“Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence and its Applications”

Thursday, January 24, 2013
3:30 p.m.
S016 Memorial Union
Asymptotic Theory for Dynamic Heterogeneous Panels
with Cross-Sectional Dependence and Its Applications

Minkee Song*
Columbia University
(Job Market Paper)

Abstract

This paper considers dynamic heterogeneous panels with cross-sectional dependence (DHP+CSD), where the dependence is modeled using a factor structure. Dynamics, heterogeneity and cross-sectional dependence are pervasive characteristics of most data sets and it is therefore essential for empirically realistic models to allow for the three features. It is also well-known that the persistence of aggregate series may not reflect the true persistence of its underlying series when the disaggregated data exhibit the three characteristics. In this regard, the estimation of DHP+CSD models is indispensable for examining the reliability of the existing analysis based on aggregate series. The main contribution of this paper is that it addresses the three issues in estimation all at the same time. To cope with the challenges in estimation arising from the greater flexibility of the model, we adopt an iterative principal component method and develop an asymptotic theory under large $N$ and large $T$. The proposed estimator is shown to be $\sqrt{T}$-consistent under non-stringent conditions and to perform well in finite samples. We apply the developed estimator to two empirical contexts. In the first application, we estimate the heterogeneous dynamics of sectoral real exchange rates to examine the role of aggregation bias in explaining the purchasing power parity puzzle. In the second application, we estimate the intrinsic persistence of the sectoral New Keynesian Phillips curves to investigate the degree of the forward-looking nature of price setting. Both applications illustrate that the DHP+CSD estimator can shed new light on the true nature of disaggregated data sets by extending the existing empirical models in meaningful ways.

Keywords: Dynamic panel models, heterogeneity, cross-sectional dependence, factor structure, purchasing power parity, inflation inertia, New Keynesian Phillips curve

*Email: ms3470@columbia.edu. I am grateful to Jushan Bai, John Donaldson, Dennis Kristensen, and Ricardo Reis for their inspiration, guidance and encouragement. I wish to thank Raffaella Giacomini, Jae Won Lee, Serena Ng, Christoph Rothe, Bernard Salanié, and Jón Steinsson for valuable discussions and comments. This paper has also benefited from discussions with Ji Hyung Lee, Hyunseung Oh, the participants of econometrics colloquium at Columbia University, and the participants of the Midwest Econometrics Group 2012 Meeting. All errors are my own.
1 Introduction

This paper studies the following panel data model with heterogeneous coefficients:

\[ y_{it} = x_{it}' \beta_{0,i} + u_{it} \]

and

\[ u_{it} = \lambda_i^0 F_t^0 + \varepsilon_{it} \]

where \( x_{it} \) is a \( k \times 1 \) vector of observed individual-specific regressors on the \( i \)th cross-section unit at time \( t \), which possibly includes lagged dependent variables. The main parameter of interest is \( \beta_{0,i} \), which is a \( k \times 1 \) vector of individual-specific coefficients. The cross-sectional dependence of the error term, \( u_{it} \), is modeled using a factor structure, in which \( F_t^0 \) is an \( r \times 1 \) vector of unobserved common factors, \( \lambda_i^0 \) is an \( r \times 1 \) vector of factor loadings, and \( \varepsilon_{it} \) is the idiosyncratic error.

There are three key ingredients of the model: (D) dynamics through lagged dependent variables, (H) the heterogeneity of the coefficients, and (CSD) cross-sectional dependence. In other words, we consider dynamic heterogeneous panels with cross-sectional dependence, hereafter referred to as DHP+CSD. Each ingredient is a highly likely characteristic of most data sets and has been studied extensively in its own literature. The problem of estimation under all three features, however, has not been studied extensively.\(^1\) The main contribution of this paper is that it addresses the three issues, D, H, and CSD, in estimation all at the same time under large \( N \) and large \( T \) setup. Given the prevalence of the three features in real world data sets, extending models in this direction is essential for more realistic empirical analysis.

The flexibility of the DHP+CSD model comes at the cost of posing a challenge to the estimation procedure.\(^2\) To cope with the challenge in estimation, we adopt an approach that is based on iterative principal component analysis. Bai (2009) already adopted the approach to develop an estimator for panel models with cross-sectional dependence under the assumption of slope homogeneity. While we closely follow the estimation procedure and the proof strategy in Bai (2009), substantial modification is required to incorporate heterogeneous coefficients, compared

\(^1\) Phillips and Sul (2003) is an exception; the authors studied the estimation and inference in the presence of the three features, D, H, and CSD. The asymptotics, however, is carried out under \( T \to \infty \) with a fixed \( N \).

\(^2\) The literature also has recognized the difficulty of addressing all three issues simultaneously. For example, Phillips and Sul (2003) noted the interdependence of the issues and the importance of taking a systematic approach.
to the case of slope homogeneity. The theory in this paper also differs substantially from the
typical dynamic heterogeneous panels without cross-sectional dependence, in which case the
individual estimator involves its own regressors and errors but no other individual’s information.
In the DHP+CSD model, we are still required to pool the data and exploit the information from
large cross-sections to control for the unobservable factors that are common to all individuals. In
sum, the three features lead to a situation in which each individual estimator is related to the
others in a complex manner. By disentangling the complex interdependence structure among the
individual estimators, we prove that the estimator for each individual coefficient is $\sqrt{T}$-consistent
under non-stringent conditions on $N$ and $T$. The Monte Carlo simulation results show that
the proposed DHP+CSD estimator works well in finite samples, and that the performance of
the estimator is comparable to that of an infeasible estimator that treats the true unobservable
factors as though they were observable. Furthermore, the overall performance of the DHP+CSD
estimator is satisfactory even though no factor structure is present. Therefore, it might be
prudent to always use the DHP+CSD estimator.

A by-product of the asymptotic theory is an expression that highlights the consequences
of pooling the cross-sections in estimation. In the rate of convergence proof, we express the
individual estimator as the sum of the infeasible estimator and a weighted average of the other
individual estimators. If the common factors are observable, the individual estimator simply
becomes identical to the infeasible estimator without an additional term. However, the other
individual estimators also appear in the expression because we pool the data to control for the
unobservable common factors. That is, pooling the cross-sections opens a channel through which
each estimator affect every other estimator. Roughly speaking, the channel between individuals $i$
and $j$ is stronger if their regressors $x_{it}$ and $x_{jt}$ are more correlated and/or if their loadings $\lambda_i^0$ and
$\lambda_j^0$ are more correlated. Under the special case in which all individuals are independent of one
another except through common factors, we can theoretically show that the DHP+CSD estimator
is asymptotically as efficient as the infeasible estimator. In other words, under cross-sectional
independence, pooling cross-sectional information allows us to control for common factors as
though they were observable.

The proposed estimator is related to several existing estimators that operate under some
combinations of the three features considered in this paper. The ordinary least squares (OLS) can estimate each of the heterogeneous coefficients of the dynamic equations under large $N$ and large $T$, thus covering $D$ and $H$. However, if cross-sectional dependence is induced by unobserved factors that are serially correlated, the OLS is no longer consistent due to the correlation between the regressor and the unobserved component. Bai (2009) and Moon and Weidner (2010) estimated dynamic panels with cross-sectional dependence, but their estimators were designed to estimate homogeneous coefficients. Thus, these estimators operate under the features $D$ and $CSD$. Pesaran (2006) proposed the Common Correlated Effects (CCE) estimator to estimate the heterogeneous coefficients under cross-sectional dependence. As we will show in the Monte Carlo results section, however, the CCE estimator does not function if lagged dependent variables are included as regressors. The CCE estimator works well in the presence of the lagged dependent variables, only if the coefficients are homogeneous. In sum, the CCE estimator can work with either $(H, CSD)$ or $(D, CSD)$. The DHP$^+$CSD estimator in this paper may be viewed as an extension of the aforementioned estimators, adding an additional feature to the existing supported combinations.

Notably, there is a trade-off between the restrictive assumption of homogeneity and the rate of convergence. Assuming the homogeneity of the coefficients, the estimators proposed by Bai (2009) and Moon and Weidner (2010) achieve a fast convergence rate of $\sqrt{NT}$ as long as $\frac{T}{N} \to 0$. However, the estimators that do not rely on the assumption of homogeneity, such as the DHP$^+$CSD estimator and the CCE estimator, achieve a slower convergence rate of $\sqrt{T}$. Those heterogeneous coefficient estimators work as long as $\frac{T}{N^2} \to 0$, which is less stringent than the conditions for the homogeneous coefficient estimators. There is a difference by a factor of $\frac{1}{N}$ because the homogeneous coefficient estimators pool $N$ individual series to boost the convergence rate.

We apply the developed DHP$^+$CSD estimator to two empirical applications. In the first application, we estimate the heterogeneous dynamics of sectoral real exchange rates to estimate their half-lives. The estimated half-lives are found to be substantially shorter than the estimates in the purchasing power parity (PPP) puzzle literature based on aggregate real exchange rates. This finding supports the argument in the literature that the aggregation procedure may have caused the PPP puzzle by exaggerating the estimated half-life. In addition
to the finding, applying the DHP+CSD model to the sectoral real exchange rates allows us to
distinguish between the effect of idiosyncratic shocks and common shocks, in contrast to the
existing literature which does not allow common shocks. In the second application, the intrinsic
persistence of sectoral inflation is examined in the context of the New Keynesian Phillips curve
(NKPC). Assuming a factor structure in the latent marginal cost processes, we can estimate
the sectoral NKPCs using the DHP+CSD model. The estimated intrinsic persistence at the
sectoral level is found to be highly heterogeneous and substantially lower than in the aggregate
NKPC literature. This finding suggests that the intrinsic persistence of aggregate inflation may
be exaggerated by the aggregation procedure. We confirm this finding in two ways. First, we
perform a brief meta-analysis of the NKPC literature with different degrees of disaggregation.
Second, we replace the sectoral inflation series with the aggregate inflation series within the
same empirical framework and find that the estimated intrinsic persistence is higher. From these
results, we conclude that the true nature of price setting at the sectoral level appears to be more
forward-looking than the existing analyses have suggested.

The two empirical applications are related to the well-known issue of aggregation in macroe-
conometrics. Granger (1980) observed that the persistence of an aggregate series, constructed
by aggregating heterogeneous dynamic processes, may not reflect the true persistence of the
underlying series.\textsuperscript{3} Zaffaroni (2004) extended the analysis of Granger (1980) to investigate the
role of cross-sectional dependence using a factor structure, albeit under the assumption of no
serial correlation of the factors. Pesaran and Chudik (2011) confirmed that aggregating dynamic
heterogeneous processes may exaggerate the persistence of the aggregate series and that the
factor structure also plays an important role. That is, the three features in our paper, $D$, $H$, and $CSD$, are also key elements in the issue of aggregation. Notably, the aforementioned papers
analyzed the consequences of aggregation at the population level, which may not be helpful if we
wish to investigate whether a specific empirical study, based on an aggregate series, is misleading.
To examine a given data set, we must investigate whether its underlying disaggregated series
exhibit the three key features, $D$, $H$, and $CSD$.\textsuperscript{4} The estimator proposed in this paper may

\textsuperscript{3}Robinson (1978) had also noted a similar issue.
\textsuperscript{4}A growing number of empirical studies, including Imbs et al. (2011), Altissimo et al. (2009) and Pesaran and
Chudik (2011), have advocated for this approach, by utilizing particular structures of their models.
provide useful information for examining whether an empirical study based on an aggregate series is able to reflect the true nature of the underlying disaggregated series, which is another important motivation of this paper.

The remainder of the paper is organized as follows. Section 2 explains the details of the model and the assumptions. Section 3 describes the estimation procedure based on the iterative principal component estimation. Section 4 presents the asymptotic results for the proposed estimator. The Monte Carlo simulation results are provided in Section 5. Section 6 illustrates an empirical application on the PPP puzzle, whereas Section 7 studies the dynamics of sectoral inflation. Finally, Section 8 provides conclusions. We use the following notations throughout the paper. The letter $M$ denotes a finite positive number, and $\|A\| = (tr [A' A])^{\frac{1}{2}}$ is the Euclidean norm of a generic matrix $A$. The expression $X_n = O_p(a_n)$ states that the random vector $X_n$ is at most of order $a_n$ in probability, and $X_n = o_p(a_n)$ states that $X_n$ is of smaller order than $a_n$ in probability. The operator $\to^p$ denotes convergence in probability, and $\to^d$ denotes convergence in distribution. The restrictions on the relative rates of $N$ and $T$ will be specified separately if needed. We define $C_{NT} = \min \left\{ \sqrt{N}, \sqrt{T} \right\}$, and we define $\mathcal{F}$ as the space of all $F$ that satisfy $\frac{F' F}{T} = I_r$.

2 Model

As described in the introduction, $y_{it}$ is the observation of the $i$th cross-section unit at time $t$, and is assumed to be generated from a heterogeneous panel data model, $y_{it} = x_{iit}' \beta_{0,i} + u_{it}$, where $u_{it}$ is assumed to have a factor structure, $u_{it} = \lambda_1^{0t} F_1^{0t} + \epsilon_{it}$. This specification potentially includes dynamic panel data models because we allow $x_{iit}$ to include lagged dependent variables.

Following Bai (2009) and Moon and Weidner (2010), we maintain the assumption that the number of factors is known. In the case of pure factor models where the individual regressor, $x_{iit}$, is absent, various methods have been proposed in the literature to determine the number of factors.\(^5\) The supplementary material of Bai (2009) provides an intuitive description of how to extend the method in Bai and Ng (2002) in the presence of individual regressors. In the

heterogeneous coefficients setup of this paper, however, it is not obvious whether the existing method remains valid, and a formal analysis goes beyond the scope of this paper.

Assumption A: Let $\Lambda^0 = (\lambda^0_1, \lambda^0_2, \cdots, \lambda^0_N)'$. We assume

A(i-1) $\exists M < \infty \text{ s.t. } \forall t \quad E \|F^0_t\|^4 \leq M$

A(ii-1) $\exists M < \infty \text{ s.t. } \forall i \quad E \|\lambda^0_i\|^4 \leq M$

A(iii) For every $i$, there exists a compact set $B_i$ such that $\beta_{0,i} \in B_i$

Assumption A contains the standard assumptions for factor models to guarantee the existence of $r$ distinct factors and loadings, asymptotically. In particular, the condition A(ii-2) implies that the factors are pervasive, meaning that the common factors affect almost all series in the limit. We do not assume that $F^0_t$ is independent over time; it can be a general covariance stationary process with a non-zero mean. A similar statement also holds true for $\lambda^0_i$; it can be correlated over cross-sections with a non-zero mean. Recalling that $x_{it}$ may include lagged dependent variables, the flexible specification that we allow for $F^0_t$ and $\lambda^0_i$ may lead to a potential correlation between $x_{it}$ and the common component $\lambda^0_i F^0_t$. We do allow $x_{it}$ to be correlated with $F^0_t$ and/or $\lambda^0_i$ in a general nonlinear fashion. The key idea that enables this flexible specification is to treat both $F^0_t$ and $\lambda^0_i$ as fixed-effects parameters to be estimated. It is possible to estimate $F^0_t$ and $\lambda^0_i$ and control for the unobserved heterogeneity because we consider the case where rich information can be obtained from large $N$ and large $T$. When we derive the asymptotic results, we will consider estimation on a compact set around the true value $\beta_{0,i}$, defined as in Assumption A(iii).

Assumption B: Let $X_i = (x_{i1}, x_{i2}, \cdots, x_{iT})'$ and $F^0 = (F^0_1, F^0_2, \cdots, F^0_T)'$. Define $M_F = I_T - F (F'F)^{-1} F'$ for any $T \times r$ matrix $F$, where $I_T$ denotes a $T \times T$ identity matrix.

B(i) $\exists M < \infty \text{ s.t. } \forall i,t \quad E \|x_{it}\|^4 \leq M$

B(ii) $\forall i \quad S_{ii} = \frac{\chi^2_{M_F,0,X_i}}{T} \rightarrow^p \Sigma_{ii} > 0 \text{ as } T \rightarrow \infty$

B(iii) $\exists \mu > 0 \text{ s.t. } \forall i \quad \inf_{F \in F} \mu_i(F) \geq \mu \text{ as } T \rightarrow \infty \text{ where } \mu_i(F) \text{ is the smallest eigenvalue of } \frac{\chi^2_{M_F,0,X_i}}{T}$
B(iv) \[ \inf_{F \in \mathcal{F}} \rho(F) > 0 \] where \( \rho(F) \) is the smallest eigenvalue of \( D = \frac{1}{N} \sum_{i=1}^{N} D_i \), \( D_i = B_i - C_i^tA_i^{-1}C_i \), \( A_i = \frac{X_i'M_FX_i}{T} \), \( B_i = (\lambda_i^0 \lambda_j^0) \otimes \frac{I}{T} \), and \( C_i = \lambda_i^0 \otimes \frac{1}{T} (X_i'M_F) \).

Assumption B contains conditions on \( x_{it} \), including the boundedness of moments. More importantly, \( x_{it} \) is assumed to exhibit sufficient variation such that the corresponding coefficient, \( \beta_i \), is identifiable, which is intuitively appealing.\(^6\) The final assumption guarantees the unique minimizer of the estimation objective function.\(^7\) The notations of \( \rho(F) \) and \( \mu_i(F) \) are used to emphasize that the minimum eigenvalues are functions of \( F \). As defined at the end of the introduction, \( \mathcal{F} \) denotes the space of all \( T \times r \) matrices \( F \) that satisfy \( F^tF = I_r \).

**Assumption C:**

C(i) \( \forall i, j, t, s \ \varepsilon_{it} \perp \beta_{0,j}, F_{s}^{0}, \lambda_{j}^{0} \)

C(ii-1) \( \forall i \neq j, \forall t, s \ \varepsilon_{it} \perp x_{js} \)

C(ii-2) \( \exists M < \infty \ \text{s.t.} \ \forall i, T \ \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(x_{it}x_{is}\varepsilon_{it}\varepsilon_{is})| \leq M \)

Assumption C describes the nature of the interaction between idiosyncratic errors and other components of the model. The idiosyncratic errors are assumed to be independent of all individual coefficients, factors, and loadings. We also assume in C(ii-1) that each idiosyncratic error term is independent of the other individual’s regressors. The idiosyncratic errors are allowed to be weakly serially correlated as long as the amount of serial correlation of the product \( x_{it}\varepsilon_{it} \) is limited to be small. If the individual regressor \( x_{it} \) includes lagged dependent variables, however, we assume that the idiosyncratic errors are independent over time. In the cross-sectional dimensions, \( x_{it}^{t} \beta_{0,i} \) and \( \lambda_{i}^{0} \) are allowed to be freely correlated. The amount of cross-sectional correlation of the idiosyncratic errors is limited by the following assumptions.

**Assumption D:**

D(i) \( \forall i, t \ \ E(\varepsilon_{it}) = 0 \ \text{and} \ \exists M < \infty \ \text{s.t.} \ \forall i, t \ \ E|\varepsilon_{it}|^{8} \leq M \)

D(ii-1) \( \forall i, j \ \exists \sigma_{ij} < \infty \ \text{s.t.} \ \forall t, s \ \ |\sigma_{ij,ts}| = |E(\varepsilon_{it}\varepsilon_{js})| \leq \sigma_{ij} \)

D(ii-2) \( \exists M < \infty \ \text{s.t.} \ \forall i, N \ \sum_{j=1}^{N} \sigma_{ij} \leq M \)

\(^6\) Similar assumptions are imposed in Pesaran (2006) to guarantee the identification of individual coefficients.

\(^7\) The assumption is a heterogeneous coefficients version of the corresponding assumption in Bai (2009).
D(ii-3) \( \forall t, s \ \exists \bar{\tau}_{ts} < \infty \) s.t. \( \forall i, j \ |\sigma_{ij,ts}| = |E(\varepsilon_it\varepsilon_js)| \leq \bar{\tau}_{ts} \)

D(ii-4) \( \exists M < \infty \) s.t. \( \forall T \ \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \bar{\tau}_{ts} \leq M \)

D(ii-5) \( \exists M < \infty \) s.t. \( \forall T_1 \sum_{t=1}^{T} \sum_{s=1}^{T} \bar{\tau}_{ts} \leq M \)

D(iii) \( \exists M < \infty \) s.t. \( \forall i, t, s \ E\left|\frac{1}{\sqrt{N}} \sum_{i=1}^{N} [\varepsilon_is\varepsilon_it - E(\varepsilon_is\varepsilon_it)]\right|^4 \leq M \)

D(iv-1) \( \exists M < \infty \) s.t. \( \forall u, N, T \ \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\text{cov}(\varepsilon_{iu}\varepsilon_{it}, \varepsilon_{ju}\varepsilon_{js})| \leq M \)

D(iv-2) \( \exists M < \infty \) s.t. \( \forall k, N, T \ \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\text{cov}(\varepsilon_{it}\varepsilon_{kt}, \varepsilon_{js}\varepsilon_{ks})| \leq M \)

Assumption D on idiosyncratic errors contains standard assumptions in the factor literature. The conditions allow for “weak” time series and cross-sectional correlation in the sense of the approximate factor model of Chamberlain and Rothschild (1984). This is in contrast to the exact factor model originally developed by Geweke (1977) and Sargent and Sims (1977), which impose strict uncorrelatedness assumptions. Heteroskedasticity is also allowed insofar as the moments are uniformly bounded by the condition D(i). More complicated technical assumptions D(iv-1) and D(iv-2) are adopted to limit the amount of correlation among the idiosyncratic errors, and these assumptions are similar to those in Bai (2009). See the discussion therein for the interpretation of the conditions.

3 Estimation

We consider a least squares objective function based on the sum of squared residuals (SSR) for estimation:

\[
\text{SSR} \left( \{\beta_i\}_{i=1}^{N}, F, \Lambda \right) = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - x_{it}'\beta_i - \lambda_i'Ft)^2 = \sum_{i=1}^{N} (Y_i - X_i\beta_i - F\lambda_i)'(Y_i - X_i\beta_i - F\lambda_i) \quad (1)
\]

subject to the normalization \( \frac{FTF}{N} = I_r \) and \( \frac{N\Lambda}{N} \) being diagonal. The objective function came originally from the literature on the principal components approach to pure factor models. Given the \( \beta_i \), the objective function amounts to that of pure factor models, with \( Y_i - X_i\beta_i \) being regarded as observed data. The objective function (1) implicitly uses an identity matrix as a
weighting matrix. Analogous to improving the OLS by using the generalized least squares (GLS),
we can potentially improve the least squares objective function (1) by using the inverse of the
variance matrix of the idiosyncratic errors as a weighting matrix.\(^8\)

In the factor literature, a typical approach to simplifying the aforementioned minimization
problem is to concentrate out the loadings from the objective function. Given any proposed
solution \(\beta_i\) and \(F\) to the minimization problem, each \(\lambda_i\) must satisfy a relationship of the form
\[
\lambda_i = (F'F)^{-1} F' (Y_i - X_i\beta_i) = (F'F)^{-1} F'W_i,
\]
to be a minimizer as well. Defining \(W_i = Y_i - X_i\beta_i\) and substituting the expression for the loadings into (1), we have the following concentrated objective function:
\[
SSR\left(\{\beta_i\}_{i=1}^N, F\right) = \sum_{i=1}^N \left( Y_i - X_i\beta_i - F (F'F)^{-1} F'W_i \right)' \left( Y_i - X_i\beta_i - F (F'F)^{-1} F'W_i \right)
\]
which, in turn, simplifies to
\[
SSR\left(\{\beta_i\}_{i=1}^N, F\right) = tr\left[ W'M_FW \right] = tr\left[ W'W \right] - tr\left[ F' (WW') F/T \right] \tag{2}
\]
where \(W = (W_1, W_2, \cdots, W_N)\) and \(M_F = I_r - F (F'F)^{-1} F' = I_r - \frac{1}{T} FF'\) denotes a projection matrix.

Given the \(\beta_i\), the minimization problem above is equivalent to the problem of maximizing
the second term of (2), which, in turn, is equivalent to
\[
\max_F tr \left[ F' \left( \frac{1}{NT} \sum_{i=1}^N (Y_i - X_i\beta_i) (Y_i - X_i\beta_i)' \right) F \right] \quad \text{subject to} \quad \frac{F'F}{T} = I_r
\]
where additional rescaling by \(\frac{1}{NT}\) is performed solely to ensure the existence of a proper limit;
it does not alter the solution. The solution to this final maximization problem is to set \(\hat{F}\)

---

\(^8\)Estimating the variance matrix of the idiosyncratic errors, which we will denote as \(\Sigma_\epsilon\), is challenging due to
its dimensions. If \(N > T\), the estimator of \(\Sigma_\epsilon\) is singular, and for comparable \(N\) and \(T\), the estimator is known to
behave poorly. In the context of pure factor models, several estimators have been proposed in the literature. Forni
et al. (2005) adopted a dynamic principal component approach to indirectly estimate \(\Sigma_\epsilon\). Boivin and Ng (2006)
suggested restricting the number of parameters to estimate by setting the off-diagonal elements of \(\Sigma_\epsilon\) to zero.
Stock and Watson (2005) used the Cochrane-Orcutt estimator to address serial correlation in the idiosyncratic
errors as well. Recently, Fan et al. (2011a) and Fan et al. (2011b) developed a novel approach using regularization
that yields a well-behaved estimator of \(\Sigma_\epsilon\).
equal to the rescaled eigenvectors of \( \frac{1}{N^T} \sum_{i=1}^N (Y_i - X_i\hat{\beta}_i) (Y_i - X_i\hat{\beta}_i)' \), corresponding to its \( r \) largest eigenvalues. Rescaling the eigenvectors by \( \sqrt{T} \) satisfies the normalization restriction \( EF' = I_r \). Now, given \( \hat{F} \), the least squares problem has a well-known solution for the \( \hat{\beta}_i \), which is 
\[
\hat{\beta}_i = (X'_i M_{\hat{F}} X_i)^{-1} X'_i M_{\hat{F}} Y_i.
\]

Therefore, the estimator \( \left( \{ \hat{\beta}_i \}_{i=1}^N , \hat{F} \right) \) should simultaneously solve a system of nonlinear equations 
\[
\hat{\beta}_i = (X'_i M_{\hat{F}} X_i)^{-1} X'_i M_{\hat{F}} Y_i
\]
and 
\[
\left[ \frac{1}{N^T} \sum_{i=1}^N \left( Y_i - X_i\hat{\beta}_i \right) (Y_i - X_i\hat{\beta}_i)' \right] \hat{F} = \hat{F} \hat{V}_{NT}
\]
where \( \hat{V}_{NT} \) is a diagonal matrix of \( r \) largest eigenvalues corresponding to \( \hat{F} \). Then, we can also obtain the estimator of each \( \lambda_i \) by computing \( \hat{\lambda}_i = \left( \hat{F}' \hat{F} \right)^{-1} \hat{F}' \left( Y_i - X_i\hat{\beta}_i \right) \).

4 Asymptotic theory

This section establishes the consistency and the rate of convergence of the DHP+CSD estimator proposed in the previous section. Three unique features of the current problem make the econometric theory challenging. First, the proposed estimator does not have a closed-form expression. Instead, we only have a set of equations that should be simultaneously satisfied by \( \hat{\beta}_i \) and \( \hat{F}_i \). Second, the unobserved factors are treated as parameters to be estimated, of which dimension grows with \( T \). Lastly, each \( i \) has its own parameter of interest, \( \beta_i \), and thus, the number of parameters grows with \( N \) as well.

When estimators have no closed-form expression, the approach in Newey and McFadden (1994) can be applied in general for consistency argument. In this paper, however, the growing dimension of the parameters prohibits the application of the typical approach. To overcome this difficulty, we follow a proof strategy based on an auxiliary objective function.\(^9\) In the appendix, we showed that the auxiliary objective function is uniformly close to the original objective function as \( (N, T) \) goes to infinity, and that the auxiliary objective function is uniquely

\(^9\)The proof strategy was first initiated by Bai (1994) and later adopted by Bai (2009) and Bonhomme and Manresa (2012).
minimized at the true parameter values. Using this framework, we can prove the consistency of \( \hat{\beta}_i \) for each \( i \) as well as the consistency of the space spanned by the estimated common factors.

**Theorem 1.** Consistency: Under Assumption A-D, as \((N, T) \to \infty\) jointly, the following statements hold:

(i) The estimator \( \hat{\beta}_i \) is consistent such that \( \hat{\beta}_i - \beta_{0,i} \to^p 0 \) for each \( i \).

(ii) The space spanned by the factors is consistent such that \( \| P_{\hat{F}} - P_{F_0} \| \to^p 0 \).

Once the consistency is established, we refine the rate of convergence of all terms that consist of the expression for the estimators \( \hat{\beta}_i \) and \( \hat{F} \). Again, the estimators are indirectly characterized by a set of simultaneous equations; thus, the convergence rate should be refined in multiple rounds in an iterative manner until we obtain the final rate of convergence. During this process, we need to take extra care with the presence of heterogeneous coefficients. Each estimator \( \hat{\beta}_i \) depends on \( \hat{F} \), and \( \hat{F} \) in turn depends on all \( \hat{\beta}_i \)s. These relationships imply that \( \hat{\beta}_i \), in fact, depends on all other \( \hat{\beta}_j \)s with \( i \neq j \). We address this issue by utilizing the symmetry of the model and the estimation procedure with respect to each individual unit. This allows us to properly characterize the stochastic order of many terms that involve all of the individual estimators in a complicated manner. Below is an interim result of the aforementioned procedure.

**Proposition 1.** Let \( \xi_i = \frac{1}{\sqrt{T}} X_i' M_{F_0} \varepsilon_i \) and \( S_{ii} = \frac{X_i' M_{F_0} X_i}{T} \). Under Assumptions A-D, as \((N, T) \to \infty\) jointly, \( \sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right) \) has the following expression as long as \( \frac{T}{N^2} \to 0 \):

\[
\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right) = S_{ii}^{-1} \xi_i + \frac{1}{N} \sum_{j=1}^{N} G_{ij} S_{jj}^{-1} \xi_j + o_p(1)
\]

where \( G_{ij} = \left( X_i' M_{F_0} X_i \right)^{-1} \left( X_i' M_{F_0} X_j \right) \left[ \lambda_i^{0'} \left( \frac{\Lambda_0' \Lambda_0}{N} \right)^{-1} \Lambda_j^{0'} \right]. \)

The expression in Proposition 1 shows how different individuals are interrelated with one another, when we estimate the unobserved common factors by pooling information from large cross-sections. Note that each individual estimator primarily consists of a term \( S_{ii}^{-1} \xi_i \) that comes from an infeasible estimator, and a weighted average of those terms from other individuals. If the common factors are observable, each individual estimator simply becomes identical to the
infeasible estimator without any additional term. However, other individual estimators also appear in the expression because we pool the data to control for the unobservable common factors. That is, pooling the cross-sections opens a channel through which each individual estimator affects the others. The weighting matrix $G_{ij}$ in front of each individual term $S^{-1}_{jj} \xi_j$ reflects the strength of the channel from unit $j$ to unit $i$. Roughly speaking, the channel between individuals $i$ and $j$ is stronger if their regressors, $x_{it}$ and $x_{jt}$, are more correlated and/or if their loadings, $\lambda^0_i$ and $\lambda^0_j$, are more correlated. This relationship highlights the consequences of controlling for cross-sectional dependence by pooling in estimations.

**Theorem 2.** Rate of Convergence: Under Assumptions A-D, as $(N, T) \to \infty$ jointly, 

$$
\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right) = O_p(1) \text{ as long as } \frac{T}{N^2} \to 0.
$$

From the expression in Proposition 1, we can easily predict the rate of convergence because the amount of correlation among individuals is limited by the assumptions in Section 2. Pesaran (2006) also established the same rate of convergence under the same condition on $N$ and $T$, although his proof did not explicitly cover dynamic panels. Notably, the condition $\frac{T}{N^2} \to 0$ is less stringent than the conditions in Bai (2009) or Moon and Weidner (2010). Their estimators for homogeneous coefficients require $\frac{T}{N} \to 0$, which is a stricter condition than that of Pesaran (2006) or ours. The stricter condition on $N$ and $T$ is required for $\sqrt{NT}$-consistency, which shows an interesting trade-off between the rate of convergence and the stringency of the condition on $N$ and $T$. When the coefficients are indeed homogeneous, we may obtain a faster rate of convergence. At the same time, however, the pooled estimator requires more stringent conditions which may not be satisfied by the data set at hand. In that case, we may choose to use the estimator for heterogeneous coefficients, which converges relatively slowly, because the estimator may remain valid due to less stringent requirements.

The expression given in Proposition 1 also suggests that we may achieve the same asymptotic efficiency as the infeasible estimator if the weighted average term is ignorable. In the case of static panel models, the independence of $(x^t_{it}, \lambda^t_i, \epsilon_{it})'$ over $i$ is sufficient for the weighted average term to be ignored. If $x_{it}$ includes lagged dependent variables, $(\tilde{x}^t_{it}, \lambda^t_i, \epsilon_{it})'$ is required to be
independent over \(i\) where \(\tilde{x}_{it} = M_{F0}x_{it}\). Here, we present our final asymptotic result for these special cases.

**Theorem 3.** Asymptotic Normality: Suppose Assumptions A-D hold and \((\tilde{x}'_{it}, \lambda'_i, \varepsilon_{it})'\) is independent over \(i\). As \((N, T) \to \infty\) jointly, \(\sqrt{T}(\hat{\beta}_i - \beta_0) - \mathcal{N}(0, \Omega_i)\) as long as \(T/N^2 \to 0\) where \(\Omega_i = \Sigma^{-1}_i \Xi_i \Sigma^{-1}_i\), \(\Sigma_i = \lim_{T \to \infty} (X'_i M_{F0} X_i)/T\), \(\Xi_i = \lim_{T \to \infty} (X'_i M_{F0} E(\varepsilon_i \varepsilon'_i) M_{F0} X_i)/T\), and \(\tilde{x}_{it} = M_{F0}x_{it}\).

5 Monte Carlo results

We evaluate the finite sample performance of the proposed DHP+CSD estimator using Monte Carlo experiments based on 1000 repetitions. We consider a simple data generating process (DGP) with a single factor:

\[
y_{i,t} = \rho_i y_{i,t-1} + \lambda_i F_t + \varepsilon_{i,t}
\]

The DGP is a restricted case of the general model studied in this paper, and it is relevant to the empirical application in a later section. The heterogeneous autoregressive coefficients \(\rho_i\) are drawn from \(U[0, 1]\). We draw the loadings \(\lambda_i\) independently from \(\mathcal{N}(1, 0.5)\). The single factor \(F_t\) is assumed to follow a simple dynamics of AR(1) with an independent white noise \(u_t \sim \mathcal{N}(0, 1)\) as its innovation process:

\[
F_t = \rho_F F_{t-1} + u_t
\]

We use \(\rho_F = 0.7\) to introduce serial correlation in \(F_t\), and the first 1,000 observations of the generated data are burned-in. Serial correlation in \(F_t\) leads to the correlation between the regressor \(y_{i,t-1}\) and the factor \(F_t\). Therefore, it should be noted that any estimator without special treatment of the unobserved factor \(F_t\) will not be consistent. Regarding the idiosyncratic errors \(\varepsilon_{i,t}\), we follow the scheme used by Doz et al. (2011) to allow for both cross-sectional dependence and heteroskedasticity. The variance \(\sigma^2_{i}\) of each idiosyncratic error term is drawn independently from \(U[0.5, 1.5]\) to introduce heteroskedasticity. For cross-sectional dependence among the idiosyncratic disturbances, a typical entry of the variance-covariance matrix of the idiosyncratic errors has the form \(\tau_{ij} = \tau^{(i-j)} \sqrt{\sigma^2_i \sigma^2_j}\) such that the entire variance-covariance matrix is a Toeplitz matrix. The correlation between \(\varepsilon_{i,t}\) and \(\varepsilon_{j,t}\) gets smaller as the distance
between the two series increases, and the tuning parameter $\tau$ controls the rate of decay and, hence, the amount of cross-sectional dependence. Note that no specific cross-sectional ordering of the series matters, and any cross-sectional permutation of the series will also work.

Given the DGP, we generate artificial data with various combinations of $(N, T)$ while burning-in the first 1,000 observations. We then use the generated artificial data to estimate all $\rho_i$ parameters. The procedure is repeated $B = 10,000$ times to evaluate the performance of four types of estimators: the OLS estimator, the Common Correlated Effects (CCE) estimator from Pesaran (2006), the DHP+CSD estimator and the infeasible estimator. The DHP+CSD estimator is the estimator of primary interest. The infeasible estimator uses the realization of the true factor processes as though they were observable, and its performance is used as a benchmark for comparison. The CCE estimator is included in the analysis to examine whether the method of controlling for unobservable common factors by cross-sectional average remains valid in the case of dynamic heterogeneous panels. The OLS estimator ignores the presence of the unobserved factor $F_t$ and, hence, is inconsistent. The result for the OLS estimator is included to illustrate the extent to which it can go wrong without the use of the DHP+CSD estimator. We evaluate the performance of each type of estimator in two ways. The first approach follows Pesaran (2006) and compares the performance of the mean group (MG) estimator $\hat{\rho} = \frac{1}{N} \sum_{i=1}^{N} \hat{\rho}_i$, in which case the parameter of interest is the mean $E(\rho_i)$ of the heterogeneous coefficients. The MG estimator is often of interest in empirical analysis, as we will see in the empirical application sections. The second approach calculates and compares the integrated mean squared error (IMSE) $\frac{1}{B} \sum_{k=1}^{B} \left( \frac{1}{N} \sum_{i=1}^{N} |\hat{\rho}_i - \rho_i|^2 \right)$, which is analogous to the IMSE in nonparametrics. The IMSE better reflects the performance of each individual estimator because the measure is based on the deviation from the true parameter of each estimates.

Table 1 shows the performance of the MG estimator based on the four types of estimators when both cross-sectional dependence and heteroskedasticity are present in the idiosyncratic errors. The first point to note from the results is the inconsistency of the OLS estimator, which can be expected from the design of the DGP. Most of the large root mean squared error (RMSE) of the OLS estimator comes from bias, which does not vanish even though the sample size increases. The poor performance of the OLS estimator confirms the necessity of developing a
new estimator that controls for unobservable common factors. The CCE estimator also suffers from non-negligible bias that does not vanish as the sample size increases. The bias appears to be coming from the averaging of the dynamic equations to recover the space spanned by the factors. By the mechanism of Granger (1980) and Zaffaroni (2004), averaging a large number of heterogeneous dynamic processes exaggerates the persistence parameter, thus distorting the space spanned by the factors. The performance of the new estimator proposed in this paper is reported under the column labeled DHP+CSD. We can see the bias and standard deviation hence, the RMSE of the DHP+CSD estimator decrease as the sample size increases. As in the case of approximate factor models, the cross-sectional correlation and heteroskedasticity do not jeopardize the validity of the DHP+CSD estimator as long as the amount of correlation is limited to be small. The DHP+CSD estimator achieves a comparable level of performance to that of the infeasible estimator, and the qualitative pattern over various sample sizes \((N,T)\) is also broadly consistent with that of the infeasible estimator. Obviously, the infeasible estimator
performs better than the DHP+CSD estimator because the DHP+CSD estimator must estimate unobserved factors in addition to other parameters. The pattern of relative performance of the four types of estimators is also confirmed in Table 2, in which performance is measured in terms of the IMSE of all individual coefficient estimators.

Table 2: IMSE performance of the individual coefficient estimators under cross-sectional dependence and heteroskedasticity

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>OLS</th>
<th>CCE</th>
<th>DHP+CSD</th>
<th>Infeasible</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>20</td>
<td>0.0739</td>
<td>0.0548</td>
<td>0.0288</td>
<td>0.0252</td>
</tr>
<tr>
<td>200</td>
<td>50</td>
<td>0.0648</td>
<td>0.0202</td>
<td>0.0086</td>
<td>0.0081</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>0.0635</td>
<td>0.0144</td>
<td>0.0039</td>
<td>0.0038</td>
</tr>
<tr>
<td>200</td>
<td>200</td>
<td>0.0633</td>
<td>0.0121</td>
<td>0.0018</td>
<td>0.0018</td>
</tr>
<tr>
<td>20</td>
<td>200</td>
<td>0.0631</td>
<td>0.0166</td>
<td>0.0021</td>
<td>0.0018</td>
</tr>
<tr>
<td>50</td>
<td>200</td>
<td>0.0632</td>
<td>0.0138</td>
<td>0.0019</td>
<td>0.0018</td>
</tr>
<tr>
<td>100</td>
<td>200</td>
<td>0.0633</td>
<td>0.0127</td>
<td>0.0019</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

It is interesting to investigate what happens if we use the DHP+CSD estimator when no factor structure is present in the errors. Without the presence of unobserved factors, one can simply use the OLS estimator for each series to estimate the heterogeneous coefficients. However, in an effort to reduce the risk of inconsistency, one might still use the DHP+CSD estimator and avoid using the OLS estimator. In this case, it is important for the DHP+CSD estimator to exhibit satisfactory performance that is robust to the presence of a factor structure in the errors. We consider a similar DGP to that used previously; the only difference is that the factors and loadings are excluded from the DGP. Table 3 reports the result for the DHP+CSD estimator as well as the result for the OLS estimator as a benchmark. As expected, the OLS estimator consistently estimates the true parameter. We observe that the DHP+CSD estimator also consistently estimates the true parameter as the sample size increases. More importantly, although the OLS estimator performs better in all aspects, the overall performance of the DHP+CSD estimator is comparable to that of the OLS estimator. This result indicates that it might be prudent to always use the DHP+CSD estimator. The DHP+CSD estimator always delivers consistent estimates, and the price paid in terms of efficiency is not large even if no factor structure is present. The pattern of relative performance is also confirmed in Table 4, in which performance is measured in terms of the IMSE of all individual coefficient estimators.
### Table 3: Performance of the Mean Group estimator under no common factor structure

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>DHP+CSD</th>
<th>OLS</th>
<th>N</th>
<th>T</th>
<th>DHP+CSD</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>-0.2346</td>
<td>-0.0767</td>
<td>100</td>
<td>10</td>
<td>Bias</td>
<td>-0.0096</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0746</td>
<td>0.0434</td>
<td>10</td>
<td>100</td>
<td>SD</td>
<td>0.0947</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.2462</td>
<td>0.0882</td>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.0952</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.0729</td>
<td>-0.0437</td>
<td>100</td>
<td>20</td>
<td>Bias</td>
<td>-0.0094</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0471</td>
<td>0.0356</td>
<td>20</td>
<td>100</td>
<td>SD</td>
<td>0.0681</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0868</td>
<td>0.0563</td>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.0687</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.0209</td>
<td>-0.0190</td>
<td>100</td>
<td>50</td>
<td>Bias</td>
<td>-0.0098</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.0321</td>
<td>0.0316</td>
<td>50</td>
<td>100</td>
<td>SD</td>
<td>0.0427</td>
</tr>
<tr>
<td></td>
<td>RMSE</td>
<td>0.0383</td>
<td>0.0368</td>
<td></td>
<td></td>
<td>RMSE</td>
<td>0.0438</td>
</tr>
</tbody>
</table>

### Table 4: IMSE performance of the individual coefficient estimators under no common factor structure

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>DHP+CSD</th>
<th>OLS</th>
<th>N</th>
<th>T</th>
<th>DHP+CSD</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>10</td>
<td>0.1848</td>
<td>0.0934</td>
<td>10</td>
<td>100</td>
<td>0.0070</td>
<td>0.0070</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.0509</td>
<td>0.0407</td>
<td>20</td>
<td>100</td>
<td>0.0071</td>
<td>0.0070</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.0152</td>
<td>0.0147</td>
<td>50</td>
<td>100</td>
<td>0.0071</td>
<td>0.0070</td>
</tr>
</tbody>
</table>

### 6 Empirical application I: Real exchange rate dynamics

In this section, we apply the new DHP+CSD estimator to the sectoral real exchange rates in an effort to examine the role of aggregation bias in explaining the purchasing power parity puzzle.

#### 6.1 PPP puzzle

The theory of PPP predicts that the real exchange rate between any two countries, defined as the relative price levels converted into a common currency, is constant. In a more exacting version of the PPP theory, price levels, once converted to the same currency, should be equal. In contrast to the theoretical predictions, empirical research has found large and persistent deviations from the PPP in real exchange rate data. In a survey, Rogoff (1996) reported an estimated half-life of three to five years as the consensus in the empirical literature. This failure to reconcile the theoretical predictions with the empirical findings is referred to as the PPP puzzle, one of the six major puzzles in international economics identified by Obstfeld and Rogoff (2001).

Among the numerous attempts to solve the PPP puzzle, Imbs et al. (2005) noted a potential upward bias in the half-life estimated in the literature. These authors argued that, as observed by
Granger (1980), aggregating a large number of heterogeneous dynamic processes may exaggerate the persistence and, hence, the half-life of deviations from the PPP. Therefore, Imbs et al. (2005) advocated directly analyzing disaggregated data. By computing the half-life using the mean group estimates of the dynamics of the sectoral real exchange rates, the authors obtained half-life estimates in the range of 11 to 26 months, depending on the specifications. These estimates are much shorter than the consensus estimates and thus favor the implications of the PPP theory. Some authors criticized the findings of Imbs et al. (2005) by arguing that the upward bias from the aggregation procedure may not, in practice, be large enough to explain the PPP puzzle (see, for example, Chen and Engel (2005) and Gadea and Mayoral (2009)).

This paper revisits this debate by distinguishing between the two different types of shocks: the common shocks that affect all real exchange rates and the idiosyncratic shocks that are sector-specific. Notably, the aforementioned studies did not distinguish between potential differences in impulse responses with respect to common and idiosyncratic shocks. Several papers, including Murray and Papell (2005) and Choi et al. (2006), explicitly addressed the cross-sectional dependence that can arise from common factors but assumed that the common factors are not serially correlated. Ignoring the presence of common factors may lead to a mismeasurement of the estimated half-lives of the sectoral real exchange rates, especially when the common factors are serially correlated. The DHP+CSD estimator proposed in this paper allows estimating the heterogeneous dynamics of sectoral real exchange rates while also considering the dynamics of the common factors. We model the dynamics of sectoral real exchange rates as follows:

\[
y_{i,t} = c_i + a_{i1}y_{i,t-1} + a_{i2}y_{i,t-2} + \cdots + a_{ip}y_{i,t-p} + \lambda_i F_t + \varepsilon_{it}
\]

where the single factor \(F_t\) captures the effect of the reduced-form common shock. We define \(A_i(L) = 1 - a_{i1}L - a_{i2}L^2 - \cdots - a_{ip}L^p\) for later use. To compute the half-life with respect to the common shock, we must model the dynamics of the common factor. We assumed that the common factor evolves following an autoregression of \(B(L)F_t = u_t\).

Two approaches have been proposed in the literature regarding how to compare the estimates based on disaggregated data to the estimate based on an aggregate series. The first
approach calculates the half-life based on the MG estimator

\[ \hat{A}(L) = 1 - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{a}_{i1} \right) L - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{a}_{i2} \right) L^2 - \cdots - \left( \frac{1}{N} \sum_{i=1}^{N} \hat{a}_{ip} \right) L^p \]

as a representative half-life of the disaggregated data. Carvalho and Nechio (2011) (hereafter CN) showed that the dynamics based on MG estimator corresponds to the dynamics of a counterfactual one-sector economy. CN further defined total heterogeneity effect as the difference between the half-life based on an aggregate series and the half-life based on the counterfactual one-sector economy. The second approach calculates the half-lives of all sectoral real exchange rates, and then compares the mean of the heterogeneous half-lives to the half-life of an aggregate real exchange rate. CN defined aggregation effect as the difference between the two measures in the second approach. CN further argued that the role of aggregation has been controversial in the literature primarily because some studies examined the total heterogeneity effect while the others examined the aggregation effect.

6.2 Estimation results

We used the Eurostat data used by Imbs et al. (2005) and Carvalho and Nechio (2011) to estimate the dynamics of sectoral real exchange rates. The data cover the real exchange rates of 11 European countries versus the U.S. and consider 19 categories of goods and services per country. See Appendix 3 in Imbs et al. (2005) for the lists of countries and categories. Excluding the series with missing values, we analyzed a sample of 165 series from January 1981 to August 1994. For our benchmark case, we followed Imbs et al. (2005) and chose \( p = 19 \) for the order of the autoregressive polynomial \( A_i(L) \). We performed a robustness check with \( p = 12 \) according to the additional specifications of Imbs et al. (2005) and with \( p = 2 \) according to the specifications of Carvalho and Nechio (2011). This robustness check yielded qualitatively similar conclusions. Following Imbs et al. (2005), we assumed that the order of autoregression \( B(L) \) for the common factor dynamics was 19 for our benchmark case. We also obtained similar results from a robustness check with orders of 2 and 12. We used the half-life measure used by Imbs et al. (2005) and Carvalho and Nechio (2011).
Table 5: Estimated half-life of a counterfactual one-sector economy

<table>
<thead>
<tr>
<th></th>
<th>Using aggregates (Imbs et al., 2005)</th>
<th>MG based on OLS (Carvalho and Nechio, 2011)</th>
<th>MG based on DHP+CSD (This paper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half-life</td>
<td>46 months</td>
<td>26 months</td>
<td>idio. shock 18 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>common shock 27 months</td>
</tr>
</tbody>
</table>

We report the half-lives with respect to the common and idiosyncratic shocks in Table 5. The third and fourth columns correspond to the half-lives with respect to the idiosyncratic and common shocks, respectively. For comparison purposes, we included the existing estimates in the first and second columns. The estimated half-life using the aggregate series is from Table II in Imbs et al. (2005) and is reported in the first column. The estimated half-life of the counterfactual one-sector economy based on the OLS estimates, without considering common factors, is from Table 2 in CN and is reported in the second column. The differences between the half-life estimated using the aggregate series and that estimated using the disaggregated series correspond to the total heterogeneity effect described in CN. Table 5 clearly indicates that the half-life estimated using disaggregated series is shorter than a half-life estimated using an aggregate series, irrespective of whether the shock is common or idiosyncratic in nature. This finding agrees with the findings of Imbs et al. (2005) and Carvalho and Nechio (2011); the half-life estimated using aggregate real exchange rates appears to contradict the PPP, primarily due to upward bias from the aggregation procedure. The magnitude of estimated half-life, however, varies depending on the nature of the shocks. The adjustment after the sector-specific shock is relatively rapid, whereas the effect of the common shock is persistent. In particular, the half-life with respect to idiosyncratic shock is only two thirds of the half-life with respect to common shock. Note also that the half-life estimated by CN is between our two estimates, whereas it is closer to the half-life of common shocks. That is, the innovation term in the specification of CN appears to capture the mixed effect of the two types of shocks, due to the misspecification caused by ignoring the role of the common factors. When the nature of the shocks is distinguished, we find more evidence for the PPP, especially if the shock is sector-specific. These findings illustrate that the application of a flexible model, such as the DHP+CSD model, allows us to clearly distinguish the effects of shocks of different natures, thus delivering richer implications.
Next, we decomposed the total heterogeneity effect reported in Table 5. In this decomposition scheme according to CN, the *aggregation effect* is defined as the difference between the persistence of the aggregate series and the mean persistence of the disaggregated series. The first and the second columns in Table 6 report the two corresponding persistence estimates measured in terms of half-life. The mean half-life of CN is based on OLS estimates without consideration of common factors. Based on the observation that a mean persistence of 43.2 months is close to the persistence of the aggregate series at 46 months, CN concluded that upward bias arising purely from the aggregation may be small. These authors further argued that the decomposition reconciles the seemingly conflicting findings of Imbs et al. (2005) and Chen and Engel (2005); Imbs et al. (2005) found large total aggregation effect, whereas Chen and Engel (2005) found small total aggregation effect. Our estimates, however, indicate that the magnitude of the aggregation effect is not small if we consider common factors. The third and the fourth columns report the mean half-lives of the disaggregated real exchange rates with respect to idiosyncratic and common shocks, respectively. Both estimates are clearly shorter than the estimate of CN, which ignored the role of common factors. In other words, the aggregation effect is found to be larger if the common factors are also considered, a finding consistent with the observations of Zaffaroni (2004) and Pesaran and Chudik (2011). The potential misspecification caused by ignoring the common factors apparently led to an underestimation of the degree of the aggregation effect.

<table>
<thead>
<tr>
<th>Half-life</th>
<th>Using aggregates (Imbs et al., 2005)</th>
<th>Mean HL based on OLS (Carvalho and Nechio, 2011)</th>
<th>Mean HL based on DHP+CSD (This paper)</th>
</tr>
</thead>
<tbody>
<tr>
<td>idio. shock</td>
<td>46 months</td>
<td>43.2 months</td>
<td>20.6 months</td>
</tr>
<tr>
<td>common shock</td>
<td>46 months</td>
<td>43.2 months</td>
<td>20.6 months</td>
</tr>
</tbody>
</table>

Table 6 also shows that common shocks have a far more persistent effect on average than do idiosyncratic shocks. We further investigate the finding using the entire histograms for the estimated half-lives, as plotted in Figure 1. The top panel of Figure 1 plots the estimated half-lives with respect to idiosyncratic shock, and the bottom panel presents half-lives with respect to common shock. We confirm that common shocks have a more persistent effect than do sector-specific shocks. Heterogeneity is another conspicuous characteristic of Figure 1. With
respect to common shocks, in particular, some sectors exhibit two or three times persistent deviations from the PPP compared to other series. Consistent with the existing literature, we believe that heterogeneity has contributed to the gap between estimated half-lives from the analyses based on the aggregate and disaggregated real exchange rates.

7 Empirical application II: Inflation dynamics

7.1 Inflation persistence

Persistence is one of the key characteristics of inflation, and thus lagged inflation has played an important role in many empirical models of inflation dynamics. In the accelerationist Phillips curve, such as the triangle model of Gordon (1982), the lags of inflation are included in the model, in addition to the driving processes that correspond to potential supply-shifters and the deviation from the nonaccelerating inflation rate of unemployment (NAIRU). In the New Keynesian Phillips curve, the lags of inflation appear in the reduced-form model if backward-looking term is present.
in the original NKPC equation.

In contrast to the consensus view that inflation is persistent, several authors have found a significantly lower degree of persistence using disaggregated inflation data (see, for example, Bils and Klenow, 2004, Clark, 2006 and Altissimo et al., 2009). The evidence from disaggregated data may be more directly relevant to the models of inflation dynamics, because the dominant models for the micro-foundations of inflation dynamics concern the price-setting decisions of individual agents at the micro level (Fuhrer, 2010). Given the relevance of analyzing disaggregated data, we extend the aforementioned empirical studies on inflation dynamics at the sectoral level. Both Bils and Klenow (2004) and Clark (2006) estimated the persistence of each sectoral inflation series without considering the potential cross-sectional dependence in the innovation terms. If the common factors that induce cross-sectional dependence are serially correlated, the series-by-series estimator may suffer from omitted variable bias. In that case, it is essential to exploit the information in the cross-sectional dependence to control for the unobservable common factors. Altissimo et al. (2009) incorporated the presence of common factors in their specifications, but they confined the analysis to a single factor case with a special structure for its dynamics to facilitate estimation.

We relax the restrictions of the existing studies and estimate a more general model, with heterogeneous dynamics and a multi-factor structure that induces cross-sectional dependence, using the DHP+CSD model of the form

\[
\pi_{it} = \delta_{1i} \pi_{i,t-1} + \lambda_i F_t + \varepsilon_{it}.
\]  

(4)

It is worth noting that the specification of the DHP+CSD model is sufficiently general to encompass the disaggregated version of both the accelerationist Phillips curve and the New Keynesian Phillips curve. The sectoral version of the accelerationist Phillips curve immediately becomes a DHP+CSD model if the deviation from the NAIRU has heterogeneous effects and/or if the supply-shifter at the sectoral level has a common factor structure. The sectoral New Keynesian Phillips curves can also be transformed into a DHP+CSD model by assuming a common factor structure in the sectoral real marginal costs and by solving the rational expectations model. In
the next subsection, we make a short detour to show that a DHP+CSD model is a reduced-form representation of the sectoral NKPCs. A reader who is only interested in the reduced-form analysis may skip the next subsection.

7.2 Intrinsic persistence of the sectoral New Keynesian Phillips curves

Once the reduced-form models are estimated, researchers typically distinguish two sources of persistence: intrinsic persistence and inherited persistence. The former refers to persistence that is “intrinsic” to itself through the lags, whereas the latter refers to persistence that is “inherited” from the driving processes. Although this decomposition is relevant in the contexts of both the accelerationist Phillips curve and the New Keynesian Phillips curve, we choose to focus on the latter. In the New Keynesian Phillips curve, the intrinsic persistence is captured by the coefficient in front of the backward-looking term. By utilizing the structural model behind the New Keynesian Phillips curve, we can further recover the intrinsic persistence parameter from the reduced-form coefficient in front of the lagged inflation. Below, we explain the details of the procedure in the context of the sectoral New Keynesian Phillips curves.

Imbs et al. (2011) derived the sectoral New Keynesian Phillips curves which describe the dynamics of inflation in each sector in terms of lagged inflation, expectation of future inflation, and real marginal cost:

\[ \pi_{it} = \gamma_b^i \pi_{it-1} + \gamma_f^i E_t \pi_{i,t+1} + \kappa_i \hat{mc}_{it} + \xi_{it} \]

where \( \pi_{it} \) is sector \( i \)'s inflation at time \( t \), \( \hat{mc}_{it} \) is the log-deviation of the real marginal cost, and \( \xi_{it} \) is typically interpreted as markup shocks. The parameters \( \gamma_b^i \) and \( \gamma_f^i \) capture the degree to which the pricing decisions are either backward-looking or forward-looking, respectively. The effect of the real marginal costs is captured by the parameter \( \kappa_i \). As this paper focuses on intrinsic persistence, \( \gamma_b^i \) is the primary parameter of interest.

It should be noted that the NKPC parameters above are functions of the parameters of the underlying structural model. For each sector \( i = 1, 2, \cdots, N \), a continuum of monopolistically competitive firms produce goods and set their prices subject to nominal rigidities following Calvo (1983). Following the partial indexation scheme in Smets and Wouters (2003), a fraction of the firms that do not re-optimize their prices are assumed to adjust the prices to keep up with the
inflation from the previous period. The assumption of partial indexation leads to the presence of intrinsic persistence, which is captured by $\gamma_i^b = \frac{\omega_i}{1 + \beta \omega_i}$, where the structural parameter $\omega_i$ is the fraction of firms that index to past inflation. Given the household discount factor $\beta$, there is a one-to-one mapping between $\gamma_i^b$ and $\omega_i$. The coefficient $\gamma_i^f$ of the forward-looking term has an expression of $\gamma_i^f = \frac{\beta}{1 + \beta \omega_i}$.

See Imbs et al. (2011) for further details of the model. Each sector’s NKPC equation is a rational expectations model that can be solved as

$$\pi_{it} = \frac{1}{\delta_2 i \gamma_i^f} \sum_{k=0}^{\infty} (\delta_2 i \gamma_i^f)^k E_t [\kappa_i \hat{mc}_{i,t+k} + \xi_{i,t+k}]$$

where $\delta_1 i \equiv \frac{1 - \sqrt{1 - 4 \gamma_i^b \gamma_i^f}}{2 \gamma_i^f} < 1$ and $\delta_2 i \equiv \frac{1 + \sqrt{1 - 4 \gamma_i^b \gamma_i^f}}{2 \gamma_i^f} > 1$. Further substituting the expression of $\gamma_i^b$ and $\gamma_i^f$, we show that $\delta_1 i$ is, in fact, equal to the degree of indexation $\omega_i$.

Regarding the sectoral real marginal costs, we assume a factor structure $\hat{mc}_{it} = a_i \hat{mc}_t + v_{it}$ where $\hat{mc}_t$ is the economy-wide real marginal cost process, and $v_{it}$ are idiosyncratic disturbances that may be cross-sectionally correlated and heteroskedastic. That is, each sector’s real marginal cost is not only affected by what happens to the entire economy but also affected by sector-specific shocks. The effect from the economy-wide real marginal costs may be different across sectors because we allow the loadings $a_i$ to vary across $i$. The factor structure of the real marginal costs is supported by numerous existing empirical studies that have found a factor structure in the sectoral inflation data (see, for example, Clark, 2006 and Reis and Watson, 2010). Given the specification of sectoral NKPCs, denying a factor structure in the sectoral real marginal costs would conflict with the empirical evidence for the factor structure in the sectoral inflation.

Assuming a factor structure in $\hat{mc}_{it}$ has the additional benefit of avoiding the measurement issue of the proxies for the real marginal costs. See, for example, Rudd and Whelan (2005) and Mazumder (2010) for discussions on the proxy issue. Instead of using unsatisfactory proxy measures, we extract the information on the real marginal cost process from large cross-sections. Note that it is essential in this exercise to consider a large number of sectors at a sufficiently disaggregated level because researchers are concerned about exaggerated persistence due to the aggregation procedure. The number of sectors, however, has been limited in the existing empirical studies, primarily due to the availability of proxy measures for the real marginal costs at the
sectoral level. In contrast, our approach treats the real marginal costs as latent and, hence, free from this constraint. Notably, we can also treat the real marginal costs as latent by relying on dynamic stochastic general equilibrium (DSGE) models. The likelihood of a fully specified DSGE model allows us to control for the latent real marginal cost processes. The validity of the DSGE approach, however, relies on a strong assumption that the model of the entire economy, as well as the model of price setting, is correctly specified. Furthermore, the number of sectors in the DSGE model is likely to be limited because the estimation of a multi-sector DSGE model typically involves a heavy computational burden. In contrast, our approach relies on the simple assumption of a factor structure that is consistent with the empirical evidence, and the estimation procedure is much more computationally efficient.

Under the assumption of a factor structure in the real marginal costs, we have

\[
\pi_{it} = \delta_1 \pi_{i,t-1} + \left( \frac{1}{\delta_2 \gamma_i} \right) \sum_{k=0}^{\infty} (\delta_2^{-1})^k E_t [\kappa_i a_i \hat{mc}_{t+k} + \kappa_i v_{i,t+k} + \xi_{i,t+k}].
\]

To cope with the expectations of the future terms, we need to model the dynamics of the markup shock and the real marginal cost process. Typically in the literature, the markup shock is assumed to be independent over time such that \( E_t \xi_{i,t+k} = 0 \) for all future terms. Consequently, only the current term \( \xi_{it} \) survives. Schorfheide (2008) discussed the difficulty of identifying the intrinsic persistence \( \gamma_i \) when the markup shock is allowed to be serially correlated. In a similar vein, we assume \( E_t v_{i,t+k} = 0 \) for the sector-specific components of the real marginal costs. Note, however, that we do allow for cross-sectional correlation and the heteroskedasticity of both sectoral markup shocks and sectoral real marginal costs. Next, we assume an AR(p) dynamics of the economy-wide real marginal cost process \( \hat{mc}_t \). Suppose, for illustration purposes, that \( \hat{mc}_t \) follows a simple AR(1) dynamics of \( \hat{mc}_t = \phi \hat{mc}_{t-1} + u_t \), where the innovation process \( u_t \) is orthogonal to the sector-specific markup shocks. The specification leads to a representation of sectoral inflation without the expectation of the future terms as follows:

\[
\pi_{it} = \delta_1 \pi_{i,t-1} + \left( \frac{\kappa_i a_i}{\delta_2 \gamma_i} \right) \left\{ \sum_{k=0}^{\infty} \left( \frac{\phi}{\delta_2} \right)^k \right\} \hat{mc}_t + \left( \frac{\kappa_i v_{it} + \xi_{it}}{\delta_2 \gamma_i} \right).\]
As $|\phi| < 1$ and $\delta_{2i} > 1$, the infinite sum converges to $(1 - \phi/\delta_{2i})^{-1}$, and we have a DHP+CSD model

$$\pi_{it} = \delta_{1i} \pi_{i,t-1} + \lambda_i F_t + \varepsilon_{it}$$

where $F_t \equiv \hat{mc}_t$, $\lambda_i \equiv \frac{\kappa_i a_i}{\delta_{2i} \gamma_i' (1 - \phi/\delta_{2i})}$ and $\varepsilon_{it} \equiv \frac{\kappa_i \nu_i + \xi_i}{\delta_{2i} \gamma_i'}$. The proposed DHP+CSD estimator delivers the estimate of $\delta_{1i}$ for each sector $i$, which is in fact the estimate of $\omega_i$. Given the calibrated $\beta = 0.99$, the intrinsic persistence parameter $\gamma_i^b$ can be calculated for every sector. Note that the economy-wide real marginal costs play the role of the common factor. In the general case of AR($p$) for the dynamics of $\hat{mc}_t$, we still have the static factor representation, where the static factor $F_t$ spans the space spanned by $\hat{mc}_t$ and its lags. In such cases, the static factor $F_t$ is no longer scalar but is an $r \times 1$ vector. For the empirical analysis, we use $r = 4$ as our benchmark case to allow for sufficient dynamics of the economy-wide real marginal cost process. We perform a robustness check for different numbers of common factors and find that the conclusion about the intrinsic persistence remains qualitatively identical.

7.3 Estimation results

We use the data set of Reis and Watson (2010) that contains the price indices for personal consumption expenditures by major product and expenditure categories, which is published in the NIPA accounts by the Bureau of Economic Analysis. The annualized quarterly inflation of 187 sectors is calculated using $\pi_{it} = 400 \times \ln \left( \frac{P_{it}}{P_{i,t-1}} \right)$ for the period from 1959Q1 to 2006Q2. Figure 2 presents a histogram of the estimated $\hat{\delta}_{1i}$ for all sectors. The substantial heterogeneity of the autoregressive coefficients stands out as a key feature of the estimates. Slightly fewer than 50% of the sectors exhibit low reduced-form persistence with values of $\hat{\delta}_{1i}$ below 0.1. Conversely, there are also those sectors with high persistence that reach approximately 0.8. Under this pattern, the persistence of the aggregate series does not reflect the true persistence of the underlying series, as observed by Granger (1980) and Zaffaroni (2004). Thus, we confirm the existing literature’s findings regarding the reduced-form persistence of disaggregated inflation; inflation exhibits low persistence at the sectoral level, and the persistence of the aggregate inflation appears to be
exaggerated by the aggregation procedure.

Figure 3 presents the estimated intrinsic parameter $\gamma^b_i$ for every sector for the benchmark case with $r = 4$. The intrinsic parameter estimates are calculated from the estimates of $\delta_{1i}$ using the relationship described above. The histogram of the estimated intrinsic persistence reveals that there is substantial heterogeneity. Approximately half of the sectors “inherit” their persistence primarily from the driving processes, as they exhibit low intrinsic persistence. Low intrinsic persistence in those sectors implies that their price-setting behavior is almost purely forward-looking in nature. Notably, given the partial indexation scheme of Smets and Wouters (2003), the intrinsic persistence parameter $\gamma^b_i$ is a monotonic transformation of the reduced-form autoregressive parameter $\delta_{1i}$. Therefore, the argument regarding the aggregation problem carries over to the intrinsic persistence; the intrinsic persistence of the aggregate inflation may have been exaggerated by the aggregation procedure itself, as the underlying series exhibit substantial heterogeneity.

To investigate whether the aggregation procedure is the main cause of the exaggerated persistence, we conduct two additional analyses. The first is a meta-analysis of the existing NKPC literature based on different numbers of sectors. Although the majority of the studies are based on the aggregate inflation, at least two empirical studies have analyzed the sectoral NKPCs. One of the studies was conducted by Leith and Malley (2007), who estimated the NKPC
for 19 U.S. manufacturing sectors. The other is an empirical study by Imbs et al. (2011), who used French data to estimate the NKPC for 16 sectors. Table 7 reports the intrinsic parameter estimates from the selected empirical studies. One conspicuous pattern in the table is that the degree of intrinsic persistence decreases with an increasing number of sectors. In our study, the mean group estimate of the intrinsic persistence is only 0.13, which is substantially lower than the estimates in the aggregate NKPC literature. With estimates of approximately 0.30, the analyses of Leith and Malley (2007) and Imbs et al. (2011) suggest that the level of disaggregation at \( N \simeq 20 \) may not be sufficient to reveal the true nature of the sectoral inflation.

Table 7: Meta-Analysis of Intrinsic Persistence in the Literature

<table>
<thead>
<tr>
<th>Aggregate ( N = 1 )</th>
<th>mean of ( N \simeq 20 )</th>
<th>mean of ( N \simeq 200 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gali and Gertler (1999)</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>Rabanal and Rubio-Ramirez (2005)</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Cho and Moreno (2006)</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Del Negro et al. (2007)</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>Justiniano and Primiceri (2008)</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Leith and Malley (2007)</td>
<td>0.30</td>
<td>DHP+CSD 0.13</td>
</tr>
<tr>
<td>Imbs et al. (2011)(^{11})</td>
<td>0.30</td>
<td></td>
</tr>
</tbody>
</table>

Acknowledging the limitation of the meta-analysis, we adopt an alternative approach to confirm the aggregation problem. The rational expectations solution of the aggregate NKPC

---

\(^{10}\)Note in advance that the estimates may not be directly comparable due to substantial differences in their underlying specifications and data. However, given the convention in the literature to directly compare the quasi-structural parameter \( \gamma_b \), the meta-analysis can be one possible means of shedding new light on the issue.
includes a discounted sum of future real marginal costs at the economy level.

\[
\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa \hat{\Pi} + \xi_t
\]

\[
= \delta_1 \pi_{t-1} + \left( \frac{\kappa}{\delta_2 \gamma_f} \right) \sum_{k=0}^{\infty} \left( \delta_2^{-1} \right)^k E_t \hat{\Pi} + \left( \frac{1}{\delta_2 \gamma_f} \right) \xi_t
\]

Recall that, in the previous analysis using disaggregated inflation, we controlled for the summation in the second term using the common factor estimates extracted from large cross-sections. Thus, we can use the extracted factors in the previous analysis to control for the second term in the aggregate NKPC. We used the aggregate PCE inflation data from the NIPA accounts along with the extracted factors to estimate the reduced parameter \(\delta_1\). Then, we recovered the estimate of the parameter of interest \(\gamma_b\), using the same relationship among the parameters previously described. The intrinsic persistence of the aggregate inflation is estimated to be \(\hat{\gamma}_b = 0.48\). Note that this number is comparable to the estimates from the aggregate NKPC literature listed in Table 7. Here, we can argue that the difference between the mean group estimator of the sectoral intrinsic persistence (0.13) and the estimated aggregate intrinsic persistence (0.48) comes solely from the difference in the level of aggregation, because the estimates are based on a common framework.

8 Conclusions

In this paper, we developed an asymptotic theory for dynamic heterogeneous panel models with cross-sectional dependence. Allowing dynamics, heterogeneity and cross-sectional dependence is essential for empirically realistic models, and is useful for examining whether the persistence of an aggregate series properly reflects the persistence of its underlying disaggregated series. The main contribution of this paper is that it addresses the three issues in estimation all at the same time. We showed that the DHP+CSD estimator is \(\sqrt{T}\)-consistent under non-stringent conditions on \(N\) and \(T\), and that it performs well in finite samples. We also provided two empirical applications to illustrate that the new estimator is useful for analyzing the dynamics of disaggregated data sets. In the context of the PPP puzzle, we confirmed the existing finding that the puzzle may
have been exaggerated due to aggregation bias. In the context of the New Keynesian Phillips curve, we found low intrinsic persistence at the sectoral level, implying that the true nature of price setting is more forward-looking than has been suggested by the existing analysis based on aggregate inflation.

Given the growing number of large dimensional panel data sets and the pervasiveness of dynamics, heterogeneity, and cross-sectional dependence in most data sets, the DHP+CSD estimator has numerous potential applications. The dynamics of many important disaggregated economic series, including sectoral consumption growth rate, liquidity measures of individual bank, and bank level profits, among others, can be modeled using the DHP+CSD to have common as well as idiosyncratic dynamics. Another interesting application is modeling the evolution of a distribution in terms of the dynamics of quantiles, similar to the conditional autoregressive value at risk (CAViaR) model by Engle and Manganelli (2004). If the dynamics of the quantiles follow a DHP+CSD model, instead of the simple autoregressions in CAViaR, the evolution of a distribution can be either due to common shocks that affect the entire distribution, or due to idiosyncratic shocks that affect only a part of the distribution.

References


Appendix

A  Proofs

Lemma A.1: Under Assumptions A-D,

(i) \( \frac{1}{NT^2} \sum_{i=1}^{N} \|X_i\|^4 = O_p(1) \)

(ii) \( \frac{1}{\sqrt{T}} \|X'_i M_\hat{F}\| = O_p(1) \)

(iii) \( \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \hat{\tau}_{ts}^2 = O(1) \) where \( \hat{\tau}_{ts} \) denotes the bound in Assumption D(ii-3) and D(ii-4)

Proof:

(i) By the definition of the norm and by the Cauchy-Schwarz inequality,

\[
E \left| \frac{1}{NT^2} \sum_{i=1}^{N} \|X_i\|^4 \right| = E \left[ \frac{1}{NT^2} \sum_{i=1}^{N} \left( \sum_{t=1}^{T} \|x_{it}\|^2 \right)^2 \right] \\
= \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left( \|x_{it}\|^2 \|x_{is}\|^2 \right) \\
\leq \frac{1}{NT^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( E \|x_{it}\|^4 \right) \left( E \|x_{is}\|^4 \right) \frac{1}{2}
\]

But, the entire summation is bounded because \( x_{it} \) has bounded 4th moment for all \( i \) and \( t \) by Assumption B(i).

(ii) From the definition of \( M_\hat{F} = I - \frac{1}{T} \hat{F} \hat{F}' \),

\[
\frac{1}{\sqrt{T}} \|X'_i M_\hat{F}\| = \frac{1}{\sqrt{T}} \left\| X'_i \left( I_T - \frac{1}{T} \hat{F} \hat{F}' \right) \right\| \\
\leq \frac{1}{\sqrt{T}} \left( \|X_i\| + \frac{1}{T} \|X_i \hat{F} \hat{F}'\| \right) \\
\leq \frac{1}{\sqrt{T}} \|X_i\| + \left( \frac{1}{\sqrt{T}} \|X_i\| \right) \left( \frac{1}{\sqrt{T}} \|\hat{F}\| \right) \left( \frac{1}{\sqrt{T}} \|\hat{F}'\| \right)
\]

where the second statement follows from the triangle inequality, and the last inequality comes from the properties of norms. Note that \( \frac{1}{\sqrt{T}} \|X_i\| \) is bounded by Assumption B(i) and that
Combining these results, we obtain the desired result.

(iii) Assumption D(ii-4) implies that $\tilde{\tau}_{ts}$ is bounded by a finite positive number $M$ for all $t$ and $s$. Thus, we have

$$\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{\tau}_{ts}^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} (\tilde{\tau}_{ts} \tilde{\tau}_{ts}) \leq M \left( \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \tilde{\tau}_{ts} \right)$$

Again, by Assumption D(ii-4), the second term of the product is bounded, thus having the desired result.

**Lemma A.2:** Under Assumptions A-D,

(i) $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{it} \varepsilon_{it} \right\|^2 = O_p(1)$

(ii) $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_{t}^0 \varepsilon_{it} \right\|^2 = O_p(1)$

(iii) $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} x_{it} \varepsilon_{it} \right\|^2 = O_p \left( \frac{1}{CTN} \right) \quad \forall F \in \mathcal{F}$

**Proof:**

(i) From the definition of norms,

$$E \left[ \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{it} \varepsilon_{it} \right\|^2 \right] = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} E \left( x_{it}^2 \varepsilon_{it} \varepsilon_{is} \right)$$

The result in part (i) follows readily from the fact that the summation inside the bracket is uniformly bounded by Assumption C(ii-2).
(ii) Using the properties of norms,

\[
E \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_{t}^{0} \varepsilon_{it} \right\|^{2} \right) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left( F_{t}^{0} F_{s}^{0} \varepsilon_{it} \varepsilon_{is} \right)
\]

\[
\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| E \left( \| F_{t}^{0} \| \| F_{s}^{0} \| \right) \right| \times \left| E \left( \varepsilon_{it} \varepsilon_{is} \right) \right|
\]

\[
\leq M \left\{ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| E \left( \varepsilon_{it} \varepsilon_{is} \right) \right| \right\}
\]

where we used the assumption of independence between the true factors and idiosyncratic errors to separate out the expectation. Note that \( |E(\| F_{t}^{0} \| \| F_{s}^{0} \|)| \leq \left( E(\| F_{t}^{0} \|^{2}) E(\| F_{s}^{0} \|^{2})^{\frac{1}{2}} \right) \) by the Cauchy-Schwarz inequality, where the right hand side is bounded by Assumption A(i-1). That is, the boundedness of (5) is determined by the following term:

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \left| E \left( \varepsilon_{it} \varepsilon_{is} \right) \right| \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \bar{\tau}_{ts}
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left( \frac{1}{N} \sum_{i=1}^{N} \bar{\tau}_{ts} \right)
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \bar{\tau}_{ts}
\]

where the inequality follows from the fact that \( |E(\varepsilon_{it} \varepsilon_{is})| \) is uniformly bounded over all \( i \) by Assumption D(ii-3). But, the right hand side is bounded by Assumption D(ii-4). This leads to the conclusion that (5) is bounded.

The part (iii) is already proved in Lemma A.1 of Bai (2009).

**Lemma A.3:** Under Assumptions A-D,

(i) \( \sup_{\beta_{i} \in B_{i}, F \in F} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_{i} - \beta_{0,i})' \left( \frac{X'_{i} M_{F} \varepsilon_{i}}{T} \right) \right\| = o_{p}(1) \)

(ii) \( \sup_{\beta_{i} \in B_{i}, F \in F} \left\| \frac{1}{NT} \sum_{i=1}^{N} X_{i}^{0} F^{0} \varepsilon_{i} \right\| = o_{p}(1) \)

(iii) \( \sup_{\beta_{i} \in B_{i}, F \in F} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_{i} P_{F} \varepsilon_{i} \right\| = o_{p}(1) \)

(iv) \( \sup_{\beta_{i} \in B_{i}, F \in F} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_{i} P_{F0} \varepsilon_{i} \right\| = o_{p}(1) \)
Proof:

Recall that each $B_i$ is a compact set around the true parameter $\beta_{0,i}$ defined in Assumption A(iii) and that $F$ is the space defined at the end of the introduction.

(i) By the properties of norms and supremum operators,

$$\sup_{\beta_i \in B_i, F \in F} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' M_F \varepsilon_i}{T} \right) \right\| \leq \sup_{\beta_i \in B_i, F \in F} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' \varepsilon_i}{T} \right) \right\| + \sup_{\beta_i \in B_i, F \in F} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' PF \varepsilon_i}{T} \right) \right\| \leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \beta_i - \beta_{0,i} \|^2 \right)^{\frac{1}{2}} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \| X_i t \varepsilon_{it} \|^2 \right)^{\frac{1}{2}} \leq \left( \frac{1}{NT^2} \sum_{i=1}^{N} \| X_i \|^4 \right)^{\frac{1}{4}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} F_i t \varepsilon_{it} \right\|^2 \right)^{\frac{1}{2}}$$

Regarding the first term of (6), by the Cauchy-Schwarz inequality, we have

$$\left\| 1 \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' \varepsilon_i}{T} \right) \right\| \leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \beta_i - \beta_{0,i} \|^2 \right)^{\frac{1}{2}} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \| X_i t \varepsilon_{it} \|^2 \right)^{\frac{1}{2}}$$

But, $\frac{1}{N} \sum_{i=1}^{N} \| \beta_i - \beta_{0,i} \|^2$ is uniformly bounded when we consider $\beta_i$ on a bounded set $B_i$ around the true $\beta_{0,i}$, and $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_i t \varepsilon_{it} \right\|^2$ is bounded by Lemma A.2(i). Thus, it follows that $\sup_{\beta_i \in B_i} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' \varepsilon_i}{T} \right) \right\| = o_p(1)$. Moreover, this result holds uniformly over $F \in F$ as well because the expression does not involve $F$ at all.

Next, by applying the Cauchy-Schwarz inequality twice to the second term of (6),

$$\left\| 1 \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' PF \varepsilon_i}{T} \right) \right\| \leq \left( \frac{1}{N} \sum_{i=1}^{N} \| X_i \|^4 \right)^{\frac{1}{4}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} F_i t \varepsilon_{it} \right\|^2 \right)^{\frac{1}{2}}$$

Note that the first three components of the product are all bounded. That is, $\frac{1}{N} \sum_{i=1}^{N} \| \beta_i - \beta_{0,i} \|^4$ is bounded when we consider $\beta_i$ on a bounded set $B_i$ around the true $\beta_{0,i}$, $\frac{1}{NT^2} \sum_{i=1}^{N} \| X_i \|^4$ is bounded by Lemma A.1(i), and $\frac{1}{\sqrt{T}} \| F \| = \sqrt{F}$ for all $F \in F$ by the definition of the set $F$. The last component of (8), $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} F_i t \varepsilon_{it} \right\|^2$, is of order $O_p \left( \frac{1}{\sqrt{NT}} \right)$ by Lemma A.2(iii). From these results, $\sup_{\beta_i \in B_i, F \in F} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left( \frac{X_i' PF \varepsilon_i}{T} \right) \right\| = o_p(1)$ as well.

Combining above results that the two terms in (6) are all uniformly $o_p(1)$, we complete
the proof of the part (i).

The proof of the part (ii) and (iii) can be found in Lemma A.1 of Bai (2009).

For (iv),

\[
\left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' P_{F_0} \varepsilon_i \right\| \leq \frac{1}{T} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_0^t \varepsilon_{it} \right\|^2 \right) \left\| \left( \frac{F_0' F_0}{T} \right)^{-1} \right\|
\]

by the Cauchy-Schwarz inequality. But \( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_0^t \varepsilon_{it} \right\|^2 \) is bounded by Lemma A.2(ii), and \( \left\| \left( \frac{F_0' F_0}{T} \right)^{-1} \right\| \) is also bounded by Assumption A(i-2). Moreover, this result holds uniformly as the expression involves neither \( \beta_i \) nor \( F \).

**Proof of Theorem 1:** Here we extend the consistency proof of Bai (2009) to incorporate heterogeneous coefficients. Note that the usual consistency result for extremum estimators by Newey and McFadden (1994) is not directly applicable to this setup due to the growing dimension of the parameters. Instead, the argument here relies on an auxiliary objective function that is uniformly close to the original objective function as \((N,T)\) goes to infinity. Moreover, we will show that the auxiliary objective function is uniquely minimized at the true parameter values. This approach has been first initiated by Bai (1994) and later adopted by Bai (2009) and Bonhomme and Manresa (2012). We also closely follow the main argument of Bai (2009).

Consider a rescaled and re-centered objective function

\[
S_{NT} \left( \{\beta_i\}_{i=1}^{N}, F \right) \equiv \frac{1}{NT} \sum_{i=1}^{N} (Y_i - X_i \beta_i)' M_F (Y_i - X_i \beta_i) - \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' M_{F_0} \varepsilon_i \quad (9)
\]

Note that the loadings are already concentrated out from the objective function, exploiting the relationship between the estimators for loadings and factors which should be satisfied at the
solution. By substituting $Y_i = X_i \beta_{0,i} + F^0 \lambda_i^0 + \varepsilon_i$ into the objective function, we have

$$S_{NT} \left( \{ \beta_i \}_{i=1}^N, F \right) = \frac{1}{N} \sum_{i=1}^N (\beta_i - \beta_{0,i})' \left( \frac{X_i' M F X_i}{T} \right) (\beta_i - \beta_{0,i}) + \frac{1}{NT} \sum_{i=1}^N \lambda_i^0 F^0 M F^0 \lambda_i^0$$

$$+ 2 \frac{1}{NT} \sum_{i=1}^N (\beta_i - \beta_{0,i})' X_i' M F^0 \lambda_i^0 + 2 \frac{1}{N} \sum_{i=1}^N (\beta_i - \beta_{0,i})' \left( \frac{X_i' M F \varepsilon_i}{T} \right)$$

$$+ \frac{2}{NT} \sum_{i=1}^N \lambda_i^0 F^0 M F \varepsilon_i + \frac{1}{NT} \sum_{i=1}^N \varepsilon_i' P F \varepsilon_i - \frac{1}{NT} \sum_{i=1}^N \varepsilon_i' P F^0 \varepsilon_i,$$

(10)

Note that Lemma A.3 implies that the last four terms in (10) are $o_p(1)$ uniformly over the entire space of $\beta_i$ and $F$. By defining the sum of the first three terms as an auxiliary objective function

$$\tilde{S}_{NT} \left( \{ \beta_i \}_{i=1}^N, F \right) = \frac{1}{N} \sum_{i=1}^N (\beta_i - \beta_{0,i})' \left( \frac{X_i' M F X_i}{T} \right) (\beta_i - \beta_{0,i})$$

$$+ \frac{1}{NT} \sum_{i=1}^N \lambda_i^0 F^0 M F^0 \lambda_i^0 + 2 \frac{1}{NT} \sum_{i=1}^N (\beta_i - \beta_{0,i})' X_i' M F^0 \lambda_i^0$$

$$= \frac{1}{N} \sum_{i=1}^N \left\{ (\beta_i - \beta_{0,i})' \left( \frac{X_i' M F X_i}{T} \right) (\beta_i - \beta_{0,i}) \right.$$  

$$+ \lambda_i^0 \left( \frac{F^0 M F^0}{T} \right) \lambda_i^0 + 2 (\beta_i - \beta_{0,i})' \left( \frac{X_i' M F^0}{T} \right) \lambda_i^0 \left\} \right.$$  

(11)

we have $S_{NT} \left( \{ \beta_i \}_{i=1}^N, F \right) = \tilde{S}_{NT} \left( \{ \beta_i \}_{i=1}^N, F \right) + o_p(1)$, meaning that the auxiliary objective function is uniformly close to the original objective function.

Now we show that the auxiliary objective function, $\tilde{S}_{NT} \left( \{ \beta_i \}_{i=1}^N, F \right)$, is uniquely minimized at the true parameter values. To be precise, $\tilde{S}_{NT} \left( \{ \beta_{0,i} \}_{i=1}^N, F^0 H \right)$ attains the unique minimum value of zero, where $H$ is an invertible $r \times r$ matrix. That is, we cannot show the consistency for the true factors themselves. Nevertheless, we can show the consistency of the space spanned by the true factors. The minimizer of $\tilde{S}_{NT} \left( \{ \beta_i \}_{i=1}^N, F \right)$ can be characterized if we transform the auxiliary objective function into a quadratic form as follows.
Rewrite the second term inside the curly braces of (11) as

\[ \lambda_{i}^{0} \left( \frac{F_{i}^{0} M_{i} F_{i}^{0}}{T} \right) \lambda_{i}^{0} = \frac{1}{T} \times tr \left[ F_{i}^{0} M_{i} F_{i}^{0} \lambda_{i}^{0} \lambda_{i}^{0} \right] \]

\[ = tr \left[ (F_{i}^{0} M_{i}) \left( \frac{I_{T}}{T} \right) (M_{i} F_{i}^{0}) \lambda_{i}^{0} \lambda_{i}^{0} \right] \]

\[ = vec (M_{i} F_{i}^{0})' \left[ (\lambda_{i}^{0} \lambda_{i}^{0}) \otimes \frac{I_{T}}{T} \right] vec (M_{i} F_{i}^{0}) \]

\[ = \eta' B_i \eta \]

where the first equality comes from the fact that \( \lambda_{i}^{0} \left( \frac{F_{i}^{0} M_{i} F_{i}^{0}}{T} \right) \lambda_{i}^{0} \) is a scalar and the properties of trace, and the third equality follows from the known relation \( tr [ABCD] = vec(A')' [D' \otimes B] vec(C) \). The last equality is just a relabeling for notational simplicity by defining \( \eta = vec (M_{i} F_{i}^{0}) \) and \( B_i = (\lambda_{i}^{0} \lambda_{i}^{0}) \otimes \frac{I_{T}}{T} \).

By the same argument, the third term inside the curly braces of (11) becomes

\[ (\beta_i - \beta_{0,i})' \left( \frac{X_{i}^{0} M_{i} F_{i}^{0}}{T} \right) \lambda_{i}^{0} = \frac{1}{T} \times tr \left[ (\beta_i - \beta_{0,i})' X_{i}^{0} M_{i} F_{i}^{0} \lambda_{i}^{0} \right] \]

\[ = \frac{1}{T} \times tr \left[ (\beta_i - \beta_{0,i})' (X_{i}^{0} M_{i}) (M_{i} F_{i}^{0}) \lambda_{i}^{0} \right] \]

\[ = (\beta_i - \beta_{0,i})' \left[ \lambda_{i}^{0} \lambda_{i}^{0} \right. \otimes \frac{1}{T} (X_{i}^{0} M_{i}) \left. vec (M_{i} F_{i}^{0}) \right] \]

\[ = (\beta_i - \beta_{0,i})' C_i \eta \]

where we define \( C_i = \lambda_{i}^{0} \otimes \frac{1}{T} (X_{i}^{0} M_{i}) \).

Using the two expressions derived above,

\[ \tilde{S}_{NT} \left( \{\beta_i\}_{i=1}^{N}, F \right) \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \left\{ (\beta_i - \beta_{0,i})' \left( \frac{X_{i}^{0} M_{i} X_{i}}{T} \right) (\beta_i - \beta_{0,i}) + \eta' B_i \eta + 2 (\beta_i - \beta_{0,i})' C_i \eta \right\} \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \left\{ (\beta_i - \beta_{0,i})' A_i (\beta_i - \beta_{0,i}) + \eta' B_i \eta + 2 (\beta_i - \beta_{0,i})' C_i \eta \right\} \]

(12)
where we define $A_i = \frac{X'_i M_i X_i}{T}$ for the last equality. Now, by completing the squares, we obtain

$$
\hat{S}_{NT}\left(\{\beta_i\}_{i=1}^N, F\right) = \eta' \left( \frac{1}{N} \sum_{i=1}^N D_i \right) \eta + \frac{1}{N} \sum_{i=1}^N \left[ (\beta_i - \beta_{0,i}) + A_i^{-1} C_i \eta \right]' \left( (\beta_i - \beta_{0,i}) + A_i^{-1} C_i \eta \right]
$$

$$
= \eta' D \eta + \frac{1}{N} \sum_{i=1}^N \theta_i'A_i \theta_i
$$

(13)

where we define $D = \frac{1}{N} \sum_{i=1}^N D_i$, $D_i = B_i - C_i'A_i^{-1} C_i$, and $\theta_i = (\beta_i - \beta_{0,i}) + A_i^{-1} C_i \eta$.

Notice that the auxiliary objective function is in a quadratic form with $D$ and $A_i$ all being positive definite matrices by Assumption B(iii) and B(iv). Thus $\hat{S}_{NT}\left(\{\beta_i\}_{i=1}^N, F\right) \geq 0$, and it attains the unique minimum of zero at $\left(\{\beta_{0,i}\}_{i=1}^N, F^0H\right)$. For any combination of $\left(\{\beta_i\}_{i=1}^N, F\right)$ that is different from $\left(\{\beta_{0,i}\}_{i=1}^N, F^0H\right)$, we have $\hat{S}_{NT}\left(\{\beta_i\}_{i=1}^N, F\right) > 0$. Moreover, if $\|\beta_i - \beta_{0,i}\| \geq c > 0 \forall i$, we have $\hat{S}_{NT}\left(\{\beta_i\}_{i=1}^N, F\right) \geq \mu c^2 > 0$ by Assumption B(iii). Thus, it follows that each $\hat{\beta}_i$ is consistent for its corresponding true parameter $\beta_{0,i}$. This proves the part (i) of Theorem 1. Given the consistency of $\hat{\beta}_i$ and the results in Lemma A.3, the same argument of Bai (2009) leads to $\hat{S}_{NT}\left(\{\hat{\beta}_i\}, \hat{F}\right) = o_p(1)$ as well as $\frac{1}{T} F^0M_{\hat{F}}F^0 = o_p(1)$. Then, it is easy to see that

$$
\|P_{\hat{F}} - P_{F^0}\|^2 = 2tr \left[ \left( \frac{F^0M_{\hat{F}}F^0}{T} \right) \left( \frac{F^0F^0}{T} \right)^{-1} \right] = o_p(1),
$$

thus proving the consistency of the space spanned by the factors in part (ii).

**Proposition A.1:** Under Assumptions A-D,

$$
\frac{1}{\sqrt{T}} \left\| \hat{F} V_{NT} \left( \frac{F^0\hat{F}}{T} \right)^{-1} \left( \frac{A^0A^0}{N} \right)^{-1} - F^0 \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right)
$$

where $V_{NT}$ denotes a diagonal matrix of eigenvalues of $\frac{1}{N} \sum_{i=1}^N \left( Y_i - X_i \hat{\beta}_i \right) \left( Y_i - X_i \hat{\beta}_i \right)'$ and $B_{NT}$ denotes the stochastic order of a generic $\|\hat{\beta}_i - \beta_{0,i}\|$. 

44
Proof: We extend the proof of Proposition A.1 in Bai (2009) while taking extra care to deal with individual estimator \( \hat{\beta}_i \)'s. From the definition of \( V_{NT} \),

\[
\left[ \frac{1}{N_T} \sum_{i=1}^{N} \left( Y_i - X_i \hat{\beta}_i \right) \left( Y_i - X_i \hat{\beta}_i \right) \right] \hat{F} = \hat{F} V_{NT} \tag{14}
\]

By substituting \( Y_i = X_i \beta_{0,i} + F^0 \lambda_i^0 + \varepsilon_i \) into the expression above and by expanding terms, we can decompose \( \hat{F} V_{NT} \) into nine terms:

\[
\hat{F} V_{NT} = \frac{1}{N_T} \sum_{i=1}^{N} X_i \left( \beta_{0,i} - \hat{\beta}_i \right) \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i \hat{F} + \frac{1}{N_T} \sum_{i=1}^{N} X_i \left( \beta_{0,i} - \hat{\beta}_i \right) X_i \hat{F} \sum_{i=1}^{N} \frac{F^0 \lambda_i^0 \left( \beta_{0,i} - \hat{\beta}_i \right) X_i \hat{F} +}{\Lambda^0_0} N \frac{1}{\sqrt{T}} \left\| I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8 + F^0 \left( \frac{A^0_0}{N} \right) \right\| F^0_0 \hat{F} \sum_{i=1}^{N} \frac{F^0 \lambda_i^0 \lambda_i^0 F^0_0 \hat{F}}{\Lambda^0_0} N \frac{1}{\sqrt{T}} \left\| \right\| F^0_0 \hat{F} \sum_{i=1}^{N} \frac{F^0 \lambda_i^0 \lambda_i^0 F^0_0 \hat{F}}{\Lambda^0_0} N \frac{1}{\sqrt{T}} \left\| \right\| = I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8 + F^0 \left( \frac{A^0_0}{N} \right) \left( \frac{F^0_0 \hat{F}}{T} \right) \tag{15}
\]

where we labeled the first eight terms of (15) as \( I_1 \) to \( I_8 \), respectively. By rearranging terms and rescaling both sides by \( \frac{1}{\sqrt{T}} \), we have

\[
\frac{1}{\sqrt{T}} \left\| \hat{F} V_{NT} \left( \frac{F^0_0 \hat{F}}{T} \right)^{-1} - F^0 \right\| = \frac{1}{\sqrt{T}} \left\| I_1 + I_2 + \cdots + I_8 \right\| \left( \frac{F^0_0 \hat{F}}{T} \right)^{-1} \left( \frac{A^0_0}{N} \right)^{-1} \right\| \leq \left( \frac{1}{\sqrt{T}} \left\| I_1 \right\| + \frac{1}{\sqrt{T}} \left\| I_2 \right\| + \cdots + \frac{1}{\sqrt{T}} \left\| I_8 \right\| \right) \times \left( \frac{F^0_0 \hat{F}}{T} \right)^{-1} \left\| \left( \frac{A^0_0}{N} \right)^{-1} \right\| \tag{16}
\]

where the last inequality follows from the properties of norms. The existence and boundedness of \( \left( \frac{A^0_0}{N} \right)^{-1} \) are guaranteed by Assumption A(ii-2). The invertibility of \( \frac{F^0_0 \hat{F}}{T} \) is argued in the
proof of Proposition A.1 of Bai (2009). Using the definition of norms and the properties of trace, we obtain

\[ \left\| \left( \frac{F_0 \hat{F}}{T} \right)^{-1} \right\|^2 = tr \left[ \left( \frac{F_0 F_0^T - F_0 \hat{M} \hat{F} F_0^T}{T} \right)^{-1} \right] \] (17)

Note that \( \frac{F_0 \hat{M} F_0}{T} = o_p(1) \) as already discussed above and that the existence and boundedness of \( \left( \frac{F_0 \hat{F}}{T} \right)^{-1} \) are guaranteed by Assumption A(i-2). Thus, we conclude that \( \left\| \left( \frac{F_0 \hat{F}}{T} \right)^{-1} \right\| = O_p(1) \).

As both \( \left\| \left( \frac{F_0 \hat{F}}{T} \right)^{-1} \right\| \) and \( \left\| \left( \frac{\Lambda_0^0 \Lambda_0^0}{N} \right)^{-1} \right\| \) are bounded, the order of (16) is determined by the sum of eight terms.

Now we examine the stochastic order of the eight terms from \( \frac{1}{\sqrt{T}} \| I1 \| \) to \( \frac{1}{\sqrt{T}} \| IS \| \). For the first term, using Cauchy-Schwarz inequality,

\[
\frac{1}{\sqrt{T}} \| I1 \| = \frac{1}{\sqrt{T}} \left\| \frac{1}{NT} \sum_{i=1}^{N} X_i \left( \beta_{0,i} - \hat{\beta}_i \right) \left( \beta_{0,i} - \hat{\beta}_i \right)^T \hat{F} \right\| \\
\leq \left( \frac{1}{NT^2} \sum_{i=1}^{N} \| X_i \|^4 \right)^{\frac{1}{2}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\beta}_i - \beta_{0,i} \|^4 \right)^{\frac{1}{2}} \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right)
\]

Note that the model considered in this paper is symmetric with respect to every individual unit. Furthermore, the proposed estimator processes all individual data in a symmetric manner as well. That is, each \( \| \hat{\beta}_i - \beta_{0,i} \| \) has the same stochastic order in terms of \( N \) and \( T \) although the exact rate may differ up to a constant. If we denote the stochastic order of a generic term \( \| \hat{\beta}_i - \beta_{0,i} \| \) by \( O_p(B_{NT}) \), the order of \( \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\beta}_i - \beta_{0,i} \|^4 \right)^{\frac{1}{2}} \) becomes \( O_p \left( B_{NT}^2 \right) \) because an average has the same order of magnitude with that of summands. As the other two terms on the right hand side are bounded by Lemma A.1(i) and by the normalizing assumption \( \frac{\hat{F} \hat{F}}{T} = I_r \), we conclude that \( \frac{1}{\sqrt{T}} \| I1 \| = O_p \left( B_{NT}^2 \right) \).
Next, for \( \frac{1}{\sqrt{T}} \| I2 \| \)

\[
\frac{1}{\sqrt{T}} \| I2 \| = \frac{1}{\sqrt{T}} \left\| \frac{1}{NT} \sum_{i=1}^{N} X_{i} \left( \beta_{0,i} - \hat{\beta}_{i} \right) \lambda_{i}^{0r} F^{0r} \hat{F} \right\|
\leq \left( \frac{1}{NT} \sum_{i=1}^{N} \| X_{i} \|^2 \right)^{\frac{1}{2}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\beta}_{i} - \beta_{0,i} \|^4 \right)^{\frac{1}{4}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \lambda_{i}^{0} \|^4 \right)^{\frac{1}{4}} \times \left( \frac{1}{\sqrt{T}} \| F^0 \| \right) \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right)
\]

where the last inequality follows from the Cauchy-Schwarz inequality. As in the case of \( \frac{1}{\sqrt{T}} \| I1 \| \), it follows that \( \left( \frac{1}{N} \sum_{i=1}^{N} \| \hat{\beta}_{i} - \beta_{0,i} \|^4 \right)^{\frac{1}{4}} = O_{p} \left( B_{NT} \right) \) while all other terms are simply \( O_{p}(1) \). Therefore, we obtain \( \frac{1}{\sqrt{T}} \| I2 \| = O_{p} \left( B_{NT} \right) \). By the same argument, we can show that \( \frac{1}{\sqrt{T}} \| I3 \| \), \( \frac{1}{\sqrt{T}} \| I4 \| \) and \( \frac{1}{\sqrt{T}} \| I5 \| \) are also \( O_{p} \left( B_{NT} \right) \). The results for \( \frac{1}{\sqrt{T}} \| I6 \| \), \( \frac{1}{\sqrt{T}} \| I7 \| \) and \( \frac{1}{\sqrt{T}} \| I8 \| \) are already provided in Theorem 1 of Bai and Ng (2002). All three terms are shown to be of order \( O_{p} \left( \frac{1}{CN_{T}} \right) \). These terms do not involve any individual estimator \( \| \hat{\beta}_{i} - \beta_{0,i} \| \) that the results still remain true in this setup. Summing over eight terms in (16), we can show that

\[
\frac{1}{\sqrt{T}} \left\| \hat{F} V_{NT} \left( F^{0r} \hat{F} \right)^{-1} \left( \frac{\Lambda^{0r} \Lambda^{0}}{N} \right)^{-1} - F^0 \right\| = O_{p} \left( B_{NT} \right) + O_{p} \left( \frac{1}{\sqrt{NT}} \right)
\]

**Lemma A.4:** Under Assumptions A-D,

(i) \( \frac{1}{T} F^{0r} \left( I1 + \cdots + I8 \right) = o_{p}(1) \)

(ii) \( \| V_{NT} \| = O_{p}(1) \)

(iii) \( V_{NT} \) is invertible and \( \| V_{NT}^{-1} \| = O_{p}(1) \)

(iv) \( \| H \| = O_{p}(1) \) where \( H = \left( \frac{\Lambda^{0r} \Lambda^{0}}{N} \right) \left( F^{0r} \hat{F} \right) V_{NT}^{-1} \)

(v) \( H \) is invertible and \( \| H^{-1} \| = O_{p}(1) \)
Proof:

(i) By the properties of norms and the triangle inequality, we have

\[ \left\| \frac{1}{T} F^0 (I_1 + \cdots + I_8) \right\| \leq \left( \frac{1}{\sqrt{T}} \left\| F^0 \right\| \right) \left( \frac{1}{\sqrt{T}} \left\| (I_1 + \cdots + I_8) \right\| \right) \]

\[ \leq \left( \frac{1}{\sqrt{T}} \left\| F^0 \right\| \right) \left( \frac{1}{\sqrt{T}} \left\| I_1 \right\| + \cdots + \frac{1}{\sqrt{T}} \left\| I_8 \right\| \right) \]

Note that the first term is bounded by Assumption A(i-2). The second term which consists of a sum of eight terms are already shown to be of order \( O_p \left( \frac{1}{\sqrt{NT}} \right) \) in Proposition A.1. Therefore, we conclude that \( \left\| \frac{1}{T} F^0 (I_1 + \cdots + I_8) \right\| = o_p(1) \).

Once the part (i) is proven, the proof of the remaining parts (ii)-(v) are identical to those in Bai (2009), so are omitted.

Corollary A.1: Under Assumptions A-D,

(i) \( \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{C_{NT}}} \right) \)

(ii) \( \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{C_{NT}}} \right) \)

Proof:

From the invertibility of \( H \) in Lemma A.4(v), \( H \) in part (i) can be factored out

\[ \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| = \frac{1}{\sqrt{T}} \left\| \left( \hat{F} H^{-1} - F^0 \right) H \right\| \]

\[ \leq \frac{1}{\sqrt{T}} \left\| \hat{F} H^{-1} - F^0 \right\| \left\| H \right\| \]

where the last inequality follows from the properties of norms. Note that, from the definition of \( H \), the first term, \( \frac{1}{\sqrt{T}} \left\| \hat{F} H^{-1} - F^0 \right\| \), is already investigated in Proposition A.1 and shown to be of order \( O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{\sqrt{C_{NT}}} \right) \). The last term, \( \left\| H \right\| \), is bounded by Lemma A.5(iv). Combining the results, we prove the part (i). The part (ii) readily follows from part (i).
Lemma A.5: Under Assumptions A-D,

(i) \( \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{kt} F_{t}^{0} \right\| = O_p(1) \)

(ii) \( \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \varepsilon_{k} \lambda_{k}^{0} \right\| = O_p(1) \)

(iii) \( \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} F^{0} \varepsilon_{i} \lambda_{i}^{0} \right\| = O_p(1) \)

Proof:

We omit the proof for part (i) and (ii) because those can be proved in an exactly same fashion with part (iii) of which proof is provided below. From the definition of norms,

\[
E \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} F^{0} \varepsilon_{i} \lambda_{i}^{0} \right\|^2 = \frac{1}{NT} \times E \left( tr \left[ \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \lambda_{i}^{0} \varepsilon_{it} F_{t}^{0} F_{s}^{0} \varepsilon_{js} \lambda_{j}^{0} \right] \right) \quad (18)
\]

By the properties of norms, the above expression is bounded by

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\lambda_{j}^{0} \lambda_{i}^{0} F_{t}^{0} F_{s}^{0})| \times |E(\varepsilon_{it} \varepsilon_{js})| \quad (19)
\]

where we used the assumption that idiosyncratic errors are independent of loadings and factors to split the expectation. We observe that

\[
|E(\lambda_{j}^{0} \lambda_{i}^{0} F_{t}^{0} F_{s}^{0})| \leq \left( E \left\| \lambda_{j}^{0} \right\|^4 E \left\| \lambda_{i}^{0} \right\|^4 E \left\| F_{t}^{0} \right\|^4 E \left\| F_{s}^{0} \right\|^4 \right)^{\frac{1}{2}} \quad (20)
\]

by repetitive application of the Cauchy-Schwarz inequality. All four terms on the right hand side of (20) are bounded by Assumption A(i-1) and A(ii-1). Thus, the order of (19) is determined by

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\varepsilon_{it} \varepsilon_{js})|,
\]

which is bounded by Assumption D(ii-5). This completes the proof of part (iii).
Lemma A.6: Under Assumptions A-D,

(i) \( \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} E \left( \varepsilon_{kt} \varepsilon_{it} \right) \right|^2 = O(1) \)

(ii) \( \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \varepsilon_{kt} \varepsilon_{it} - E \left( \varepsilon_{kt} \varepsilon_{it} \right) \right] \right|^2 = O_p(1) \)

(iii) \( \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it} \varepsilon_{kt} \lambda_{it} \right\| = O_p \left( \frac{1}{C_{NT} \sqrt{N}} \right) \)

(iv) \( \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_{t} \varepsilon_{i} F_{0} \right\| = O_p \left( \frac{1}{C_{NT} \sqrt{T}} \right) \)

(v) \( \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_{i} \varepsilon_{i} \right\| = O_p \left( \frac{1}{C_{NT}} \right) \)

Proof:

(i) From Assumption D(ii-1),

\[
\sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} E \left( \varepsilon_{kt} \varepsilon_{it} \right) \right|^2 \leq \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} \bar{\sigma}_{ki} \right|^2 \quad (21)
\]

\[
= \sum_{i=1}^{N} \bar{\sigma}_{ki}^2 \quad (22)
\]

where the last equality holds because \( \bar{\sigma}_{ki} \) in (21) does not depend on the subscript \( t \). Note that (22) is bounded by Assumption D(ii-2), thus having the desired result.

(ii) For notational simplicity, let \( \zeta_{it} = \varepsilon_{kt} \varepsilon_{it} - E \left( \varepsilon_{kt} \varepsilon_{it} \right) \). Then, the object of interest can be rewritten as

\[
E \left( \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \varepsilon_{kt} \varepsilon_{it} - E \left( \varepsilon_{kt} \varepsilon_{it} \right) \right] \right|^2 \right) = E \left( \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \zeta_{it} \right|^2 \right) \quad (23)
\]

\[
= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left( \zeta_{it} \zeta_{is} \right)
\]

using the definition of norms. Note that the above expression is bounded by

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E \left( \zeta_{it} \zeta_{js} \right)|
\]
which takes additional summation over subscript $j$. But, from the definition of $\zeta_{it}$,

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\zeta_{it}\zeta_{js})| = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\text{COV}(\epsilon_{kt}\epsilon_{it}, \epsilon_{ks}\epsilon_{js})|$$

which is bounded by Assumption D(iv-2). Thus, (23) is bounded.

(iii) By adding and subtracting $E(\epsilon_{it}\epsilon_{kt})$,

$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{it}\epsilon_{kt} \lambda_{i}^{0} \right\| = \frac{1}{NT} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it}\epsilon_{kt} - E(\epsilon_{it}\epsilon_{kt}) + E(\epsilon_{it}\epsilon_{kt})) \lambda_{i}^{0} \right\| \leq \frac{1}{\sqrt{NT}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it}\epsilon_{kt} - E(\epsilon_{it}\epsilon_{kt})) \lambda_{i}^{0} \right\| + \frac{1}{N} \left\| \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\epsilon_{it}\epsilon_{kt}) \lambda_{i}^{0} \right\|$$

(24)

where the last inequality follows from the triangle inequality. We examine the first term of (24):

$$E \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} (\epsilon_{it}\epsilon_{kt} - E(\epsilon_{it}\epsilon_{kt})) \lambda_{i}^{0} \right\|^{2} = E \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \zeta_{it} \lambda_{i}^{0} \right\|^{2}$$

(25)

$$= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E(\lambda_{j}^{0} \lambda_{i}^{0} \zeta_{it}\zeta_{js}) \leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\lambda_{j}^{0} \lambda_{i}^{0})| \times |E(\zeta_{it}\zeta_{js})|$$

where the first statement follows from the use of $\zeta_{it}$ for notational simplicity, and the second statement comes from the definition of norms. The last inequality is obtained by utilizing the assumption that idiosyncratic errors are independent of loadings as stated in Assumption C(i). Note that $|E(\lambda_{j}^{0} \lambda_{i}^{0})| \leq \left( E \|\lambda_{j}^{0}\|^{2} E \|\lambda_{i}^{0}\|^{2} \right)^{1/2}$ by the Cauchy-Schwarz inequality, where the right hand side of the inequality is bounded by Assumption A(ii-1). That is, our object of interest in (25) is bounded by $\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\zeta_{it}\zeta_{js})|$, which is, in turn, bounded by Assumption D(iv-2). In sum, (25) is bounded.
Now, we examine the second term of (24)

\[
\left\| \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} E (\varepsilon_i \varepsilon_{kt}) \lambda_i^0 \right\| \leq \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \left| E (\varepsilon_i \varepsilon_{kt}) \right| \right) \left\| \lambda_i^0 \right\|
\]

\[
\leq \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \bar{\sigma}_{ik} \right) \left\| \lambda_i^0 \right\|
\]

\[
= \sum_{i=1}^{N} \bar{\sigma}_{ik} \left\| \lambda_i^0 \right\|
\]

where the first inequality follows from the triangle inequality and the properties of norms, and the second inequality holds true by Assumption D(ii-1). The last statement simply comes from the fact that \( \bar{\sigma}_{ik} \) does not depend on subscript \( t \). Now we can easily show that (26) is bounded using the assumptions that \( \left\| \lambda_i^0 \right\| \) is bounded by Assumption A(ii-1) and that \( \sum_{i=1}^{N} \bar{\sigma}_{ik} \) is bounded by Assumption D(ii-2).

In summary, the original object of interest

\[
\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_i \varepsilon_{kt} \lambda_i^0 \right\| = O_p \left( \frac{1}{\sqrt{NT}} \right) + O_p \left( \frac{1}{N} \right)
\]

where we used triangle inequality to obtain the last inequality. We examine the two terms in the above expression.

(iv) By adding and subtracting \( E (\varepsilon_i \varepsilon_i') \), we have

\[
\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon_i' \varepsilon_i' F^0 \right\| = \frac{1}{NT^2} \left\| \sum_{i=1}^{N} \varepsilon_i' \left[ \varepsilon_i \varepsilon_i' - E (\varepsilon_i \varepsilon_i') + E (\varepsilon_i \varepsilon_i') \right] F^0 \right\|
\]

\[
\leq \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{NT^3}} \sum_{i=1}^{N} \varepsilon_i' \left[ \varepsilon_i \varepsilon_i' - E (\varepsilon_i \varepsilon_i') \right] F^0 \right\|
\]

\[
+ \frac{1}{T} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' E (\varepsilon_i \varepsilon_i') F^0 \right\|
\]

where we used triangle inequality to obtain the last inequality. We examine the two terms in the above expression.

52
First, using the Cauchy-Schwarz inequality,

\[
\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varepsilon'_{k} \left[ \varepsilon_{i} \varepsilon'_{i} - E(\varepsilon_{i} \varepsilon'_{i}) \right] F^{0} \right\| = \frac{1}{\sqrt{NT}} \left\| \sum_{t=1}^{T} \sum_{s=1}^{S} \sum_{i=1}^{N} \varepsilon_{kt} \left[ \varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}) \right] F^{0}_{s} \right\|
\]

\[
\leq \left( \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{s=1}^{S} \varepsilon_{it} \varepsilon_{is} - E(\varepsilon_{it} \varepsilon_{is}) \right\|^{2} \right)^{\frac{1}{2}} \times \left( \frac{1}{T} \sum_{t=1}^{T} \left\| \varepsilon_{kt} \right\|^{2} \right)^{\frac{1}{2}}
\]

(28)

Note that the last term of (28) is bounded by Assumption D(i). To examine the order of the first term of (28), consider

\[
E \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{s=1}^{S} \varepsilon_{iu} \varepsilon_{is} - E(\varepsilon_{iu} \varepsilon_{is}) \right\|^{2} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E( F^{0}_{t} \varepsilon_{is} \varepsilon_{jt})
\]

\[
\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} \left| E( F^{0}_{t} F^{0}_{s}) \right| \times |E(\varepsilon_{is} \varepsilon_{jt})|
\]

(29)

where we used new symbol \(\zeta_{is} = \varepsilon_{iu} \varepsilon_{is} - E(\varepsilon_{iu} \varepsilon_{is})\) for notational simplicity, and the last statement follows from the independence between idiosyncratic errors and factors. Note that

\[
\left| E( F^{0}_{t} F^{0}_{s}) \right| \leq \left( E \left\| F^{0}_{t} \right\|^{2} E \left\| F^{0}_{s} \right\|^{2} \right)^{\frac{1}{2}} \times \left( E \left\| \varepsilon_{kt} \right\|^{2} \right)^{\frac{1}{2}}
\]

by the Cauchy-Schwarz inequality, where the right hand side of inequality is bounded by Assumption A(i-1). Therefore, overall bound of (29) is determined by

\[
\frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} \left| E(\varepsilon_{is} \varepsilon_{jt})\right|,
\]

which is assumed to be bounded by Assumption D(iv-1).

In sum, we conclude that the first term of (27) is \(\frac{1}{\sqrt{NT}} \times O_{p}(1)\).

Next, related to the second term of (27),

\[
E \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon'_{k} E(\varepsilon_{i} \varepsilon'_{i}) F^{0} \right\| = E \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} \varepsilon_{kt} E(\varepsilon_{it} \varepsilon_{is}) F^{0}_{s} \right\|
\]

\[
\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{S} E \left\| \varepsilon_{kt} F^{0}_{s} \right\| \times |E(\varepsilon_{it} \varepsilon_{is})|
\]

(30)

where the last statement follows from the properties of norms and the fact that \(E(\varepsilon_{it} \varepsilon_{is})\) is a non-random object. Note that \(E \left\| \varepsilon_{kt} F^{0}_{s} \right\| \leq \left( E \left\| \varepsilon_{kt} \right\|^{2} E \left\| F^{0}_{s} \right\|^{2} \right)^{\frac{1}{2}}\) by the Cauchy-Schwarz
inequality, where the right hand side is bounded by Assumption A(i-1) and D(i). Now the order of (30) is determined by 
\[ \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\varepsilon_{it}\varepsilon_{is})|. \] By Assumption D(ii-3), this term is bounded by 
\[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \bar{\tau}_{ts} \right), \] which is in turn bounded using Assumption D(ii-4).

Using the results above, we obtain the order of (27):
\[ \| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varepsilon_{i}^{t} \| = O_p \left( \frac{1}{\sqrt{NT}} \right). \]

(v) By adding and subtracting \( E(\varepsilon_{i}\varepsilon_{i}') \), we have
\[ \| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \varepsilon_{i}^{t} \| = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \left[ E(\varepsilon_{i}\varepsilon_{i}') - E(\varepsilon_{i}\varepsilon_{i}') \right] \]
\[ \leq \frac{1}{\sqrt{N}} \left( \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \left[ E(\varepsilon_{i}\varepsilon_{i}') - E(\varepsilon_{i}\varepsilon_{i}') \right] \right) + \frac{1}{\sqrt{T}} \left( \frac{1}{\sqrt{N^2T}} \sum_{i=1}^{N} E(\varepsilon_{i}\varepsilon_{i}') \right) \]
where the last inequality follows from the triangle inequality. For the first term of (31),
\[ E \left( \left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \left[ E(\varepsilon_{i}\varepsilon_{i}') - E(\varepsilon_{i}\varepsilon_{i}') \right] \right\| \right) \]
\[ \leq \left( \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} E \left( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left[ E(\varepsilon_{i}\varepsilon_{i}') - E(\varepsilon_{i}\varepsilon_{i}') \right] \right) \right)^{1/2} \]
where the last inequality follows from Jensen’s inequality. The above expression is bounded by Assumption D(iii).

Next, related to the second term of (31),
\[ \left\| \frac{1}{\sqrt{N^2T}} \sum_{i=1}^{N} E(\varepsilon_{i}\varepsilon_{i}') \right\| \leq \frac{1}{\sqrt{N^2T}} \sum_{i=1}^{N} \| E(\varepsilon_{i}\varepsilon_{i}') \|
\[ = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} |E(\varepsilon_{it}\varepsilon_{is})|^2 \right)^{1/2} \]
where the inequality in the first statement follows from the triangle inequality, and the last statement comes from the definition of norms. Combining Assumption D(ii-3) and D(ii-4), we can show that (32) is bounded.

From the results above, we can calculate the stochastic order of (31):
\[ \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_{i}^{t} \right\| = O_p \left( \frac{1}{NT} \right). \]
Lemma A.7: Under Assumptions A-D,

(i) \( \frac{1}{NT^2} \sum_{i=1}^{N} \left\| \varepsilon_i' \hat{F} \right\|^2 = O_p \left( B_{NT}^2 \right) + O_p \left( \frac{1}{C_{NT}} \right) \)

(ii) \( \left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_i' \hat{F} \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \)

Proof:

(i) By adding and subtracting \( F^0 H \),

\[
\frac{1}{NT^2} \sum_{k=1}^{N} \left\| \varepsilon_k' \hat{F} \right\|^2 = \frac{1}{NT^2} \sum_{k=1}^{N} \left\| \varepsilon_k' \left( \hat{F} - F^0 H + F^0 H \right) \right\|^2 \\
\leq \left( \frac{1}{NT} \sum_{k=1}^{N} \left\| \varepsilon_k \right\|^2 \right) \left( \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 \right) + \frac{1}{T} \left( \frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_k F_t^0 \right\|^2 \right) \left\| H \right\|^2
\]

where the last inequality follows from the triangle inequality and the properties of norms. Note that we already have all the results for the terms in the above expression: \( \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 = O_p \left( B_{NT}^2 \right) + O_p \left( \frac{1}{C_{NT}} \right) \) by Corollary A.1(i), and both \( \frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_k F_t^0 \right\|^2 \) and \( \left\| H \right\| \) are bounded by Lemma A.5(i) and Lemma A.4(iv), respectively. Combining these results, we have

\( \frac{1}{NT^2} \sum_{k=1}^{N} \left\| \varepsilon_k' \hat{F} \right\|^2 = O_p \left( B_{NT}^2 \right) + O_p \left( \frac{1}{C_{NT}} \right) \).

(ii) Again, by adding and subtracting \( F^0 H \), we have

\[
\left\| \frac{1}{\sqrt{NT^2}} \sum_{k=1}^{N} \lambda_i^0 \varepsilon_k' \left( \hat{F} - F^0 H \right) \right\| \leq \left( \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| \right) \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \varepsilon_k \lambda_k^0 \right\| \left\| H \right\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} F_t^0 \varepsilon_k \lambda_k^0 \right\|
\]

where the last inequality follows from the properties of norms. By Lemma A.5 (ii) and (iii), both \( \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \varepsilon_k \lambda_k^0 \right\| \) and \( \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} F_t^0 \varepsilon_k \lambda_k^0 \right\| \) are bounded. Lemma A.4(iv) proved that \( \left\| H \right\| \) is bounded. Lastly, \( \frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \) by Corollary A.1(i). Thus, we conclude that

\( \left\| \frac{1}{\sqrt{NT}} \sum \lambda_i^0 \varepsilon_i' \hat{F} \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \).
Lemma A.8: Under Assumptions A-D,

(i) \( \frac{1}{T} F^0 \left( \hat{F} - F^0 H \right) = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \)

(ii) \( \frac{1}{T} \hat{F}' \left( \hat{F} - F^0 H \right) = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \)

(iii) \( \frac{1}{T} X_k' \left( \hat{F} - F^0 H \right) = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \)

(iv) \( \frac{1}{T} \varepsilon_k' \left( \hat{F} - F^0 H \right) = O_p \left( \frac{B_{NT}}{C_{NT}} \right) + O_p \left( \frac{1}{C_{NT}} \right) \)

Proof:

(i) From (15), we have

\[
\frac{1}{T} \left\| F^0 \left( \hat{F} - F^0 H \right) \right\| = \frac{1}{T} \left\| F^0 \left( I_1 + \cdots + I_8 \right) V_{NT}^{-1} \right\|
\leq \frac{1}{T} \left\| F^0 I_1 V_{NT}^{-1} \right\| + \cdots + \frac{1}{T} \left\| F^0 I_8 V_{NT}^{-1} \right\|
\]

where the last statement follows from the triangle inequality. We shall characterize the stochastic orders of the eight terms in (33). For the first five terms, we have

\[
\frac{1}{T} \left\| F^0 I X V_{NT}^{-1} \right\| \leq \left( \frac{1}{\sqrt{T}} \left\| F^0 \right\| \right) \left( \frac{1}{\sqrt{T}} \left\| I X \right\| \right) \left\| V_{NT}^{-1} \right\|
\]

where a generic notation \( I X \) is used to denote each of \( I_1, I_2, I_3, I_4 \) and \( I_5 \). We have already shown that \( \frac{1}{\sqrt{T}} \left\| I X \right\| = O_p \left( B_{NT} \right) \) in Proposition A.1.

Next, from the definition of \( I_6 \) and the properties of norms,

\[
\frac{1}{T} \left\| F^0 I_6 V_{NT}^{-1} \right\| = \frac{1}{T} \left\| F^0 \left( \frac{1}{NT} \sum_{i=1}^{N} F^0 \lambda_i^0 \varepsilon_i \hat{F} \right) V_{NT}^{-1} \right\|
\leq \frac{1}{\sqrt{N}} \left\| \frac{F^0 \varepsilon_1}{T} \right\| \left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_i \hat{F} \right\| \left\| V_{NT}^{-1} \right\|
\]

where the last inequality follows from the properties of norms. Each of terms in the above expression is investigated already: \( \left\| \frac{F^0 \varepsilon_1}{T} \right\| \) is bounded by Assumption A(i-2), \( \left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \lambda_i^0 \varepsilon_i \hat{F} \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right) \) by Lemma A.7(ii), and \( \left\| V_{NT}^{-1} \right\| \) is bounded by Lemma A.4(iii). From all these results, we conclude that

\[
\frac{1}{T} \left\| F^0 I_6 V_{NT}^{-1} \right\| = O_p \left( \frac{B_{NT}}{\sqrt{N}} \right) + O_p \left( \frac{1}{C_{NT} \sqrt{N}} \right).
\]
Next term $\frac{1}{T} \|F^0 I T V_{NT}^{-1}\| = \frac{1}{T} \left\| F^0 \left( \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \lambda_i^0 F^0 \hat{F} \right) V_{NT}^{-1} \right\|$ does not involve any $\beta_i$. It is already shown to be of order $O_p \left( \frac{1}{C_{NT}} \right)$ in Bai (2003).

Lastly, using the definition of $I8$ and the Cauchy-Schwarz inequality, we see

$$\frac{1}{T} \|F^0 I T V_{NT}^{-1}\| = \frac{1}{T} \left\| F^0 \left( \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i' \right) V_{NT}^{-1} \right\|$$

$$\leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F^0_t \varepsilon_{it} \right\|^2 \right)^{\frac{1}{2}} \left( \frac{1}{NT^2} \sum_{i=1}^{N} \left\| \varepsilon_i' \hat{F} \right\|^2 \right)^{\frac{1}{2}} \|V_{NT}^{-1}\|$$

We have already shown that both $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} F^0_t \varepsilon_{it} \right\|^2$ and $\|V_{NT}^{-1}\|$ are bounded by Lemma A.2(ii) and by Lemma A.4(iii), respectively. Lemma A.7(i) showed that $\frac{1}{NT^2} \sum_{i=1}^{N} \left\| \varepsilon_i' \hat{F} \right\|^2 = O_p \left( \frac{B_{NT}^2}{C_{NT}} \right) + O_p \left( \frac{1}{C_{NT}^2} \right)$. Combining all the results, we have $\frac{1}{T} \|F^0 I T V_{NT}^{-1}\| = O_p \left( \frac{B_{NT}}{\sqrt{T}} \right) + O_p \left( \frac{1}{C_{NT} \sqrt{T}} \right)$.

Finally, summing over the eight terms in (33), we obtain the result of part (i).

(ii) Given the result of part (i),

$$\left\| \frac{1}{T} \hat{F}' \left( \hat{F} - F^0 H \right) \right\| = \left\| \frac{1}{T} \left( \hat{F} - F^0 H + F^0 H \right)' \left( \hat{F} - F^0 H \right) \right\|$$

$$\leq \frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 + \|H\| \left\| \frac{1}{T} F^0 \left( \hat{F} - F^0 H \right) \right\|$$

where the last statement follows from applying the triangle inequality. We have shown that $\frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 = O_p \left( \frac{B_{NT}^2}{C_{NT}} \right) + O_p \left( \frac{1}{C_{NT}} \right)$ and $\left\| \frac{1}{T} F^0 \left( \hat{F} - F^0 H \right) \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right)$ while $\|H\|$ is bounded. It follows that $\left\| \frac{1}{T} \hat{F}' \left( \hat{F} - F^0 H \right) \right\| = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{C_{NT}} \right)$.

(iii) We proceed with a similar approach as in the part (i)

$$\frac{1}{T} \left\| X' \left( \hat{F} - F^0 H \right) \right\| = \frac{1}{T} \left\| X' \left( I + \cdots + I8 \right) V_{NT}^{-1} \right\|$$

$$\leq \frac{1}{T} \left\| X' I V_{NT}^{-1} \right\| + \cdots + \frac{1}{T} \left\| X'I8 V_{NT}^{-1} \right\|$$ (34)
Again, the first five terms can be shown to be of order $O_p(B_{NT})$ because

$$
\frac{1}{T} \| X_k^t I X V_{NT}^{-1} \| \leq \left( \frac{1}{\sqrt{T}} \| X_k \| \right) \left( \frac{1}{\sqrt{T}} \| I X \| \right) \| V_{NT}^{-1} \|
$$

$$
= O_p(1) O_p(B_{NT}) O_p(1)
$$

$$
= O_p(B_{NT})
$$

where a generic notation $IX$ is used to denote each of $I1, I2, I3, I4$ and $I5$. The results for $\frac{1}{\sqrt{T}} \| I X \|$ follows from the results in Proposition A.1, and the boundedness of $\frac{1}{\sqrt{T}} \| X_k \|$ and $\| V_{NT}^{-1} \|$ comes from Assumption B(i) and Lemma A.4(iii), respectively.

Next, using the definition of $I6$ and the properties of norms,

$$
\frac{1}{T} \| X_k^t I6V_{NT}^{-1} \| = \frac{1}{T} \left\| X_k^t \left( \frac{1}{NT} \sum_{i=1}^{N} F_0^0 \lambda_i^0 \hat{F}_i \right) V_{NT}^{-1} \right\|
$$

$$
\leq \frac{1}{\sqrt{N}} \left( \frac{1}{\sqrt{T}} \| X_k \| \right) \left( \frac{1}{\sqrt{T}} \| F_0^0 \| \right) \left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \lambda_i^0 \hat{F}_i \right\| \| V_{NT}^{-1} \|
$$

Note that $\left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \lambda_i^0 \hat{F}_i \right\| = O_p(B_{NT}) + O_p \left( \frac{1}{\sqrt{NT}} \right)$ as shown in Lemma A.7(ii) while all other terms are bounded. Therefore, it follows that $\frac{1}{T} \| X_k^t I6V_{NT}^{-1} \| = O_p(B_{NT}) + O_p \left( \frac{1}{\sqrt{NT}} \right)$.

Next term $\frac{1}{T} \| X_k^t I7V_{NT}^{-1} \| = \frac{1}{T} \left\| X_k^t \left( \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \lambda_i^0 F_0^0 \hat{F}_i \right) V_{NT}^{-1} \right\|$, again, does not involve any $\beta_i$. In the exactly same manner with the part (i), we can show that $\frac{1}{T} \| X_k^t I7V_{NT}^{-1} \| = O_p \left( \frac{1}{\sqrt{NT}} \right)$

Lastly, using the Cauchy-Schwarz inequality, we have

$$
\frac{1}{T} \| X_k^t I8V_{NT}^{-1} \| = \frac{1}{T} \left\| X_k^t \left( \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \hat{F}_i \right) V_{NT}^{-1} \right\|
$$

$$
\leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \varepsilon_i \hat{F}_i \right\|^2 \right)^{\frac{1}{2}} \left( \frac{1}{NT^2} \sum_{i=1}^{N} \| F_0^0 \| \right)^{\frac{1}{2}} \| V_{NT}^{-1} \|
$$

Note that $\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{kt} \varepsilon_{it} \right\|^2$ and $\| V_{NT}^{-1} \|$ are bounded by Lemma A.2(i) and Lemma A.4(iii), respectively. We already have shown in Lemma A.7(i) that $\left( \frac{1}{NT^2} \sum_{i=1}^{N} \| \hat{F} \|^2 \right)^{\frac{1}{2}} = 58$
\( \O_p(B_{NT}) + \O_p\left(\frac{1}{T^{1/2}}\right) \). Thus, we conclude that \( \frac{1}{T} \| X'_k I 8 V_{NT}^{-1} \| = \O_p\left(\frac{B_{NT}}{\sqrt{T}} \right) + \O_p\left(\frac{1}{T^{1/2}}\right) \).

Finally, summing over the eight terms in (34), we obtain the desired result.

(iv) Using (15) and the triangle inequality,

\[
\left\| \frac{1}{T} \varepsilon'_k \left( \hat{F} - F^0 H \right) \right\| = \frac{1}{T} \left\| \varepsilon'_k (I1 + \cdots + I8) V_{NT}^{-1} \right\| \\
\leq \frac{1}{T} \left\| \varepsilon'_k I1 V_{NT}^{-1} \right\| + \cdots + \frac{1}{T} \left\| \varepsilon'_k I8 V_{NT}^{-1} \right\| \tag{35}
\]

From the definition of \( I1 \) and the Cauchy-Schwarz inequality,

\[
\frac{1}{T} \left\| \varepsilon'_k I1 V_{NT}^{-1} \right\| = \frac{1}{T} \left\| \varepsilon'_k \left[ \frac{1}{NT} \sum_{i=1}^{N} X_i \left( \beta_{0,i} - \hat{\beta}_i \right) \left( \beta_{0,i} - \hat{\beta}_i \right)' X'_i \hat{F} \right] V_{NT}^{-1} \right\| \\
\leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{kt} X'_{it} \right\|^2 \right)^{1/2} \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \beta_{0,i} - \hat{\beta}_i \right\|^8 \right)^{1/4} \\
\times \left( \frac{1}{NT^2} \sum_{i=1}^{N} \| X_i \|^4 \right)^{1/4} \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \| V_{NT}^{-1} \|
\]

Note that \( \left( \frac{1}{N} \sum_{i=1}^{N} \| \beta_{0,i} - \hat{\beta}_i \|^8 \right)^{1/4} = \O_p\left(\frac{B_{NT}^2}{NT} \right) \) while all other terms are already shown to be bounded by Lemma A.2(i), Lemma A.1(i), the normalizing assumption and Lemma A.4(iii), respectively in order. Thus, we have \( \frac{1}{T} \| \varepsilon'_k I 1 V_{NT}^{-1} \| = \O_p\left(\frac{B_{NT}^2}{\sqrt{T}} \right) \).

The stochastic order of next three terms in (35) can be shown in the same manner. For
\[
\frac{1}{T} \| \epsilon_k' I 2V_{NT}^{-1} \|
\]

\[
\frac{1}{T} \| \epsilon_k' I 2V_{NT}^{-1} \| = \frac{1}{T} \| \epsilon_k' \left[ \frac{1}{NT} \sum_{i=1}^{N} X_i \left( \beta_{0,i} - \hat{\beta}_i \right) \lambda_i' F_0 \hat{F} \right] V_{NT}^{-1} \|
\]

\[
\leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_{kt} x_{it}' \right\}^2 \right)^{\frac{1}{2}} \left( \frac{1}{N} \sum_{i=1}^{N} \| \beta_{0,i} - \hat{\beta}_i \|^4 \right)^{\frac{1}{4}}
\]

\[
\times \left( \frac{1}{N} \sum_{i=1}^{N} \| \lambda_i' \|^4 \right)^{\frac{1}{4}} \left( \frac{1}{\sqrt{T}} \| F_0 \| \right) \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \| V_{NT}^{-1} \|
\]

\[
= \frac{1}{\sqrt{T}} \times O_p(1) O_p(B_{NT}) O_p(1) O_p(1) O_p(1) O_p(1)
\]

\[
= O_p \left( \frac{B_{NT}}{\sqrt{T}} \right)
\]

where the second statement follows from applying the Cauchy-Schwarz inequality. The proof for
\[
\frac{1}{T} \| \epsilon_k' I 3V_{NT}^{-1} \| = O_p \left( \frac{B_{NT}}{\sqrt{T}} \right)
\]

and
\[
\frac{1}{T} \| \epsilon_k' I 4V_{NT}^{-1} \| = O_p \left( \frac{B_{NT}}{\sqrt{T}} \right)
\]

are omitted.

Next,

\[
\left\| \frac{1}{T} \epsilon_k' I 5V_{NT}^{-1} \right\| = \left\| \frac{1}{T} \epsilon_k' \left[ \frac{1}{NT} \sum_{i=1}^{N} \epsilon_i \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i' \hat{F} \right] V_{NT}^{-1} \right\|
\]

\[
\leq \frac{1}{NT \sqrt{T}} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{kt} \epsilon_{it} \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i' \right\| \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \| V_{NT}^{-1} \| \quad (36)
\]

Note that \( \frac{1}{\sqrt{T}} \| \hat{F} \| \) and \( \| V_{NT}^{-1} \| \) are bounded by the normalizing assumption and Lemma A.4(iii), respectively, that the order of (36) is determined by

\[
\frac{1}{NT \sqrt{T}} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \epsilon_{kt} \epsilon_{it} \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i' \right\|.
\]
By adding and subtracting \( E(\varepsilon_{kt}\varepsilon_{it}) \),

\[
\frac{1}{NT\sqrt{T}} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{kt}\varepsilon_{it} \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i' \right\| = \frac{1}{NT\sqrt{T}} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \left[ \varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it}) + E(\varepsilon_{kt}\varepsilon_{it}) \right] \right\|
\]

\[
\times \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i'
\]

\[
\leq \frac{1}{NT\sqrt{T}} \left\{ \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \left[ \varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it}) \right] \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i' \right\|
\]

\[
+ \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} E(\varepsilon_{kt}\varepsilon_{it}) \left( \beta_{0,i} - \hat{\beta}_i \right)' X_i' \right\|
\]

\[
\leq \left\{ \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[ \varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it}) \right] \right)^2 \right)^{\frac{1}{2}} \right\}
\]

\[
+ \frac{1}{\sqrt{N}} \left( \sum_{i=1}^{N} \left( \frac{1}{T} \sum_{t=1}^{T} E(\varepsilon_{kt}\varepsilon_{it}) \right)^2 \right)^{\frac{1}{2}} \right\}
\]

\[
\times \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \beta_{0,i} - \hat{\beta}_i \right\|^{4} \right)^{\frac{1}{4}} \left( \frac{1}{NT^2} \sum_{i=1}^{N} \left\| X_i \right\|^{4} \right)^{\frac{1}{4}}
\]

(37)

where the inequality in the second statement comes from the triangle inequality, and the last inequality follows from applying the Cauchy-Schwarz inequality. From Lemma A.6 (ii) and (i), the entire term in the curly braces in (37) is of order \( \left\{ \frac{1}{\sqrt{T}} \times O_p(1) + \frac{1}{\sqrt{N}} \times O_p(1) \right\} = O_p \left( \frac{1}{\sqrt{NT}} \right) \). The other terms can be shown to be \( \left( \frac{1}{N} \sum_{i=1}^{N} \left\| \beta_{0,i} - \hat{\beta}_i \right\|^{4} \right)^{\frac{1}{4}} = O_p \left( B_{NT} \right) \) and \( \left( \frac{1}{NT^2} \sum_{i=1}^{N} \left\| X_i \right\|^{4} \right)^{\frac{1}{4}} = O_p(1) \), respectively. In sum, \( \left\| \frac{1}{T} \varepsilon_k' I 6V_{NT}^{-1} \right\| = O_p \left( \frac{B_{NT}}{\sqrt{NT}} \right) \).

The proof for \( \left\| \frac{1}{T} \varepsilon_k' I 6V_{NT}^{-1} \right\| \) is much simpler. Using the Cauchy-Schwarz inequality,

\[
\frac{1}{T} \left\| \varepsilon_k' I 6V_{NT}^{-1} \right\| = \frac{1}{T} \left\| \varepsilon_k' \left[ \frac{1}{NT} \sum_{i=1}^{N} F_0 \lambda_i \varepsilon_i' \hat{F} \right] V_{NT}^{-1} \right\|
\]

\[
\leq \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{kt} F_0 \right\| \left\| \frac{1}{\sqrt{NT^2}} \sum_{i=1}^{N} \lambda_i \varepsilon_i' \hat{F} \right\| \left\| V_{NT}^{-1} \right\|
\]

We have already shown that \( \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{kt} F_0 \right\| \) and \( \left\| V_{NT}^{-1} \right\| \) are bounded by Lemma A.5(i) and
Lemma A.4(iii), respectively. Lemma A.7(ii) provides the result for $\|\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \lambda_i^0 \hat{\epsilon}_i \hat{F}\| = O_p(B_{NT}) + O_p\left(\frac{1}{\sqrt{NT}}\right)$. Therefore, it follows that $\frac{1}{T} \| \varepsilon'_k I 6V_{NT}^{-1} \| = O_p\left(\frac{B_{NT}}{\sqrt{NT}}\right) + O_p\left(\frac{1}{C_{NT} \sqrt{NT}}\right)$.

Next,

$$\frac{1}{T} \| \varepsilon'_k I 7V_{NT}^{-1} \| = \frac{1}{T} \left\| \varepsilon'_k \left[ \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i^0 \lambda_i^0 F^0 \hat{F} \right] V_{NT}^{-1} \right\|$$

$$\leq \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{kt} \varepsilon_i \lambda_i^0 \right\| \left( \frac{1}{\sqrt{T}} \| F^0 \| \right) \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \| V_{NT}^{-1} \|$$

$$= O_p\left( \frac{1}{C_{NT} \sqrt{N}} \right) O_p(1) O_p(1) O_p(1)$$

where the inequality follows from the properties of norms, and the rate result for the first term, $\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{kt} \varepsilon_i \lambda_i^0 \right\|$, comes from Lemma A.6(iii). Thus, we have $\frac{1}{T} \| \varepsilon'_k I 7V_{NT}^{-1} \| = O_p\left( \frac{1}{C_{NT} \sqrt{N}} \right)$.

Lastly,

$$\frac{1}{T} \| \varepsilon'_k I 8V_{NT}^{-1} \| = \frac{1}{T} \left\| \varepsilon'_k \left[ \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \varepsilon'_i \hat{F} \right] V_{NT}^{-1} \right\|$$

$$= \frac{1}{NT^2} \left\| \sum_{i=1}^{N} \varepsilon'_k \varepsilon_i \varepsilon'_i \left( \hat{F} - F^0 H + F^0 H \right) V_{NT}^{-1} \right\|$$

$$\leq \left\{ \left( \frac{1}{\sqrt{T}} \| \varepsilon_k \| \right) \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \varepsilon'_i \right\| \left( \frac{1}{\sqrt{T}} \| \left( \hat{F} - F^0 H \right) \| \right) \right. $$

$$\left. + \left\| \frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon'_k \varepsilon_i \varepsilon'_i F^0 \| H \| \right\| \right\} \| V_{NT}^{-1} \|$$

where the second statement comes from adding and subtracting $F^0 H$, and the last inequality follows from applying the triangle inequality and using the properties of norms. Combining the results $\left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \varepsilon'_i \right\| = O_p\left( \frac{1}{C_{NT}} \right)$ from Lemma A.6(v), $\left\| \frac{1}{\sqrt{T}} \| \left( \hat{F} - F^0 H \right) \| \right\| = O_p\left( B_{NT} \right)$ + $O_p\left( \frac{1}{C_{NT}} \right)$ from Corollary A.1(i), $\left\| \frac{1}{NT^2} \sum_{i=1}^{N} \varepsilon'_k \varepsilon_i \varepsilon'_i F^0 \right\| = O_p\left( \frac{1}{C_{NT} \sqrt{NT}} \right)$ from Lemma A.6(iv) and the fact that other terms are bounded, we can show that $\frac{1}{T} \| \varepsilon'_k I 8V_{NT}^{-1} \| = O_p\left( \frac{B_{NT}}{C_{NT}} \right) + O_p\left( \frac{1}{C_{NT}} \right)$.

Summing over the eight terms in (35), we finally have the desired result $\frac{1}{T} \varepsilon'_k \left( \hat{F} - F^0 H \right) = O_p\left( \frac{B_{NT}}{C_{NT}} \right) + O_p\left( \frac{1}{C_{NT}} \right)$.
Corollary A.2: Under Assumptions A-D, \( \frac{1}{T} \| X_i'M_F F^0 H \| = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}} \right) \)

Proof: Exploiting the fact that \( M_F \hat{F} = 0 \), the object of interest \( \frac{1}{T} \| X_i'M_F F^0 H \| \) is identical to \( \frac{1}{T} \| X_i'M_F (F^0 H - \hat{F}) \| \) or \( \frac{1}{T} \| X_i'M_F (\hat{F} - F^0 H) \| \). Now substituting \( M_F = I - \frac{1}{T} \hat{F} \hat{F}' \), we have

\[
\frac{1}{T} \| X_i'M_F F^0 H \| = \frac{1}{T} \| X_i'M_F (\hat{F} - F^0 H) \|
\]

\[
= \frac{1}{T} \| X_i' \left( I - \frac{1}{T} \hat{F} \hat{F}' \right) (\hat{F} - F^0 H) \|
\]

\[
\leq \frac{1}{T} \| X_i' (\hat{F} - F^0 H) \| + \left( \frac{1}{\sqrt{T}} \| X_i \| \right) \left( \frac{1}{T} \| \hat{F} \| \right) \frac{1}{T} \| \hat{F}' (\hat{F} - F^0 H) \|
\]

where the last inequality follows from the triangle inequality and the properties of norms. We already know the order of each term: \( \frac{1}{T} \| X_i' (\hat{F} - F^0 H) \| = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}} \right) \) by Lemma A.8(iii), \( \frac{1}{\sqrt{T}} \| X_i \| = O_p(1) \) by Assumption B(i-2), \( \frac{1}{\sqrt{T}} \| \hat{F} \| = O_p(1) \) by the normalization, and \( \frac{1}{T} \| \hat{F}' (\hat{F} - F^0 H) \| = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}} \right) \) by Lemma A.8(ii). Combining all these results, we obtain the desired corollary.

Lemma A.9: Under Assumptions A-D,

(i) \( HH' - \left( \frac{F^0 F^0}{T} \right)^{-1} = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}} \right) \)

(ii) \( \frac{1}{T} X_i'M_F X_j = \frac{1}{T} X_i'M_{F^0} X_j + o_p(1) \)

(iii) \( \frac{1}{\sqrt{T}} X_i'M_F \varepsilon_i = \frac{1}{\sqrt{T}} X_i'M_{F^0} \varepsilon_i + O_p (B_{NT}) + O_p \left( \frac{B_{NT} \sqrt{T}}{C_{NT}} \right) + O_p \left( \frac{\sqrt{T}}{C_{NT}} \right) \)

Proof:

(i) By factoring \( \left( \frac{F^0 F^0}{T} \right)^{-1} \) out, we can rewrite the object of interest as

\[
\left\| HH' - \left( \frac{F^0 F^0}{T} \right)^{-1} \right\| = \left\| HH' \left( \frac{F^0 F^0}{T} \right) - I \right\| \left( \frac{F^0 F^0}{T} \right)^{-1}
\]

\[
\leq \left\| HH' \left( \frac{F^0 F^0}{T} \right) - I \right\| \left( \frac{F^0 F^0}{T} \right)^{-1}
\]

where the last inequality follows from the properties of norms. Note that \( \left( \frac{F^0 F^0}{T} \right)^{-1} \) is bounded that the order of (38) depends on the order of \( HH' \left( \frac{F^0 F^0}{T} \right) - I \). By pre-multiplying and
post-multiplying $HH^{-1} = I$, we have

$$\|HH' \left( \frac{F^0F^0}{T} \right) - I \| = \|HH^{-1} \left[ HH' \left( \frac{F^0F^0}{T} \right) - I \right] HH^{-1}\| \leq \|H\| \left\| H' \left( \frac{F^0F^0}{T} \right) H - I \right\| \|H^{-1}\|$$

Again, the order of the above object in (39) is determined by $\|H' \left( \frac{F^0F^0}{T} \right) H - I\|$ because both $\|H\|$ and $\|H^{-1}\|$ are shown to be bounded in Lemma A.4. By adding and subtracting $H' \left( \frac{F^0F}{T} \right)$ and using the fact that $I = \frac{F^0F}{T}$ by the normalization assumption, we have

$$\|H' \left( \frac{F^0F^0}{T} \right) H - I\| = \|H' \left( \frac{F^0F^0}{T} \right) H - H' \left( \frac{F^0F}{T} \right) + H' \left( \frac{F^0F}{T} \right) - \frac{F^0F}{T}\|$$

$$= \left\| H' \left[ \frac{1}{T} F^0 \left( F^0H - \hat{F} \right) \right] + \left[ \frac{1}{T} \left( H' F^0 - \hat{F}' \right) \hat{F} \right]\right\|$$

$$\leq \|H\| \left\| \frac{1}{T} F^0 \left( \hat{F} - F^0H \right) \right\| + \left\| \frac{1}{T} \hat{F}' \left( \hat{F} - F^0H \right) \right\|$$

where the last statement follows from the triangle inequality. Note that we have already derived the orders of all terms above: $\|H\| = O_p(1)$ by Lemma A.4(iv), $\left\| \frac{1}{T} F^0 \left( \hat{F} - F^0H \right) \right\| = O_p \left( B_{NT} + O_p \left( \frac{1}{C_{NT}} \right) \right)$ by Lemma A.8(i), and $\left\| \frac{1}{T} \hat{F}' \left( \hat{F} - F^0H \right) \right\| = O_p \left( B_{NT} + O_p \left( \frac{1}{C_{NT}} \right) \right)$ by Lemma A.8(ii). Combining all the results above, we obtain the desired lemma: $HH' - \left( \frac{F^0F^0}{T} \right)^{-1} = O_p \left( B_{NT} + O_p \left( \frac{1}{C_{NT}} \right) \right)$. 

64
(ii) From the definitions of $M_F = I - \frac{1}{T} \hat{F} \hat{F}'$ and $M_{F0} = I - F^0 (F^0 F^0)^{-1} F^0'$, we have

$$\frac{1}{T} X'_i (M_{F0} - M_F) X_j = \frac{1}{T^2} X'_i \hat{F} \hat{F}' X_j - \frac{1}{T^2} X'_i F^0 \left( \frac{F^0 F^0}{T} \right)^{-1} F^0' X_j$$

$$= \frac{1}{T^2} X'_i \left( \hat{F} - F^0 H + F^0 H \right) \left( \hat{F} - F^0 H + F^0 H \right)' X_j$$

$$- \frac{1}{T^2} X'_i F^0 \left( \frac{F^0 F^0}{T} \right)^{-1} F^0' X_j$$

$$= \frac{1}{T} X'_i \left( \hat{F} - F^0 H \right) H' \left( \frac{1}{T} F^0 X_j \right)$$

$$+ \left[ \frac{1}{T} X'_i \left( \hat{F} - F^0 H \right) \right] \left[ \frac{1}{T} \left( \hat{F} - F^0 H \right)' X_j \right]$$

$$+ \left( \frac{1}{T} X'_i F^0 \right) H \left[ \frac{1}{T} \left( \hat{F} - F^0 H \right)' X_j \right]$$

$$+ \left( \frac{1}{T} X'_i F^0 \right) \left[ H H' - \left( \frac{F^0 F^0}{T} \right)^{-1} \right] \left( \frac{1}{T} F^0 X_j \right)$$ \( (40) \)

where the second statement follows from adding and subtracting $F^0 H$, and the last equality comes from expanding all terms. We can easily examine the order of four terms in (40) because we have already derived the order of each subcomponent. For the first term in (40), we know

$$\frac{1}{T} X'_i \left( \hat{F} - F^0 H \right) = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}^2} \right)$$

while other subcomponents are all bounded. The order of the second term \( \frac{1}{T} X'_i \left( \hat{F} - F^0 H \right) \left[ \frac{1}{T} \left( \hat{F} - F^0 H \right)' X_j \right] \) is determined by the product of

$$\frac{1}{T} X'_i \left( \hat{F} - F^0 H \right) = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}^2} \right)$$

and

$$\frac{1}{T} \left( \hat{F} - F^0 H \right)' X_j = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}^2} \right)$$

hence of order $O_p (B_{NT}^2) + O_p \left( \frac{B_{NT}}{C_{NT}} \right) + O_p \left( \frac{1}{C_{NT}^2} \right)$. The third term simply has the same rate with

$$\frac{1}{T} \left( \hat{F} - F^0 H \right) = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}^2} \right).$$

The order of the last term is determined by $H H' - \left( \frac{F^0 F^0}{T} \right)^{-1} = O_p (B_{NT}) + O_p \left( \frac{1}{C_{NT}^2} \right)$ form the part (i) of this lemma. Summing over the four terms in (40), we obtain

$$\frac{1}{T} X'_i (M_{F0} - M_F) X_j = o_p (1).$$

That is, we have the desired lemma:

$$\frac{1}{T} X'_i M_{F0} X_j = \frac{1}{T} X'_i M_{F0} X_j + o_p (1).$$

(iii) As in the proof of part (ii), using the definitions of $M_F = I - \frac{1}{T} \hat{F} \hat{F}'$ and $M_{F0} = I -
\[ F^0 \left( F^0 F^0 \right)^{-1} F^0 \], we have

\[
\frac{1}{\sqrt{T}} X_i' \left( M_{F^0} - M_{F} \right) \varepsilon_i = \frac{1}{\sqrt{T}} \times \frac{1}{T} X_i' \hat{F} \hat{F}' \varepsilon_i - \frac{1}{\sqrt{T}} \times \frac{1}{T} X_i' F^0 \left( \frac{F^0 F^0}{T} \right)^{-1} F^0' \varepsilon_i \\
= \frac{1}{\sqrt{T}} \times \frac{1}{T} X_i' \left( \hat{F} - F^0 H + F^0 H \right) \left( \hat{F} - F^0 H + F^0 H \right)' \varepsilon_i \\
- \frac{1}{\sqrt{T}} \times \frac{1}{T} X_i' F^0 \left( \frac{F^0 F^0}{T} \right)^{-1} F^0' \varepsilon_i \\
= \frac{1}{T} X_i' \left( \hat{F} - F^0 H \right) H' \left( \frac{1}{\sqrt{T}} F^0 \varepsilon_i \right) \\
+ \sqrt{T} \left[ \frac{1}{T} X_i' \left( \hat{F} - F^0 H \right) \right] \left[ \frac{1}{T} \left( \hat{F} - F^0 H \right)' \varepsilon_i \right] \\
+ \sqrt{T} \left( \frac{1}{T} X_i' F^0 \right) H \left[ \frac{1}{T} \left( \hat{F} - F^0 H \right)' \varepsilon_i \right] \\
+ \left( \frac{1}{T} X_i' F^0 \right) \left[ HH' - \left( \frac{F^0 F^0}{T} \right)^{-1} \right] \left( \frac{1}{\sqrt{T}} F^0 \varepsilon_i \right)
\]

where, again, the second statement follows from adding and subtracting \( F^0 H \), and the last equality comes from expanding all terms. The order of the four terms in (41) can be determined by the existing results: \( \frac{1}{T} X_i' \left( \hat{F} - F^0 H \right) = O_p \left( B_{NT} \right) + O_p \left( \frac{1}{T} \varepsilon_k \left( \hat{F} - F^0 H \right) \right) = O_p \left( \frac{B_{NT}}{T} \right) + O_p \left( \frac{1}{T} \varepsilon_k \right) \). Combining all those results, we can easily show that \( \frac{1}{\sqrt{T}} X_i' \left( M_{F^0} - M_{F} \right) \varepsilon_i = O_p \left( B_{NT} \right) + O_p \left( \frac{B_{NT} \sqrt{T}}{T} \right) + O_p \left( \frac{\sqrt{T}}{C_{NT}} \right) \), thus proving the lemma. The exact order of \( O_p \left( B_{NT} \right) + O_p \left( \frac{B_{NT} \sqrt{T}}{T} \right) + O_p \left( \frac{\sqrt{T}}{C_{NT}} \right) \) will be needed later in the proof of Proposition 1.

**Proof of Proposition 1:**

Recall that the proposed estimator has an expression \( \hat{\beta}_i = \left( X_i' M_{F} X_i \right)^{-1} \left( X_i' M_{F} Y_i \right) \). Substituting \( Y_i = X_i \beta_0 + F^0 \lambda_i^0 + \varepsilon_i \) and rearranging the terms, we obtain

\[
\left( \frac{1}{T} X_i' M_{F} X_i \right) \left( \hat{\beta}_i - \beta_{0,i} \right) = \frac{1}{T} X_i' M_{F} \varepsilon_i + \frac{1}{T} X_i' M_{F} F^0 \lambda_i^0 \\
= \frac{1}{T} X_i' M_{F} \varepsilon_i \\
+ \frac{1}{T} X_i' M_{F} \left[ \hat{F} H^{-1} - (I1 + \cdots + I8) \left( \frac{F^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^0 \Lambda^0}{N} \right)^{-1} \right] \lambda_i^0 \\
= \frac{1}{T} X_i' M_{F} \varepsilon_i + J1 + J2 + J3 + J4 + J5 + J6 + J7 + J8
\]
where the second statement follows from (15), and the last statement is just relabeling of the eight terms after expansion.

We examine the stochastic order of the eight terms in (42) one at a time. First, by the definition of $J1$,

$$
\|J1\| = \left\| (-1)^{\frac{1}{T}} X_i^T M_F I1 \left( \frac{F^{0f} \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \right\|
$$

\[
\leq \left( \frac{1}{\sqrt{T}} \|X_i^T M_F\| \right) \left( \frac{1}{\sqrt{T}} \|I1\| \right) \left\| \left( \frac{F^{0f} \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \right\| \|\lambda_i^0\| \tag{43}
\]

where the inequality comes from the properties of norms. We have already derived the order of each term in (43): $\frac{1}{\sqrt{T}} \|X_i^T M_F\| = o_p(1)$ by Lemma A.1, $\frac{1}{\sqrt{T}} \|I1\| = O_p(B_{NT}^2)$ by Proposition A.1(ii), $\left\| \left( \frac{F^{0f} \hat{F}}{T} \right)^{-1} \right\| = O_p(1)$ by (17) in Proposition A.1, $\left\| \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \right\| = O_p(1)$ by Assumption A(ii-2), and, lastly, $\|\lambda_i^0\| = O_p(1)$ by Assumption A(ii-1). Combining all the results, we conclude that $\|J1\| = o_p(B_{NT})$.

Next, by the definition of $J2$ and $I2$, we have

$$
J2 = -\frac{1}{T} X_i^T M_F I2 \left( \frac{F^{0f} \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \lambda_i^0
$$

\[
= -\frac{1}{T} X_i^T M_F \left[ -\frac{1}{NT} \sum_{j=1}^{N} X_{j} (\beta_{0,j} - \hat{\beta}_j) \lambda_j^{0f} F^{0f} \hat{F} \right] \left( \frac{F^{0f} \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \lambda_i^0
\]

\[
= -\frac{1}{N} \sum_{j=1}^{N} \left( \frac{X_i^T M_F X_j}{T} \right) (\beta_{0,j} - \hat{\beta}_j) \left[ \lambda_j^{0f} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \lambda_i^0 \right]
\]

\[
= \frac{1}{N} \sum_{j=1}^{N} \left( \frac{X_i^T M_F X_j}{T} \right) \left[ \lambda_j^{0f} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \lambda_i^0 \right] (\hat{\beta}_j - \beta_{0,j})
\]

where the second statement comes from simply substituting $I2 = \frac{1}{NT} \sum_{j=1}^{N} X_{j} (\beta_{0,j} - \hat{\beta}_j) \lambda_j^{0f} F^{0f} \hat{F}$, and the third statement holds as $\frac{F^{0f} \hat{F}}{T}$ cancels out with $\left( \frac{F^{0f} \hat{F}}{T} \right)^{-1}$. The last equality follows from the fact that $\lambda_j^{0f} \left( \frac{\Lambda^{0f} \Lambda^0}{N} \right)^{-1} \lambda_i^0$ is a scalar so that it can switch its position with $(\hat{\beta}_j - \beta_{0,j})$. Then, we can show that $J2 = O_p(B_{NT})$, and this will not be dominated by other terms in the later proof. Therefore, we keep above expression of $J2$ for later use.
For $J3$,

$$\|J3\| = \left\| -\frac{1}{T} X_t^t M_{\tilde{F}} I3 \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^0_0 \Lambda^0_0}{N} \right)^{-1} \lambda_i^0 \right\|$$

$$= \left\| -\frac{1}{T} X_t^t M_{\tilde{F}} \left[ \frac{1}{NT} \sum_{k=1}^N X_k \left( \beta_{0,k} - \hat{\beta}_k \right) \epsilon_k \hat{F} \right] \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^0_0 \Lambda^0_0}{N} \right)^{-1} \lambda_i^0 \right\|$$

$$\leq \left( \frac{1}{\sqrt{T}} \| X_t^t M_{\tilde{F}} \| \right) \left( \frac{1}{NT^2} \sum_{k=1}^N \| X_k \|^4 \right)^{\frac{1}{4}} \left( \frac{1}{N} \sum_{k=1}^N \| \beta_{0,k} - \hat{\beta}_k \|^4 \right)^{\frac{1}{4}}$$

$$\times \left( \frac{1}{NT^2} \sum_{k=1}^N \| \epsilon_k \hat{F} \|^2 \right)^{\frac{1}{2}} \left( \| \frac{F_0^0 \hat{F}}{T} \| \right)^{-1} \left( \left\| \frac{\Lambda^0_0 \Lambda^0_0}{N} \right\| \right)^{-1} \| \lambda_i^0 \|$$

where the first statement is the definition of $J3$, and the second statement comes from substituting the definition of $I3$. The last inequality follows from the Cauchy-Schwarz inequality. Again, from the previous results, we know that $\left( \frac{1}{N} \sum_{k=1}^N \| \beta_{0,k} - \hat{\beta}_k \|^4 \right)^{\frac{1}{4}} = O_p(B_{NT})$ and that $\left( \frac{1}{NT^2} \sum_{k=1}^N \| \epsilon_k \hat{F} \|^2 \right)^{\frac{1}{2}} = O_p(B_{NT}) + O_p\left( \frac{1}{\sqrt{NT}} \right)$ while other terms are bounded. Therefore, we conclude that $J3 = o_p(B_{NT})$.

Next, consider $J4$

$$\|J4\| = \left\| -\frac{1}{T} X_t^t M_{\tilde{F}} I4 \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^0_0 \Lambda^0_0}{N} \right)^{-1} \lambda_i^0 \right\|$$

$$= \left\| \frac{1}{T} X_t^t M_{\tilde{F}} \left[ \frac{1}{NT} \sum_{k=1}^N F_0^0 \lambda_k^0 (\beta_{0,k} - \hat{\beta}_k)^t X_k \hat{F} \right] \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda^0_0 \Lambda^0_0}{N} \right)^{-1} \lambda_i^0 \right\|$$

$$\leq \left( \frac{1}{T} \| X_t^t M_{\tilde{F}} F_0^0 H \| \right) \| H^{-1} \| \left( \frac{1}{N} \sum_{k=1}^N \| \lambda_k^0 \|^4 \right)^{\frac{1}{4}} \left( \frac{1}{N} \sum_{k=1}^N \| \beta_{0,k} - \hat{\beta}_k \|^4 \right)^{\frac{1}{4}}$$

$$\times \left( \frac{1}{NT} \sum_{k=1}^N \| X_k \|^2 \right)^{\frac{1}{2}} \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \left( \| \frac{F_0^0 \hat{F}}{T} \| \right)^{-1} \left( \left\| \frac{\Lambda^0_0 \Lambda^0_0}{N} \right\| \right)^{-1} \| \lambda_i^0 \|$$

where the first and second statements follows from the definition of $J4$ and $I4$, and the last inequality comes from the use of Cauchy-Schwarz inequality. Note that $\frac{1}{T} \| X_t^t M_{\tilde{F}} F_0^0 H \| = \cdots$
\( O_p(B_{NT}) + O_p \left( \frac{1}{C_N T} \right) \) by Corollary A.2 and that \( \left( \frac{1}{N} \sum_{k=1}^{N} \| \beta_{0,k} - \hat{\beta}_k \|^4 \right)^{\frac{1}{4}} \). Other terms are already shown to be bounded. Combining all these results, we can easily see that \( J_4 = o_p(B_{NT}) \).

Now

\[
\| J_5 \| = \left\| -\frac{1}{T} X_i^T M_{F'} I_5 \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda_0^0 \Lambda_0^0}{N} \right)^{-1} \lambda_i \right\|
\]

\[
= \frac{1}{T} \left\| X_i^T M_{F'} \left[ \frac{1}{NT} \sum_{k=1}^{N} \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \hat{F} \right] \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda_0^0 \Lambda_0^0}{N} \right)^{-1} \lambda_i \right\|
\]

\[
\leq \frac{1}{NT \sqrt{T}} \left\| \sum_{k=1}^{N} X_i^T M_{F'} \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \right\| \left( \frac{1}{\sqrt{T}} \left\| \hat{F} \right\| \right) \left( \frac{F_0^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda_0^0 \Lambda_0^0}{N} \right)^{-1} \left\| \lambda_i \right\| \tag{44}
\]

where the last inequality hold true by the Cauchy-Schwarz inequality. Note that the last four terms of (44) are bounded that the order of \( J_5 \) solely depends on \( \frac{1}{NT \sqrt{T}} \left\| \sum_{k=1}^{N} X_i^T M_{F'} \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \right\| \).

Substituting the definition of \( M_{F'} = I - \frac{1}{T} \hat{F} \hat{F}' \), we have

\[
\frac{1}{NT \sqrt{T}} \left\| \sum_{k=1}^{N} X_i^T M_{F'} \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \right\| = \frac{1}{NT \sqrt{T}} \left\| \sum_{k=1}^{N} X_i^T \left[ I_T - \frac{1}{T} \hat{F} \hat{F}' \right] \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \right\|
\]

\[
\leq \frac{1}{NT \sqrt{T}} \left\{ \left\| \sum_{k=1}^{N} X_i \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \right\| + \frac{1}{T} \left\| X_i^T \sum_{k=1}^{N} \hat{F}' \varepsilon_k \left( \beta_{0,k} - \hat{\beta}_k \right)' X_k \right\| \right\}
\]

\[
\leq \left\{ \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} x_{it} \varepsilon_{kt} \right\|^2 \right)^{\frac{1}{2}} \right\} + \left( \frac{1}{\sqrt{T}} \left\| X_i \right\| \right) \left( \frac{1}{\sqrt{T}} \left\| \hat{F} \right\| \right) \left( \frac{1}{NT^2} \sum_{k=1}^{N} \left\| \hat{F}' \varepsilon_k \right\|^2 \right)^{\frac{1}{2}} \right\} \times \left( \frac{1}{N} \sum_{k=1}^{N} \left\| \beta_{0,k} - \hat{\beta}_k \right\|^4 \right)^{\frac{1}{4}} \left( \frac{1}{NT^2} \sum_{k=1}^{N} \left\| X_k \right\|^4 \right)^{\frac{1}{4}} \tag{45}
\]

where the inequality in the second statement comes from the triangle inequality, and the last
inequality follows from the Cauchy-Schwarz inequality. From the previous lemmas, the terms inside the curly braces in (45) sum up to yield \( O_p(B_{NT}) + O_p \left( \frac{1}{T_N} \right) \). Note that \( \left( \frac{1}{N} \sum_{k=1}^{N} \| \beta_{0,k} - \beta_k \|^4 \right)^{\frac{1}{4}} = O_p(B_{NT}) \) and that \( \left( \frac{1}{N T^2} \sum_{k=1}^{N} \| X_k \|^4 \right)^{\frac{1}{4}} \) is simply bounded. Therefore, we conclude that 

\[
\frac{1}{NT \sqrt{T}} \left| \sum_{k=1}^{N} X_i' M_F \varepsilon_k \left( \beta_{0,k} - \beta_k \right)' X_i' \right| = o_p(B_{NT}) \]

which in turn implies that \( J5 = o_p(B_{NT}) \).

Next,

\[
\| J6 \| \equiv \left\| -\left( \frac{1}{T} X_i' M_F I6 \left( \frac{F_{OH}}{T} \right) \right)^{-1} \left( \frac{\Lambda_0 A_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
= \left\| -\frac{1}{T} X_i' M_F \left[ \frac{1}{NT} \sum_{k=1}^{N} F_0 \lambda_0^0 \varepsilon_k F_{k} \right] \left( \frac{F_{OH}}{T} \right)^{-1} \left( \frac{\Lambda_0 A_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
\leq \frac{1}{\sqrt{N}} \left( \frac{1}{T} \| X_i' M_F F_0 H \| \right) \| H^{-1} \| \left( \frac{1}{\sqrt{N T^2}} \sum_{k=1}^{N} \lambda_0^0 \varepsilon_k F_{k} \right) \left( \frac{F_{OH}}{T} \right)^{-1} \| \lambda_i^0 \|
\]

where the last inequality follows from the Cauchy-Schwarz inequality. Note that \( \frac{1}{T} \| X_i' M_F F_0 H \| = O_p(B_{NT}) + O_p \left( \frac{1}{T_N} \right) \) by Corollary A.2 and that \( \left( \frac{1}{\sqrt{N T^2}} \sum_{k=1}^{N} \lambda_0^0 \varepsilon_k F_{k} \right) = O_p(B_{NT}) + O_p \left( \frac{1}{T_N} \right) \) by Lemma A.7(ii) while remaining three terms are bounded. Combining the results, we can easily show that \( J6 = o_p(B_{NT}) + O_p \left( \frac{1}{T_N \sqrt{N}} \right) \).

Consider \( J7: \)

\[
\| J7 \| \equiv \left\| -\left( \frac{1}{T} X_i' M_F I7 \left( \frac{F_{OH}}{T} \right) \right)^{-1} \left( \frac{\Lambda_0 A_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
= \left\| -\frac{1}{T} X_i' M_F \left[ \frac{1}{NT} \sum_{k=1}^{N} \varepsilon_k \lambda_0^0 F_{k} \right] \left( \frac{F_{OH}}{T} \right)^{-1} \left( \frac{\Lambda_0 A_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
= \frac{1}{NT} \left\| \sum_{k=1}^{N} X_i' M_F \varepsilon_k \lambda_0^0 \left( \frac{\Lambda_0 A_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
\leq \frac{1}{NT} \left\| \sum_{k=1}^{N} X_i' M_F \varepsilon_k \lambda_0^0 \right\| \left( \frac{\Lambda_0 A_0}{N} \right)^{-1} \| \lambda_i^0 \| \quad (46)
\]

where the third statement follows from the fact that \( \frac{\varepsilon_k}{T} \) cancels out with \( \left( \frac{F_{OH}}{T} \right)^{-1} \), and the
last inequality comes from the properties of norms. Given the boundedness of \( \left\| \left( \frac{\Lambda_0^0 \Lambda_0}{N} \right)^{-1} \right\| \) and \( \| \lambda_i^0 \| \), it is \( \frac{1}{NT} \left\| \sum_{k=1}^{N} X_i^t M_{\varepsilon_k^0} \lambda_k^0 \right\| \) that determines the order of (46). Using the definition of \( M_{\hat{F}} = I - \frac{1}{T} \hat{F} \hat{F}' \),

\[
\frac{1}{NT} \left\| \sum_{k=1}^{N} X_i^t M_{\varepsilon_k^0} \lambda_k^0 \right\| = \frac{1}{NT} \left\| \sum_{k=1}^{N} X_i^t \left( I - \frac{1}{T} \hat{F} \hat{F}' \right) \varepsilon_k^0 \lambda_k^0 \right\|
\]

\[
\leq \frac{1}{\sqrt{NT}} \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} X_i^t \varepsilon_k^0 \lambda_k^0 \right\|
\]

\[
+ \frac{1}{\sqrt{T}} \left( \frac{1}{\sqrt{T}} \| X_i^t \| \right) \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \left( \frac{1}{\sqrt{NT^2}} \sum_{k=1}^{N} \hat{F}' \varepsilon_k^0 \lambda_k^0 \right)
\]

\[
= \frac{1}{\sqrt{NT}} \times O_p(1) + \frac{1}{\sqrt{T}} \times O_p(1) \times O_p(1) \left[ O_p(B_{NT}) + O_p \left( \frac{1}{C_{NT}} \right) \right]
\]

\[
= o_p(B_{NT}) + O_p \left( \frac{1}{C_{NT} \sqrt{T}} \right)
\]

where the second statement follows from the triangle inequality and the properties of norms. We have already shown that \( \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} X_i^t \varepsilon_k^0 \lambda_k^0 \right\| \) is bounded and that \( \left\| \frac{1}{\sqrt{NT^2}} \sum_{k=1}^{N} \hat{F}' \varepsilon_k^0 \lambda_k^0 \right\| = O_p(B_{NT}) + O_p \left( \frac{1}{C_{NT}} \right) \). In sum, we have \( J7 = o_p(B_{NT}) + O_p \left( \frac{1}{C_{NT} \sqrt{T}} \right) \).

Lastly, consider \( J8 \):

\[
\| J8 \| = \left\| \frac{1}{T} X_i^t M_{\hat{F}} I8 \left( \frac{F_{0}^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda_0^0 \Lambda_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
= \left\| \frac{1}{T} X_i^t M_{\hat{F}} \left[ \frac{1}{NT} \sum_{k=1}^{N} \varepsilon_k \varepsilon_k' \hat{F} \right] \left( \frac{F_{0}^0 \hat{F}}{T} \right)^{-1} \left( \frac{\Lambda_0^0 \Lambda_0}{N} \right)^{-1} \lambda_i^0 \right\|
\]

\[
\leq \frac{1}{NT^2} \left\| \sum_{k=1}^{N} X_i^t M_{\varepsilon_k^0 \varepsilon_k'} \hat{F} \right\| \left\| \left( \frac{F_{0}^0 \hat{F}}{T} \right)^{-1} \right\| \left\| \left( \frac{\Lambda_0^0 \Lambda_0}{N} \right)^{-1} \right\| \| \lambda_i^0 \| \quad (47)
\]

where the first two statements follow from the definition of \( J8 \) and \( I8 \), and the last inequality comes from the properties of norms. Note that last three terms in (47) are already shown to be bounded. Thus, we examine \( \frac{1}{NT^2} \left\| \sum_{k=1}^{N} X_i^t M_{\varepsilon_k^0 \varepsilon_k'} \hat{F} \right\| \) which determines the order of \( J8 \). Using
the definition of $M_{\hat{F}} = I - \frac{1}{T} \hat{F} \hat{F}'$,

$$\frac{1}{NT^2} \left\| \sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \varepsilon_k' \hat{F} \right\| = \frac{1}{NT^2} \left\| \sum_{k=1}^{N} X_i' \left( I_T - \frac{1}{T} \hat{F} \hat{F}' \right) \varepsilon_k \varepsilon_k' \hat{F} \right\|

\leq \frac{1}{\sqrt{T}} \left( \frac{1}{N} \sum_{k=1}^{N} \left\| \sum_{t=1}^{T} x_{it} \varepsilon_{kt} \right\| \right)^{\frac{1}{2}} \left( \frac{1}{NT^2} \sum_{k=1}^{N} \| \varepsilon_k \hat{F} \|^2 \right)^{\frac{1}{2}}

+ \left( \frac{1}{\sqrt{T}} \| X_i \| \right) \left( \frac{1}{\sqrt{T}} \| \hat{F} \| \right) \left( \frac{1}{NT^2} \sum_{k=1}^{N} \| \varepsilon_k \hat{F} \|^2 \right)

$$

where the last inequality follows from the combination of the triangle inequality and the Cauchy-Schwarz inequality. Note that $\left( \frac{1}{NT^2} \sum_{k=1}^{N} \| \varepsilon_k \hat{F} \|^2 \right)^{\frac{1}{2}} = O_p(NT) + O_p \left( \frac{1}{C_{NT}} \right)$ from Lemma A.7(i) while other terms are bounded. Thus, we conclude

$$\frac{1}{NT^2} \sum_{k=1}^{N} \| X_i' M_{\hat{F}} \varepsilon_k \varepsilon_k' \hat{F} \| = o_p(NT) + O_p \left( \frac{1}{C_{NT}} \right),$$

which in turn leads to $J8 = o_p(NT) + O_p \left( \frac{1}{C_{NT}^2} \right)$.

Collecting the results for $J1$ to $J8$, we have

$$\left( \frac{1}{T} X_i' M_{\hat{F}} X_i \right) \left( \hat{\beta}_i - \beta_{0,i} \right) = \frac{1}{T} X_i' M_{\hat{F}} \varepsilon_i + J1 + J2 + J3 + J4 + J5 + J6 + J7 + J8

= \frac{1}{T} X_i' M_{\hat{F}} \varepsilon_i + \frac{1}{N} \sum_{j=1}^{N} \left( \frac{X_i' M_{\hat{F}} X_j}{T} \right) \left[ \lambda_j^0 \left( \frac{A_0^0 A_0^0}{N} \right)^{-1} \lambda_i^0 \right] \left( \hat{\beta}_j - \beta_{0,j} \right)

+ O_p \left( \frac{1}{C_{NT}^2} \right) + o_p \left( \frac{1}{\sqrt{T}} \right) + o_p (B_{NT}) \tag{48}

$$

Multiplying by $\sqrt{T}$ on both sides,

$$\left( \frac{1}{T} X_i' M_{\hat{F}} X_i \right) \sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right) = \frac{1}{\sqrt{T}} X_i' M_{\hat{F}} \varepsilon_i

+ \frac{1}{N} \sum_{j=1}^{N} \left( \frac{X_i' M_{\hat{F}} X_j}{T} \right) \left[ \lambda_j^0 \left( \frac{A_0^0 A_0^0}{N} \right)^{-1} \lambda_i^0 \right] \sqrt{T} \left( \hat{\beta}_j - \beta_{0,j} \right)

+ O_p \left( \frac{\sqrt{T}}{C_{NT}^2} \right) + o_p(1)

$$

Note that the last term $o_p \left( \sqrt{T} B_{NT} \right)$ is dominated by $\sqrt{T} \left( \hat{\beta}_j - \beta_{0,j} \right)$ which is of order $O_p \left( \sqrt{T} B_{NT} \right)$.
Now we replace $M_F$ in the above expression by $M_{F^0}$ using Lemma A.9 (ii) and (iii).

\[
\left[\frac{1}{T}X_i'M_{F^0}X_i + o_p(1)\right] + o_p(1) = \left[\frac{1}{\sqrt{T}}X_i'M_{F^0}\xi_i + o_p(1)\right] + O_p\left(\frac{B_{NT}}{C_{NT}}\right) + \frac{\sqrt{T}}{C_{NT}}\left(\hat{\beta}_i - \beta_{0,i}\right) \times \lambda_0^\prime \left(\frac{\Lambda_0^\prime \Lambda_0}{N}\right) \lambda_0^0 \sqrt{T} \left(\hat{\beta}_j - \beta_{0,j}\right)
\]

\[
= \frac{1}{\sqrt{T}}X_i'M_{F^0}\xi_i + \frac{1}{N} \sum_{j=1}^N \left(\frac{X_i'M_{F^0}X_j}{T} + o_p(1)\right) \times \lambda_0^\prime \left(\frac{\Lambda_0^\prime \Lambda_0}{N}\right) \lambda_0^0 \sqrt{T} \left(\hat{\beta}_j - \beta_{0,j}\right) + o_p(1)
\]

where the last equality holds under an additional condition on $N$ and $T$, that is $\frac{T}{N^2} \to 0$. We will keep this additional assumption throughout. Taking the inverse of $\frac{1}{T}X_i'M_{F^0}X_i + o_p(1)$ on both sides, we obtain the desired proposition:

\[
\sqrt{T} \left(\hat{\beta}_i - \beta_0^i\right) = \left[\frac{1}{T}X_i'M_{F^0}X_i + o_p(1)\right]^{-1} \left\{\frac{1}{\sqrt{T}}X_i'M_{F^0}\xi_i\right\} + \frac{1}{N} \sum_{j=1}^N \left(\frac{X_i'M_{F^0}X_j}{T} + o_p(1)\right) \times \lambda_0^\prime \left(\frac{\Lambda_0^\prime \Lambda_0}{N}\right) \lambda_0^0 \sqrt{T} \left(\hat{\beta}_j - \beta_{0,j}\right) + o_p(1)
\]

For notational simplicity, define $\xi_i = \frac{1}{\sqrt{T}}X_i'M_{F^0}\xi_i$, $S_{ii} \equiv \frac{X_i'M_{F^0}X_i}{T}$, and

\[
G_{ij} \equiv \left(\frac{X_i'M_{F^0}X_i}{T}\right)^{-1} \left(\frac{X_i'M_{F^0}X_j}{T}\right) \times \lambda_0^\prime \left(\frac{\Lambda_0^\prime \Lambda_0}{N}\right) \lambda_0^0 \sqrt{T} \left(\hat{\beta}_j - \beta_{0,j}\right)
\]
Then the expression (49) above can be rewritten as

\[
\sqrt{T} (\hat{\beta}_i - \beta_{0,i}) = S_{ii}^{-1} \xi_i + \frac{1}{N} \sum_{j=1}^{N} G_{ij} \sqrt{T} (\hat{\beta}_j - \beta_{0,j}) + o_p(1).
\]

Stacking all \(\sqrt{T} (\hat{\beta}_i - \beta_{0,i})\) as a vector leads to

\[
\begin{bmatrix}
\sqrt{T} (\hat{\beta}_1 - \beta_{0,1}) \\
\sqrt{T} (\hat{\beta}_2 - \beta_{0,2}) \\
\vdots \\
\sqrt{T} (\hat{\beta}_N - \beta_{0,N})
\end{bmatrix} =
\begin{bmatrix}
S_{11}^{-1} \xi_1 \\
S_{22}^{-1} \xi_2 \\
\vdots \\
S_{NN}^{-1} \xi_N
\end{bmatrix} + \frac{1}{N} \begin{bmatrix}
G_{11} & G_{12} & \cdots & G_{1N} \\
G_{21} & G_{22} & \cdots & G_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
G_{N1} & G_{N2} & \cdots & G_{NN}
\end{bmatrix}
\begin{bmatrix}
\sqrt{T} (\hat{\beta}_1 - \beta_{0,1}) \\
\sqrt{T} (\hat{\beta}_2 - \beta_{0,2}) \\
\vdots \\
\sqrt{T} (\hat{\beta}_N - \beta_{0,N})
\end{bmatrix} + o_p(1)
\]

Defining

\[
\sqrt{T} (\hat{\beta} - \beta) =
\begin{bmatrix}
\sqrt{T} (\hat{\beta}_1 - \beta_{0,1}) \\
\sqrt{T} (\hat{\beta}_2 - \beta_{0,2}) \\
\vdots \\
\sqrt{T} (\hat{\beta}_N - \beta_{0,N})
\end{bmatrix}, \quad S =
\begin{bmatrix}
S_{11}^{-1} & 0 & \cdots & 0 \\
0 & S_{22}^{-1} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & S_{NN}^{-1}
\end{bmatrix},
\]

\[
\xi = 
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\vdots \\
\xi_N
\end{bmatrix}, \quad G =
\begin{bmatrix}
G_{11} & G_{12} & \cdots & G_{1N} \\
G_{21} & G_{22} & \cdots & G_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
G_{N1} & G_{N2} & \cdots & G_{NN}
\end{bmatrix},
\]

we have \(\sqrt{T} (\hat{\beta} - \beta) = S^{-1} \xi + \left(\frac{1}{N} G\right) \sqrt{T} (\hat{\beta} - \beta) + o_p(1)\). Now, solving for \(\sqrt{T} (\hat{\beta} - \beta)\) yields

\[
\sqrt{T} (\hat{\beta} - \beta) = \left( I - \frac{1}{N} G \right)^{-1} S^{-1} \xi + o_p(1)
\]

\[
= \left( I + \frac{1}{N} G + \frac{1}{N^2} G^2 + \frac{1}{N^3} G^3 + \cdots \right) S^{-1} \xi + o_p(1).
\]
Then, a generic individual $i$’s estimator has an expression

$$\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right) = S_{ii}^{-1} \xi_i + \frac{1}{N} \sum_{j=1}^{N} G_{ij} S_{jj}^{-1} \xi_j + o_p(1).$$

(50)

Note that higher order terms related to $\frac{1}{N^k} G^k S^{-1} \xi$ can be ignored as $o_p(1)$ thanks to the increasing order of $\frac{1}{N^k}$. They contribute to $\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right)$ in the form of a weighted average of $N$ bounded terms, but the denominator $\frac{1}{N^k}$ increases at a much faster rate.

**Proof of Theorem 2:**

Theorem 2 readily follows from the expression (50) in Proposition 1. Under Assumption A-D, every term $G_{ij} S_{jj}^{-1} \xi_j$ is bounded, and the average of all the bounded terms is also bounded. Note also that every $S_{ii}^{-1} \xi_i$ term from the infeasible estimator is bounded. Therefore, we obtain the desired result that the term $\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right) = O_p(1)$.

**Proof of Theorem 3:**

Under cross-sectional independence, an average of independent terms $\frac{1}{N} \sum_{j=1}^{N} G_{ij} S_{jj}^{-1} \xi_j$ in (50) converges faster that $S_{ii}^{-1} \xi_i$ becomes the only leading term of $\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right)$. By cross-sectional independence, we mean a vector of $(\bar{x}'_{it}, L' \bar{\epsilon}_it)'$ is independent from $(\bar{x}'_{jt}, L' \bar{\epsilon}_js)'$ for all $i, j, t$ and $s$, where we define $\bar{x}_{it} = M_{F^0} x_{it}$. If no lagged dependent variables are included as regressors, replacing $\bar{x}_{it}$ by $x_{it}$ provides sufficient conditions. Note that $S_{ii}^{-1} \xi_i = \left( \frac{X_i'M_{F^0} X_i}{\bar{T}} \right)^{-1} \frac{1}{\sqrt{\bar{T}}} X_i'M_{F^0} \bar{\epsilon}_i$ corresponds to the infeasible estimator where the latent common factor $F^0$ is treated as if observable. That is, under cross-sectional dependence, the individual estimator $\sqrt{T} \left( \hat{\beta}_i - \beta_{0,i} \right)$ asymptotically behaves like the infeasible estimator that its asymptotic distribution can be further characterized as a normal distribution with mean zero and variance $\Omega_i$. Here, the asymptotic variance $\Omega_i$ is identical to that of infeasible estimator, which has the usual sandwich formula $\Omega_i = \Sigma_i^{-1} \Xi_i \Sigma_i^{-1}$ where $\Sigma_i = p \lim_{T \to \infty} X_i'M_{F^0} X_i$ and $\Xi_i = p \lim_{T \to \infty} \frac{X_i'M_{F^0} E(\bar{\epsilon}_i\bar{\epsilon}_i')M_{F^0} X_i}{T}$. 

75