies that $\hat{d}_1, \ldots, \hat{d}_{G+1}$ are all consistent. The consistency of $\hat{d}_1, \ldots, \hat{d}_G$ is important for obvious reasons. The consistency of $\hat{d}_{G+1}$ ensures that the stopping rule in Step 2 of the estimation procedure is asymptotically valid, which in turn implies that $\hat{G}$ is consistent. Hence, under the conditions of Lemma 3.3, we have

$$P(\hat{G} = G) \to 1$$

as $N, T \to \infty$. The consistency of $\hat{G}$ further implies that $d_f$ can be consistently estimated using $\hat{d}_f = \hat{d}_1 + \cdots + \hat{d}_G$.

As is well known, $F^0$ and $\gamma_i^0$ are only identified up to a positive definite matrix rotation (see, for example, Bai and Ng, 2002). However, we cannot claim that $\hat{F}$ is rotationally consistent for $F^0$, as the number of rows of both objects is growing with $T$. We therefore have to resort to alternative consistency concepts. This is where Lemma 3.3 (b) comes in. It shows that the spaces spanned by $\hat{F}$ and $F^0$ are asymptotically the same. This establishes that all the Step 2 estimates are consistent. We therefore now move on to investigate the asymptotic properties of the IPC estimator $\hat{\beta}$ of $\beta^0$ obtained in Step 3.

Assumption 6 (Strong orthogonality). $p < 1/2$ and $T^{\nu_1}/N^{1/2-p} \to c_2 < \infty$.

Assumption 6 strengthens the Assumption 5 requirement that $p < 1$ to $p < 1/2$, which means that $F^0_g$ and $\Gamma^0_h$ should be relatively more orthogonal than before. We say that they
are “strongly” orthogonal, because $p < 1/2$ rules out the usual central limit law rate when $\| \Gamma_0^p \Gamma_0^{\prime} \| = O_P(\sqrt{N})$. The condition that $T^{\nu_1}/N^{1/2-p} \to c_2 < \infty$ is there to ensure that the orthogonality of $\Gamma_0^p$ and $\Gamma_0^{\prime}$ is strong enough to offset the impact of the factors.

**Assumption 7** (Rates).

(a) $N/T^{\nu_G} \to \rho_1 \in [0, \infty)$.

(b) $T^{2-\nu_G}/N \to \rho_2 \in [0, \infty)$.

It is important to note that if all the factors in $f^0_i$ are stationary such that $G = 1$ and $\nu_G = \nu_1 = 1$, then Assumption 7 requires that $N/T \to \rho_1 = 1/\rho_2 \in (0, \infty)$, which is the condition considered in Bai (2009), and Moon and Weidner (2015). Assumption 7 is therefore more general than the condition considered in these other papers.

**Assumption 8** (Asymptotic normality).

$$
1/\sqrt{NT} \sum_{i=1}^{N} D_T Z_i (F^0)^{\prime} \varepsilon_i \to_D MN(0_{d \times 1}, \Omega) \tag{3.5}
$$

as $N, T \to \infty$, where

$$
\Omega = \lim_{N,T \to \infty} 1/NT \sum_{i=1}^{N} \sum_{j=1}^{N} D_T E[Z_i (F^0)^{\prime} \varepsilon_i \varepsilon_j Z_j (F^0)|C] D_T, \tag{3.6}
$$

with $C$ being the sigma algebra generated by $F^0$.

Assumption 8 is a central limit theorem that is analogous to Assumption E of Bai (2009). The reason for requiring that the asymptotic distribution is mixed normal as opposed to normal is that by doing so we can accommodate stochastically integrated factors (see, for example, Theorem 1 in Bai et al., 2009, or Theorem 3.5 in Huang et al., 2020). In absence of such integrated factors, the mixed normal distribution is normal. Either way, Assumption 8 ensures that standard normal and chi-squared inference based on $\hat{\beta}$ is possible.

**Theorem 3.1** (Asymptotic distribution of $\hat{\beta}$). Under Assumptions 1–8, as $N, T \to \infty$,

$$
\sqrt{NT} D_T^{-1}(\hat{\beta} - \beta^0) \to_D MN(B_0^{-1}(\sqrt{\rho_1 A_1} + \sqrt{\rho_2 A_2}), B_0^{-1} \Omega B_0^{-1}), \tag{3.7}
$$

where

$$
B_0 = \lim_{N,T \to \infty} E[B(F^0)|C], \tag{3.8}
$$
\begin{align}
A_1 &= - \lim_{N,T \to \infty} \frac{1}{T(1-w_G)/2} \sum_{i=1}^{N} D_T E[X_i' M_{F^0} \Sigma_f F^0 (F^0 F^0)^{-1}(F^0 F^0)^{-1} \gamma_{G,i}^0 |C], \quad (3.9) \\
A_2 &= - \lim_{N,T \to \infty} \frac{1}{T(3-w_G)/2} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T E[Z_i(0)' F^0 (F^0 F^0)^{-1}(F^0 F^0)^{-1} \gamma_{G,j}^0 \varepsilon_j |C], \quad (3.10)
\end{align}

with \( \Sigma_f = N^{-1} \sum_{i=1}^{N} \Sigma_f, \) and \( Z_i(0) = X_i - \sum_{i=1}^{N} X_{ij} \).

We begin by noting that the asymptotic bias in Theorem 3.1 is driven by the factors and loadings of group \( G \), which is intuitive as the factors of this group are smallest in order of magnitude. They therefore dominate the asymptotic bias. By bounding \( \nu_G \) from below Assumption 7 ensures that \( B_0^{-1}(\sqrt{\rho_1} A_1 + \sqrt{\rho_2} A_2) \) is not diverging. It is important to note that \( \rho_1 \) and \( \rho_2 \) may both be zero, which will be the case if \( \nu_G \) is large enough. Hence, if the magnitude of the factors is sufficiently large the asymptotic bias of the IPC estimator is completely eliminated without imposing any other restriction. In order to put this into perspective, suppose again that all the factors are stationary, such that \( G = 1, \nu_G = \nu_1 = 1, F^0_G = F^0_1 = F^0, \gamma_{G,i}^0 = \gamma_{i}^0, \Gamma^0_G = \Gamma^0_1 = \Gamma^0 \) and \( \rho_1 = 1/\rho_2 \in (0, \infty) \), and that the regressors are stationary, too, such that \( D_T = I_d \). In this case, the bias in Theorem 3.1 reduces to

\[ B_0^{-1}(\sqrt{\rho_1} A_1 + \sqrt{\rho_2} A_2) = B_0^{-1}(\sqrt{\rho_1} A_1 + \rho_1^{-1/2} A_2), \quad (3.11) \]

where

\begin{align}
A_1 &= - \lim_{N,T \to \infty} \sum_{i=1}^{N} E[X_i' M_{F^0} \Sigma_f F^0 (F^0 F^0)^{-1}(F^0 F^0)^{-1} \gamma_{i}^0 |C], \quad (3.12) \\
A_2 &= - \lim_{N,T \to \infty} \frac{1}{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E[Z_i(0)' F^0 (F^0 F^0)^{-1}(F^0 F^0)^{-1} \gamma_{j}^0 \varepsilon_j |C], \quad (3.13)
\end{align}

which is identically the bias reported in Theorem 3 of Bai (2009). It is important to note that while \( \sqrt{\rho_1} A_1 \) can be made arbitrarily small (large) by just taking \( \rho_1 \) to zero (infinity), this will make \( \rho_1^{-1/2} A_2 \) divergent (negligible). It follows that unless \( \rho_1 \in (0, \infty) \) the bias will diverge, which in turn means that there is no way to make the bias disappear by just manipulating \( \rho_1 \), which in practical terms means restricting \( T/N \). By allowing \( \nu_G > 1 \), we break the inverse relationship between \( \rho_1 \) and \( \rho_2 \), which means that one can be zero without for that matter forcing the other to infinity. Moreover, the larger \( \nu_G \) is, the less restrictive the condition on \( T/N \) is. For example, if \( \nu_G = 2 \), then we only require \( N/T^2 \to 0 \) for the bias to disappear. In this sense, trending factors are a blessing.

**Remark 3.2.** Bai (2009, Corollary 1) shows that PC can be unbiased even if \( f_i^0 \) and \( x_{i,t} \) are
stationary. However, this requires either that \( \epsilon_{i,t} \) is independent and identically distributed across both \( i \) and \( t \), or that it is uncorrelated and homoskedastic in \( t \) (i) with \( T/N \to 0 \) \((N/T \to 0)\).

**Remark 3.3.** It is important to note that while biased, \( \hat{\beta} \) is still consistent at the best achievable rate. Again, if \( x_{i,t} \) is stationary, \( D_T = I_{d_x} \) and the rate of convergence is given by \( 1/\sqrt{NT} \), which is the same as in Bai (2009). This is in contrast to Lemma 3.1 and the relatively slow rate of convergence reported there. The reason for this difference is that unlike Theorem 3.1, which requires that Assumptions 1–8 all hold, Lemma 3.1 only requires Assumptions 1 and 2, and under these very relaxed conditions the Theorem 3.1 rate is not attainable.

**Corollary 3.1 (Unbiased asymptotic distribution).** Suppose that the conditions of Theorem 3.1 are met and that \( \rho_1 = \rho_2 = 0 \). Then, as \( N, T \to \infty \),

\[
\sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) \to_D MN(0_{d_x \times 1}, B_0^{-1}\Omega B_0^{-1}).
\]

According to Corollary 3.1, the condition that \( \rho_1 = \rho_2 = 0 \) ensures that the asymptotic distribution of \( \sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) \) is correctly centered at zero. The condition is not necessary, though, as \( B_0^{-1}(\sqrt{\rho_1}A_1 + \sqrt{\rho_2}A_2) \) can be zero also if \( A_1 \) and \( A_2 \) are zero. In Section A.2 of the appendix, we therefore provide some alternative conditions that ensure \( A_1 = A_2 = 0_{d_x \times 1} \).

**Remark 3.4.** The asymptotic distribution of \( \sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) \) is mixed normal even if \( \rho_1 \neq 0 \) or \( \rho_2 \neq 0 \) in Theorem 3.1. The problem is, as already explained, that without the condition that \( \rho_1 = \rho_2 = 0 \), the mean need not be zero, which then requires some form of correction if the purpose it to conduct inference. One possibility is to follow Bai (2009), and Moon and Weidner (2015), and to use analytical correction; however, one can also use jackknifing, as is done in Fernández-Val and Weidner (2016), and Westerlund (2018).

Theorem 3.1 imposes only minimal conditions on the correlation and heteroskedasticity of \( \epsilon_{i,t} \), and is in this sense very general. Such generality is, however, not possible if we also want to ensure consistent estimation of \( \Omega \). Let us therefore assume for a moment that \( E(\epsilon_{i,t}\epsilon_{j,s}) = 0 \) for all \( i \neq j \) and \( t \neq s \), so that \( \epsilon_{i,t} \) is serially and cross-sectionally uncorrelated. In this case,

\[
\sum_{i=1}^{N} \sum_{j=1}^{N} E[Z_i(F^0)'\epsilon_i\epsilon_jZ_j(F^0)|\mathcal{C}] = \sum_{i=1}^{N} \sigma_{\epsilon,i}^2E[Z_i(F^0)'Z_i(F^0)|\mathcal{C}].
\]

A natural estimator of this matrix is given by

\[
\sum_{i=1}^{N} \sigma_{\epsilon,i}^2 \tilde{Z}_i\tilde{Z}_i,
\]
where \( \hat{\sigma}^2_{\varepsilon,i} = T^{-1} \sum_{t=1}^T (y_i - X_i \hat{\beta})' M \hat{\beta} (y_i - X_i \hat{\beta}) \) and \( \hat{Z}_i \) is as in the definition of \( \hat{\beta} \). It is not difficult to show that under the conditions of Theorem 3.1,

\[
\left\| \frac{1}{NT} \sum_{i=1}^N \hat{\sigma}^2_{\varepsilon,i} D_T \hat{Z}_i' \hat{Z}_i D_T - \Omega \right\| = o_P(1).
\] (3.17)

Of course, in this paper we do not assume knowledge of the order of the regressors, which in practice means that the appropriate normalization matrix \( D_T \) to use is unknown. This is not a problem, however, as the usual Wald and \( t \)-test statistics are self-normalizing. As an illustration, consider testing the null hypothesis of \( H_0 : R \beta^0 = r \), where \( R \) is a \( r_0 \times d_x \) matrix of rank \( r_0 \leq d_x \) and \( r \) is a \( r_0 \times 1 \) vector. The Wald test statistic for testing this hypothesis is given by

\[
W = (R \hat{\beta} - r)' \left( \sum_{i=1}^N \hat{Z}_i' \hat{Z}_i \right)^{-1} \left( \sum_{i=1}^N \hat{Z}_i' \hat{Z}_i \right)^{-1} (R \hat{\beta} - r),
\] (3.18)

which has a limiting chi-squared distribution with \( r_0 \) degrees of freedom under \( H_0 \), as is clear from

\[
W = \sqrt{NT} [D_T^{-1} (R \hat{\beta} - r)]' \left[ \frac{1}{NT} \sum_{i=1}^N D_T \hat{Z}_i' \hat{Z}_i D_T \right]^{-1} \left[ \frac{1}{NT} \sum_{i=1}^N \hat{\sigma}^2_{\varepsilon,i} D_T \hat{Z}_i' \hat{Z}_i D_T \right]^{-1} \sqrt{NT} D_T^{-1} (R \hat{\beta} - r) \to_d \chi^2(r).
\] (3.19)

The above results are for the case when \( \varepsilon_{i,t} \) is serially and cross-sectionally uncorrelated. If \( \varepsilon_{i,t} \) is serially and/or cross-sectionally correlated, we recommend following Bai (2009), who discusses the issue of consistent covariance matrix estimation at length. The same arguments can be applied without change in current context.

4 Monte Carlo results

In this section, we report the results from a small-scale Monte Carlo simulation exercise based on six data generating processes (DGPs). In all DGPs, \( y_{i,t} \) is generated according to a restricted version of (2.1) that sets \( d_x = 3, \beta^0 = 1_{3 \times 1} \) for DGPs 1-5 and \( d_x = 2, \beta^0 = 1_{2 \times 1} \) for DGP 6, and \( \varepsilon_{i,t} \sim N(0,1) \). Let \( x_{j,i,t}, f_{j,t}^0 \) and \( \gamma_{j,i}^0 \) be the \( j \)-th rows of \( x_{i,t}, f_t^0 \) and \( \gamma_i^0 \), respectively. The following specification of \( x_{j,i,t} \) allows the factors and loadings to be correlated with the
Moreover, each combination of \( N \) and \((\nu, \nu_f)\) seen as an estimator of \( \nu \) while the former frequency captures the accuracy of the estimation of both \( \nu \) and \( \nu_f \) measured by the square root of the average of \( \sum_{n=1}^{df} (|f_{n,t}^0| + |\gamma_{g,i}^0|) \).

The first five DGPs that we consider are based on (4.1) and can be described as follows.

**DGP 1** \((G = 2)\). We generate \( f_{g,t}^0 \sim t^{(\nu_g - 1)/2} \cdot N(\mathbf{0}_{d_g \times 1}, \mathbf{I}_{d_g}) \) and \( \gamma_{g,i}^0 \sim 2 \cdot \mathbb{I}(m = 1) \cdot \mathbf{1}_{d_g \times 1} + N(\mathbf{0}_{d_g \times 1}, \mathbf{I}_{d_g}) \). Also, in this DGP, \( G = d_f = 2 \), \((\nu_1, \nu_2) = (2.5, 1.8)\) and \((d_1, d_2) = (1, 1)\).

**DGP 2** \((G = 3)\). This DGP is the same as DGP 1, but now \( G = 3 \), \( d_f = 5 \), \((\nu_1, \nu_2, \nu_3) = (3, 2, 1)\) and \((d_1, d_2, d_3) = (1, 2, 1)\).

**DGP 3** \((G = 4)\). DGP 2 is again the same as DGP 1, except that now \( G = 4 \), \( d_f = 5 \), \((\nu_1, \nu_2, \nu_3, \nu_4) = (3, 2.3, 1.7, 1.1)\) and \((d_1, d_2, d_3, d_4) = (1, 2, 1, 1)\).

**DGP 4** (Linear trend). In contrast to DGPs 1–3, in this DGP \( f_{t}^0 = (t, 1)' \) and \( \gamma_{t}^0 \sim (1, N(1, 2))' \), implying that \((d_1, d_2) = (1, 1)\) and \((\nu_1, \nu_2) = (3, 1)\). Hence, unlike before, now \( f_{t}^0 \) has a linear trend.

**DGP 5** (Weak factor). In this DGP, \( f_{t}^0 \sim (1, N(T^{-1/2}, T^{-1/3}))' \) and \( \gamma_{t}^0 \sim (N(1, 2), N(0, 2))' \), which means that \((d_1, d_2) = (1, 1)\) and \((\nu_1, \nu_2) = (1, 2/3)\). One of the factors is therefore weak.

**DGP 6** (Stochastic trend). Unlike in DGPs 1–5, in DGP 6

\[
x_{1,i,t} = N(0, 1) + \frac{1}{d_f} \sum_{n=1}^{d_f} (|f_{n,t}^0| + |\gamma_{g,i}^0|),
\]

\[
x_{2,i,t} = x_{2,i,t-1} + N(0, 1).
\]

Moreover, \( f_{1,t}^0 = f_{1,t-1}^0 + N(0, 1) \) and \( f_{2,t}^0 \sim N(\mathbf{0}_{2 \times 1}, \mathbf{I}_{2}) \), which means that \( G = 2 \), \((d_1, d_2) = (1, 2)\) and \((\nu_1, \nu_2) = (2, 1)\).

For each DGP, we make 1,000 replications of panels of size \( N, T \in \{40, 80, 160, 320\} \). For each combination of \( N \) and \( T \), we report the correct selection frequency for \((\hat{d}_1, \ldots, \hat{d}_G)\) when seen as an estimator of \((d_1, \ldots, d_G)\) and for \( \hat{d}_g \) individually for each group \( g = 1, \ldots, G \). Hence, while the former frequency captures the accuracy of the estimation of both \((d_1, \ldots, d_G)\) and \( G \), the latter frequency only captures the accuracy of the estimation of \( d_g \) for each \( g = 1, \ldots, G \) taking \( G \) as known. We also report the root mean squared error (RMSE) of \( \hat{\beta} \) and \( \mathbf{P}_{\hat{F}} \), as measured by the square root of the average of \( \|\hat{\beta} - \beta^0\| \) and \( \|\mathbf{P}_{\hat{F}} - \mathbf{P}_{F^0}\| \), respectively, over the replications. Moreover, in interest of comparison, we report the RMSE of the infeasible OLS
estimator of $\beta^0$ based on taking $F^0$ as known, $\hat{\beta}(F^0)$. The proposed IPC estimator is based on setting $\delta = 1$ and $d_{\text{max}} = \lfloor \min\{\sqrt{N}, \sqrt{T}\} \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of $x$.

Tables 1 and 2 contain the correct selection frequencies for all and each individual group of factors, respectively. The first thing to note is that the accuracy is very high. Careful inspection of the results for the individual groups suggests that the accuracy is generally degreasing in $g$, which is partly expected because the signal strength of the factors ($\nu_g$) decreases with increasing values of $g$. We also see that the accuracy of ($\hat{d}_1, \ldots, \hat{d}_G$) is typically lower than that of $\hat{d}_G$, suggesting that $G$ is generally over-estimated. This is reassuring given that the main concern here is not to underestimate $G$ (see, for example, Fan et al., 2013, and Moon and Weidner, 2015). The correct selection frequencies for DGP 5 are generally lower than those for DGPs 1 and 4, which is partly expected because $G = 2$ is the same in these DGPs and consistency in the presence of weak factors places stronger conditions on $N$ and $T$ than when all the factors are strong (Assumption 4).

Looking next at the RMSE results reported in Table 3, we see that there is a clear improvement as $N$ and/or $T$ increases. The rate at which this happens is roughly the same across the six DGPs. Because of the way we have defined it, the order of the RMSE will be dictated by the slowest growing regressor, which is given by $x_{1,i,t}$ (in DGP 6 $x_{1,i,t}$ and $x_{2,i,t}$ grow at the same rate). The order of this regressor is in turn driven by that of the slowest growing factor, $f^0_G$. Hence, since $\nu_G = 1$ in all DGPs but DGPs 1 and 5 in which $\nu_G = \nu_2 = 1.8$ and $\nu_G = \nu_2 = 2/3$, respectively, the rate at which the RMSE values come down should be roughly the same. The only exception is in DGP 1 where it should go faster, and we see that the performance is best in this DGP. Looking across the two estimators of $\beta^0$, we see the best overall performance is generally obtained when taking $F^0$ as known, which is in accordance with our a priori expectations. However, the improvement is not very large and it decreases with increases in $N$ and/or $T$. The reason for this is the accuracy of the estimated factors, which high and increasing in $N$ and/or $T$. The relatively good performance in DGP 1 is consistent with the relatively high value of $\nu_G$ in this case.

5 Empirical illustrations

5.1 Returns to scale

In this section, we estimate the returns to scale (RTS) of US banks, an area that has attracted considerable attention in the empirical production literature (see, for example, Feng and Serletis, 2008). Let us therefore denote by $CT_{i,t}$ the observed total cost of bank $i$ during period $t$. The total cost is a function $C$ of a $J \times 1$ vector of input prices $p_{i,t} = (p_{1,i,t}, \ldots, p_{J,i,t})'$ and a
A \times 1 \) vector of output quantities \( \mathbf{q}_{i,t} = (q_{1,i,t}, \ldots, q_{L,i,t})' \). Hence, \( CT_{i,t} = C(\mathbf{p}_{i,t}, \mathbf{q}_{i,t}) \), or when expressed in logs, \( \ln CT_{i,t} = \ln C(\mathbf{p}_{i,t}, \mathbf{q}_{i,t}) \). The RTS is simply the reciprocal of the sum of the output elasticities, which can be estimated for a given \( C \). In order to ensure that \( C \) is linearly homogenous in prices, however, it is very common to first normalize costs and prices by one of the prices, \( p_{J,i,t} \) say. Hence, \( \ln(CT_{i,t}/p_{J,i,t}) = \ln C(\mathbf{p}_{i,t}/p_{J,i,t}, \mathbf{q}_{i,t}) \), which in turn implies that the RTS is given by

\[
\text{RTS} = \left( \sum_{j=1}^{L} \frac{\partial \ln C(\mathbf{p}_{i,t}/p_{J,i,t}, \mathbf{q}_{i,t})}{\partial \ln q_{j,i,t}} \right)^{-1}.
\] (5.1)

In order to estimate this quantity, estimates of the partial derivatives are needed. The standard way to obtain such estimates in the literature is to assume that \( \ln C(\mathbf{p}_{i,t}/p_{J,i,t}, \mathbf{q}_{i,t}) \) is linear, to estimate the resulting linear relationship between \( \ln(CT_{i,t}/p_{J,i,t}) \), \( \ln(\mathbf{p}_{i,t}/p_{J,i,t}) \) and \( \ln \mathbf{q}_{i,t} \) by OLS, and to estimate the partial derivatives by simply replacing the parameters by their OLS estimates. A number of modelling issues then arise. First, as pointed out by Feng and Zhang (2012, page 1884), “[t]here are many ... types of ... unobserved heterogeneity in the US banking industry, which include, but are not limited to, location-related corporate income tax rate, property tax rate, and personal income tax rate.” Hence, there is a need to control for unobserved heterogeneity. Second, the literature typically assumes that the regressors in \( \ln(\mathbf{p}_{i,t}/p_{J,i,t}) \) and \( \ln \mathbf{q}_{i,t} \) are stationary, which, as pointed out by Dong et al. (2020), is highly unlikely to be the case in practice. There is therefore a need to consider approaches that do not require the regressors to be stationary. Third, to account for the fact that costs are typically trending, many researchers fit their models with linear and quadratic trend terms. Of course, deterministic trends can account for some trending behaviour, but not all. Moreover, the results tend to be highly sensitive to how the trending is modelled, suggesting that this is an important issue (see Feng and Serletis, 2008, for a discussion).

The discussion of the last paragraph suggests that there is a need for an approach that is general enough to accommodate not only unobserved heterogeneity but also trending behaviour. The proposed IPC approach fits this bill and we will therefore use it in this empirical illustration to the RTS of US commercial banks. The data that we will use for this purpose, which are the same as in Feng et al. (2017), are quarterly and cover the period 1986–2005, which means that \( T = 80 \). To avoid the impact of entry and exit, we only include continuously operating large banks with assets of at least one billion USD. This gives us a total of \( N = 466 \) banks. The data set contains observations on three input prices and output quantities. They are the wage rate of labour \( (p_{1,i,t}) \), the price of deposits and purchased funds \( (p_{2,i,t}) \), the price of physical capital \( (p_{3,i,t}) \), consumer loans \( (q_{1,i,t}) \), non-consumer loans, which is composed of
industrial, commercial and real estate loans \((q_{2,i,t})\), and securities, including non-loan financial assets \((q_{3,i,t})\). All outputs are deflated by the GDP deflator. Following the previous literature (see, for example, Stiroh, 2000), total cost \((CT_{i,t})\) is computed as the sum of total salaries and benefits divided by the number of full-time employees, the price of deposits and purchased funds equals total interest expense divided by total deposits and purchased funds, and the price of capital equals expenses on premises and equipment divided by premises and fixed assets. Hence, after normalization by \(p_{3,i,t}\), there are six variables included in the model; \(\ln(CT_{i,t}/p_{3,i,t})\), \(\ln(p_{1,i,t}/p_{3,i,t})\), \(\ln(p_{2,i,t}/p_{3,i,t})\), \(\ln q_{1,i,t}\), \(\ln q_{2,i,t}\) and \(\ln q_{3,i,t}\).

Figure 1: The variables in the RTS illustration.

In order to get a feeling for the trending behaviour of the data, in Figure 1 we plot all \(N\) series for each variable. A few observations are worthy of some discussion. First, there is strong co-movement among the series. This is true for all six variables, including the dependent variable, \(\ln(CT_{i,t}/p_{3,i,t})\), which we take as evidence in support of our interactive effects specification.
Second, the variables are highly persistent and most are trending over time, suggesting that they are non-stationary. This last observation is supported by a formal augmented Dickey–Fuller (ADF) test (with intercept and linear trend) applied to each series. The highest rejection frequency across all $N$ series is obtained for $\ln q_{2,i,t}$ and is given by 6.7%, which in turn suggests that the evidence against unit root null hypothesis is weak. In fact, since the ADF test does not account for the multiplicity of the testing problem, the true proportion of (trend-)stationary series is likely to be even lower. Third, it is unclear if the trend in $\ln(CT_{i,t}/p_{3,i,t})$ is the same as those in the regressors, which means that it is important to allow the factors to be trending, as otherwise any unaccounted for trend in $\ln(CT_{i,t}/p_{3,i,t})$ will be pushed into the regression errors.

Figure 2: The estimated factors in the RTS illustration.

We now go on to discuss the actual estimation results, which are reported in Table 4. The estimated model is given by (2.1) with $y_{i,t} = \ln(CT_{i,t}/p_{3,i,t})$, $\beta^0 = (\beta_1^0, \ldots, \beta_5^0)'$ and $x_{i,t} = (\ln(p_{1,i,t}/p_{3,i,t}), \ln(p_{2,i,t}/p_{3,i,t}), \ln q_{1,i,t}, \ln q_{2,i,t}, \ln q_{3,i,t})'$. Hence, in this notation,

$$\text{RTS} = \frac{1}{\beta_3^0 + \beta_4^0 + \beta_5^0}$$

(5.2)
is the object of interest, which we estimate using

\[ \hat{\text{RTS}} = \frac{1}{\hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_5}, \quad (5.3) \]

where \( \hat{\beta}_3, \hat{\beta}_4 \) and \( \hat{\beta}_5 \) are from \( \hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_5)' \). The values of \( \delta \) and \( d_{max} \) are set as in Section 4.

In order to enable inference not only for \( \beta^0 \) but also for the RTS, we follow Feng et al. (2017) and report 95% bootstrap confidence intervals.\(^4\) The estimated factor groups are given by \( \hat{d}_1 = \hat{d}_2 = 1 \), implying that the estimated number of factors equals \( \hat{d}_f = \hat{d}_1 + \hat{d}_2 = 2 \). In Figure 2, we plot the estimated factors, denoted \( \hat{f}_{1,t} \) and \( \hat{f}_{2,t} \). We see that while both factor estimates are trending, \( \hat{f}_{2,t} \) is growing much faster. As a measure of this difference in the trends, we look at

\[ \sum_{i=1}^{N} \hat{\gamma}_{1,i}^2 / \sum_{i=1}^{N} \hat{\gamma}_{2,i}^2. \]

By using the results provided in the appendix, we can show that this ratio should be \( O_P(T^{-1/2}) \), implying \( \ln(\sum_{i=1}^{N} \hat{\gamma}_{1,i}^2 / \sum_{i=1}^{N} \hat{\gamma}_{2,i}^2)/\ln T = \nu_1 - \nu_2 + o_P(1) \). By plugging in the known value of \( T \) and \( \sum_{i=1}^{N} \hat{\gamma}_{1,i}^2 / \sum_{i=1}^{N} \hat{\gamma}_{2,i}^2 = 123.4 \), we get \( \ln(\sum_{i=1}^{N} \hat{\gamma}_{1,i}^2 / \sum_{i=1}^{N} \hat{\gamma}_{2,i}^2)/\ln T = 1.1 \), which is thus an estimate of \( \nu_1 - \nu_2 \). Hence, there is indeed a substantial difference in the degree of trending of the two factors. But it is not as large as it would have been if \( \hat{f}_{1,t} \) and \( \hat{f}_{2,t} \) were made up of a quadratic and a linear trend, say, in which case \( \nu_1 = 5 \) and \( \nu_2 = 3 \), such that \( \nu_1 - \nu_2 = 2 \). Hence, just as expected given Figure 2, while trending, the behaviour of the factors is not just a simple deterministic function of time, which of course casts doubt on much of the existing research.

While our main interest is in the RTS, for completeness in Table 4, we also report the estimated elasticities. The first thing to note is that the RTS is estimated to be significantly larger than one, suggesting that the large US banks considered here have had increasing returns to scale during the period of investigation. This is consistent with the results reported by, for example, Feng et al. (2017) and Wheelock and Wilson (2012). One explanation for this finding is that the banks have been adopting productivity-enhancing internet technologies, leading to increased RTS. The confidence intervals are very narrow not only for the RTS, but also for the elasticities, which are all significant.

The above results are based on the proposed IPC approach. For comparison, we also estimated a fixed effects model in which \( f_{i,t}^0 = 1 \), and \( x_{i,t} \) is augmented by \( t \) and \( t^2 \) as additional regressors. We then compared the fit of the model relative to that of our estimated interactive effects model in terms of RMSE. Not surprisingly given the results reported above for the estimated factors, the RMSE of the interactive effects model, 0.0701, is smaller than that of the

\(^4\)In particular, we consider the following wild bootstrap procedure. We begin by estimating (2.1) using the IPC estimation procedure and calculating \( \hat{e}_{i,t} = y_{i,t} - \hat{y}_{i,t} \), where \( \hat{y}_{i,t} \) is the fitted value. For each bootstrap replication, we then generate \( y_{i,t}^* = \hat{y}_{i,t} + \hat{e}_{i,t} u_{i,t} \), where \( u_{i,t} \) is drawn randomly from \( N(0, 1) \). Finally, we estimate (2.1) based on \( y_{i,t}^* \) and \( x_{i,t} \), and store the values of the estimated coefficients and RTS for each replication.
trend-augmented fixed effects model, 0.0977, which reinforces the evidence in favor of interactive effects.

5.2 House prices and income

Economists have become concerned that recently house prices have grown too quickly, and that prices are now too high relative to per capita incomes. If this is correct and there is any truth to the theory on the matter, prices should stagnate or fall until they are better aligned with income, which in statistical terms mean that house prices should be cointegrated with income. The validity of this assumption has important implications for policy, because a failure could be due to a housing bubble.

In this section, we revisit the real house price data set of Holly et al. (2010), which comprises data on log real house prices ($p_{i,t}$) and log real per capita income ($w_{i,t}$) for 49 US states across the 1975–2003 period. According to theory, $p_{i,t}$ and $w_{i,t}$ should be cointegrated with cointegrating vector $(1, -1)'$. The previous empirical evidence of this prediction has, however, been mixed and far from convincing (see, for example, Gallin, 2006). Holly et al. (2010) argue that this lack of empirical support can be attributed in part to a failure to account for cross-sectional dependence, leading to deceptive conclusions. The authors therefore apply the “CIPS” panel unit root test of Pesaran (2007), which allow for cross-section dependence in the form of a common factor. The test is applied both to $p_{i,t}$ and $w_{i,t}$ separately, and to $p_{i,t} - w_{i,t}$. According to the results, while the variables are unit root non-stationary, their difference is not. Holly et al. (2010) also report CCE results suggesting that the estimated income elasticity is indeed close to one. They therefore conclude that $p_{i,t}$ and $w_{i,t}$ are cointegrated with cointegrating vector $(1, -1)'$, just as predicted by theory.

Our interest in the work of Holly et al. (2010) stems from their preference to apply the CIPS test, which tests for a unit root in the defactored data. This means that if $p_{i,t}$ and $w_{i,t}$ are not cointegrated by themselves, but only when conditioning on unit root common factors, because of the way that the data are defactored prior to the testing, the unit root null hypothesis is likely to be rejected by the CIPS test. That is, the test is likely to lead to the conclusion of cointegration when in fact there is none. In this section, we use IPC as a means to investigate this possibility.

As in the RTS illustration, we begin by plotting the variables. This is done in Figure 3. As expected, both variables are highly persistent and the ADF test provides no evidence against the unit root null. This corroborates the unit root test results reported by Holly et al. (2010). However, we also see that the trending behaviours of $p_{i,t}$ and $w_{i,t}$ are very different, suggesting that their stochastic trends are not the same, which they should be under cointegration. We
also see that the trending behaviour is very similar across states, which is suggestive of non-stationary common factors. Of course, the IPC procedure does not require cointegration and it does allow for very general types of factors. We therefore proceed with the estimation of the model. Hence, in this illustration, \( y_{i,t} = p_{i,t} \) and \( x_{i,t} = w_{i,t} \). For comparison purposes, the IPC results are presented together with the results obtained by applying the PC estimator of Bai (2009), as well as the usual OLS estimator with time and state fixed effects. The results reported in Table 5. The first thing to note is that the estimated slopes vary a lot depending on the estimator used. Interestingly, the point estimates are increasing in the generality of the estimator with fixed effects OLS (IPC) leading to the lowest (highest) estimate. We therefore begin by considering the IPC results. The point estimate of 2.1024 is far from the theoretically predicted value of one, which is also not included in the reported 95% confidence interval. As
for the factors, we estimate $\hat{d}_1 = \hat{d}_2 = \hat{d}_3 = 1$, implying that $\hat{d}_f = 3$. The estimated factors, denoted $\hat{f}_{1,t}$, $\hat{f}_{2,t}$ and $\hat{f}_{3,t}$, are plotted in Figure 4. As expected given Figure 3, all three factors are highly persistent, although with clearly distinct trends. We take this as evidence against cointegration between $p_{i,t}$ and $w_{i,t}$, since under cointegration the factors should be stationary.

Because we estimate three time-varying factors, fixed effects OLS is invalid, as it only allows for a common time effect. Moreover, since the factors come from three distinct groups, and are not all stationary, PC is invalid, too. This leaves us with the IPC estimator, which again provides strong evidence against the theoretically predicted one-to-one cointegrated relationship between $p_{i,t}$ and $w_{i,t}$. Holly et al. (2010, page 172) conclude that “[o]ur results support the hypothesis that real house prices have been rising in line with fundamentals (real incomes), and there seems little evidence of house price bubbles at the national level.” The results reported here reveal a completely different picture with housing prices being long run disconnected with real income.
6 Conclusion

The PC approach of Bai (2009) has attracted considerable interest in recent years, so much so that it has given rise to a separate PC literature. A key assumption in this literature is that both the unknown factors and regressors are stationary, which is rarely the case in practice. In the present paper, we relax this assumption by considering a very general specification in which the factors and regressors are essentially unrestricted. In spite of this generality, the proposed IPC estimator can be applied without any input from the practitioner, except for the maximum number of factors to be considered. The fact that in IPC there is no need to distinguish between deterministic and stochastic factors means that the usual problem in applied work of deciding on which deterministic terms to include in the model does not arise, as these are estimated along with the other factors of the model. There also no need to pre-test the regressors for unit roots, which is otherwise standard practice when using procedures that do not require data to be stationary. In other words, the proposed IPC is not only very general but also extremely user-friendly. It should therefore be a valuable addition to the already existing menu of techniques for panel regression models with interactive effects.
References


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


Table 1: Correct selection frequency for $(\hat{d}_1, \ldots, \hat{d}_{G})$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>DGP 1</th>
<th>DGP 2</th>
<th>DGP 3</th>
<th>DGP 4</th>
<th>DGP 5</th>
<th>DGP 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>40</td>
<td>1.000</td>
<td>0.744</td>
<td>0.726</td>
<td>0.945</td>
<td>0.909</td>
<td>0.898</td>
</tr>
<tr>
<td>80</td>
<td>1.000</td>
<td>0.945</td>
<td>0.892</td>
<td>0.986</td>
<td>0.946</td>
<td>0.977</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.987</td>
<td>0.970</td>
<td>0.934</td>
<td>0.996</td>
<td>0.936</td>
<td>0.996</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>0.870</td>
<td>0.691</td>
<td>0.704</td>
<td>0.947</td>
<td>0.894</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>40</td>
<td>1.000</td>
<td>0.826</td>
<td>0.789</td>
<td>0.998</td>
<td>0.957</td>
<td>0.936</td>
</tr>
<tr>
<td>80</td>
<td>1.000</td>
<td>0.988</td>
<td>0.925</td>
<td>1.000</td>
<td>0.982</td>
<td>0.993</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>0.999</td>
<td>0.994</td>
<td>0.990</td>
<td>1.000</td>
<td>0.992</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>0.968</td>
<td>0.826</td>
<td>0.821</td>
<td>1.000</td>
<td>0.987</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>40</td>
<td>1.000</td>
<td>0.959</td>
<td>0.803</td>
<td>1.000</td>
<td>0.983</td>
<td>0.962</td>
</tr>
<tr>
<td>80</td>
<td>1.000</td>
<td>0.998</td>
<td>0.967</td>
<td>1.000</td>
<td>0.999</td>
<td>0.995</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>0.999</td>
<td>0.997</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>1.000</td>
<td>0.957</td>
<td>0.917</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>40</td>
<td>1.000</td>
<td>0.987</td>
<td>0.772</td>
<td>1.000</td>
<td>0.959</td>
<td>0.957</td>
</tr>
<tr>
<td>80</td>
<td>1.000</td>
<td>1.000</td>
<td>0.968</td>
<td>1.000</td>
<td>0.999</td>
<td>0.997</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>320</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The table reports the correct selection frequency for the IPC method of step 2 of the estimation procedure. This means that both the groups and their number, $G$, are treated as unknown. DGPs 1–6 refer to the models with $G = 2$, $G = 3$, $G = 4$, linear trend, weak factors and stochastic trend, respectively. See Section 4 for a more detailed description of the Monte Carlo setup.
Table 2: Correct selection frequency for $\hat{d}_g$ for each group $g$.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$g = 1$</th>
<th>$g = 2$</th>
<th>$g = 3$</th>
<th>$g = 1$</th>
<th>$g = 2$</th>
<th>$g = 3$</th>
<th>$g = 4$</th>
<th>$g = 1$</th>
<th>$g = 2$</th>
<th>$g = 3$</th>
<th>$g = 4$</th>
<th>$g = 1$</th>
<th>$g = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>40</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.995</td>
<td>0.745</td>
<td>1.000</td>
<td>0.922</td>
<td>0.750</td>
<td>0.783</td>
<td>1.000</td>
<td>0.945</td>
<td>0.941</td>
</tr>
<tr>
<td>80</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.946</td>
<td>1.000</td>
<td>0.984</td>
<td>0.905</td>
<td>0.903</td>
<td>1.000</td>
<td>0.986</td>
<td>0.979</td>
<td>0.946</td>
</tr>
<tr>
<td>160</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.979</td>
<td>1.000</td>
<td>1.000</td>
<td>0.982</td>
<td>0.977</td>
<td>1.000</td>
<td>0.996</td>
<td>0.990</td>
<td>0.936</td>
</tr>
<tr>
<td>320</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.779</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.837</td>
<td>1.000</td>
<td>0.947</td>
<td>0.994</td>
<td>0.894</td>
</tr>
<tr>
<td>80</td>
<td>40</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.988</td>
<td>1.000</td>
<td>0.997</td>
<td>0.926</td>
<td>0.927</td>
<td>1.000</td>
<td>1.000</td>
<td>0.991</td>
<td>0.982</td>
</tr>
<tr>
<td>160</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>320</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.912</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.943</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.987</td>
</tr>
<tr>
<td>160</td>
<td>40</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.959</td>
<td>1.000</td>
<td>0.960</td>
<td>0.816</td>
<td>0.830</td>
<td>1.000</td>
<td>1.000</td>
<td>0.984</td>
<td>0.983</td>
</tr>
<tr>
<td>320</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td>1.000</td>
<td>0.967</td>
<td>0.967</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>320</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.959</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Notes: The table reports the correct selection frequencies for each of the estimated factor groups taking the number of groups, $G$, as known. See Table 1 for an explanation of the rest.
<table>
<thead>
<tr>
<th>DGP</th>
<th>N \ T</th>
<th>40</th>
<th>80</th>
<th>160</th>
<th>320</th>
<th>40</th>
<th>80</th>
<th>160</th>
<th>320</th>
<th>40</th>
<th>80</th>
<th>160</th>
<th>320</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>0.0381</td>
<td>0.0263</td>
<td>0.0175</td>
<td>0.0119</td>
<td>0.0358</td>
<td>0.0238</td>
<td>0.0149</td>
<td>0.0097</td>
<td>0.0660</td>
<td>0.0345</td>
<td>0.0448</td>
<td>0.0380</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0260</td>
<td>0.0180</td>
<td>0.0121</td>
<td>0.0081</td>
<td>0.0247</td>
<td>0.0163</td>
<td>0.0107</td>
<td>0.0068</td>
<td>0.0552</td>
<td>0.0342</td>
<td>0.0328</td>
<td>0.0224</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.0183</td>
<td>0.0123</td>
<td>0.0083</td>
<td>0.0056</td>
<td>0.0176</td>
<td>0.0114</td>
<td>0.0075</td>
<td>0.0050</td>
<td>0.0436</td>
<td>0.0246</td>
<td>0.0191</td>
<td>0.0154</td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>0.0129</td>
<td>0.0086</td>
<td>0.0059</td>
<td>0.0039</td>
<td>0.0124</td>
<td>0.0078</td>
<td>0.0053</td>
<td>0.0035</td>
<td>0.0251</td>
<td>0.0209</td>
<td>0.0138</td>
<td>0.0117</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>0.0425</td>
<td>0.0279</td>
<td>0.0185</td>
<td>0.0141</td>
<td>0.0364</td>
<td>0.0240</td>
<td>0.0153</td>
<td>0.0099</td>
<td>0.3355</td>
<td>0.4364</td>
<td>0.2896</td>
<td>0.3710</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0288</td>
<td>0.0187</td>
<td>0.0121</td>
<td>0.0087</td>
<td>0.0256</td>
<td>0.0167</td>
<td>0.0104</td>
<td>0.0067</td>
<td>0.2725</td>
<td>0.2491</td>
<td>0.2480</td>
<td>0.2204</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.0189</td>
<td>0.0131</td>
<td>0.0086</td>
<td>0.0057</td>
<td>0.0180</td>
<td>0.0121</td>
<td>0.0075</td>
<td>0.0049</td>
<td>0.1376</td>
<td>0.1637</td>
<td>0.1541</td>
<td>0.1039</td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>0.0133</td>
<td>0.0090</td>
<td>0.0060</td>
<td>0.0039</td>
<td>0.0127</td>
<td>0.0083</td>
<td>0.0052</td>
<td>0.0034</td>
<td>0.1252</td>
<td>0.1169</td>
<td>0.1051</td>
<td>0.1226</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>0.0407</td>
<td>0.0279</td>
<td>0.0190</td>
<td>0.0098</td>
<td>0.0364</td>
<td>0.0243</td>
<td>0.0156</td>
<td>0.0089</td>
<td>0.2465</td>
<td>0.2380</td>
<td>0.1449</td>
<td>0.1752</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0276</td>
<td>0.0183</td>
<td>0.0122</td>
<td>0.0085</td>
<td>0.0261</td>
<td>0.0167</td>
<td>0.0105</td>
<td>0.0067</td>
<td>0.1625</td>
<td>0.1350</td>
<td>0.1417</td>
<td>0.1358</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.0191</td>
<td>0.0127</td>
<td>0.0088</td>
<td>0.0058</td>
<td>0.0182</td>
<td>0.0118</td>
<td>0.0078</td>
<td>0.0049</td>
<td>0.1189</td>
<td>0.0835</td>
<td>0.0865</td>
<td>0.0948</td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>0.0132</td>
<td>0.0094</td>
<td>0.0060</td>
<td>0.0040</td>
<td>0.0129</td>
<td>0.0087</td>
<td>0.0053</td>
<td>0.0034</td>
<td>0.0857</td>
<td>0.0671</td>
<td>0.0598</td>
<td>0.0757</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>0.0545</td>
<td>0.0343</td>
<td>0.0244</td>
<td>0.0166</td>
<td>0.0448</td>
<td>0.0304</td>
<td>0.0215</td>
<td>0.0150</td>
<td>0.2917</td>
<td>0.3024</td>
<td>0.3544</td>
<td>0.3197</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0315</td>
<td>0.0224</td>
<td>0.0155</td>
<td>0.0107</td>
<td>0.0306</td>
<td>0.0220</td>
<td>0.0151</td>
<td>0.0102</td>
<td>0.1891</td>
<td>0.2060</td>
<td>0.1763</td>
<td>0.2492</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.0227</td>
<td>0.0153</td>
<td>0.0107</td>
<td>0.0076</td>
<td>0.0224</td>
<td>0.0151</td>
<td>0.0105</td>
<td>0.0074</td>
<td>0.1367</td>
<td>0.1946</td>
<td>0.1582</td>
<td>0.1709</td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>0.0158</td>
<td>0.0109</td>
<td>0.0077</td>
<td>0.0054</td>
<td>0.0157</td>
<td>0.0107</td>
<td>0.0075</td>
<td>0.0053</td>
<td>0.1408</td>
<td>0.1264</td>
<td>0.1187</td>
<td>0.1175</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>0.0448</td>
<td>0.0303</td>
<td>0.0212</td>
<td>0.0144</td>
<td>0.0412</td>
<td>0.0277</td>
<td>0.0193</td>
<td>0.0132</td>
<td>0.3260</td>
<td>0.2952</td>
<td>0.3368</td>
<td>0.4711</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0297</td>
<td>0.0211</td>
<td>0.0139</td>
<td>0.0096</td>
<td>0.0286</td>
<td>0.0200</td>
<td>0.0133</td>
<td>0.0093</td>
<td>0.2530</td>
<td>0.2101</td>
<td>0.3326</td>
<td>0.2907</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.0210</td>
<td>0.0144</td>
<td>0.0095</td>
<td>0.0068</td>
<td>0.0202</td>
<td>0.0140</td>
<td>0.0092</td>
<td>0.0066</td>
<td>0.1539</td>
<td>0.2116</td>
<td>0.1835</td>
<td>0.2007</td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>0.0148</td>
<td>0.0101</td>
<td>0.0069</td>
<td>0.0046</td>
<td>0.0145</td>
<td>0.0099</td>
<td>0.0067</td>
<td>0.0045</td>
<td>0.1274</td>
<td>0.1230</td>
<td>0.1358</td>
<td>0.1658</td>
</tr>
<tr>
<td>6</td>
<td>40</td>
<td>0.0227</td>
<td>0.0154</td>
<td>0.0103</td>
<td>0.0067</td>
<td>0.0176</td>
<td>0.0113</td>
<td>0.0076</td>
<td>0.0048</td>
<td>0.1906</td>
<td>0.2303</td>
<td>0.1868</td>
<td>0.1865</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>0.0159</td>
<td>0.0102</td>
<td>0.0069</td>
<td>0.0046</td>
<td>0.0120</td>
<td>0.0078</td>
<td>0.0049</td>
<td>0.0033</td>
<td>0.1574</td>
<td>0.1537</td>
<td>0.1411</td>
<td>0.1425</td>
</tr>
<tr>
<td></td>
<td>160</td>
<td>0.0113</td>
<td>0.0075</td>
<td>0.0048</td>
<td>0.0034</td>
<td>0.0089</td>
<td>0.0057</td>
<td>0.0037</td>
<td>0.0025</td>
<td>0.1209</td>
<td>0.0933</td>
<td>0.1026</td>
<td>0.1047</td>
</tr>
<tr>
<td></td>
<td>320</td>
<td>0.0079</td>
<td>0.0052</td>
<td>0.0033</td>
<td>0.0023</td>
<td>0.0062</td>
<td>0.0039</td>
<td>0.0026</td>
<td>0.0017</td>
<td>0.0887</td>
<td>0.0738</td>
<td>0.0646</td>
<td>0.0771</td>
</tr>
</tbody>
</table>

Notes: The reported RMSE results for $\hat{\beta}$ and $P_{\hat{F}}$ refer to the square root of the average of $\|\hat{\beta} - \beta^0\|$ and $\|P_{\hat{F}} - P^0_F\|$, respectively, across the Monte Carlo replications. The results for $\hat{\beta}(F^0)$ are the corresponding results for the infeasible OLS estimator based on taking $F^0$ as known. See Table 1 for an explanation of the rest.
Table 4: Empirical results for the RTS illustration.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Point estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_1$</td>
<td>0.4063</td>
<td>(0.3797, 0.4395)</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>0.0199</td>
<td>(0.0182, 0.0215)</td>
</tr>
<tr>
<td>$\hat{\beta}_3$</td>
<td>0.3805</td>
<td>(0.3752, 0.3850)</td>
</tr>
<tr>
<td>$\hat{\beta}_4$</td>
<td>0.0912</td>
<td>(0.0877, 0.0953)</td>
</tr>
<tr>
<td>$\hat{\beta}_5$</td>
<td>0.4726</td>
<td>(0.4671, 0.4801)</td>
</tr>
<tr>
<td>RTS</td>
<td>1.0590</td>
<td>(1.0520, 1.0636)</td>
</tr>
</tbody>
</table>

Notes: $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$ and $\hat{\beta}_5$ refer to the estimated IPC slopes of $\ln(p_{i,t}/p_{3,i,t})$, $\ln(p_{2,i,t}/p_{3,i,t})$, $\ln q_{1,i,t}$, $\ln q_{2,i,t}$ and $\ln q_{3,i,t}$, respectively, when the dependent variable is $\ln(CT_{i,t}/p_{3,i,t})$. “CI” and RTS refer to the bootstrap confidence intervals, and the RTS, respectively.

Table 5: Empirical results for the house price and income illustration.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Point estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>FE</td>
<td>0.3453</td>
<td>(0.2937, 0.3969)</td>
</tr>
<tr>
<td>PC</td>
<td>1.1602</td>
<td>(1.0238, 1.2944)</td>
</tr>
<tr>
<td>IPC</td>
<td>2.1024</td>
<td>(2.0374, 2.2112)</td>
</tr>
</tbody>
</table>

Notes: “FE”, “PC” and “IPC” refer to the two-way fixed effects OLS estimator, the PC estimator of Bai (2009) and the proposed IPC estimator, respectively, in a regression of $p_{i,t}$ onto $w_{i,t}$. “CI” refers to the bootstrap confidence intervals.
We begin this appendix by laying out the notation that will be used throughout. This is done in Section A.1, which in Section A.2 is followed by a discussion of the conditions needed in order to ensure that the asymptotic distribution of \( \hat{\beta} \) is correctly centered at zero. Sections A.3 and A.4 provide some auxiliary lemmas and their proofs, respectively. The proofs of the main results are provided in Section A.5.

### A.1 Notation

The matrices \( \Sigma_{F_0} \) and \( \Sigma_{I_0} \) have been defined in Assumption 1. In this appendix, we use \( \Sigma_{F_0} \) and \( \Sigma_{I_0} \) to denote the sub-matrices of \( \Sigma_{F_0} \) and \( \Sigma_{I_0} \) corresponding to \( T^{-\nu_g}F_0^gF_0^g \) and \( N^{-1}\Gamma_0^g\Gamma_0^g \), respectively, for \( g = 1, \ldots, G \). We also define \( V_g = \text{diag}(\lambda_{g,1}, \ldots, \lambda_{g,d_{max}}) \), where \( \lambda_{g,d} \) has been defined in Step 2 of the IPC estimation procedure. We partition \( F_0 = (F_0^1, \ldots, F_0^g, F_0^{g+g}) \) and \( C_T = \text{diag}(T^{-\nu_1/2}I_{d_1}, \ldots, T^{-\nu_g/2}I_{d_g}, C_{+g,T}) \), where \( F_0^g = (F_0^g, \ldots, F_0^g) \) is \( T \times (d_{g+1} + \cdots + d_G) \), and \( C_{+g,T} = \text{diag}(T^{-\nu_{g+1}/2}I_{d_{g+1}}, \ldots, T^{-\nu_G/2}I_{d_G}) \) is \( (d_{g+1} + \cdots + d_G) \times (d_{g+1} + \cdots + d_G) \). We partition \( \hat{F}, \gamma_0^g \) and \( \Gamma_0^g \) conformably as \( \hat{F} = (\hat{F}_1, \ldots, \hat{F}_g, \hat{F}_{g+g}) \), \( \gamma_0^g = (\gamma_0^{g,1}, \ldots, \gamma_0^{g,i}, \gamma_0^{g,j})^T \) and \( \Gamma_0^g = (\Gamma_0^g_1, \ldots, \Gamma_0^g_g, \Gamma_0^{g+g}) \), respectively.

We introduce \( \lambda_{g,d} = T^{-\nu_g}h_{g,d}^0F_0^g\Sigma_{g_0}^gF_0^g h_{g,d}^0 \), where \( \Sigma_{g_0}^g = N^{-1}F_0^g\Gamma_0^gF_0^g \) and \( h_{g,d}^0 \) is the \( d \)-th column of \( H_g^0 = N^{-1}T(\nu_g - 1/2)\Gamma_0^gF_0^g(V_g^0)^{-1} \) with \( \bar{F}_g^0 \) being the \( T \times d_g \) matrix consisting of the first \( d_g \) columns of \( \bar{F}_g \) and \( V_g^0 \) being the leading \( d_g \times d_g \) principal submatrix of \( V_g \). In other words, \( \bar{F}_g^0 \) and \( V_g^0 \) are \( \bar{F}_g \) and \( V_g \) based on treating the number of factors for each group, \( d_g \), as known. We also define \( H_g = T^{-\nu_g/2}H_g^0 \). In order to appreciate the implication of the difference in normalization with respect to \( T \), let us consider \( H_g^0 \). By Assumption 1, \( N^{-1}\Gamma_0^g\Gamma_0^g \) is asymptotically of full rank, and hence \( \|N^{-1}\Gamma_0^g\Gamma_0^g\| = O_p(1) \). Hence, since

\[
\|T^{-\nu_g/2}F_0^g\hat{F}_g\|^2 \leq \|\hat{F}_g\|^2 T^{-\nu_g/2}\|F_0^g\|^2 = T^{-\nu_g}\|F_0^g\|^2 = O_p(1).
\] (A.1.1)
and \( \|(T^{-\nu_1}V_0^0)^{-1}\| = O_P(1) \) by the proof of Lemma A.3, we can show that
\[
\|H^0_g\| \leq \|N^{-1}T^0_gT^0_g\|\|T^{-(\nu_g+\delta)/2}\|F^0_gF^0_g\|\|(T^{-\nu_0}V_0^0)^{-1}\| = O_P(1),
\]
which in turn implies
\[
\|H_g\| = T^{-(\nu_g-\delta)/2}\|H^0_g\| = O_P(T^{-(\nu_g-\delta)/2}).
\] (A.1.3)

We further use \( \hat{F}^0_{g,d} \) to refer to the \( d \)-th column of \( \hat{F}^0_g \). In this notation, \( \hat{\lambda}_{g,d} = T^{-\delta}\hat{F}^0_{g,d}\hat{\Sigma}_{g}\hat{F}^0_{g,d} \) for \( d = 1, \ldots, d_g \).

We also partition \( X_i \) as \( X_i = (X_{1,i}, \ldots, X_{d_i,i}) \) with \( X_{j,i} \) being the \( j \)-th column of \( X_i \). The \( j \)-th column of \( X_iD_T \) is therefore given by \( T^{-\kappa_j/2}X_{j,i} \).

Moreover, \( \text{vec} A, \text{rank} A, \text{span} A \) and \( \lambda(A) \) denote the vectorized version, rank, span and eigenvalues of \( A \), respectively, \( a \land b = \min\{a, b\} \) and \( a \lor b = \max\{a, b\} \).

We assume throughout that \( G \geq 2 \). The proofs for the cases when \( G = 0,1 \) are much simpler and can be obtained by manipulating the proofs for \( G \geq 2 \).

## A.2 Conditions that ensure asymptotic unbiasedness

In this section, we provide a set of assumptions that ensure that the asymptotic distribution of \( \sqrt{NTD_T}(\hat{\beta} - \beta^0) \) given in Theorem 3.1 is free of bias without for that matter requiring that \( \varepsilon_{i,t} \) is serially and cross-sectionally independent. One way to accomplish this is to assume that \( \rho_1 = \rho_2 = 0 \), as in Corollary 3.1. The assumptions considered here, which are stated in Assumption A.1, can be seen as alternatives to this last condition. In terms of the notation of Theorem 3.1, they ensure that \( A_1 = A_2 = 0_{d_x \times 1} \).

### Assumption A.1 (No asymptotic bias)

One of the following set of conditions is met:

- (a) \( T^{1-\nu_y}E(f^0_{g,t}f^0_{g,s}|x_t, x_s) = \phi_{ts} \) w.p.a.1 and \( \sum_{t=1}^T \sum_{s=1}^T |\phi_{ts}| = O(T) \), where \( x_t = (x_{1,t}, \ldots, x_{N,t})' \). If \( G \geq 2 \), then \( q < (\nu_G + \nu_{G-1})/2 - 1/4 \) and \( \nu_{g-1} - \nu_g > 1/2 \) for \( g = 2, \ldots, G \).

- (b) \( T/N \to c_4 \in (0, \infty) \) and \( \nu_G > 1 \). If \( G \geq 2 \), then \( q < (\nu_G + \nu_{G-1} - 1)/2 \) and \( \nu_{g-1} - \nu_g > 1/2 \).

- (c) \( T^{1-\nu_y-\kappa_j}E(\sum_{t=1}^T \sum_{s=1}^T x_{j,i,t}x_{j,k,s}f^0_{g,t}f^0_{g,s}) = O(T^{2-r}) \), where \( r < 2, r + \nu_G - 1 > 0 \) and \( x_{j,i,t} \) is the \( j \)-th row of \( x_{i,t} \). If \( G \geq 2 \), then \( \nu_{g-1} - \nu_g > 1/2 \).

Assumption A.1 ensures that the asymptotic distribution of \( \sqrt{NTD_T}(\hat{\beta} - \beta^0) \) is bias-free. It is, however, not necessary, and (a)–(b) should therefore be viewed as examples of conditions.
under which there is no asymptotic distribution bias. These conditions all have their strengths and weaknesses, and so their suitability will in general depend on the context. Take as an example condition (b), which has the advantage of not requiring any more moment conditions than those that are already in Assumption 1. It does, however, require that \( \nu_G > 1 \), which rules out both weak and stationary factors (\( \nu_G \leq 1 \)). Conditions (a) and (c) are more general in this regard, but then at the expense of requiring additional moment conditions. Note in particular how (a) rules out many types of trends in \( f_0^g, t \). For example, if \( f_0^g, t = t \), in which case \( \nu_G = 3 \), then \( \phi_{ts} = T^{-2} ts \) and hence \( \sum_{t=1}^{T} \sum_{s=1}^{T} |\phi_{ts}| = (T^{-1} \sum_{t=1}^{T} t)^2 = O(T^2) \). Condition (c) is a functional central limit theorem style moment condition.

The next corollary to Theorem 3.1 verifies that the asymptotic distribution of \( \sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) \) is indeed bias-free under Assumption A.1.

**Corollary A.1** (Unbiased asymptotic distribution). Suppose that Assumptions 1–6, 8, and A.1 are met and that \( N/T^{\nu_G} \to 0 \). Then, as \( N, T \to \infty \),

\[
\sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) \to_D MN(0_{d_x \times 1}, B_0^{-1}\Omega B_0^{-1}). \tag{A.2.1}
\]

The asymptotic distribution in Corollary A.1 is the same as the one given in Corollary 3.1. The difference is that while the latter requires \( \rho_1 = \rho_2 = 0 \), the former requires that Assumption A.1 is met, which in turn ensures that \( A_1 = A_2 = 0_{d_x \times 1} \).

### A.3 Auxiliary lemmas

**Lemma A.1.** Suppose that \( A \) and \( A + E \) are \( n \times n \) symmetric matrices and that \( Q = (Q_1, Q_2) \), where \( Q_1 \) is \( n \times r \) and \( Q_2 \) is \( n \times (n-r) \), is an orthogonal matrix such that span \( Q_1 \) is an invariant subspace for \( A \). Decompose \( Q'AQ \) and \( Q'EQ \) as \( Q'AQ = \text{diag}(D_1, D_2) \) and

\[
Q'EQ = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}. \tag{A.3.1}
\]

Let

\[
\text{sep}(D_1, D_2) = \min_{\lambda_1 \in \lambda(D_1), \lambda_2 \in \lambda(D_2)} |\lambda_1 - \lambda_2|. \tag{A.3.2}
\]

If \( \text{sep}(D_1, D_2) > 0 \) and \( \|E\|_2 \leq \text{sep}(D_1, D_2)/5 \), then there exists a \( (n-r) \times r \) matrix \( P \) with \( \|P\|_2 \leq 4\|E_{21}\|_2/\text{sep}(D_1, D_2) \), such that the columns of \( Q_1' = (Q_1 + Q_2P)(I_r + P'P)^{-1/2} \) define an orthonormal basis for a subspace that is invariant for \( A + E \).
Lemma A.2.

(a) \( \sup_{\mathbf{F} \in \mathbb{F}} (NT)^{-1} \sum_{i=1}^{N} \mathbf{F} \mathbf{e}_i = O_P (N^{-1} \sqrt{T^{-1}}) \);

(b) \( \sup_{\mathbf{F} \in \mathbb{F}} \| (NT)^{-1} \sum_{i=1}^{N} \mathbf{D}_T \mathbf{X}_i \mathbf{F} \mathbf{e}_i \| = O_P (N^{-1/2} \sqrt{T^{-1/2}}) \);

(c) \( (NT)^{-1} \| \mathbf{e} \mathbf{e}' \| = O_P (N^{-1/2} \sqrt{T^{-1}}) \) and \( (NT)^{-1} \| \mathbf{e}' \mathbf{e} \| = O_P (N^{-1/2} \sqrt{T^{-1}}) \);

(d) \( \| \mathbf{F}^0 \mathbf{e} \| = O_P (\sqrt{NT}) \),

where (a) and (b) hold under Assumption 1, and (c) and (d) hold under Assumptions 1, 3 and 4.

Lemma A.3. Let Assumptions 1-4 hold. Then, as \( N, T \to \infty \),

\[
T^{-\nu_1} \mathbf{V}_1 \to_P \mathbf{Y}_1, \tag{A.3.3}
\]

where \( \mathbf{Y}_1 \) is a \( d_{\max} \times d_{\max} \) matrix of zeroes except for the first \( d_1 \) elements on the main diagonal, which are given by the eigenvalues of \( \Sigma_{\mathbf{F}_1^0} \Sigma_{\mathbf{F}_1^0} \).

Lemma A.4. The following hold under the conditions of Lemma A.3:

(a) \( T^{-\nu_1} |\lambda_{1,d} - \lambda_{1,d}| = O_P (T^{-(\nu_1 - \nu_2)/2}) \) for \( d = 1, \ldots, d_1 \);

(b) \( T^{-\nu_1} |\lambda_{1,d}| = O_P (T^{-(\nu_1 - \nu_2)}) \) for \( d = d_1 + 1, \ldots, d_{\max} \).

Lemmas A.3–A.4 contain the required auxiliary results for the first factor group \((g = 1)\). We now move on to subsequent groups and in so doing we focus on the \( G = 2 \) case. The results for \( G > 2 \) follow by the same arguments, although the derivations naturally become more tedious. All derivations simplify considerably if \( q = -\infty \) so that \( \| \mathbf{F}^0_{g} \mathbf{T}^0_{h} \| = 0 \) for \( g \neq h \).

Lemma A.5. The following hold under Assumptions 1-5:

(a) \( T^{-\delta} \| \mathbf{F}^0_{g} \mathbf{F}_1 \| = O_P (T^{-(\delta + \nu_1 - \nu_2)/2} + N^{-1/2} T^{-(\delta + \nu_1 - \nu_2 - 1)/2} + N^{-(1-p)} T^{-(\delta + \nu_1 - 2\nu_2)/2} + T^{-(\nu_1 + \delta)/2 - q}) \);

(b) \( \sum_{i=1}^{N} \| \mathbf{F}^0_{1} \gamma_{1,i}^0 - \mathbf{F}_1 \hat{\gamma}_{1,i} \| = O_P (N \vee T + NT^{2q - \nu_1} + N^{-(1-2p)} T^{\nu_2}) \).

Lemma A.6. Let the conditions of Lemma A.5 hold. Then, as \( N, T \to \infty \),

\[
T^{-\nu_2/2} \mathbf{V}_2 \to_P \mathbf{Y}_2, \tag{A.3.4}
\]

where \( \mathbf{Y}_2 \) is a \( d_{\max} \times d_{\max} \) matrix of zeroes except for the first \( d_2 \) elements on the main diagonal, which are given by the eigenvalues of \( \Sigma_{\mathbf{F}_2^0} \Sigma_{\mathbf{F}_2^0} \).

4
Lemma A.7. Let \( \tau_{NT} = N^{-1/2}T^{-(\nu_2-1)/2} + T^{-\nu_2/2} + T^{-(\nu_2-\nu_3)} + T^{-(\nu_1+\nu_2)/2} + N^{-(1-p)} \). The following hold under the conditions of Lemma A.5:

(a) \( T^{-\nu_2} |\hat{\lambda}_{2,d} - \lambda_{2,d}| = O_P(\tau_{NT}) \) for \( d = 1, \ldots, d_2 \);

(b) \( T^{-\nu_2} |\hat{\lambda}_{2,d}| = O_P(\tau_{NT}^2) \) for \( d = d_2 + 1, \ldots, d_{\text{max}} \).

Lemma A.8. Under Assumptions 1–7 (a),

\[
\|D_T^{-1}(\hat{\beta}_1 - \hat{\beta}_0)\| = O_P((NT)^{-1/2} \vee \|D_T^{-1}(\hat{\beta}_0 - \beta^0)\|).
\] (A.3.5)

A.4 Proofs of auxiliary lemmas

Proof of Lemma A.1.

This is Lemma 3 of Lam et al. (2011). The proof is therefore omitted. ■

Proof of Lemma A.2.

Consider (a). We have

\[
\sup_{F \in \mathcal{D}_F} \frac{1}{NT} \sum_{i=1}^N \varepsilon_i' P_F \varepsilon_i = \sup_{F \in \mathcal{D}_F} (NT)^{-1} \text{tr}(P_F \varepsilon' \varepsilon) \\
\leq O(1) \sup_{F \in \mathcal{D}_F} (NT)^{-1} \|P_F\|_2 \|\varepsilon' \varepsilon\|_2 \\
\leq O(1) \sup_{F \in \mathcal{D}_F} (NT)^{-1} \|\varepsilon\|_2^2 = O_P(N^{-1} \vee T^{-1}),
\] (A.4.1)

where the first inequality follows from the fact that \( |\text{tr} A| \leq \text{rank} A \|A\|_2 \), and the second inequality follows from the fact that \( \|P_F\|_2 = 1 \), and the second equality hold by Assumption 1 (b).

The result in (b) is due to

\[
\sup_{F \in \mathcal{D}_F} \left\| \frac{1}{NT} \sum_{i=1}^N D_T X_j' P_F \varepsilon_i \right\| \\
\leq \frac{1}{NT} \sum_{j=1}^{d_x} \sup_{F \in \mathcal{D}_F} \left\| \sum_{i=1}^N T^{-\kappa_j/2} X_j' P_F \varepsilon_i \right\| \\
\leq \frac{1}{NT} \sum_{j=1}^{d_x} \sup_{F \in \mathcal{D}_F} |\text{tr} (T^{-\kappa_j/2} X_j P_F \varepsilon')| \\
\leq O(1) \frac{1}{NT} \sum_{j=1}^{d_x} \sup_{F \in \mathcal{D}_F} \|T^{-\kappa_j/2} X_j\|_2 \|P_F\|_2 \|\varepsilon\|_2
\]
\[ (NT)^{-1}O_P(\sqrt{NT})O_P(\sqrt{N} \lor \sqrt{T}) = O_P(N^{-1/2} \lor T^{-1/2}), \]  
(A.4.2)

where, with a slight abuse of notation and in this proof only, \( X_j = (X_{j,1}, \ldots, X_{j,N})' \). The third inequality here follows from \( |\text{tr} \ A| \leq \text{rank} \ A \| A \|_2 \), while the first equality is due to Assumption 1.

For (c), we use

\[
(NT)^{-2}E\|\epsilon' \epsilon\|^2 = \frac{1}{(NT)^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \sum_{i=1}^{N} E[\epsilon_{i,t}^2 \epsilon_{i,s}^2] + \sum_{i=1}^{N} \sum_{j \neq i} E[\epsilon_{i,t} \epsilon_{i,s} \epsilon_{j,t} \epsilon_{j,s}] \right) 
\]

\[
= \frac{1}{(NT)^2} \sum_{t=1}^{T} \left( \sum_{i=1}^{N} E[\epsilon_{i,t}^4] + \sum_{i=1}^{N} \sum_{j \neq i} E[(\epsilon_{i,t} \epsilon_{i,s} - \sigma_{\epsilon,ij})^2] \right) 
\]

\[
+ \frac{1}{(NT)^2} \sum_{t=1}^{T} \sum_{s \neq t} \left( \sum_{i=1}^{N} E[\epsilon_{i,t}^2 \epsilon_{i,s}^2] + \sum_{i=1}^{N} \sum_{j \neq i} E[(\epsilon_{i,t} \epsilon_{j,t} - \sigma_{\epsilon,ij})(\epsilon_{i,s} \epsilon_{j,s} - \sigma_{\epsilon,ij})] \right) 
\]

\[
+ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j \neq i} \sigma_{\epsilon,ij}^2 
\]

\[ = O(N^{-1}) + O(T^{-1}), \]  
(A.4.3)

where the third equality follows from using the mixing condition on \( \epsilon_{i,t} \epsilon_{j,t} \) across \( t \). The above result implies that

\[ (NT)^{-1}\|\epsilon' \epsilon\| = O_P(N^{-1/2}) + O_P(T^{-1/2}), \]  
(A.4.4)

as required for the first result in (c). The second follows from

\[
(NT)^{-2}E(\|\epsilon' \epsilon\|^2) = \frac{1}{N^2 T^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E(\epsilon_{i,t} \epsilon_{j,t} \epsilon_{i,s} \epsilon_{j,s}) 
\]

\[
= \frac{1}{(NT)^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \sum_{i=1}^{N} E(\epsilon_{i,t}^2 \epsilon_{i,s}^2) + \sum_{i=1}^{N} \sum_{j \neq i} E(\epsilon_{i,t} \epsilon_{i,s} \epsilon_{j,t} \epsilon_{j,s}) \right) 
\]

\[ = O(N^{-1}) + O(T^{-1}), \]  
(A.4.5)

where the last step follows by the same arguments used to establish the first result.
It remains to prove (d), which is a direct consequence of Assumptions 1 and 3, as seen from

\[
E(\|\Gamma_0^\nu \epsilon\|^2) = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{N} E(\gamma_i^\nu \gamma_j^\nu \epsilon_{i,t} \epsilon_{j,t}) \leq O(T) \sum_{i=1}^{N} \sum_{j=1}^{N} |\sigma_{\epsilon_{i,j}}| = O(NT). \tag{A.4.6}
\]

This establishes (d) and hence the proof of the lemma is complete. \qed

**Proof of Lemma A.3.**

As in Appendix A.1, decompose \( \mathbf{F}^0 = (\mathbf{F}_1^0, \mathbf{F}_{+1}^0) \) and \( \mathbf{C}_T = \text{diag}(T^{-\nu_1/2} \mathbf{I}_{d_1}, \mathbf{C}_{+1,T}) \), where \( \mathbf{F}_{+1}^0 = (\mathbf{F}_2^0, \ldots, \mathbf{F}_G^0) \) is \( T \times (d_f - d_1) \) and \( \mathbf{C}_{+1,T} = \text{diag}(T^{-\nu_2/2} \mathbf{I}_{d_1}, \ldots, T^{-\nu_G/2} \mathbf{I}_{d_G}) \) is \( (d_f - d_1) \times (d_f - d_1) \).

We partition \( \hat{\mathbf{F}}, \gamma_i^0 \) and \( \mathbf{F}^0 \) conformably as \( \hat{\mathbf{F}} = (\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_{+1}) \), \( \gamma_i^0 = (\gamma_{1,i}^0, \gamma_{+1,i}^0)' \) and \( \mathbf{F}^0 = (\mathbf{F}_1^0, \mathbf{F}_{+1}^0) \), respectively.

By the definition of the eigenvectors and eigenvalues, \( \hat{\Sigma}_1 \hat{\mathbf{F}}_1 = \hat{\mathbf{F}}_1 \mathbf{V}_1 \). By using this and the definition of \( \hat{\Sigma}_1 \),

\[
T^{-(\nu_1+\delta)/2} \hat{\mathbf{F}}_1 \mathbf{V}_1 \]
\[
= \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \mathbf{X}_i (\beta^0 - \hat{\beta}_i) (\beta^0 - \hat{\beta}_i)' \mathbf{X}_i \hat{\mathbf{F}}_1
\]
\[
+ \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \mathbf{X}_i (\beta^0 - \hat{\beta}_i) \gamma_{1,i}^0 \mathbf{F}_1 \hat{\mathbf{F}}_1 + \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \mathbf{F}_1 \gamma_{1,i}^0 (\beta^0 - \hat{\beta}_i)' \mathbf{X}_i \hat{\mathbf{F}}_1
\]
\[
+ \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \mathbf{X}_i (\beta^0 - \hat{\beta}_i) (\mathbf{F}_{+1} \gamma_{+1,i}^0 + \epsilon_i)' \hat{\mathbf{F}}_1
\]
\[
+ \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} (\mathbf{F}_{+1} \gamma_{+1,i}^0 + \epsilon_i) (\mathbf{F}_{+1} \gamma_{+1,i}^0 + \epsilon_i)' \hat{\mathbf{F}}_1
\]
\[
+ \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \mathbf{F}_1 \gamma_{1,i}^0 (\mathbf{F}_{+1} \gamma_{+1,i}^0 + \epsilon_i)' \hat{\mathbf{F}}_1 + \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} (\mathbf{F}_{+1} \gamma_{+1,i}^0 + \epsilon_i) \gamma_{1,i}^0 \mathbf{F}_1 \hat{\mathbf{F}}_1
\]
\[
+ \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \mathbf{F}_1 \gamma_{1,i}^0 \gamma_{1,i}^0 \mathbf{F}_1 \hat{\mathbf{F}}_1
\]
\[
= \sum_{j=1}^{9} \mathbf{J}_j, \tag{A.4.7}
\]
with implicit definitions of $J_1, \ldots, J_9$. Note that

$$J_9 = F_1^0(N^{-1} \Gamma_1^0 \Gamma_1^0)(T^{-(\nu_1+\delta)/2}F_1^0 \hat{F}_1). \tag{A.4.8}$$

Hence, moving this term over to the left-hand side, the above expression for $T^{-(\nu_1+\delta)/2} \hat{F}_1 V_1$ becomes

$$T^{-(\nu_1+\delta)/2} \hat{F}_1 V_1 - J_9 = \sum_{j=1}^{8} J_j. \tag{A.4.9}$$

We now evaluate each of the terms on the right-hand side.

Because $T^{-\delta/2} \|\hat{F}_1\|^2 = d_{\text{max}}$ and $(NT)^{-1} \sum_{i=1}^{N} \|X_i D_T\|^2 = O_p(1)$ by Assumption 1, the order of $J_1$ is given by

$$T^{-\delta/2} \|J_1\| \leq \left\| \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} X_i (\beta_0^0 - \hat{\beta}_0)(\beta_0^0 - \hat{\beta}_0)^T X_i^T \right\| T^{-\delta/2} \|\hat{F}_1\|$$

$$\leq O(1) \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \|X_i D_T D_T^{-1} (\beta_0^0 - \hat{\beta}_0)\|^2$$

$$= O_P(T^{-1-\nu_1+\delta/2} \|D_T^{-1}(\beta_0^0 - \hat{\beta}_0)\|^2). \tag{A.4.10}$$

Moreover, since

$$\frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} \|F_1^0 \gamma_{1,i}^0\|^2 \leq \frac{1}{N} \sum_{i=1}^{N} \|\gamma_{1,i}^0\|^2 T^{-\nu_1} \|F_1^0\|^2 = O_P(1) \tag{A.4.11}$$

by Assumption 1, we can show that

$$T^{-\delta/2} \|J_2\| \leq O(1) \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} \|X_i (\beta_0^0 - \hat{\beta}_0) \gamma_{1,i}^0 F_1^0\|^2$$

$$\leq O(1) \left( \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} \|X_i D_T D_T^{-1} (\beta_0^0 - \hat{\beta}_0)\|^2 \right)^{1/2} \left( \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} \|F_1^0 \gamma_{1,i}^0\|^2 \right)^{1/2}$$

$$= O_P(T^{-1-\delta/2} \|D_T^{-1}(\beta_0^0 - \hat{\beta}_0)\|), \tag{A.4.12}$$

and by exactly the same arguments,

$$T^{-\delta/2} \|J_3\| = O_P(T^{-1-\delta/2} \|D_T^{-1}(\beta_0^0 - \hat{\beta}_0)\|). \tag{A.4.13}$$
For $J_4$, we use
\[
\frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^N \| X_i D_T D_T^{-1}(\beta_0 - \hat{\beta}_0)(F_0^0 \gamma_0^{\nu_1+1,i})' \| \leq O(1) \left( \frac{1}{NT^{\delta}} \sum_{i=1}^N \| X_i D_T D_T^{-1}(\beta_0 - \hat{\beta}_0) \|^2 \right)^{1/2} \left( \frac{1}{NT^{\nu_1}} \sum_{i=1}^N \| F_0^0 \gamma_0^{\nu_1+1,i} \|^2 \right)^{1/2}
\]
\[
\leq O(1)O_P(T^{(1-\delta)/2}\| D_T^{-1}(\beta_0 - \hat{\beta}_0) \|)O_P(T^{-\nu_1/2}\| C_+^{-1,t} \|)
\]
\[
= O_P(T^{(1-\delta-\nu_1+\nu_2)/2}\| D_T^{-1}(\beta_0 - \hat{\beta}_0) \|),
\]
where the last equality makes use of the fact that $\| C_+^{-1,t} \| = O(T^{\nu_2/2})$, as $\nu_2 > \cdots > \nu_G$ by Assumption 1. We can further show that
\[
\frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^N \| X_i D_T D_T^{-1}(\beta_0 - \hat{\beta}_0) \epsilon_i' \|
\]
\[
= \left\| \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^N \text{vec} [ X_i D_T D_T^{-1}(\beta_0 - \hat{\beta}_0) \epsilon_i'] \right\|
\]
\[
\leq \left\| \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^N (\epsilon_i \otimes X_i D_T) \right\| \| D_T^{-1}(\beta_0 - \hat{\beta}_0) \|
\]
\[
= O_P(N^{-1/2}T^{(2-\nu_1-\delta)/2}\| D_T^{-1}(\beta_0 - \hat{\beta}_0) \|),
\]
where the last equality holds, because by Assumption 3 and $(\text{tr } A'B)^2 \leq (\text{tr } A'A)(\text{tr } B'B)$, we have
\[
E \left\| \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^N (\epsilon_i \otimes X_i D_T) \right\|^2
\]
\[
= \frac{1}{N^2T^{\nu_1+\delta}} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N \sum_{j=1}^N E(x_{j,t}'D_T^2x_{i,t} \epsilon_i \epsilon_j \epsilon_{i,s} \epsilon_{j,s})
\]
\[
= \frac{1}{N^2T^{\nu_1+\delta}} \sum_{s=1}^T \sum_{i=1}^N \sum_{j=1}^N E(x_{j,t}'D_T^2x_{i,t}) \sigma_{\epsilon,ij}
\]
\[
\leq \frac{1}{N^2T^{\nu_1+\delta}} \sum_{i=1}^N \sum_{j=1}^N E \text{tr}(D_T X_j'X_i D_T) | \sigma_{\epsilon,ij}|
\]
\[
\leq \frac{1}{N^2T^{\nu_1+\delta-2}} \sum_{i=1}^N \sum_{j=1}^N \sqrt{T^{-1}E \| D_T X_j \|^2} \sqrt{T^{-1}E \| X_i D_T \|^2} | \sigma_{\epsilon,ij}|
\]
\[ T^{-\delta/2} \| J_4 \| \leq O(1) \frac{1}{NT^{(\nu_1+\delta)/2}} \left\| \sum_{i=1}^{N} X_i(\beta^0 - \widehat{\beta}_0)(F^0_{+1}\gamma^0_{+1,i} + \varepsilon_i)' \right\| \]
\[ \leq O(1) \frac{1}{NT^{(\nu_1+\delta)/2}} \left\| \sum_{i=1}^{N} X_iD_TD_T^{-1}(\beta^0 - \widehat{\beta}_0)(F^0_{+1}\gamma^0_{+1,i})' \right\| \]
\[ + O(1) \frac{1}{NT^{(\nu_1+\delta)/2}} \left\| \sum_{i=1}^{N} X_iD_TD_T^{-1}(\beta^0 - \widehat{\beta}_0)\varepsilon_i' \right\| \]
\[ = O(1)O_P(T^{(1-\delta-\nu_1+\nu_2)/2} \| D_T^{-1}(\beta^0 - \widehat{\beta}_0) \|) \]
\[ + O(1)O_P(N^{-1/2}T^{(2-\nu_1-\delta)/2} \| D_T^{-1}(\beta^0 - \widehat{\beta}_0) \|) \]
\[ = o_P(T^{-(1-\delta)/2} \| D_T^{-1}(\beta^0 - \widehat{\beta}_0) \|), \quad (A.4.17) \]

where we have used Assumptions 1 and 4 \((T/N^2 = O(1) \text{ under } \nu_G < 1)\) to show that \(T^{(1-\delta-\nu_1+\nu_2)/2} \text{ and } N^{-1/2}T^{(2-\nu_1-\delta)/2}\) are \(o(1)\). The same steps can be used to show that

\[ T^{-\delta/2} \| J_5 \| \leq o_P(T^{-(1-\delta)/2} \| D_T^{-1}(\beta^0 - \widehat{\beta}_0) \|). \quad (A.4.18) \]

For \( J_6 \), we use

\[ T^{-\delta/2} \| J_6 \| \leq O_P(1) \left\| \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} (F^0_{+1}\gamma^0_{+1,i} + \varepsilon_i)(F^0_{+1}\gamma^0_{+1,i} + \varepsilon_i)' \right\| \]
\[ \leq O_P(1)(N^{-1}T^{-(\nu_1+\delta)/2} \| F^0_{+1}\Gamma^0_{+1}\Gamma^0_{+1}F^0_{+1} \| + N^{-1}T^{-(\nu_1+\delta)/2} \| \varepsilon' \varepsilon \|) \]
\[ + 2N^{-1}T^{-(\nu_1+\delta)/2} \| F^0_{+1}\Gamma^0_{+1} \|). \quad (A.4.19) \]

By Assumption 1,

\[ N^{-1}T^{-(\nu_1+\delta)/2} \| F^0_{+1}\Gamma^0_{+1} \| \leq T^{-\nu_2} \| F^0_{+1} \|^2 N^{-1} \| \Gamma^0_{+1} \| = O_P(T^{\nu_2-(\nu_1+\delta)/2}). \quad (A.4.20) \]

Another application of Assumption 1 and Lemma A.2 gives

\[ N^{-1}T^{-(\nu_1+\delta)/2} \| \varepsilon' \varepsilon \| = O_P(T^{1-(\nu_1+\delta)/2}(N^{-1/2} \lor T^{-1/2})), \quad (A.4.21) \]
\[ N^{-1}T^{-(\nu_1+\delta)/2} \| F^0_{+1}\Gamma^0_{+1} \| \leq N^{-1/2}T^{-(\nu_1+\delta-\nu_2-1)/2}T^{-\nu_2/2} \| F^0_{+1} \| (NT)^{-1/2} \| \Gamma^0_{+1} \|. \]
\[ = O_P(N^{-1/2}T^{-(\nu_1+\delta-\nu_2-1)/2}). \]  

(A.4.22)

These results can be inserted into the expression for \( T^{-\delta/2}\|J_6\| \), giving

\[ T^{-\delta/2}\|J_6\| = O_P(T^{\nu_2-(\nu_1+\delta)/2}) + O_P(N^{-1/2}T^{1-(\nu_1+u)/2}). \]  

(A.4.23)

Next up is \( J_7 \). By using Assumptions 1 and 4, and Lemma A.2, and the arguments use in evaluating \( J_6 \),

\[ T^{-\delta/2}\|J_7\| \leq O(1)\left\| \frac{1}{NT^{(\nu_1+\delta)/2}} \sum_{i=1}^{N} F_1^0 T_i^0 (F_{+1}^0 \gamma_{1,i} + \epsilon_i) \right\| \]

\[ \leq O(1)N^{-1}T^{-(\nu_1+\delta)/2}\|F_1^0 T_1^0 F_{+1}^0 \| + O(1)N^{-1}T^{-(\nu_1+\delta)/2}\|F_1^0 F_{+1}^0 \epsilon \| \]

\[ = O_P(T^{(\nu_2-\delta)/2}) + O_P(N^{-1/2}T^{(1-\delta)/2}) = O_P(T^{(\nu_2-\delta)/2}), \]  

(A.4.24)

and we can similarly show that

\[ T^{-\delta/2}\|J_8\| = O_P(T^{(\nu_2-\delta)/2}). \]  

(A.4.25)

By putting everything together, (A.4.9) becomes

\[ T^{-\delta/2}T^{-(\nu_1+\delta)/2} \tilde{F}_1 V_1 - J_9 \| \leq \sum_{j=1}^{8} T^{-\delta/2}\|J_j\| = O_P(T^{-(\delta-1)/2}\|D_T^{-1}(\beta^0 - \tilde{\beta}_0)\|) + O_P(N^{-1/2}T^{1-(\nu_1+\delta)/2}) \]

+ \( O_P(T^{-(\delta-\nu_2)/2}) \).  

(A.4.26)

We now left multiply (A.4.9) by \( T^{-(\nu_1+\delta)}F_1^{0r} \) and make use of this last result to obtain

\[ \|T^{-(\nu_1+\delta)/2}F_1^{0r}\tilde{F}_1(T^{-(\nu_1)}V_1) - T^{-(\nu_1+\delta)/2}F_1^{0r}J_9\| \]

\[ \leq T^{-(\nu_1-\delta)/2}(T^{-(\nu_1)/2}\|F_1^{0r}\|) \sum_{j=1}^{8} T^{-\delta/2}\|J_j\| \]

\[ = T^{-(\nu_1-\delta)/2}O_P(1)[O_P(T^{-(\delta-1)/2}\|D_T^{-1}(\beta^0 - \tilde{\beta}_0)\|) + O_P(N^{-1/2}T^{1-(\nu_1+\delta)/2}) + O_P(T^{-(\delta-\nu_2)/2})] \]

\[ = O_P(T^{-(\nu_1-1)/2}\|D_T^{-1}(\beta^0 - \tilde{\beta}_0)\|) + O_P(T^{-(\nu_1-\nu_2)/2}) \]

\[ = O_P(T^{-(\nu_1-\nu_2)/2}), \]  

(A.4.27)

where the third equality follows from Assumption 4 and Lemma 3.1. Further use of Assumption
shows that the term on the left can be written as

\[
T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1(T^{-\nu_1}V_1) = T^{-\nu_1}F_{1j}^0
\]

\[
= (T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1)(T^{-\nu_1}V_1) - (T^{-\nu_1}F_{1j}^0)\left(N^{-1}\Gamma_{1}^0\Gamma_{1}^0\right)(T^{-(\nu_1+\delta)/2}\hat{F}_1^0)\hat{F}_1
\]

\[
= (T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1)(T^{-\nu_1}V_1) - \Sigma_{\hat{F}_1^0}\Sigma_{\Gamma_{1}}(T^{-(\nu_1+\delta)/2}\hat{F}_1^0) + o_P(1).
\] (A.4.28)

Hence, since \(T^{-(\nu_1-\nu_2)/2} = o(1)\) by Assumption 1,

\[
\Sigma_{\hat{F}_1^0}\Sigma_{\Gamma_{1}}(T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1) = (T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1)(T^{-\nu_1}V_1) + o_P(1),
\] (A.4.29)

which is another (limiting) eigenvalue-eigenvector relation with \(T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1\) being the eigenvectors of \(\Sigma_{\hat{F}_1^0}\Sigma_{\Gamma_{1}}\), while \(T^{-\nu_1}V_1\) contains the associated eigenvalues. Hence, since \(\Sigma_{\hat{F}_1^0}\Sigma_{\Gamma_{1}}\) is full rank by Assumption 1 and \(T^{-\nu_1}V_1 = \text{diag}(T^{-\nu_1}V_1^0, 0_{(d_{max}-d_1)\times(d_{max}-d_1)}), T^{-\nu_1}V_1^0\) is invertible. It follows that

\[
T^{-\nu_1}V_1 \to P \hat{\Psi}_1
\] (A.4.30)
as \(N, T \to \infty\). This completes the proof of the lemma.

**Proof of Lemma A.4.**

Consider (a). This proof starts with (A.4.9) in the proof of Lemma A.3. Because here \(d \leq d_1\), we can without loss of generality assume that \(d_1\) is known such that \(\hat{F}_1^0\) and \(V_1^0\) can be replaced by \(\hat{F}_1^0\) and \(V_1^0\), respectively. Let us write the left-hand side of the resulting equation as

\[
T^{-(\nu_1+\delta)/2}\hat{F}_1^0V_1 - J_0^0 = T^{(\nu_1-\delta)/2}\hat{F}_1^0T^{-\nu_1}V_1 - F_{1j}^0(N^{-1}\Gamma_{1}^0\Gamma_{1}^0)(T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1)
\]

\[
= [T^{(\nu_1-\delta)/2}\hat{F}_1^0 - F_{1j}^0(N^{-1}\Gamma_{1}^0\Gamma_{1}^0)(T^{-(\nu_1+\delta)/2}\hat{F}_1^0\hat{F}_1)](T^{-\nu_1}V_1^{-1})(T^{-\nu_1}V_1^{-1})
\]

\[
= (T^{(\nu_1-\delta)/2}\hat{F}_1^0 - F_{1j}^0H_{1j}^0)(T^{-\nu_1}V_1^{-1}),
\] (A.4.31)

implying that (A.4.9) can be written

\[
T^{(\nu_1-\delta)/2}\hat{F}_1^0 - F_{1j}^0H_{1j}^0 = \sum_{j=1}^{8} J_j^0(T^{-\nu_1}V_1^{-1}),
\] (A.4.32)

where \(J_1^0, \ldots, J_9^0\) are \(J_1, \ldots, J_9\) as defined in Proof of Lemma A.3, except that now \(d_1\) is taken
as known. Note that

\[ \sum_{j=1}^{8} T^{-\nu_1/2} \| \mathbf{J}_j^0 \| \]

\[ = T^{-(\nu_1-\delta)/2} \sum_{j=1}^{8} T^{-\delta/2} \| \mathbf{J}_j^0 \| \]

\[ = T^{-(\nu_1-\delta)/2} \left[ O_P(T^{-(\delta-1)/2} \| \mathbf{D}_T^{-1} (\mathbf{\beta}^0 - \widehat{\mathbf{\beta}}_0) \| ) + O_P(N^{-1/2} T^{1-(\nu_1+\delta)/2}) + O_P(T^{-(\delta-\nu_2)/2}) \right] \]

\[ = O_P(T^{-(\nu_1-1)/2} \| \mathbf{D}_T^{-1} (\mathbf{\beta}^0 - \widehat{\mathbf{\beta}}_0) \| ) + O_P(T^{-(\nu_1-\nu_2)/2}) \]

\[ = O_P(T^{-(\nu_1-\nu_2)/2}), \quad (A.433) \]

where the last equality holds since \( \| \mathbf{D}_T^{-1} (\mathbf{\beta}^0 - \widehat{\mathbf{\beta}}_0) \| = O_P(N^{-1/2} \vee T^{-1/2}) \) by Lemma 3.1. Moreover, since \( T^{-\nu_1} \mathbf{V}_1^0 \) converges to a full rank matrix by the proof of Lemma A.3, we have \( \|(T^{-\nu_1} \mathbf{V}_1^0)^{-1}\| = O_P(1) \), which in turn implies

\[ T^{-\nu_1/2} \| T^{(\nu_1-\delta)/2} \widehat{\mathbf{F}}_1^0 - \mathbf{F}_1^0 \mathbf{H}_1^0 \| \leq \sum_{j=1}^{8} T^{-\nu_1/2} \| \mathbf{J}_j^0 \| \| (T^{-\nu_1} \mathbf{V}_1^0)^{-1} \| = O_P(T^{-(\nu_1-\nu_2)/2}). \quad (A.434) \]

This is an important result and in what follows we will use it frequently.

Let us now consider \( T^{-\nu_1} (\widehat{\lambda}_{1,d} - \lambda_{1,d}) \). By the definitions of \( \lambda_{1,d} \) and \( \widehat{\lambda}_{1,d} \) given in Section A.1,

\[ T^{-\nu_1} (\widehat{\lambda}_{1,d} - \lambda_{1,d}) \]

\[ = T^{-(\nu_1+\delta)} \widehat{\mathbf{F}}_{1,d}^0 \mathbf{\Sigma}_{1,d} \widehat{\mathbf{F}}_{1,d}^0 - T^{-2\nu_1} \mathbf{h}_{1,d}^0 \mathbf{F}_1^0 \mathbf{\Sigma}_1^0 \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \]

\[ = (T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 + T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) T^{-\nu_1} (\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^0 + \mathbf{\Sigma}_1^0) \]

\[ \times (T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 + T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) - T^{-2\nu_1} \mathbf{h}_{1,d}^0 \mathbf{F}_1^0 \mathbf{\Sigma}_1^0 \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \]

\[ = (T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) T^{-\nu_1} (\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^0) T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \]

\[ + 2(T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) T^{-\nu_1} (\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^0) T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \]

\[ + (T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) T^{-\nu_1} \mathbf{\Sigma}_1^0 (T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) \]

\[ + 2(T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0) T^{-3\nu_1/2} \mathbf{\Sigma}_1^0 \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \]

\[ + T^{-2\nu_1} \mathbf{h}_{1,d}^0 \mathbf{F}_1^0 (\mathbf{\Sigma}_1 - \mathbf{\Sigma}_1^0) \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \]

\[ = J_1 + 2J_2 + J_3 + 2J_4 + J_5, \quad (A.435) \]

with obvious definitions of \( J_1, \ldots, J_5 \). From (A.434),

\[ T^{-\nu_1/2} \| T^{(\nu_1-\delta)/2} \widehat{\mathbf{F}}_{1,d}^0 - \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \| = \| T^{-\delta/2} \widehat{\mathbf{F}}_{1,d}^0 - T^{-\nu_1/2} \mathbf{F}_1^0 \mathbf{h}_{1,d}^0 \| = O_P(T^{-(\nu_1-\nu_2)/2}), \quad (A.436) \]
which is $o_P(1)$ under Assumption 1. This implies $|J_1| = o_P(|J_5|)$, $|J_2| = o_P(|J_5|)$ and $|J_3| = o_P(|J_4|)$. It remains to consider $J_4$ and $J_5$. The order of the first of these terms is given by

$$|J_4| \leq \|T^{-\delta/2} \hat{F}_{1,d}^0 - T^{-\nu_1/2} \hat{F}_1^0 \| \|T^{-3\nu_1/2} \Sigma_1^0 \hat{F}_1^0 \|$$

$$\leq \|T^{-\delta/2} \hat{F}_{1,d}^0 - T^{-\nu_1/2} \hat{F}_1^0 \| \|T^{-\nu_1/2} \| F_1^0 \| \| \Sigma_1^0 \hat{F}_1^0 \| \| h_{1,d}^0 \|$$

$$= O_P(T^{-(\nu_1 - \nu_2)/2}),$$

(A.4.37)

where we have made use of the fact that $\|H_0^0\| = O_P(1)$ (see Section A.1), which implies that $\|h_{1,d}^0\|$ is of the same order.

The order of $J_5$ is the same as that of $J_4$. In order to appreciate this, we begin by noting how

$$T^{-\nu_1}(\hat{\Sigma}_1 - \Sigma_1^0) = \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} X_i (\beta^0 - \hat{\beta}_0)(\beta^0 - \hat{\beta}_0)' X_i'$$

$$+ \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} X_i (\beta^0 - \hat{\beta}_0)\gamma_{1,i}^0 F_1^0 + \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} F_1^0 \gamma_{1,i}^0 (\beta^0 - \hat{\beta}_0)' X_i'$$

$$+ \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} X_i (\beta^0 - \hat{\beta}_0)(F_4^0 + \gamma_{1,i}^0 + \epsilon_i)'$$

$$+ \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} (F_4^0 + \gamma_{1,i}^0 + \epsilon_i)'$$

$$+ \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} F_1^0 \gamma_{1,i} (F_4^0 + \gamma_{1,i}^0 + \epsilon_i)'$$

$$+ \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} F_1^0 \gamma_{1,i} (F_4^0 + \gamma_{1,i}^0 + \epsilon_i)'$$

$$+ \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} F_1^0 \gamma_{1,i} (F_4^0 + \gamma_{1,i}^0 + \epsilon_i)'$$

$$= T^{-(\nu_1 - \delta)/2} \sum_{j=1}^{8} J_j,$$

(A.4.38)

which holds, because $\Sigma_1^0 = N^{-1} F_1^0 T_1^0 T_1^0 F_1^0 = N^{-1} \sum_{i=1}^{N} F_1^0 \gamma_{1,i}^0 \gamma_{1,i}^0 F_1^0$. Hence, since by the proof of Lemma A.3,

$$T^{-(\nu_1 - \delta)/2} \sum_{j=1}^{8} \| T^{-\delta} J_j^0 \| = T^{-(\nu_1 - \delta)/2} [O_P(T^{-(\delta - 1)/2} \| D_T^{-1}(\beta^0 - \hat{\beta}_0)\|)]$$

14
we can show that

\[ T^{-\nu_1}\|\hat{\Sigma}_1 - \Sigma_1^0\| = O_P(T^{-(\nu_1-\nu_2)/2}), \quad (A.4.39) \]

we can show that

\[ T^{-\nu_1}\|\hat{\Sigma}_1 - \Sigma_1^0\| = O_P(T^{-(\nu_1-\nu_2)/2}), \quad (A.4.40) \]

and so

\[ |J_5| \leq \|h_{1,d}^0\|2T^{-\nu_1}\|F_1^0\|^2T^{-\nu_1}\|\hat{\Sigma}_1 - \Sigma_1^0\| = O_P(T^{-(\nu_1-\nu_2)/2}). \quad (A.4.41) \]

Hence, by putting everything together,

\[ T^{-\nu_1}|\hat{\lambda}_{1,d} - \lambda_{1,d}| \leq |J_1| + 2|J_2| + |J_3| + 2|J_4| + |J_5| = O_P(T^{-(\nu_1-\nu_2)/2}), \quad (A.4.42) \]

which establishes (a).

Consider (b). This proof is based on Lemma A.1. We therefore start by introducing some notation in order to make the problem here fit the one in Lemma A.1. Let us therefore denote by \( F_1^+ \) a \( T \times (d_{\text{max}} - d_1) \) matrix such that \( T^{-\nu_1}(F_1^+, F_1^0R)(F_1^+, F_1^0R) = \text{diag}(I_{d_{\text{max}}-d_1}, I_{d_1}) \), where \( R \) is a \( d_1 \times d_1 \) rotation matrix. The matrices \( T^{-\nu_1/2}F_1^+, T^{-\nu_1/2}F_1^0R \), \( \Sigma_1^0 \) and \( \hat{\Sigma}_1 - \Sigma_1^0 \) correspond to \( Q_1, Q_2, A \) and \( E \) of Lemma A.1. Our counterpart of the matrix \( Q_1^0 \) appearing in this other lemma is thus given by

\[ \hat{F}_1^+ = T^{-\nu_1/2}(F_1^+, F_1^0RP)(I_{d_{\text{max}}-d_1} + P'P)^{-1/2}, \quad (A.4.43) \]

where

\[ \|P\|_2 \leq \frac{4}{\text{sep}(0, T^{-2\nu_1}F_1^0\Sigma_1^0F_1^0)}T^{-\nu_1}\|\hat{\Sigma}_1 - \Sigma_1^0\| \leq O_P(1)T^{-\nu_1}\|\hat{\Sigma}_1 - \Sigma_1^0\| = O_P(T^{-(\nu_1-\nu_2)/2}). \quad (A.4.44) \]

Since \( \hat{F}_1^+ \) is an orthonormal basis for a subspace that is invariant for \( \hat{\Sigma}_1 \), we have \( \hat{\lambda}_{1,d_1+d} = \hat{F}_d^+\hat{\Sigma}_1\hat{F}_d^+ \), where \( d = 1, \ldots, d_{\text{max}} - d_1 \) and \( \hat{F}_d^+ \) is the \( d \)-th column of \( \hat{F}_1^+ \). Consider \( \|\hat{F}_1^+ - T^{-\nu_1/2}F_1^+\|_2 \). By the definition of \( \hat{F}_1^+ \),

\[ \|\hat{F}_1^+ - T^{-\nu_1/2}F_1^+\|_2 = T^{-\nu_1/2}\|F_1^+ + F_1^0RP - F_1^+(I_{d_{\text{max}}-d_1} + P'P)^{1/2}(I_{d_{\text{max}}-d_1} + P'P)^{-1/2}\|_2 \]
where the second and third inequalities follow from (Magnus and Neudecker, 2007, Exercise 1 on page 231). This last result can be used to show that

\[ T^{-\nu_1/2} \| F_1^\perp (I_{d_{\text{max}}-d_1} - (I_{d_{\text{max}}-d_1} + P'P)^{1/2}) (I_{d_{\text{max}}-d_1} + P'P)^{-1/2} \|_2 \]

\[ + T^{-\nu_1/2} \| F_0^0 R P (I_{d_{\text{max}}-d_1} + P'P)^{-1/2} \|_2 \]

\[ \leq \| (I_{d_{\text{max}}-d_1} - (I_{d_{\text{max}}-d_1} + P'P)^{1/2}) (I_{d_{\text{max}}-d_1} + P'P)^{-1/2} \|_2 + \| P (I_{d_{\text{max}}-d_1} + P'P)^{-1/2} \|_2 \]

\[ \leq \| I_{d_{\text{max}}-d_1} - (I_{d_{\text{max}}-d_1} + P'P)^{1/2} \|_2 + \| P \|_2 \leq 2 \| P \|_2 \]

\[ = O_P(T^{-(\nu_1-\nu_2)/2}), \tag{A.4.45} \]

where \( F_1^\perp \) is the \( d \)-th column of \( F_1^\perp \), and the last equality follows from (A.4.45) and the proof of part (a). This completes the proof of the lemma. \( \blacksquare \)

Lemmas A.5–A.7 suppose that Assumptions 1–5 hold, and under these conditions we know from Lemmas 3.2 and 3.3 that \( \hat{d}_1, \ldots, \hat{d}_G \) are consistent. We therefore assume that \( d_1, \ldots, d_G \) are known. In order to keep the notation as tidy as possible, we suppress the zero superscript in \( \hat{F}_g^0 \) and \( V_g^0 \).

**Proof of Lemma A.5.**

For (a), we take the same starting point as in the proof of part (a) in Lemma A.4, which is (A.4.9) with \( \hat{F}_1 \) and \( V_1 \) based on treating \( d_1 \) as known. The rationale for doing so is, as already explained, that \( \hat{d}_1 \) is consistent. Pre-multiplying this equation through by \( T^{-(\nu_1+\delta)/2} F_2^{0g} \) gives

\[ T^{-(\nu_1+\delta)} F_2^{0g} \hat{F}_1 V_1 - T^{-(\nu_1+\delta)/2} F_2^{0g} J_9 \]

\[ = T^{-(\nu_1+\delta)} F_2^{0g} \hat{F}_1 V_1 - T^{-(\nu_1+\delta)/2} F_2^{0g} F_1^{0g} (N^{-1} G_1 T_1^{0g})(T^{-(\nu_1+\delta)/2} F_2^{0g} \hat{F}_1) \]

\[ = \sum_{j=1}^{8} T^{-(\nu_1+\delta)/2} F_2^{0g} J_j, \tag{A.4.47} \]
where the last equality follows from Assumption 5 and Lemma A.2. We can similarly show that

\[
T^{-(\delta + \nu_2)/2} \sum_{j=1}^{8} T^{-(\delta + \nu_2)/2} F_{2}^{0} J_{j} (T^{\nu_1} V_{1})^{-1},
\]

(A.4.48)

Under Assumption 5, the orders of \( J_{1}, \ldots, J_{6} \) are the same as in the proof of Lemma A.3. The stated orders of \( J_{7} \) and \( J_{8} \) are, however, not sharp and can be improved upon. The order of \( T^{-(\delta + \nu_2)/2} \| F_{2}^{0} J_{7} \| \) is given by

\[
T^{-(\delta + \nu_2)/2} \| F_{2}^{0} J_{7} \| \leq O_{P}(1) T^{\nu_2/2} \left\| \frac{1}{NT^{(\nu_1 + \delta)/2}} \sum_{i=1}^{N} F_{2}^{0} \gamma_{1, i} \left( F_{i+1}^{0} \gamma_{i+1, i} + \varepsilon_{i} \right) \right\|
\]

\[
\leq O_{P}(1) \frac{1}{NT^{(\nu_2 + \nu_1 + \delta)/2}} \| F_{2}^{0} F_{1}^{0} \| T^{0} F_{0}^{0} + 1 \| F_{2}^{0} F_{1}^{0} \| \varepsilon \|
\]

\[
= O_{P}(N^{-1/2} T^{\nu_2 - (\nu_2 + \nu_1 + \delta - 1)/2}) + O_{P}(N^{-1/2} T^{\nu_2 - (\nu_2 + \nu_1 + \delta - 1)/2}),
\]

(A.4.49)

where the last equality follows from Assumption 5 and Lemma A.2. We can similarly show that

\[
T^{-(\delta + \nu_2)/2} \| F_{2}^{0} J_{8} \| \leq O_{P}(1) T^{\nu_2/2} \left\| \frac{1}{NT^{(\nu_1 + \delta)/2}} \sum_{i=1}^{N} F_{2}^{0} \left( F_{2}^{0} \gamma_{2, i} + F_{i+2}^{0} \gamma_{i+2, i} + \varepsilon_{i} \right) \gamma_{i, i}^{0} F_{1}^{0} \right\|
\]

\[
\leq O_{P}(1) \frac{1}{NT^{(\nu_2 + \nu_1 + \delta)/2}} \| F_{2}^{0} F_{2}^{0} \| T^{0} F_{0}^{0} + 1 \| F_{2}^{0} F_{1}^{0} \| \varepsilon \|
\]

\[
+ O_{P}(1) N^{-1} T^{-(\nu_2 + \nu_1 + \delta)/2} \| F_{2}^{0} F_{2}^{0} \| T^{0} F_{0}^{0} + 1 \| F_{2}^{0} F_{1}^{0} \| \varepsilon \|
\]

\[
= O_{P}(N^{-1/2} T^{\nu_2 - (\delta - \nu_2)/2}) + O_{P}(N^{-1/2} T^{\nu_2 - (\delta - \nu_2)/2}),
\]

(A.4.50)

where \( F_{i+2}^{0} \) and \( T_{i+2}^{0} \) are defined analogously to \( F_{i+1}^{0} \) and \( T_{i+1}^{0} \) in the proof of Lemma A.3. By using these last two results together with the orders of \( J_{1}, \ldots, J_{6} \) given in the proof of Lemma A.3,
\[ \begin{align*} &= O_P(T^{\nu_2-(\nu_1+\delta)/2}) + O_P(T^{-\delta/2}) + O_P(N^{-1/2}T^{1-(\nu_1+\delta)/2}) + O_P(N^{-1/2}T^{(\delta-1)/2}) \\
&+ O_P(N^{-(1-p)}T^{-(\delta-\nu_2)/2}) \\
&= O_P(T^{-\delta/2}) + O_P(N^{-1/2}T^{-(\delta-1)/2}) + O_P(N^{-(1-p)}T^{-(\delta-\nu_2)/2}), \quad (A.4.51) \end{align*} \]

where the first equality follows from the fact that \( \|D_T^{-1}(\beta^0 - \hat{\beta}_0)\| = O_P(N^{-1/2} \vee T^{-1/2}) \) by Lemma 3.1, while the second is due to Assumptions 3 and 5. Hence,

\[ \begin{align*} &\|T^{-\delta}F_2^0\hat{F}_1 - T^{-(\nu_1+\delta)/2}F_2^0J_9(T^{-\nu_1}V_1)^{-1}\| \\
&\leq T^{-(\nu_1-\nu_2)/2} \left\| \sum_{j=1}^{8} T^{-(\delta+\nu_2)/2}F_2^0J_j \right\| \| (T^{-\nu_1}V_1)^{-1} \| \\
&= T^{-(\nu_1-\nu_2)/2}[O_P(T^{-\delta/2}) + O_P(N^{-1/2}T^{-(\delta-1)/2}) + O_P(N^{-(1-p)}T^{-(\delta-\nu_2)/2})] \\
&= O_P(T^{-(\delta+\nu_1-\nu_2)/2}) + O_P(N^{-1/2}T^{-(\delta+\nu_1-\nu_2-1)/2}) + O_P(N^{-(1-p)}T^{-(\delta+\nu_1-2\nu_2)/2}), \quad (A.4.52) \end{align*} \]

which together with Assumption 5 yields

\[ \begin{align*} &\|T^{-\delta}F_2^0\hat{F}_1\| \leq \|T^{-(\nu_1+\delta)/2}F_2^0J_9\| \| (T^{-\nu_1}V_1)^{-1} \| \\
&+ T^{-(\nu_1-\nu_2)/2} \left\| \sum_{j=1}^{8} T^{-(\delta+\nu_2)/2}F_2^0J_j \right\| \| (T^{-\nu_1}V_1)^{-1} \| \\
&= O_P(T^{0-(\nu_1+\delta)/2}) + O_P(T^{-(\delta+\nu_1-\nu_2)/2}) + O_P(N^{-1/2}T^{-(\delta+\nu_1-\nu_2-1)/2}) \\
&+ O_P(N^{-(1-p)}T^{-(\delta+\nu_1-2\nu_2)/2}), \quad (A.4.53) \end{align*} \]

as was to be shown for (a).

Let us now consider (b). Analogously to the proof of (a), by invoking Assumption 5 we can improve the orders of \( J_7 \) and \( J_8 \). For \( J_7 \),

\[ \begin{align*} T^{-\delta/2}\|J_7\| &\leq O_P(1)N^{-1}T^{-(\nu_1+\delta)/2}\|F_1^0\hat{F}_1 - F_1^0F_1^0\| + O_P(1)N^{-1}T^{-(\nu_1+\delta)/2}\|F_1^0\hat{F}_1\| \\
&= O_P(N^{-1-p}T^{-(\delta-\nu_2)/2}) + O_P(N^{-1/2}T^{-(\delta-1)/2}), \quad (A.4.54) \end{align*} \]

where the equality follows from Assumption 5 and Lemma A.2. For \( J_8 \),

\[ \begin{align*} T^{-\delta/2}\|J_8\| &\leq O_P(N^{-1-p}T^{-(\delta-\nu_2)/2}) + O_P(N^{-1/2}T^{-(\delta-1)/2}). \quad (A.4.55) \end{align*} \]

This implies that the result in (A.4.34) changes to (after replacing \( T^{-\nu_1/2} \) by \( T^{-\delta/2} \))

\[ T^{-\delta/2}\|T^{(\nu_1-\delta)/2}\hat{F}_1 - F_1^0H_1^0\| \]
Hence, since $H_1$ and Lemma 3.1, we get
\[
\sum_{j=1}^{8} T^{-\delta/2} \|J_j\| \| (T^{-\nu_1} V_1)^{-1} \|
\leq O_P(T^{-\delta/2}) + O_P(N^{-1/2} T^{-(\delta-1)/2}) + O_P(N^{-(1-p)} T^{-(\delta-\nu_2)/2}). \tag{A.4.56}
\]

We now evaluate each of the terms on the right-hand side one by one. Making use of Assumption 3.1 and by another application of Lemma A.2,
\[
T^{-\delta/2} \|\hat{F}_1 H_1^{-1} - F_1^0\| = O_P(T^{-\delta/2}) + O_P(N^{-1/2} T^{-(\delta-1)/2}) + O_P(N^{-(1-p)} T^{-(\delta-\nu_2)/2}). \tag{A.4.57}
\]

We are now ready to consider $\sum_{i=1}^{N} \| F_1^0 \gamma_{1,i}^0 - \hat{F}_i \hat{\gamma}_{1,i} \|^2$.
\[
\sum_{i=1}^{N} \| F_1^0 \gamma_{1,i}^0 - \hat{F}_1 \hat{\gamma}_{1,i} \|^2
= \sum_{i=1}^{N} \| F_1^0 \gamma_{1,i}^0 - T^{-\delta} \hat{F}_1 \hat{F}_1'(y_i - X_i \hat{\beta}_0) \|^2
\leq O(1) \sum_{i=1}^{N} [\| P_{\hat{F}_1} X_i (\beta^0 - \hat{\beta}_0) \|^2 + \| M_{\hat{F}_1} F_1^0 \gamma_{1,i}^0 \|^2 + \| P_{\hat{F}_1} F_1^0 \gamma_{1,i}^0 \|^2 + \| P_{\hat{F}_1} \epsilon_i \|^2]. \tag{A.4.58}
\]

We now evaluate each of the terms on the right-hand side one by one. Making use of Assumption 1 and Lemma 3.1, we get
\[
\sum_{i=1}^{N} \| P_{\hat{F}_1} X_i (\beta^0 - \hat{\beta}_0) \|^2 \leq \sum_{i=1}^{N} \| X_i D_T \|^2 \| D_T^{-1} (\beta^0 - \hat{\beta}_0) \|^2
= O_P(1) O_P(NT) O_P(N^{-1} \vee T^{-1}) = O_P(N \vee T), \tag{A.4.59}
\]
and by another application of Lemma A.2,
\[
\sum_{i=1}^{N} \| P_{\hat{F}_1} \epsilon_i \|^2 \leq \sum_{i=1}^{N} \| \epsilon_i \|^2 = O_P(N \vee T). \tag{A.4.60}
\]

For $\sum_{i=1}^{N} \| M_{\hat{F}_1} F_1^0 \gamma_{1,i}^0 \|^2$, we use (A.4.57) from which it follows that
\[
\| M_{\hat{F}_1} F_1^0 \|^2 = \| M_{\hat{F}_1} (F_1^0 - \hat{F}_1 H_1^{-1}) \|^2 \leq T^\delta (T^{-\delta} \| F_1^0 - \hat{F}_1 H_1^{-1} \|^2)
\leq O_P(1) + O_P(N^{-1}T) + O_P(N^{-2(1-p)} T^{\nu_2}), \tag{A.4.61}
\]
which in turn implies
\[
\sum_{i=1}^{N} \|M_{Z_i} \mathbf{F}_1^0 \gamma_{1,i}^0 \|^2 = O_P(N) + O_P(T) + O_P(N^{-1-2p} T^{\nu_2}).
\] (A.4.62)

For \(\sum_{i=1}^{N} \| P_{Z_i} \mathbf{F}_1^0 \gamma_{1,i}^0 \|^2\), we use the result given in part (a), giving
\[
\sum_{i=1}^{N} \| P_{Z_i} \mathbf{F}_1^0 \gamma_{1,i}^0 \|^2 = \sum_{i=1}^{N} \| T^{-\delta} \mathbf{F}_1 (T^{-\delta} \mathbf{F}_2 \gamma_{2,i}) - \mathbf{F}_1 \gamma_{1,i}^0 \|^2
\]
\[\leq T^{-\delta/2} \| \mathbf{F}_1 \|^2 \sum_{i=1}^{N} \| T^{-\delta/2} \mathbf{F}_2 \gamma_{2,i}^0 \|^2
\]
\[= O_P(N) T^\delta \| T^{-\delta} \mathbf{F}_2 \gamma_{2,i}^0 \|^2
\]
\[= O_P(N T^\delta) [O_P(T^{-\delta/2})] + O_P(T^{-\delta/2})
\]
\[+ O_P(N^{-1/2} T^{-\delta/2}) + O_P(N^{-1/2} T^{-\delta/2})]
\[= O_P(N) [O_P(T^{-\delta/2})] + O_P(N^{-1/2} T^{-\delta/2})
\]
\[+ O_P(N^{-1/2} T^{-\delta/2}) + O_P(T^{-\delta/2})]
\] (A.4.63)

The order of \(\sum_{i=1}^{N} \| P_{Z_i} \mathbf{F}_1^0 \gamma_{1,i}^0 \|^2\) is the same. Hence, by adding the results,
\[
\sum_{i=1}^{N} \| P_{Z_i} \mathbf{F}_1^0 \gamma_{1,i}^0 \|^2 \leq \sum_{i=1}^{N} \| P_{Z_i} \mathbf{F}_2 \gamma_{2,i}^0 \|^2 + \sum_{i=1}^{N} \| P_{Z_i} \mathbf{F}_1^0 \gamma_{1,i}^0 \|^2
\]
\[= O_P(N \vee T) + O_P(N^{-1-2p} T^{\nu_2}) + O_P(N) [O_P(T^{-\nu_2})]
\]
\[+ O_P(N^{-1-2p} T^{-\nu_2}) + O_P(N^{-1-2p} T^{-\nu_2}) + O_P(T^{-\nu_2})]
\[= O_P(N \vee T) + O_P(N T^{\nu_2}) + O_P(N^{-1-2p} T^{\nu_2}),
\] (A.4.64)

as required for (b).

\[\begin{align*}
\text{Proof of Lemma A.6.} \\
\text{Let } U_i = F_{1+\nu_2}^0 \gamma_{1,i}^0 + F_{1}^0 \gamma_{1,i}^0 - F_{1} \gamma_{1,i}. \text{ In this notation,}
\end{align*}\]
\[
T^{-(\nu_2+\delta)/2} \mathbf{F}_2 \mathbf{V}_2
\]
\[
= \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} [(X_i(\beta^0 - \hat{\beta}_0) + F_{2}^0 \gamma_{2,i} + U_i + \varepsilon_i))
\]

20
\[
\times (X_i(\beta^0 - \hat{\beta}_0) + F^o_2\gamma_{2,i} + U_i + \varepsilon_i)' \hat{F}_2
\]

\[
= \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} X_i(\beta^0 - \hat{\beta}_0)(\beta^0 - \hat{\beta}_0)'X_i' \hat{F}_2
\]

\[
+ \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} X_i(\beta^0 - \hat{\beta}_0) \gamma'^r_{2,i} F^o_2 \hat{F}_2 + \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} F^o_2 \gamma'^r_{2,i} (\beta^0 - \hat{\beta}_0)'X_i' \hat{F}_2
\]

\[
+ \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i \varepsilon_i' \hat{F}_2 + \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i \gamma'^r_{2,i} \varepsilon_i' \hat{F}_2 + \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_i \gamma'^r_{2,i} F^o_2 \hat{F}_2
\]

\[
+ \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} U_i'U_j' \hat{F}_2 + \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} U_i' \gamma'^r_{2,i} U_j' \hat{F}_2 + \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} U_i' \gamma'^r_{2,i} F^o_2 \hat{F}_2
\]

\[
= \sum_{j=1}^{16} K_j, \quad (A.4.65)
\]

where \( K_1, \ldots, K_{16} \) are implicitly defined. Analogously to the proof of Lemma A.3 we move \( K_{16} = F^o_2(N^{-1}F^o_2T^o_2)(T^{-(\nu_2 + \delta)/2}F^o_2 \hat{F}_2) \) over to the left, giving

\[
\quad T^{-(\nu_2 + \delta)/2} \hat{F}_2 V_2 - F^o_2(N^{-1}F^o_2T^o_2)(T^{-(\nu_2 + \delta)/2}F^o_2 \hat{F}_2) = \sum_{j=1}^{15} K_j. \quad (A.4.66)
\]

By using the same steps employed in the proof of Lemma A.3 (a), we can show that

\[
T^{-\delta/2}\|K_1 + K_2 + K_3\| = O_P(T^{-(\delta-1)/2}\|D^{-1}_T(\beta^0 - \hat{\beta}_0)\|). \quad (A.4.67)
\]

For \( K_4, \)

\[
T^{-\delta/2}\|K_4\|
\]

\[
= T^{-\delta/2} \left\| \frac{1}{NT^{(\nu_2 + \delta)/2}} \sum_{i=1}^{N} X_i(\beta^0 - \hat{\beta}_0)U_i' \hat{F}_2 \right\|
\]

21
where the second equality follows Lemma A.5, the third follows from Assumptions 1 and Assumption 5, and the fourth follows from Assumptions 3 and Assumption 5. The same arguments can be used to show that

\[
T^{-\delta/2}\|\mathbf{K}_6\| = o_P(T^{-\delta/2}\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|). \tag{A.4.69}
\]

For \(\mathbf{K}_6\),

\[
T^{-\delta/2}\|\mathbf{K}_6\| = T^{-\delta/2}\left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \mathbf{X}_i (\beta^0 - \hat{\beta}_0) \varepsilon_i \hat{\mathbf{F}}_2 \right\| 
\leq T^{-\delta/2}\|\hat{\mathbf{F}}_2\| \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \mathbf{X}_i (\beta^0 - \hat{\beta}_0) \varepsilon_i \right\|
= O_P(1)O_P(N^{-1/2}T^{-\delta+\nu_2-2}/2\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|)
= o_P(T^{-\delta/2}\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|), \tag{A.4.70}
\]

where the development is similar to (A.4.15). The order of \(T^{-\delta/2}\|\mathbf{K}_7\|\) is the same.

For \(\mathbf{K}_8\),

\[
T^{-\delta/2}\|\mathbf{K}_8\| = T^{-\delta/2}\left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i \hat{\mathbf{F}}_2 \right\| \leq T^{-\delta/2}\|\hat{\mathbf{F}}_2\| N^{-1}T^{-(\nu_2+\delta)/2}\|\varepsilon'\varepsilon\|
= O_P(T^{1-(\nu_2+\delta)/2}(N^{-1/2} \vee T^{-1/2})), \tag{A.4.71}
\]

where the last equality holds by Lemma A.2.
Further use of Lemma A.2 gives

\[ T^{-\delta/2} \|K_9\| = T^{-\delta/2} \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} F_1^0 y_{1,i} \mathbb{I}_i \hat{F}_2 \right\| \leq N^{-1} T^{-(\nu_2+\delta)/2} \|F_1^0 \| \left\| F_2^0 \right\| \|T^{-\delta/2} \| \hat{F}_2 \| \]

\[ = O_P(N^{-1/2} T^{-(\delta-1)/2}), \quad (A.4.72) \]

and we can show that \( T^{-\delta/2} \|K_{10}\| \) is of the same order.

\( K_{11} \) requires more work. We begin by expanding it in the following way:

\[ T^{-\delta/2} \|K_{11}\| \]
\[ = T^{-\delta/2} \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i (F_1^0 y_{1,i} - \hat{F}_1 \tilde{y}_{1,i} + F_{+2}^0 y_{2,i}) \hat{F}_2 \right\| \]
\[ \leq \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i (F_1^0 y_{1,i} - \hat{F}_1 T^{-\delta} \hat{F}_1 (y_i - X_i \beta_0) + F_{+2}^0 y_{2,i}) \right\| T^{-\delta/2} \| \hat{F}_2 \| \]
\[ \leq O_P(1) \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i (\beta^0 - \hat{\beta}_0)' X_i \mathbb{P}_1 \right\| \]
\[ + O_P(1) \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i \gamma_{+1,i}^0 F_{+1}^0 \mathbb{F}_1 \right\| + O_P(1) \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i \gamma_{+2,i}^0 F_{+2}^0 \mathbb{F}_1 \right\| \]
\[ = O_P(1)(K_{111} + K_{112} + K_{113} + K_{114} + K_{115}), \quad (A.4.73) \]

where, similarly to the analysis of \( K_6 \) and using \( \|\mathbb{P}_1 \| = 1 \),

\[ K_{111} \leq \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} \varepsilon_i (\beta^0 - \hat{\beta}_0)' X_i \right\| \]
\[ = O_P(N^{-1/2} T^{-(\delta+\nu_2-2)/2} \|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|) \]
\[ = o_P(T^{-(\delta-1)/2} \|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|). \quad (A.4.74) \]

Also, making use of (A.4.61), we can show that

\[ K_{112} \leq N^{-1} T^{-(\nu_2+\delta)/2} \|\mathbb{F}_1 \| \|\mathbb{M}_1 \| \|\mathbb{F}_1 \| \]
\[ = O_P(1) N^{-1} T^{-(\nu_2+\delta)/2} O_P(1) + O_P(N^{-1/2} \sqrt{T}) + O_P(N^{-1-p} T^{\nu_2/2}) \]
\[ = O_P(N^{-3/2} T^{-(\delta-1)/2}) + O_P(N^{-1} T^{1-(\nu_2+\delta)/2}) + O_P(N^{-3/2} T^{-(\delta-1)/2}). \quad (A.4.75) \]
For $K_{113}$,
\[
K_{113} \leq N^{-1}T^{-(\nu_2+\delta)/2}\|T_+^{i+1}\mathbf{e}\|\|\mathbf{F}_+\mathbf{F}_+^0\| \
= O_P(1)N^{-1}T^{-(\nu_2+\delta)/2}O_P(\sqrt{NT}) \times O_P(T^{\delta/2})O_P(T^{-(\delta+\nu_1-\nu_2)/2}) + O_P(N^{-1/2}T^{-(\delta+\nu_1-\nu_2-1)/2}) \nonumber \\
+ O_P(N^{-(1-p)}T^{-(\delta+\nu_1-2\nu_2)/2}) + O_P(T^{2-(\nu_1+\delta)/2}) \leq O_P(K_{112}) + O_P(N^{-1}T^{-(\nu_2+\delta)/2}\sqrt{N TT^{q-\nu_1}/2}) \leq o_P(K_{112}) = O_P(NT^{q-\nu_1}/2) \right),
\] (A.4.77)

where the first equality is due to Lemmas A.2 and A.5, while the second is due to Assumption 1. Further use of Lemma A.2 shows that $K_{114}$ is of the following order:
\[
K_{114} \leq N^{-1}T^{-(\nu_2+\delta)/2}\|\mathbf{e}^T\mathbf{F}_+^0\mathbf{F}_+^0\| = O_P(T^{-(\nu_2+\delta-2)/2}(N^{-1/2} \vee T^{-1/2})),
\] (A.4.78)

while the order of $K_{115}$ is
\[
K_{115} = N^{-1}T^{-(\nu_2+\delta)/2}\|\mathbf{e}\| T_0^{i+2} = O_P(N^{-1/2}T^{-(\nu_2+\delta-1-\nu_3)/2}).
\] (A.4.79)

By inserting the above results into (A.4.73), we obtain
\[
T^{-\delta/2}\|\mathbf{K}_1\| = O_P(N^{-1/2}T^{-(\nu_2+\delta-1)/2}) + O_P(N^{-1}T^{1-(\nu_2+\delta)/2}) + O_P(N^{-3/2}T^{-(\delta-1)/2}) \nonumber \\
+ O_P(N^{-1/2}T^{q-(\nu_1+\nu_2+\delta-1)/2}) + O_P(T^{-(\delta-1)/2}\|\mathbf{D}_T^{-1}(\beta^0 - \beta_0)\|) \nonumber \\
= O_P(N^{-(3/2-p)}T^{-(\delta-1)/2}) + O_P(N^{-1/2}T^{q-(\nu_1+\nu_2+\delta-1)/2}) + O_P(T^{-(\delta-1)/2}\|\mathbf{D}_T^{-1}(\beta^0 - \beta_0)\|).
\] (A.4.80)

The order of $T^{-\delta/2}\|\mathbf{K}_1\|$ is the same, which can be shown using the above steps.

We move on to $K_{13}$, whose order is given by
\[
T^{-\delta/2}\|\mathbf{K}_{13}\| \nonumber \\
= T^{-\delta/2}\left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} (\mathbf{F}_2^0 + \mathbf{F}_2 \mathbf{g}_2 + \mathbf{F}_2 \mathbf{g}_1 - \mathbf{F}_2 \mathbf{g}_1) \right\| \nonumber \\
\leq \left\| \frac{1}{NT^{(\nu_2+\delta)/2}} \sum_{i=1}^{N} (\mathbf{F}_2^0 + \mathbf{F}_2 \mathbf{g}_2 + \mathbf{F}_2 \mathbf{g}_1 - \mathbf{F}_2 \mathbf{g}_1) \right\| T^{-\delta/2}\|\mathbf{F}_2\|,
\]
\[ \leq O_P(1) N^{-1} T^{-(\nu_2 + \delta)/2} \| F_{0,2}' \Gamma_{0,2}' F_{0,2}' \| \]

\[ + O_P(1) \left\| \frac{1}{NT^{\delta}} \sum_{i=1}^{N} (F_{0,2}' \gamma_{0,1,i} - \hat{F}_{1,i} \gamma_{1,i} ) \gamma_{2,i} F_{0,2}' \right\| \]

\[ = O_P(1) \left\| \frac{1}{NT^{\delta}} \sum_{i=1}^{N} (F_{0,2}' \gamma_{0,1,i} - \hat{F}_{1,i} \gamma_{1,i} ) \gamma_{2,i} F_{0,2}' \right\|^{1/2} \]

\[ + O_P(1) \left\| \frac{1}{NT^{\delta}} \sum_{i=1}^{N} \| F_{0,2}' \gamma_{0,2,i} \| \right\|^{1/2} \]

\[ = O_P(N^{-(1-p)} T^{-(\delta - \nu_2)/2}) + O_P(N^{-1/2} T^{-\delta/2} \sqrt{N} \sqrt{T}) + O_P(T^{q-(\delta + \nu_1)/2}) \]

\[ + O_P(N^{-(1-p)} T^{-(\delta - \nu_2)/2}) \]

\[ = O_P(N^{-1/2} T^{-(\delta - 1)/2}) + O_P(T^{-\delta/2}) + O_P(T^{q-(\delta + \nu_1)/2}) + O_P(N^{-(1-p)} T^{-(\delta - \nu_2)/2}), \quad (A.4.81) \]

where the second equality follows from Lemma A.5 and Assumption 1. The order of \( T^{-\delta/2} \| K_{14} \| \) is the same.

For \( K_{15} \),

\[ T^{-\delta/2} \| K_{15} \| \]

\[ = T^{-\delta/2} \left\| \frac{1}{NT^{\delta}} \sum_{i=1}^{N} U_i U_i' \hat{F}_{2} \right\| \]

\[ \leq \left\| \frac{1}{NT^{\delta}} \sum_{i=1}^{N} U_i U_i' \right\| T^{-\delta/2} \| \hat{F}_{2} \| \]

\[ \leq O_P(1) \frac{1}{NT^{\delta}} \sum_{i=1}^{N} \| F_{0,2}' \gamma_{0,1,i} - \hat{F}_{1,i} \gamma_{1,i} \|^2 + O_P(1) \frac{1}{NT^{\delta}} \sum_{i=1}^{N} \| F_{0,2}' \gamma_{0,2,i} \|^2 \]

\[ = O_P(1) N^{-1} T^{-(\nu_2 + \delta)/2} \left[ O_P(N \sqrt{T}) + O_P(NT^{2q-\nu_1}) + O_P(N^{-(1-2p)} T^{\nu_2}) + O_P(NT^{\nu_3}) \right] \]

\[ = O_P(T^{-(\nu_2 + \delta)/2}) + O_P(N^{-1} T^{-(\nu_2 + \delta)/2}) + O_P(T^{2q-\nu_1-(\nu_2 + \delta)/2}) + O_P(N^{-2(1-p)} T^{-(\delta - \nu_2)/2}) \]

\[ + O_P(T^{\nu_3-(\nu_2 + \delta)/2}), \quad (A.4.82) \]

where the second equality follows from Lemma A.5.

We now insert the above results for \( K_1, \ldots, K_{15} \) into (A.4.66). But first we left multiply by \( T^{-\nu_2} F_{0,2}' \). This gives

\[ T^{-(\nu_2 + \delta)/2} F_{0,2}' \hat{F}_{2} (T^{-\nu_2} V_2) - (T^{-\nu_2} F_{0,2}' F_{0,2}')(N^{-1} \Gamma_{2}' \Gamma_{2}') (T^{-(\nu_2 + \delta)/2} F_{0,2}' \hat{F}_{2}) \]

\[ = T^{-(\nu_2 - \delta)/2} \sum_{j=1}^{15} T^{-\nu_2/2} F_{0,2}' (T^{-\delta/2} K_{j}) \]

\[ = T^{-(\nu_2 - \delta)/2} \left[ O_P(T^{-(\delta - 1)/2}) \| D_T^{-1} (\beta_0' - \hat{\beta}_0) \| \right] + O_P(T^{-1-(\nu_2 + \delta)/2} (N^{-1/2} \sqrt{T^{-1/2}})) \]

\[ + O_P(N^{-1/2} T^{-(\delta - 1)/2}) + O_P(N^{-(3/2-p)} T^{-(\delta - 1)/2}) + O_P(N^{-1/2} T^{q-(\nu_1 + \nu_2 + \delta - 1)/2}) \]

25
\[O_N \left( N^{-1/2} \sqrt{T^{-1/2}} \right) + O_P \left( N^{-1/2} T^{-\left(\nu_2+\delta-1\right)/2} \right) + O_P \left( N^{-1/2} T^{-\left(\nu_2+\delta-1\right)/2} \right) + O_P \left( N^{-1/2} T^{-\left(\delta-1\right)/2} \right) + O_P \left( T^{-\delta/2} \right) + O_P \left( T^{q-\left(\delta+\nu_1\right)/2} \right) + O_P \left( N^{-1-\nu} T^{-\left(\delta-\nu_2\right)/2} \right) + O_P \left( T^{-\nu_2+\delta/2} \right) \]
\[= T^{-\left(\nu_2\right)/2} \left[ O_P \left( N^{-1/2} T^{-\left(\delta-1\right)/2} \right) + O_P \left( T^{-\delta/2} \right) + O_P \left( N^{-1/2} T^{-\left(\nu_2+\delta\right)/2} \right) + O_P \left( T^{q-\left(\delta+\nu_1\right)/2} \right) + O_P \left( N^{-1-\nu} T^{-\left(\delta-\nu_2\right)/2} \right) \right] \]
\[+ O_P \left( N^{-1-\nu} T^{-\left(\delta-\nu_2\right)/2} \right) + O_P \left( T^{-\nu_2+\delta/2} \right) \]
\[= O_P \left( N^{-1/2} T^{-\left(\nu_2-1\right)/2} \right) + O_P \left( T^{-\nu_2/2} \right) + O_P \left( T^{-\left(\nu_2-\nu_3\right)} \right) + O_P \left( T^{q-\left(\nu_1+\nu_2\right)/2} \right) + O_P \left( N^{-1-\nu} \right), \]
\[(A.4.83)\]

which is \(O_P(1)\) under Assumption 5. Hence, analogous to the proof of Lemma A.3, the \(d_2\) nonzero elements of the limit of \(T^{-\nu_2}V_2\) are the eigenvalues of the limit of \((T^{-\nu_2}F_2^0F_2^0)(N^{-1}F_2^0F_2^0)\). Thus,
\[T^{-\nu_2}V_2 \rightarrow_p \Psi_2 \]
\[(A.4.84)\]
as \(N, T \rightarrow \infty\).

**Proof of Lemma A.7.**

This proof follows from the same arguments used in the proof of Lemma A.3. Note in particular how (a) is a direct consequence of the fact that
\[\|T^{-\delta/2}F_{2,d} - T^{-\nu_2/2}F_2^0h_2^0\| = O_P \left( N^{-1/2} T^{-\left(\nu_2-1\right)/2} \right) + O_P \left( T^{-\nu_2/2} \right) + O_P \left( T^{-\left(\nu_2-\nu_3\right)} \right) + O_P \left( T^{q-\left(\nu_1+\nu_2\right)/2} \right) + O_P \left( N^{-1-\nu} \right), \]
\[(A.4.85)\]
which holds by applying the arguments of the proof of Lemma A.4 to the results provided in the proof of Lemma A.6. The rest of the proof is omitted.

**Proof of Lemma A.8.**

Let us assume without loss of generality that \(\beta^0 = 0_{d_\times 1}\), as in Bai (2009). It then follows that
\[(NT)^{-1} \left[ \text{SSR} \left( \hat{\beta}_1, \hat{F} \right) - \text{SSR} \left( \beta_0, \hat{F} \right) \right] = \frac{1}{NT} \sum_{i=1}^{N} (X_i \hat{\beta}_1 + F_0^0 \gamma_i^0)'M_F(X_i \hat{\beta}_1 + F_0^0 \gamma_i^0) + \hat{\beta}_1 \frac{2}{NT} \sum_{i=1}^{N} X_i' M_F \varepsilon_i + \frac{2}{NT} \sum_{i=1}^{N} \gamma_i^0 F_0^0 M_F \varepsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' M_F \varepsilon_i \]
\[= \frac{1}{NT} \sum_{i=1}^{N} (X_i \hat{\beta}_0 + F_0^0 \gamma_i^0)'M_F(X_i \hat{\beta}_0 + F_0^0 \gamma_i^0) \]

26
\[- \beta_0' \frac{2}{NT} \sum_{i=1}^N X_i' M_{\hat{F}} \varepsilon_i - \frac{2}{NT} \sum_{i=1}^N \gamma_i' F_0^0 M_{\hat{F}} \varepsilon_i - \frac{1}{NT} \sum_{i=1}^N \varepsilon_i' M_{\hat{F}} \varepsilon_i \]

\[
\begin{align*}
&= \frac{1}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} X_i (\beta_1 - \beta_0) + \frac{2}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} X_i (\hat{\beta}_0 - \beta_0) \\
&\quad + \frac{2}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} F_0 \gamma_i^0 + (\beta_1 - \beta_0)' \frac{2}{NT} \sum_{i=1}^N X_i' M_{\hat{F}} \varepsilon_i \\
&\geq \frac{1}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} X_i (\beta_1 - \beta_0) \\
&\quad - 2 \left( \frac{1}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} X_i (\beta_1 - \beta_0) \right)^{1/2} \left( \frac{1}{NT} \sum_{i=1}^N (\beta_0 - \beta_0)' X_i' M_{\hat{F}} X_i (\hat{\beta}_0 - \beta_0) \right)^{1/2} \\
&\quad + \frac{2}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} F_0 \gamma_i^0 + (\beta_1 - \beta_0)' \frac{2}{NT} \sum_{i=1}^N X_i' M_{\hat{F}} \varepsilon_i. \tag{A.4.86}
\end{align*}
\]

Note that by using the same steps as in the proof of Theorem 3.1,

\[
\left\| \frac{1}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} F_0 \gamma_i^0 \right\| \leq \left\| D_T^{-1} (\beta_1 - \beta_0) \right\| \left\| \frac{1}{NT} \sum_{i=1}^N D_T X_i' M_{\hat{F}} F_0 \gamma_i^0 \right\| = O_P(\| D_T^{-1} (\beta_1 - \beta_0) \|) [O_P((NT)^{-1/2}) + o_P(\| D_T^{-1} (\beta_0 - \beta_0) \|)]. \tag{A.4.87}
\]

It follows that

\[
0 \geq (NT)^{-1} \left[ \text{SSR}(\beta_1, \hat{F}) - \text{SSR} \left( \hat{\beta}_0, \hat{F} \right) \right] \\
= \frac{1}{NT} \sum_{i=1}^N (\beta_1 - \beta_0)' X_i' M_{\hat{F}} X_i (\beta_1 - \beta_0) \\
\quad + O_P(\| D_T^{-1} (\beta_1 - \beta_0) \|) [O_P((NT)^{-1/2}) + o_P(\sqrt{NT}\| D_T^{-1} (\beta_0 - \beta_0) \|)] \\
\geq \lambda_{\min} \left( \frac{1}{NT} \sum_{i=1}^N D_T X_i' M_{\hat{F}} D_T \right) \| D_T^{-1} (\beta_1 - \beta_0) \|^2 \\
\quad + O_P(\| D_T^{-1} (\beta_1 - \beta_0) \|) [O_P((NT)^{-1/2}) + o_P(\| D_T^{-1} (\beta_0 - \beta_0) \|)], \tag{A.4.88}
\]

where the first term on the right is quadratic in \( \| D_T^{-1} (\beta_1 - \beta_0) \| \). Hence, to ensure the right-hand side is non-positive, \( \| D_T^{-1} (\beta_1 - \beta_0) \| \) cannot converge to zero at a rate faster than \( (NT)^{-1/2} \lor \)
\[ \| D_T^{-1}(\hat{\beta}_0 - \beta^0) \|. \] It follows that
\[ \| D_T^{-1}(\hat{\beta}_1 - \hat{\beta}_0) \| = O_P((NT)^{-1/2} \vee \| D_T^{-1}(\hat{\beta}_0 - \beta^0)\|), \] (A.4.89)
as was to be shown.

\section*{A.5 Proofs of main results}

\textbf{Proof of Lemma 3.1.}

As in Proof of Lemma A.8, we assume that \( \beta^0 = 0_{d_x \times 1} \). This implies
\[
(NT)^{-1}[\text{SSR}(\beta, F) - \text{SSR}(\beta^0, F^0)]
= \frac{1}{NT} \sum_{i=1}^{N} (X_i \beta + F_i^0 \gamma^0_i)'M_F (X_i \beta + F_i^0 \gamma^0_i)
+ \beta' \frac{2}{NT} \sum_{i=1}^{N} X_i'M_F \epsilon_i + \frac{2}{NT} \sum_{i=1}^{N} \gamma^0_i F^0'M_F \epsilon_i + \frac{1}{NT} \sum_{i=1}^{N} \epsilon_i'(P_{F^0} - P_F) \epsilon_i
= \frac{1}{NT} \sum_{i=1}^{N} (X_i D_TD_T^{-1} \beta + F_i^0 \gamma^0_i)'M_F (X_i D_TD_T^{-1} \beta + F_i^0 \gamma^0_i)
+ \beta' D_T^{-1} \frac{2}{NT} \sum_{i=1}^{N} D_T X_i'M_F \epsilon_i + \frac{2}{NT} \sum_{i=1}^{N} \gamma^0_i F^0'M_F \epsilon_i + O_P(N^{-1} \vee T^{-1})
= \beta' D_T^{-1} B(F) D_T^{-1} \beta + \theta' B \theta + \beta' D_T^{-1} \frac{2}{NT} \sum_{i=1}^{N} D_T X_i'M_F \epsilon_i
+ \frac{2}{NT} \sum_{i=1}^{N} \gamma^0_i F^0'M_F \epsilon_i + O_P(N^{-1} \vee T^{-1}), \] (A.5.1)
where the second equality follows from Lemma A.2 and Assumption 1, and \( \theta = \eta + B^{-1}C \beta \) and \( B = N^{-1} \Gamma F^0 \Gamma \otimes I_T \) are defined as on page 1265 of Bai (2009). Note that for \( \beta \in \mathbb{R}^{d_x} \), we may have
\[
\sup_{\beta \in \mathbb{R}^{d_x}, F \in F} \left| \beta' D_T^{-1} \frac{2}{NT} \sum_{i=1}^{N} D_T X_i'M_F \epsilon_i \right| \neq o_P(1). \] (A.5.2)

In our proof of consistency, we consider two cases; (i) \( \| D_T^{-1} \beta \| \leq C \) and (ii) \( \| D_T^{-1} \beta \| > C \),
where $C$ is a large positive constant. Under (i),

$$\sup_{\|D_T^{-1}\beta\| \leq C, \beta \in D_F} \left| \beta' D_T^{-1} \frac{2}{NT} \sum_{i=1}^{N} D_T X_i' M_F \varepsilon_i \right| = O_P(N^{-1/2} \lor T^{-1/2}) \quad (A.5.3)$$

by Lemma A.2 and Assumption 1. The expression given in (A.5.1) for $(NT)^{-1}[SSR(\beta, F) - SSR(\beta^0, F^0)]$ therefore reduces to

$$(NT)^{-1}[SSR(\beta, F) - SSR(\beta^0, F^0)] = \beta' D_T^{-1} B(\beta) D_T^{-1} \beta + \theta' B \theta + \frac{2}{NT} \sum_{i=1}^{N} \gamma_i \varepsilon_i,$$

where $\beta' D_T^{-1} B(\beta) D_T^{-1} \beta$ does not involve $F^0$ and $\sum_{i=1}^{N} \gamma_i \varepsilon_i$ is independent of $\beta$. Hence, provided $d_{\text{max}} \geq d_f$, the consistency of $D_T^{-1} \hat{\beta}_0$ in case (i) follows from the same arguments as in Bai (2009).

Under (ii), (A.5.1) can be written as

$$(NT)^{-1}[SSR(\beta, F) - SSR(\beta^0, F^0)]$$

$$= \beta' D_T^{-1} B(\beta) D_T^{-1} \beta + \theta' B \theta + \frac{2}{NT} \sum_{i=1}^{N} D_T X_i' M_F \varepsilon_i + 2 \frac{N}{NT} \sum_{i=1}^{N} \gamma_i \varepsilon_i,$$

$$+ O_P(N^{-1/2} \lor T^{-1/2}),$$

$$\geq c_0 \|D_T^{-1} \beta\|^2 + \beta' D_T^{-1} \frac{2}{NT} \sum_{i=1}^{N} D_T X_i' M_F \varepsilon_i + \theta' B \theta + \frac{2}{NT} \sum_{i=1}^{N} \gamma_i \varepsilon_i,$$

$$+ O_P(N^{-1/2} \lor T^{-1/2})$$

$$\geq c_0 \frac{C^2}{2} + \theta' B \theta + \frac{2}{NT} \sum_{i=1}^{N} \gamma_i \varepsilon_i + O_P(N^{-1/2} \lor T^{-1/2}), \quad (A.5.5)$$

where $c_0$ is defined in Assumption 2, and the second inequality follows from the fact that the quadratic term dominates the linear one for large values of $C$. Hence, $(NT)^{-1}[SSR(\beta, F) - SSR(\beta^0, F^0)] > 0$, but from the definition of $\hat{\beta}_0$ we also know that SSR($\hat{\beta}_0$, $\hat{F}$) -- SSR($\beta^0$, $F^0$) $\leq 0$, which means that $D_T^{-1} \hat{\beta}_0$ cannot belong to (ii). Note that since $\| (NT)^{-1} \sum_{i=1}^{N} D_T X_i' M_F \varepsilon_i \| = O_P(N^{-1/2} \lor T^{-1/2})$ by Lemma A.2, all we need is $C = C^0(N^{-1/2} \lor T^{-1/2})$ for some large constant $C^0$ in order to ensure that the last inequality of (A.5.5) holds. This implies

$$D_T^{-1}(\hat{\beta}_0 - \beta^0) = O_P(N^{-1/2} \lor T^{-1/2}), \quad (A.5.6)$$
and so the proof is complete.

**Proof of Lemma 3.2.**

Suppose first that \(d_1 = 0\), such that \(d_f = 0\). In this case,

\[
\frac{1}{NT} \sum_{i=1}^{N} \| y_i - X_i \tilde{\beta}_0 \|^2 = \frac{1}{NT} \sum_{i=1}^{N} \| X_i D_T D_T^{-1}(\beta_0 - \tilde{\beta}_0) + \varepsilon_i \|^2
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \| \varepsilon_i \|^2 + o_P(1). \tag{A.5.7}
\]

This implies \(\tau_N \asymp 1 / \ln(T \lor N)\), which is much larger than \(\hat{\lambda}_{1,1} / \hat{\lambda}_{1,0}\). The result then follows immediately.

Suppose now instead that \(d_1 > 0\). Straightforward algebra reveals that

\[
\frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} \| y_i - X_i \tilde{\beta}_0 \|^2 = \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} \| X_i D_T D_T^{-1}(\beta_0 - \tilde{\beta}_0) + F_0^0 \gamma_i^0 + \varepsilon_i \|^2
\]

\[
= \frac{1}{NT^{\nu_1}} \sum_{i=1}^{N} \| F_1^0 \gamma_{1,i} \|^2 + o_P(1)
\]

\[
= (\text{vec} \Sigma_{F_1^0})' \text{vec} \Sigma_{F_1^0} + o_P(1), \tag{A.5.8}
\]

which together with Assumption 1 implies \(\tau_N \asymp 1 / \ln(T \lor N)\). Note that for \(d_1 = 1, \ldots, d_1\),

\[
\frac{T^{\nu_1}}{\lambda_{1,d}} = \frac{T^{2\nu_1}}{\lambda_{1,d}^2} = \frac{NT^{2\nu_1}}{h_{1,d}^0 F_1^0 \Sigma_0^0 F_1^0 h_{1,d}^0}
\]

\[
= \frac{1}{(T^{-\nu_1} h_{1,d}^0 F_1^0 \Sigma_0^0 F_1^0 h_{1,d}^0)(N^{-1} T^{00} T_1^0)(T^{-\nu_1} F_1^0 F_1^0 h_{1,d}^0)}
\]

\[
= \frac{1}{(T^{-\nu_1 + \delta)/2} F_1^0 F_1^0)(N^{-1} T^{00} T_1^0)(T^{-\nu_1 + \delta)/2} F_1^0 F_1^0 h_{1,d}^0)} (1 + o_P(1)) \asymp 1, \tag{A.5.9}
\]

where the fourth equality follows from (A.4.34) and the last step is due to (A.4.29) and Assumption 1. Using Lemma A.4, (A.5.8) and (A.5.9), we obtain that

\[
\frac{\hat{\lambda}_{1,1}}{\hat{\lambda}_{1,0}} = \frac{T^{-\nu_1} \hat{\lambda}_{1,1}}{(NT^{\nu_1})^{-1} \sum_{i=1}^{N} \| y_i - X_i \tilde{\beta}_0 \|^2} \asymp 1, \tag{A.5.10}
\]

\[
\frac{\hat{\lambda}_{1,d+1}}{\hat{\lambda}_{1,d}} = \frac{T^{-\nu_1} \hat{\lambda}_{1,d+1}}{T^{-\nu_1} \hat{\lambda}_{1,d}} \asymp 1 \tag{A.5.11}
\]

30
for $d = 1, \ldots, d_1 - 1$, and

$$\frac{\hat{\lambda}_{1,d_1+1}}{\hat{\lambda}_{1,d_1}} = O_p(T^{-(\nu_1 - \nu_2)}).$$  \hfill (A.5.12)

Moreover, for $d = d_1 + 1, \ldots, d_{\text{max}}$,

$$\frac{T^{-\nu_1} \hat{\lambda}_{1,d}}{T^{-\nu_1} \hat{\lambda}_{1,0}} = O_p(T^{-(\nu_1 - \nu_2)}),$$  \hfill (A.5.13)

which is less than $\tau_N$ by (A.5.8) and Lemma A.4. The required result follows from this and the definition of $\hat{d}_1$.

Here onwards we assume that $G \geq 3$. The proofs simplify considerably if $G = 2$. We also assume that $d_1$ is known, which as pointed out in Appendix A.4 is not a restriction (up to an negligible remainder), as by Lemma 3.2 $d_1$ can be consistently estimated using $\hat{d}_1$.

**Proof of Lemma 3.3.**

Each step in the sequential procedure of Step 2 introduces additional reminder terms that all converge to zero under Assumptions 3 and 5 by Lemmas A.4 and A.6. This proves (a). Part (b) follows from the (rotational) consistency of $\hat{F}_g$ established as a part of the proofs of Lemmas A.3 and A.6.  \hfill $\blacksquare$

**Proof of Theorem 3.1.**

Note that

$$\hat{\beta}_1 - \beta^0 = D_T \left( \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F}_X X_i D_T \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F}_y y_i - \beta^0$$

$$= D_T \left( \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F}_X X_i D_T \right)^{-1} \times \left( \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F} F \gamma_i^0 + \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F} \varepsilon_i \right)$$

$$= D_T B^{-1} \left( L + \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F} \varepsilon_i \right),$$  \hfill (A.5.14)

with obvious definitions of $L$ and $B$. 

31
Consider $L$. Let $Q_g = (N^{-1} \Gamma_g^0 \Gamma_g^0) (T^{-\nu_g + \delta} / 2) F_g^0 \hat{F}_g$, such that $H_g^{-1} = T^{-\nu_g + \delta / 2} V_g^0 Q_g^{-1}$. We also introduce $e_{g,i}$, which is defined to be zero for $g = 1$ and $e_{g,i} = \sum_{j=1}^{g-1} (F_j^0 \gamma_{j,i} - \hat{F}_j \hat{\gamma}_{j,i})$ for $g = 2, \ldots, G$. In this notation,

$$
L = \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i F^0 g \gamma^0_{i} = -\frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_g H_g^{-1} - F^0 g \gamma^0_{i},
$$

$$
= -(L_1 + \cdots + L_{15}),
$$

(A.5.15)

where

$$
L_1 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} X_j (\beta^0 - \hat{\beta}_{g,j}) (\beta_{g,j} - \hat{\beta}_0) X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_2 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} X_j (\beta^0 - \hat{\beta}_{g,j}) \gamma_{g,j}^0 X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_3 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} F^0_g \gamma_{g,j} \beta_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_4 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} X_j (\beta^0 - \hat{\beta}_{g,j}) e_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_5 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} e_{g,j} (\beta^0 - \hat{\beta}_{g,j}) X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_6 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} X_j (\beta^0 - \hat{\beta}_{g,j}) \gamma_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_7 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} \gamma_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_8 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} \gamma_{g,j} e_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_9 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} F^0_g \gamma_{g,j} e_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_{10} = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} \gamma_{g,j} F^0_g \gamma_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

$$
L_{11} = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M \hat{F}_i \frac{1}{NT (\nu_g + \delta) / 2} \sum_{j=1}^{N} \gamma_{g,j} e_{g,j} X_j' \hat{F}_g Q^-1_g \gamma_{g,i},
$$

32
We will use this expression for $L$. We therefore start from

Let us now move on to $L$.

Then, we have

We now evaluate each of these terms. From the analysis of $L_2$ below, it is easy to show that

$$\|L_1\| = o_P(\|D^{-1}_T(\hat{\beta}_0 - \beta^0)\|). \tag{A.5.16}$$

We therefore start from $L_2$, which we write as

We will use this expression for $L_2$ later.

Let us now move on to $L_3$.

Then, we have

$$\|L_1\| = o_P(\|D^{-1}_T(\hat{\beta}_0 - \beta^0)\|). \tag{A.5.16}$$
where, by using arguments that are similar to those used in the proofs of Lemmas A.3 and A.6,

\[
\frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{g=1}^{G} \left( F^{0}_{g} - \hat{F}_{g} H^{-1}_{g} \right) \frac{1}{NT(\nu_{g}+\delta)^{2}} \sum_{j=1}^{N} \gamma_{g,j}^{0} (\beta^{0} - \hat{\beta}_{0})^{\prime} X_{j}^{\prime} \hat{F}_{g} Q^{-1}_{g} \gamma_{g,i}^{0} \right\|^{2} \\
\leq \frac{1}{NT} \sum_{i=1}^{N} \left( \sum_{g=1}^{G} \left\| F^{0}_{g} - \hat{F}_{g} H^{-1}_{g} \right\| \frac{1}{NT(\nu_{g}+\delta)^{2}} \sum_{j=1}^{N} \gamma_{g,j}^{0} (\beta^{0} - \hat{\beta}_{0})^{\prime} D^{-1}_{T} D_{j}^{\prime} X_{j}^{\prime} \hat{F}_{g} ||Q_{g}^{-1} \gamma_{g,i}^{0}|| \right)^{2} \\
\leq O_{P}(1) \sum_{g=1}^{G} T^{-\nu_{g}} \left\| F^{0}_{g} - \hat{F}_{g} H^{-1}_{g} \right\| ^{2} \left\| D_{T}^{-1} (\beta^{0} - \hat{\beta}_{0}) \right\| ^{2} \\
= o_{P}(\left\| D_{T}^{-1} (\beta^{0} - \hat{\beta}_{0}) \right\| ^{2}), \tag{A.5.19}
\]

implying

\[
\| L_{3} \| \leq \left( \frac{1}{NT} \sum_{i=1}^{N} \left\| D_{T} X_{i}^{\prime} M_{\hat{F}} \right\| ^{2} \right)^{1/2} \times \left( \frac{1}{NT} \sum_{i=1}^{N} \left\| \sum_{g=1}^{G} \left( F^{0}_{g} - \hat{F}_{g} H^{-1}_{g} \right) \frac{1}{NT(\nu_{g}+\delta)^{2}} \sum_{j=1}^{N} \gamma_{g,j}^{0} (\beta^{0} - \hat{\beta}_{0})^{\prime} X_{j}^{\prime} \hat{F}_{g} Q^{-1}_{g} \gamma_{g,i}^{0} \right\|^{2} \right)^{1/2} \\
= o_{P}(\left\| D_{T}^{-1} (\beta^{0} - \hat{\beta}_{0}) \right\| ), \tag{A.5.20}
\]

The same steps can be used to show that \( L_{4} \) and \( L_{5} \) are of the same order.

For \( L_{6} \), write

\[
L_{6} = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_{T} X_{i}^{\prime} M_{\hat{F}} \frac{1}{NT(\nu_{g}+\delta)^{2}} \sum_{j=1}^{N} X_{j} (\beta^{0} - \hat{\beta}_{0}) \epsilon_{j} \hat{F}_{g} Q^{-1}_{g} \gamma_{g,i}^{0} \\
\leq \sum_{g=1}^{G} \frac{1}{N^{2} T(\nu_{g}+\delta)^{2}/2+1} \sum_{i=1}^{N} \sum_{j=1}^{N} D_{T} X_{i}^{\prime} M_{\hat{F}} X_{j} (\beta^{0} - \hat{\beta}_{0}) \epsilon_{j} F^{0}_{g} H_{g} Q^{-1}_{g} \gamma_{g,i}^{0} \\
+ \sum_{g=1}^{G} \frac{1}{N^{2} T(\nu_{g}+\delta)^{2}/2+1} \sum_{i=1}^{N} \sum_{j=1}^{N} D_{T} X_{i}^{\prime} M_{\hat{F}} X_{j} (\beta^{0} - \hat{\beta}_{0}) \epsilon_{j} (\hat{F}_{g} - F^{0}_{g} H_{g}) Q^{-1}_{g} \gamma_{g,i}^{0}, \tag{A.5.21}
\]

where the first term on the right is bounded by

\[
\left\| \sum_{g=1}^{G} \frac{1}{N^{2} T(\nu_{g}+\delta)^{2}/2+1} \sum_{i=1}^{N} \sum_{j=1}^{N} D_{T} X_{i}^{\prime} M_{\hat{F}} X_{j} (\beta^{0} - \hat{\beta}_{0}) \epsilon_{j} F^{0}_{g} H_{g} Q^{-1}_{g} \gamma_{g,i}^{0} \right\| \\
\leq \sum_{g=1}^{G} \frac{1}{N^{2} T(\nu_{g}+\delta)^{2}/2+1} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| D_{T} X_{i}^{\prime} M_{\hat{F}} X_{j} D_{T} \epsilon_{j} F^{0}_{g} H_{g} Q^{-1}_{g} \gamma_{g,i}^{0} \| \| D_{T}^{-1} (\beta^{0} - \hat{\beta}_{0}) \| \right.
\]

34
\[
\begin{align*}
\leq O_P(1)\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\| & \sum_{g=1}^{G} \frac{1}{N^2 T^{(v_g + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} \|D_T X_i' M_{\hat{F}}\| \|M_{\hat{F}} X_j D_T\| \\
\times \|\epsilon_j' D_g \| \|H_g\| \|\gamma_{g,i}^0\| \\
\leq O_P(1)\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\| & \sum_{g=1}^{G} \frac{1}{N^2 T^{(v_g + \delta)/2}} \left( \sum_{i=1}^{N} \|D_T X_i' M_{\hat{F}}\|^2 \right)^{1/2} \left( \sum_{i=1}^{N} \|\gamma_{g,i}^0\|^2 \right)^{1/2} \\
\times \left( \sum_{j=1}^{N} \|M_{\hat{F}} X_j D_T\|^2 \right)^{1/2} \left( \sum_{j=1}^{N} \|\epsilon_j' D_g \|^2 \right)^{1/2} \|H_g\| \\
\leq O_P(1)\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\| & \sum_{g=1}^{G} T^{-(v_g + \delta)/2} O_P(\sqrt{T}) O_P(\sqrt{T}) O_P(T^{v_g/2}) O_P(T^{-(v_g + \delta)/2}) \\
= o_P(\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|). & \quad \text{(A.5.22)}
\end{align*}
\]

The second term is of the same order. Thus,
\[
\|L_6\| = o_P(\|D_T^{-1}(\beta^0 - \hat{\beta}_0)\|). & \quad \text{(A.5.23)}
\]

The same is true for \(\|L_7\|\).

We now examine \(L_8\). Let us define \(\Sigma_\epsilon = N^{-1} \sum_{i=1}^{N} \Sigma_{\epsilon,i,i}\), in which \(\Sigma_{\epsilon,i,i}\) has been defined in Assumption 3. \(L_8\) can then be written as
\[
L_8 = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M_{\hat{F}} \frac{1}{N^2 T^{(v_g + \delta)/2}} \sum_{j=1}^{N} \epsilon_j' \epsilon_j' D_g Q_g^{-1} \gamma_{g,i}^0 \\
= \sum_{g=1}^{G} \frac{1}{N^2 T^{(v_g + \delta)/2}} \sum_{i=1}^{N} D_T X_i' M_{\hat{F}} \Sigma_{\epsilon,i} \hat{F}_g Q_g^{-1} \gamma_{g,i}^0, \\
+ \sum_{g=1}^{G} \frac{1}{N^2 T^{(v_g + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' (\epsilon_j' \epsilon_j' - \Sigma_{\epsilon,j,j}) \hat{F}_g Q_g^{-1} \gamma_{g,i}^0, \\
- \sum_{g=1}^{G} \frac{1}{N^2 T^{(v_g + \delta)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' P_\hat{F}_i (\epsilon_j' \epsilon_j' - \Sigma_{\epsilon,j,j}) \hat{F}_g Q_g^{-1} \gamma_{g,i}^0. & \quad \text{(A.5.24)}
\]

For the first term on the right,
\[
\sqrt{NT} \left\| \sum_{g=1}^{G} \frac{1}{N^2 T^{(v_g + \delta)/2}} \sum_{i=1}^{N} D_T X_i' M_{\hat{F}} \Sigma_{\epsilon,i} \hat{F}_g Q_g^{-1} \gamma_{g,i}^0 \right\| \\
\leq \sum_{g=1}^{G} \frac{1}{\sqrt{NT^{(v_g + \delta)/2}}} \sum_{i=1}^{N} \|D_T X_i' M_{\hat{F}} \Sigma_{\epsilon,i,j} \hat{F}_g Q_g^{-1} \gamma_{g,i}^0\|. 
\]

35
\[
\begin{align*}
\leq O_P(1) \sum_{g=1}^{G} N^{-1/2} T^{-(\nu_g+1)/2} O_P(NT) O_P(T^{(\delta-1)/2}) \\
\leq O_P(1) \sum_{g=1}^{G} N^{1/2} T^{-\nu_g/2} = O_P(\sqrt{NT}^{-\nu G/2}) = O_P(1),
\end{align*}
\]

where the last equality holds by Assumption 7. Let

\[
\mathcal{A}_1 = \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_{F_0} \Sigma_{\varepsilon} T^{-(\nu G-1)/2} F_G^0 (T^{-\nu G} F_G^0 F_G^0) \gamma_{G,i}^0.
\]

Note how \( \mathcal{A}_1 = \text{plim}_{N,T \to \infty} E(\mathcal{A}_1 | C) \). In this notation,

\[
\begin{align*}
\sqrt{NT} \sum_{g=1}^{G} \frac{1}{NT^{(\nu_g+\delta)/2+1}} & \sum_{i=1}^{N} D_T X_i' M_{F_0} \Sigma_{\varepsilon} \tilde{F}_g Q_g^{-1} \gamma_{g,i}^0 \\
= \sqrt{NT} \sum_{g=1}^{G} \frac{1}{NT^{(\nu_g+\delta)/2+1}} & \sum_{i=1}^{N} D_T X_i' M_{F_0} \Sigma_{\varepsilon} \tilde{F}_g (T^{-(\nu_g+\delta)/2} F_g^0 F_g^0) \gamma_{g,i}^0 (1 + o_P(1)) \\
= \sum_{g=1}^{G} \sqrt{\frac{N}{T^{\nu_g}}} & \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_{F_0} \Sigma_{\varepsilon} T^{-(\nu_g-1)/2} F_g^0 (T^{-\nu_g} F_g^0 F_g^0) \gamma_{g,i}^0 (1 + o_P(1)) \\
= \sqrt{\frac{N}{T^{\nu_g}}} & \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_{F_0} \Sigma_{\varepsilon} T^{-(\nu G-1)/2} F_G^0 (T^{-\nu G} F_G^0 F_G^0) \gamma_{G,i}^0 (1 + o_P(1)) \\
= \sqrt{NT}^{-\nu G/2} & \mathcal{A}_1 (1 + o_P(1)),
\end{align*}
\]

where the second equality follows from arguments similar to those used in (A.4.32) and (A.4.85), while the third follows from Assumption 7. Note also that \( \lim \sqrt{NT}^{-\nu G/2} = \rho_1 \). Further use of the same steps used by Jiang et al. (2017, pages 30–31) establishes that the second and third terms of \( L_8 \) are \( o_P(||D_T^{-1}(\hat{\beta}_0 - \beta_0)||) + o_P((NT)^{-1/2}) \). Hence,

\[
L_8 = T^{-(\nu G+1)/2} \mathcal{A}_1 + o_P(||D_T^{-1}(\hat{\beta}_0 - \beta_0)||) + o_P((NT)^{-1/2}).
\]

\( L_9 \) can be written as

\[
\begin{align*}
L_9 &= \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} D_T X_i' M_{F} \frac{1}{NT^{(\nu_g+\delta)/2}} \sum_{j=1}^{N} \gamma_{g,j}^0 \varepsilon_j' \tilde{F}_g Q_g^{-1} \gamma_{g,i}^0 \\
&= \sum_{g=1}^{G} \frac{1}{N^2 T^{(\nu_g+\delta)/2+1}} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' M_{F} (F_g^0 - \tilde{F}_g H_g^{-1}) \gamma_{g,j}^0 \varepsilon_j' F_g^0 H_g Q_g^{-1} \gamma_{g,i}^0.
\end{align*}
\]

36
where the first term on the right-hand side is bounded by

\[
\left\| \sum_{g=1}^{G} \frac{1}{N^2 T^2(\nu_g+\delta)/2+1} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' M F (F_g - \hat{F}_g H_g^{-1}) \gamma^0_{g,i,j} \varepsilon_j (\hat{F}_g - F_g H_g) Q_g^{-1} \gamma^0_{g,i} \right\|
\]

\[
\leq \sum_{g=1}^{G} \frac{1}{N^2 T^2(\nu_g+\delta)/2+1} \sum_{i=1}^{N} \sum_{j=1}^{N} \|D_T X_i' M F \| \|F_g - \hat{F}_g H_g^{-1}\| \|T^{-(\nu_g-1)/2} T_{g}^{0} \varepsilon F_g \|
\]

\[
\times T^{(\nu_g-1)/2} \|H_g\| \|Q_g^{-1} \gamma^0_{g,i}\|
\]

\[
\leq O_P(1) \sum_{g=1}^{G} N^{-1} T^{-(\nu_g+\delta)/2-1} O_P(\sqrt{T}) \|F_g - \hat{F}_g H_g^{-1}\| O_P(\sqrt{NT}) O_P(T^{(\nu_g-1)/2}) O_P(T^{-(\nu_g-\delta)/2})
\]

\[
\leq O_P(1) \sum_{g=1}^{G} (NT)^{-1/2} T^{-(\nu_g)/2} \|F_g - \hat{F}_g H_g^{-1}\|
\]

\[
o_P((NT)^{-1/2}), \quad (A.5.30)
\]

where the second inequality follows from \(\|T^{-(\nu_g-1)/2} T_g^{0} \varepsilon F_g \| = O_P(\sqrt{NT})\). The second term on the right-hand side of \(L_g\) is also \(o_P((NT)^{-1/2})\). We therefore conclude that

\[
\|L_0\| = o_P((NT)^{-1/2}). \quad (A.5.31)
\]

\(L_{10}\) can be written more compactly as

\[
L_{10} = \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' M F \frac{1}{N T^{(\nu_g+\delta)/2}} \sum_{j=1}^{N} \varepsilon_j \gamma^0_{g,i,j} F_g \hat{F}_g Q_g^{-1} \gamma^0_{g,i}
\]

\[
= \frac{1}{NT} \sum_{g=1}^{G} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' M F \frac{1}{N} \sum_{j=1}^{N} \varepsilon_j \gamma^0_{g,i,j} T^{-(\nu_g+\delta)/2} F_g \hat{F}_g (T^{-(\nu_g+\delta)/2} F_g \hat{F}_g)^{-1} (N^{-1} T^{-1} \gamma^0_{g,i,j})^{-1}
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' M F \varepsilon_j \gamma^0_{g,i,j}, \quad (A.5.32)
\]

which we will again make use of later.

Let us move on to \(L_{11}\). We begin by rewriting \(e_{g,i,j}\) as

\[
e_{g,j} = \sum_{d=1}^{g-1} [F_d^{0} \gamma^0_{d,j} - \hat{F}_d T^{-\delta} \hat{F}_d (y_j - X_j \hat{\beta}_0 - \hat{F}_d \hat{\gamma}_{d,j})]
\]
\[
\sum_{d=1}^{g-1} \left[ -P_{F_d}X_j(\beta^0 - \tilde{\beta}_0) + M_{F_d}F_d^0\gamma_{d,j}^0 - P_{F_d}F_{d+d}^0\gamma_{d,j}^0 - P_{F_d}e_{d,j} - P_{F_d}\epsilon_j \right], \tag{A.5.33}
\]

where \(F_{d+d}^0 = (F_{d+1}^0, \ldots, F_G^0)\) and \(\gamma_{d,j}^0 = (\gamma_{d+1,j}^0, \ldots, \gamma_{G,j}^0)'\), as in Section A.1. By inserting this into \(L_{11}\), we get

\[
L_{11} = \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_TX_i'M_{\hat{F}} \sum_{j=1}^{N} \epsilon_j \sum_{d=1}^{g-1} \gamma_{d,j}^0 F_{d+d}^0M_{\hat{F}}F_gQ_g^{-1}\gamma_{g,i}^0
\]

\[
+ M_{\hat{F}}F_0^0 - P_{\hat{F}_d}F_0^0 + \gamma_{d,j}^0 - P_{\hat{F}_d}e_{d,j} - P_{\hat{F}_d}\epsilon_j |\hat{F}_dF_gQ_g^{-1}\gamma_{g,i}^0. \tag{A.5.34}
\]

Hence, \(L_{11}\) can be written as a sum of five terms. There is no need to study the forth term, the one due to \(P_{\hat{F}_d}e_{d,j}\), as we can keep expanding \(e_{d,j}\) until we cannot. The first and fifth terms are \(o_P(||D_T^{-1}(\beta_0 - \beta^0)||)\) and \(o_P((NT)^{-1/2})\), respectively, by the same arguments used for evaluating \(L_7\) and \(L_8\). Moreover, the steps used for evaluating \(L_9\) can be used to show that

\[
\left| \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_TX_i'M_{\hat{F}} \sum_{j=1}^{N} \epsilon_j \sum_{d=1}^{g-1} \gamma_{d,j}^0 F_{d+d}^0M_{\hat{F}}F_gQ_g^{-1}\gamma_{g,i}^0 \right|
\]

\[
= \left| \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_TX_i'M_{\hat{F}} \sum_{j=1}^{N} \epsilon_j \sum_{d=1}^{g-1} \gamma_{d,j}^0 F_{d+d}^0M_{\hat{F}}F_gQ_g^{-1}\gamma_{g,i}^0 \right|
\]

\[
= \left| \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_TX_i'M_{\hat{F}} \sum_{j=1}^{N} \epsilon_j \sum_{d=1}^{g-1} \gamma_{d,j}^0 F_{d+d}^0M_{\hat{F}}F_gQ_g^{-1}\gamma_{g,i}^0 \right|
\]

\[
= o_P((NT)^{-1/2}). \tag{A.5.35}
\]

The second term in \(L_{11}\) is therefore negligible. It remains to consider the third term, which is

\[
\left| \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_TX_i'M_{\hat{F}} \sum_{j=1}^{N} \epsilon_j \sum_{d=1}^{g-1} \gamma_{d,j}^0 F_{d+d}^0P_{\hat{F}_d}F_gQ_g^{-1}\gamma_{g,i}^0 \right|
\]

\[
= \left| \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_TX_i'M_{\hat{F}} \sum_{j=1}^{N} \epsilon_j \sum_{d=1}^{g-1} \gamma_{d,j}^0 F_{d+d}^0P_{\hat{F}_d}F_gQ_g^{-1}\gamma_{g,i}^0 \right|
\]

\[
= o_P(||L_{10}||) = o_P((NT)^{-1/2}) \tag{A.5.36}
\]
where we have used Assumption 5. Therefore,

$$\|L_{11}\| = o_P((NT)^{-1/2}).$$

(A.5.37)

For $L_{12}$,

$$L_{12} = \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' \mathcal{M} \hat{F}_g \frac{1}{NT(N^{\nu_{g}} + \delta)/2} \sum_{j=1}^{N} e_{g,j} \varepsilon_j' \hat{F}_g Q_g^{-1} \gamma_{g,i}$$

$$= \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' \mathcal{M} \hat{F}_g \frac{1}{NT(N^{\nu_{g}} + \delta)/2} \sum_{j=1}^{N} F_{g,j}^{0} \gamma_{g,i} \varepsilon_j' \hat{F}_g Q_g^{-1} \gamma_{g,i} = L_9,$$

(A.5.38)

where the second equality follows from the construction of $e_{g,j}$. Hence, since $L_9$ is negligible, $L_{12}$ is also negligible.

Next up is $L_{13}$, whose order can be worked out in the following way:

$$\|L_{13}\| = \left\| \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' \mathcal{M} \hat{F}_g \frac{1}{NT(N^{\nu_{g}} + \delta)/2} \sum_{j=1}^{N} e_{g,j} \gamma_{g,j} F_{g,j}^{0} \hat{F}_g Q_g^{-1} \gamma_{g,i} \right\|$$

$$\leq \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} \left\| D_T X_i' \mathcal{M} \hat{F}_g \right\| \frac{1}{NT(N^{\nu_{g}} + \delta)/2} \sum_{j=1}^{N} \left\| e_{g,j} \gamma_{g,j} F_{g,j}^{0} \hat{F}_g Q_g^{-1} \gamma_{g,i} \right\|$$

$$\leq \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} \left\| D_T X_i' \mathcal{M} \hat{F}_g \sum_{d=1}^{g-1} \left( F_{d,d}^{0} - \hat{F}_{d,d} H_{d,d}^{-1} \right) \Gamma_{g}^{0} \hat{F}_g Q_g^{-1} \gamma_{g,i} \right\|$$

$$\leq O_P(1) \sum_{g=1}^{G} \frac{1}{NT} \sum_{i=1}^{N} \left\| D_T X_i' \mathcal{M} \hat{F}_g \right\| \left\| F_{d,d}^{0} - \hat{F}_{d,d} H_{d,d}^{-1} \right\| \left\| \Gamma_{g}^{0} \hat{F}_g Q_g^{-1} \right\| \gamma_{g,i}$$

$$\leq O_P(1) \sum_{g=1}^{G} \frac{1}{NT} \sum_{d=1}^{g-1} O_P(\sqrt{T}) \left\| F_{d,d}^{0} - \hat{F}_{d,d} H_{d,d}^{-1} \right\| O_P(N^p)$$

$$\leq O_P(1) \sum_{g=1}^{G} \frac{1}{NT} \sum_{d=1}^{g-1} \left\| F_{d,d}^{0} - \hat{F}_{d,d} H_{d,d}^{-1} \right\|$$

$$= o_P((NT)^{-1/2}),$$

(A.5.39)

where the first inequality follows from the construction of $e_{g,j}$, and the last equality follows from Assumption 6. The same arguments show that $\|L_{14}\|$ and $\|L_{15}\|$ are negligible, too.

We now put everything together. This yields

$$L = -(L_1 + \cdots + L_{15})$$

$$= -L_2 - L_8 - L_{10} + o_P(\|D_T^{-1}(\hat{\beta}_0 - \beta^0)\|) + o_P((NT)^{-1/2})$$

39
which can be solved for $D$

This expression for $D$ which in turn implies

\[
D^{-1}(\hat{\beta} - \beta^0) = B^{-1}\left( L + \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_F X_j D_T a_j + o_p(1) \right)\ 
\]

\[
= \left( -B^{-1} \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_F \varepsilon_i \right) + B^{-1}\left( \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_F \varepsilon_i - \frac{1}{NT} \sum_{i=1}^{N} D_T X_i' M_F a_j \varepsilon_j \right) 
\]

\[
- T^{-(\nu \alpha + 1)/2} B^{-1} A_1 + o_p((NT)^{-1/2}) \]  

where $Z_i(F)$ is defined in Assumption 2, and

\[
N = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i' M_F X_j D_T a_j. 
\]

This expression for $D^{-1}(\hat{\beta} - \beta^0)$ can be inserted into $D^{-1}(\hat{\beta}_0 - \beta^0)$, giving

\[
D^{-1}(\hat{\beta}_0 - \beta^0) = D^{-1}(\hat{\beta}_1 - \beta^0) - D^{-1}(\hat{\beta}_1 - \hat{\beta}_0) 
\]

\[
= (B^{-1} N + o_p(1)) D^{-1}(\hat{\beta}_0 - \beta^0) + B^{-1} \left( \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i(F)' \varepsilon_i \right) - D^{-1}(\hat{\beta}_1 - \hat{\beta}_0) - T^{-(\nu \alpha + 1)/2} B^{-1} A_1 + o_p((NT)^{-1/2}), 
\]

which can be solved for $D^{-1}(\hat{\beta}_0 - \beta^0)$;

\[
D^{-1}(\hat{\beta}_0 - \beta^0) 
\]

\[
= (I_d - B^{-1} N + o_p(1))^{-1} \left( B^{-1} \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i(F)' \varepsilon_i - D^{-1}(\hat{\beta}_1 - \hat{\beta}_0) - T^{-(\nu \alpha + 1)/2} B^{-1} A_1 \right) 
\]

40
Here we have made use of the fact that
\[ + o_p((NT)^{-1/2}) \]
\[ = (B - N + o_p(1))^{-1} \left( \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i (\hat{F})' \varepsilon_i - BD_T^{-1} (\hat{\beta}_1 - \hat{\beta}_0) - T^{-(\nu_G + 1)/2} A_1 \right) \]
\[ + o_p((NT)^{-1/2}) \]
\[ = [B(\hat{F}) + o_p(1)]^{-1} \left( \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i (\hat{F})' \varepsilon_i - BD_T^{-1} (\hat{\beta}_1 - \hat{\beta}_0) - T^{-(\nu_G + 1)/2} A_1 \right) \]
\[ + o_p((NT)^{-1/2}) \]
\[ = B(\hat{F})^{-1} \left( \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i (\hat{F})' \varepsilon_i - BD_T^{-1} (\hat{\beta}_1 - \hat{\beta}_0) - T^{-(\nu_G + 1)/2} A_1 \right) \]
\[ + o_p(\|D_T^{-1}(\hat{\beta}_1 - \hat{\beta}_0)\|) + o_p((NT)^{-1/2}). \quad (A.5.43) \]

Here we have made use of the fact that
\[ B - N = \frac{1}{NT} \sum_{i=1}^{N} D_T X_i'M_{\hat{F}}X_i D_T - \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X_i'M_{\hat{F}} X_j D_T \nu_{ji} \]
\[ = \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i (\hat{F})' Z_i (\hat{F}) D_T = B(\hat{F}), \quad (A.5.44) \]

where \( B_i(F) \) is as in Assumption 2. Also, making use of Assumption 3, it is not difficult to show that \( (NT)^{-1} \sum_{i=1}^{N} D_T Z_i (\hat{F})' \varepsilon_i = O_p((NT)^{-1/2}) \). Hence,
\[ D_T^{-1}(\hat{\beta}_0 - \beta^0) + B(\hat{F})^{-1} BD_T^{-1} (\hat{\beta}_1 - \hat{\beta}_0) \]
\[ = B(\hat{F})^{-1} \frac{1}{NT} \sum_{i=1}^{N} D_T Z_i (\hat{F})' \varepsilon_i - T^{-(\nu_G + 1)/2} B(\hat{F})^{-1} A_1 \]
\[ + o_p(\|D_T^{-1}(\hat{\beta}_1 - \hat{\beta}_0)\|) + o_p((NT)^{-1/2}). \quad (A.5.45) \]

By using the fact that
\[ B(\hat{F})^{-1} BD_T^{-1} = D_T^{-1} \left( \sum_{i=1}^{N} Z_i (\hat{F})' Z_i (\hat{F}) \right)^{-1} \sum_{i=1}^{N} X_i'M_{\hat{F}} X_i, \quad (A.5.46) \]

the left-hand side of this last equation can be written as
\[ D_T^{-1}(\hat{\beta}_0 - \beta^0) + B(\hat{F})^{-1} BD_T^{-1} (\hat{\beta}_1 - \hat{\beta}_0) \]
It follows that

\[
\sqrt{NT}D_T^{-1} \left[ \beta_0 + \left( \sum_{i=1}^N Z_i(\hat{F})'Z_i(\hat{F}) \right)^{-1} \sum_{i=1}^N X_i'M \hat{F}X_i(\hat{\beta}_1 - \hat{\beta}_0) - \beta^0 \right],
\]

(A.5.47)

which in turn implies

\[
\|D_T^{-1}(\hat{\beta}_0 - \beta^0)\| = O_P((NT)^{-1/2})(1 + o_P(1)) = O_P((NT)^{-1/2}).
\]

(A.5.50)

The rate of convergence given in Lemma 3.1 is therefore not the best one possible. The improved rate implies that

\[
\|D_T^{-1}(\hat{\beta}_1 - \hat{\beta}_0)\| = o_P((NT)^{-1/2} \|D_T^{-1}(\hat{\beta}_0 - \beta^0)\|) = o_P((NT)^{-1/2}),
\]

(A.5.51)

which can be inserted back into (A.5.48), leading to

\[
\sqrt{NT}D_T^{-1} \left[ \beta_0 + \left( \sum_{i=1}^N Z_i(\hat{F})'Z_i(\hat{F}) \right)^{-1} \sum_{i=1}^N X_i'M \hat{F}X_i(\hat{\beta}_1 - \hat{\beta}_0) - \beta^0 \right]
\]

\[
= B(\hat{F})^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N D_TZ_i(\hat{F})'\varepsilon_i - \sqrt{NT^{-\nu_G/2}B(\hat{F})^{-1}A_1} + o_P(1).
\]

(A.5.52)

Note that \(Z_i(\hat{F})\) is \(\hat{Z}_i\) with \(\hat{a}_{ij}\) replaced by \(a_{ij}\). We now show that the effect of the estimation
of $a_{ij}$ is negligible. We begin by noting how

$$
\frac{1}{NT} \sum_{i=1}^{N} [D_T Z_i(\hat{F})' Z_i(\hat{F}) D_T - D_T \hat{Z}_i \hat{Z}_i D_T]
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T \hat{a}_{ij} - \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T a_{ij}. \quad (A.5.53)
$$

Here,

$$
\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T \hat{a}_{ij}
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T \sum_{g=1}^{G} (y_j - X_j \hat{\beta}_0 - \hat{F}_g \hat{\gamma}_{-g,j})' \hat{F}_g
$$

$$
\times \left( \hat{F}_g' \hat{g}_g' \hat{F}_g \right)^{-1} \hat{F}_g (y_i - X_i \hat{\beta}_0 - \hat{F}_g \hat{\gamma}_{-g,i})
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T \sum_{g=1}^{G} (y_j - X_j \hat{\beta}_0 - \hat{F}_g \hat{\gamma}_{-g,j})' \hat{F}_g
$$

$$
\times T^{-d} V_{g}^{-1} \hat{F}_g' (y_i - X_i \hat{\beta}_0 - \hat{F}_g \hat{\gamma}_{-g,i})
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T \sum_{g=1}^{G} [X_j (\beta^0 - \hat{\beta}_0) + F_g^0 \gamma^0_{g,j} + e_{g,j} + \varepsilon_j]' \hat{F}_g
$$

$$
\times T^{-d} V_{g}^{-1} \hat{F}_g' [X_j (\beta^0 - \hat{\beta}_0) + F_g^0 \gamma^0_{g,j} + e_{g,j} + \varepsilon_j]
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T \sum_{g=1}^{G} [X_j (\beta^0 - \hat{\beta}_0) + F_g^0 \gamma^0_{g,j} + e_{g,j} + \varepsilon_j]' \hat{F}_g
$$

$$
\times T^{-\nu_d}(T^{-\nu_d} V_{g}^{-1} \hat{F}_g' [X_j (\beta^0 - \hat{\beta}_0) + F_g^0 \gamma^0_{g,j} + e_{g,j} + \varepsilon_j](1 + o_P(1))
$$

$$
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i M_{\hat{F}} X_j D_T a_{ij}(1 + o_P(1)), \quad (A.5.54)
$$

where the second equality here follows from the definition of $V_g^0$, which is again based on taking $d_g$ as known, while the last equality follows from direct calculation. The effect of the estimation of $a_{ij}$ in $\hat{Z}_i$ is therefore negligible, which in turn implies that the right-hand side of (A.5.52) becomes

$$
\sqrt{NT D_T^{-1}} \left[ \tilde{\beta}_0 + \left( \sum_{i=1}^{N} Z_i(\hat{F})' Z_i(\hat{F}) \right)^{-1} \sum_{i=1}^{N} X'_i M_{\hat{F}} X_i (\tilde{\beta}_1 - \tilde{\beta}_0) - \beta^0 \right]
$$
\[
\sqrt{NT}D_T^{-1}\left[ \beta_0 + \left( \sum_{i=1}^{N} \hat{Z}_i \hat{Z}_i \right)^{-1} \sum_{i=1}^{N} X'_i M_F X_i (\hat{\beta}_1 - \beta_0) - \beta \right] + o_P(1)
\]
\[
= \sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) + o_P(1).
\] (A.5.55)

It follows that
\[
\sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) = B(\hat{F})^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_T Z_i (\hat{F})' e_i - \sqrt{NT^{-\nu/2}} B(\hat{F})^{-1} A_1
\]
\[
+ o_P(1).
\] (A.5.56)

Consider \((NT)^{-1/2} \sum_{i=1}^{N} D_T Z_i (\hat{F})' e_i\). From the definition of \(Z_i(F)\),
\[
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_T Z_i (\hat{F})' e_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_T Z_i (F^0)' e_i + \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_T [Z_i (\hat{F}) - Z_i (F^0)]' e_i
\]
\[
= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_T Z_i (F^0)' e_i - \sqrt{NT}(R_1 - R_2),
\] (A.5.57)

where
\[
R_1 = \frac{1}{NT} \sum_{i=1}^{N} D_T X'_i (P_{\hat{F}} - P_{F^0}) e_i,
\] (A.5.58)
\[
R_2 = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} D_T X'_j (P_{\hat{F}} - P_{F^0}) e_i.
\] (A.5.59)

Here,
\[
R_1 = \frac{1}{NT} \sum_{i=1}^{N} D_T X'_i (P_{\hat{F}} - P_{F^0}) e_i
\]
\[
= \frac{1}{NT} \sum_{i=1}^{N} D_T X'_i (T^{-\delta} \hat{F} \hat{F}' - P_{F^0}) e_i
\]
\[
= \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_T X'_i (\hat{F} - F^0H)H' F^0 e_i + \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_T X'_i (\hat{F} - F^0H) (\hat{F} - F^0H)' e_i
\]
\[
+ \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_T X'_i F^0 H (\hat{F} - F^0H)' e_i + \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_T X'_i [H H' - T^\delta (F^0 F^0)' ] F^0 e_i
\]
\[
= R_{11} + R_{12} + R_{13} + R_{14}.
\] (A.5.60)
with obvious implicit definitions of $\mathbf{R}_{11}, \ldots, \mathbf{R}_{14}$ and $\mathbf{H} = \text{diag}(\mathbf{H}_1, \ldots, \mathbf{H}_G)$. Let $\mathbf{R}_{1m,j}$ be the $j$-th row of $\mathbf{R}_{1m}$ for $m \in \{1, \ldots, 4\}$. In this notation,

$$
\|\mathbf{R}_{11,j}\| \leq \frac{1}{NT^{\beta/2+1}} \sum_{i=1}^{N} (\epsilon_i' \mathbf{F}^0 \mathbf{H} \otimes (T^{-\kappa_j/2} \mathbf{X}_{j,i})') \left\| T^{-\delta/2} \text{vec}(\hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H}) \right\| \\
= O_p((NT)^{-1/2}) T^{-\delta/2} \left\| \hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H} \right\| = o_p((NT)^{-1/2}),
$$

(A.5.61)

where the equality follows from Lemma A.2 and the fact that $T^{(v_\gamma - \delta/2)} \mathbf{H}_g = O_p(1)$. Similarly, for $\mathbf{R}_{12},$

$$
\|\mathbf{R}_{12,j}\| \leq \frac{1}{NT} \sum_{i=1}^{N} \epsilon_i' \otimes (T^{-\kappa_j/2} \mathbf{X}_{j,i})' \left\| T^{-\delta} \text{vec}[(\hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H})(\hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H})'] \right\| \\
= O_p(N^{-1/2}) T^{-\delta} \left\| \hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H} \right\|^2 = o_p((NT)^{-1/2}),
$$

(A.5.62)

where the second equality is due to (A.4.57).

Consider $\mathbf{R}_{14}$. From $T^{-\delta/2} \| \hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H} \| = o_p(1)$ and $\|T^{(v_\gamma - \delta/2)} \mathbf{H}_g\| = O_p(1)$, we obtain

$$
\|\mathbf{I}_{d_j} - T^{\delta} (\mathbf{H}' \mathbf{F}^0 \mathbf{F}^0 \mathbf{H})^{-1}\| = o_p(1),
$$

(A.5.63)

$$
\|T^{v_\gamma - \delta} \mathbf{H}_g \mathbf{H}_g' - (T^{v_\gamma} \mathbf{F}_g^0 \mathbf{F}_g^0)^{-1}\| = o_p(1).
$$

(A.5.64)

Together with Lemma A.2 this implies

$$
\|\mathbf{R}_{14,j}\| = \left\| \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} T^{-\kappa_j/2} \mathbf{X}_{j,i}' \mathbf{F}^0 [\mathbf{H} \mathbf{H}' - T^{\delta} \mathbf{H}(\mathbf{H}' \mathbf{F}^0 \mathbf{F}^0 \mathbf{H})^{-1} \mathbf{H}'] \mathbf{F}^0 \epsilon_i \right\| \\
= \left\| \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} T^{-\kappa_j/2} \mathbf{X}_{j,i}' \mathbf{F}^0 \mathbf{H} [\mathbf{I}_{d_j} - T^{\delta} (\mathbf{H}' \mathbf{F}^0 \mathbf{F}^0 \mathbf{H})^{-1}] \mathbf{H} \mathbf{F}^0 \epsilon_i \right\| \\
\leq \left\| \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} (\epsilon_i' \mathbf{F}^0 \mathbf{H} \otimes (T^{-\kappa_j/2} \mathbf{X}_{j,i}' \mathbf{F}^0 \mathbf{H}')) \left\| \mathbf{I}_{d_j} - T^{\delta} (\mathbf{H}' \mathbf{F}^0 \mathbf{F}^0 \mathbf{H})^{-1} \right\| \\
= O_p((NT)^{-1/2}) \|\mathbf{I}_{d_j} - T^{\delta} (\mathbf{H}' \mathbf{F}^0 \mathbf{F}^0 \mathbf{H})^{-1}\| = o_p((NT)^{-1/2}).
$$

(A.5.65)

It remains to consider $\mathbf{R}_{13};$

$$
\mathbf{R}_{13} = \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} \mathbf{D}_T \mathbf{X}_i' \mathbf{F}^0 \mathbf{H}(\hat{\mathbf{F}} - \mathbf{F}^0 \mathbf{H})' \epsilon_i \\
= \sum_{g=1}^{G} \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} \mathbf{D}_T \mathbf{X}_g' \mathbf{F}_g \mathbf{H}_g(\hat{\mathbf{F}}_g - \mathbf{F}_g \mathbf{H}_g)' \epsilon_i
$$

45
\[
\sum_{g=1}^{G} \frac{1}{NT^1+\delta} \sum_{i=1}^{N} D_T X'_i F'_g H'_g \left( \widehat{F}_g H_g^{-1} - F'_g \right) \varepsilon_i. \tag{A.5.66}
\]

This term can be expanded in the same way as \(L\) and in analogy to before the leading terms are those that are equivalent to \(L_2, L_8\) and \(L_{10}\). Specifically,

\[
R_{13} = \sum_{g=1}^{G} \frac{1}{NT^1+\delta} \sum_{i=1}^{N} D_T X'_i F'_g H'_g \left( \frac{1}{NT^{(\nu_g+\delta)} / 2} \sum_{j=1}^{N} Q_g^{-1} \widehat{F}_g' \gamma_{g,j}^{0} (\beta^0 - \widehat{\beta}_0)' X'_{j} \varepsilon_i \right)
\]

\[
+ \sum_{g=1}^{G} \frac{1}{NT^1+\delta} \sum_{i=1}^{N} D_T X'_i F'_g H'_g \left( \frac{1}{NT^{(\nu_g+\delta)} / 2} \sum_{j=1}^{N} Q_g^{-1} \widehat{F}_g' \gamma_{g,j}' \varepsilon_i \right)
\]

\[
+ \sum_{g=1}^{G} \frac{1}{NT^1+\delta} \sum_{i=1}^{N} D_T X'_i \left( T^{-(\nu_g+1)/2} F'_g H'_g \right) \left( N^{-1} \Gamma_{g}^{0} \right)^{-1} \left( T^{-\nu_g} F'_g F'_g \right)^{-1} T^{-(\nu_g-1)/2} \left( F'_g \Sigma \varepsilon_i \right)
\]

\[
+ o_P((NT)^{-1/2})
\]

\[
= R_{131} + R_{132} + R_{133} + o_P((NT)^{-1/2}). \tag{A.5.67}
\]

It is easy to see that \(\|R_{131}\| = o_P(\|\beta^0 - \widehat{\beta}_0\|) = o_P(1)\) and \(\|R_{133}\| = o_P((NT)^{-1/2})\). As for \(R_{132}\), by using the definition of \(Q_g\) and \((A.5.64)\), we can show that

\[
R_{132} = \sum_{g=1}^{G} \frac{1}{NT^1+\delta} \sum_{i=1}^{N} D_T X'_i F'_g H'_g \left( \frac{1}{NT^{(\nu_g+\delta)} / 2} \sum_{j=1}^{N} Q_g^{-1} \widehat{F}_g' \gamma_{g,j}^{0} \varepsilon'_i \right)
\]

\[
= \sum_{g=1}^{G} \frac{1}{NT^{1+\nu}} \sum_{i=1}^{N} D_T X'_i \left( T^{\nu_g-\delta} H'_g \right) \left( \frac{1}{N} \sum_{j=1}^{N} \left( N^{-1} \Gamma_{g}^{0} \right)^{-1} \gamma_{g,j}^{0} \varepsilon'_i \right)
\]

\[
= \sum_{g=1}^{G} \frac{1}{NT^{1+\nu}} \sum_{i=1}^{N} D_T X'_i \left( T^{-\nu_g} F'_g F'_g \right)^{-1} \left( \frac{1}{N} \sum_{j=1}^{N} \left( N^{-1} \Gamma_{g}^{0} \right)^{-1} \gamma_{g,j} \varepsilon'_i \right)
\]

\[
= \sum_{g=1}^{G} \frac{1}{NT^{(\nu_g-1)/2}} \sum_{i=1}^{N} \sum_{j=1}^{N} T^{-(\nu_g+1)/2} D_T X'_i \left( T^{-\nu_g} F'_g F'_g \right)^{-1} \left( N^{-1} \Gamma_{g}^{0} \right)^{-1} \gamma_{g,j} T^{-1} \varepsilon'_i, \tag{A.5.68}
\]

and so

\[
\sqrt{NT} R_{132}
\]

\[
= \sum_{g=1}^{G} \frac{1}{\sqrt{NT^{(\nu_g-2)/2}}} \sum_{i=1}^{N} \sum_{j=1}^{N} T^{-(\nu_g+1)/2} D_T X'_i \left( T^{-\nu_g} F'_g F'_g \right)^{-1} \left( N^{-1} \Gamma_{g}^{0} \right)^{-1} \gamma_{g,j} T^{-1} \varepsilon'_i.
\]
\[ Z \text{ to the one given above for} \]

The required result is now implied by Assumption 8. 

\[ \text{Proof of Corollary A.1.} \]

---

\[ \sqrt{NT}R_1 = \sqrt{NT}(R_{11} + R_{12} + R_{13} + R_{14}) \]

\[ = T^{(2-\nu_G)/2}N^{-1/2} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} T^{-(\nu_G+1)/2}D_TX_i'F_0G(T^{-\nu_G}F_0^{0}F_0^{0})^{-1}(N^{-1}\Gamma_0^{0}\Gamma_0^{0})^{-1}\gamma_{G,j}^{0}T^{-1}\epsilon_j'\epsilon_i \]

\[ + o_P(1). \quad (A.5.69) \]

Hence, by adding the results,

\[ \sqrt{NT}R_2 \text{ can be evaluated in exactly the same way, and the limiting representation is analogous} \]

\[ \text{to the one given above for } \sqrt{NT}R_1 \text{ with } X_i \text{ replaced by } -\sum_{j=1}^{N} X_ja_{ij}. \text{ Moreover,} \]

\[ \|B(\widehat{F}) - B(F^0)\| = o_P(1). \quad (A.5.71) \]

It follows that if we define

\[ \overline{A}_2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} T^{-(\nu_G+1)/2}D_TZ_i(0)'F_0G(T^{-\nu_G}F_0^{0}F_0^{0})^{-1}(N^{-1}\Gamma_0^{0}\Gamma_0^{0})^{-1}\gamma_{G,j}^{0}T^{-1}\epsilon_j'\epsilon_i, \quad (A.5.72) \]

where \( Z_i(0) = X_i - \sum_{j=1}^{N} X_ja_{ij} \), then (A.5.57) reduces to

\[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_TZ_i(\widehat{F})'\epsilon_i = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_TZ_i(F^0)'\epsilon_i - T^{(2-\nu_G)/2}N^{-1/2}\overline{A}_2 + o_P(1), \quad (A.5.73) \]

which in turn implies that (A.5.56) becomes

\[ \sqrt{NT}D_T^{-1}(\hat{\beta} - \beta^0) = B(\widehat{F})^{-1}\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_TZ_i(\widehat{F})'\epsilon_i - \sqrt{NT}^{-\nu_G/2}B(\widehat{F})^{-1}\overline{A}_1 + o_P(1) \]

\[ = B(F^0)^{-1}\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_TZ_i(F^0)'\epsilon_i - \sqrt{NT}^{-\nu_G/2}B(F^0)^{-1}\overline{A}_1 \]

\[ - T^{(2-\nu_G)/2}N^{-1/2}B(F^0)^{-1}\overline{A}_2 + o_P(1). \quad (A.5.74) \]

The required result is now implied by Assumption 8. 

---

\[ \text{Proof of Corollary A.1.} \]
Under $NT^{-\nu_G} \to 0$, (A.5.56) in the proof of Theorem 3.1 may be written as

$$
\sqrt{NT^2}D_T^{-1}(\tilde{\beta} - \beta^0) = B(F^0)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_TZ_i(\hat{F})'\varepsilon_i + o_P(1).
$$

Consider $(NT)^{-1} \sum_{i=1}^{N} D_TX'_iM_{\tilde{\beta}}\varepsilon_i$, which we can write as

$$
\frac{1}{NT} \sum_{i=1}^{N} D_TX'_i(T^{-\delta}\hat{F}'F' - P_F)\varepsilon_i = \frac{1}{NT} \sum_{i=1}^{N} D_TX'_iM_{F^0}\varepsilon_i - \frac{1}{NT} \sum_{i=1}^{N} D_TX'_i(P_{\hat{F}} - P_{F^0})\varepsilon_i = \frac{1}{NT} \sum_{i=1}^{N} D_TX'_iM_{F^0}\varepsilon_i - R_1. 
$$

As in Proof of Theorem 3.1,

$$
R_1 = \frac{1}{NT} \sum_{i=1}^{N} D_TX'_i(T^{-\delta}\hat{F}'F' - P_F)\varepsilon_i = \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_TX'_i(\hat{F} - F^0H)H'F^0\varepsilon_i + \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_TX'_i(\hat{F} - F^0H)(\hat{F} - F^0H)'\varepsilon_i + \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_TX'_iF^0H(\hat{F} - F^0H)'\varepsilon_i + \frac{1}{NT^{1+\delta}} \sum_{i=1}^{N} D_TX'_iF^0[HH' - (F^0F^0)^{-1}]F^0\varepsilon_i = R_{11} + R_{12} + R_{13} + R_{14},
$$

where $\|R_{11}\|, \|R_{12}\|$ and $\|R_{14}\|$ are all $o_P((NT)^{-1/2})$, just as before. Let us therefore consider $R_{13}$, which is the source of the bias in the asymptotic distribution of $\sqrt{NT^2}D_T^{-1}(\tilde{\beta} - \beta^0)$. The purpose of Assumption A.1 is to control this term. Suppose first that condition (b) holds. Let $x_{j,k,t}$ denote the $j$-th row of $X_{k,t}$. In this notation,

$$
E\left(\frac{1}{NT^{3/2+1}} \sum_{i=1}^{N} \varepsilon'_i \otimes (T^{-\kappa/2}X'_{j,i}F^0_gT^{-(\nu_g - \delta)/2})^2 \right) = \frac{1}{N^2T^{\delta+2}} \sum_{s=1}^{T} E \left( \sum_{i=1}^{N} \varepsilon_{i,s} \sum_{t=1}^{T} T^{-\kappa/2}x_{j,i,t}T^{-(\nu_g - 1)/2}f_{g,t}^{0\nu}T^{(\nu_g - 1)/2}T^{-(\nu_g - \delta)/2} \right)^2.
$$

Moreover, by applying Assumption 8 (b) to the expression given for $\|T^{-\delta/2}\hat{F}_{2,d} - T^{-\nu_2/2}F^0_2h^0_{2,d}\|$
in the proof of Lemma A.7, we can show that
\[
\|T^{-\delta/2} \text{vec}(\mathbf{F} - \mathbf{F}^0 \mathbf{H})'\| = o_P(T^{-1/2}). \tag{A.5.78}
\]
Making use of these results, we obtain
\[
\|\mathbf{R}_{3,j}\| \leq \left\| \frac{1}{NT^{8/2+1}} \sum_{i=1}^{N} \mathbf{\varepsilon}'_i \otimes (T^{-\kappa_j/2} \mathbf{X}'_{j,i} \mathbf{F}^0 \mathbf{H}) \right\| \|T^{-\delta/2} \text{vec}(\mathbf{F} - \mathbf{F}^0 \mathbf{H})'\| = o_P((NT)^{-1/2}). \tag{A.5.79}
\]
Alternatively, we may invoke Assumption 8 (a) to arrive at the same result. In this case, \[
\|T^{-\delta/2} \text{vec}(\mathbf{F} - \mathbf{F}^0 \mathbf{H})'\| = o_P(1), \] but we also have
\[
E \left\| \frac{1}{NT^{8/2+1}} \sum_{i=1}^{N} \mathbf{\varepsilon}'_i \otimes (T^{-\kappa_j/2} \mathbf{X}'_{j,i} \mathbf{F}^0 \mathbf{H}) \right\|^2 = \frac{1}{(NT)^2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{k=1}^{N} E[T^{-\kappa_j} \mathbf{X}_{j,i,t} X_{j,k,t} T^{-\nu_g-1/2} f_{g,t}^0] \sigma_{\varepsilon,ik} \]
\[
+ \frac{2}{(NT)^2} \sum_{t=2}^{T} \sum_{s<t}^{T} \sum_{i=1}^{N} \sum_{k=1}^{N} E[T^{-\kappa_j} \mathbf{X}_{j,i,t} X_{j,k,s} T^{-\nu_g-1} f_{g,t}^0 f_{g,s}^0] \sigma_{\varepsilon,ik} = O((NT)^{-1}), \tag{A.5.80}
\]
and so \(\|\mathbf{R}_{3,j}\|\) is of the same order as before. The proof under Assumption 8 (c) is simpler and is therefore omitted.

Hence, by adding the results,
\[
\sqrt{NT}\|\mathbf{R}_1\| \leq \sqrt{NT}(\|\mathbf{R}_{11}\| + \|\mathbf{R}_{12}\| + \|\mathbf{R}_{13}\| + \|\mathbf{R}_{14}\|) = o_P(1). \tag{A.5.81}
\]
We have therefore shown that
\[
\sqrt{NT} \left\| \frac{1}{NT} \sum_{i=1}^{N} \mathbf{D}_T \mathbf{X}'_i (\mathbf{P}_\mathbf{F} - \mathbf{P}_{F^0}) \mathbf{\varepsilon}_i \right\| = o_P(1), \tag{A.5.82}
\]
and we can similarly show that

\[ \sqrt{NT} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} D_T X'_i (P_{\hat{F}} - P_{F^0}) X_j D_T a_{ij} \right\| = o_P(1). \]  

(A.5.83)

These results can be inserted into (A.5.56), giving

\[ \sqrt{NT} D^{-1}_T (\hat{\beta} - \beta^0) = B(F^0)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} D_T Z_i (F^0)' \varepsilon_i + o_P(1). \]  

(A.5.84)

The sought result now follows from Assumption 8.  

■