Essays on Large Panel Data Analysis

Minkee Song

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2013

ABSTRACT

Essays on Large Panel Data Analysis

Minkee Song

A growing number of studies in macroeconomics and finance have attempted to utilize large panel data sets. Large panel data sets contain rich information on the dynamics of many cross-sectional units over long time periods. These data sets often consist of numerous series in different categories that reflect the multifaceted aspects of an economy. In other circumstances, data sets are constructed from a large number of series at a highly disaggregated level within the same category so that they can reveal dynamics in greater detail. Numerous studies have proven the usefulness of large panel data sets in improving forecast performance, distinguishing common shocks from idiosyncratic shocks, and uncovering the discrepancies in dynamics between aggregate series and disaggregated series.

To gain the most from large panel data sets, econometric models should allow all the key characteristics of these rich data sets without distortion. Among the pervasive and important characteristics of large panels are dynamics, heterogeneity, and cross-sectional dependence. While there has been a great deal of research on each of these three features, the consequences of jointly incorporating them into a single model have not been extensively studied in the existing literature. Chapter 1 of this dissertation considers dynamic heterogeneous panels with cross-sectional dependence (DHP+CSD) that allow for all three key characteristics at the same time. Cross-sectional dependence is modeled through the use of a common factor structure in the error terms. We propose an estimator for the DHP+CSD model and develop an asymptotic theory under a large N and large T setup. The estimator relies on an iterative principal component method to cope with the challenges in estimation arising from the greater generality of the DHP+CSD model.

The proposed estimator is shown to be $\sqrt{T}$-consistent under non-stringent conditions and performs well in finite samples. Furthermore, the overall performance of the estimator is satisfactory even if no factor structure is present. Consequently, the DHP+CSD approach facilitates prudent estimation without requiring an additional procedure of pre-testing cross-sectional dependence.

The econometric tool developed in Chapter 1 can be particularly useful in analyzing possible discrepancies in persistence between an aggregate series and its underlying disaggregated series. It is well-known that an aggregate series can exhibit drastically different dynamics from its underlying processes. Early literature focuses on the role of heterogeneity in the dynamics of disaggregated series, whereas recent studies note that the dynamics of common factors also play an important role. Therefore, it is essential to use a model that incorporates dynamics, heterogeneity, and cross-sectional dependence (that arises from common factors) for analyzing the dynamics of disaggregated series. We apply the DHP+CSD estimator to investigate the dynamics of disaggregated data sets in two important empirical contexts: the purchasing power parity (PPP) hypothesis and the intrinsic persistence of inflation. Most studies have relied on models that utilized dynamics and heterogeneity without considering common factors. Given the important role of common factor dynamics, revisiting the issue of aggregation with the DHP+CSD model in these empirical contexts can meaningfully extend the existing studies.

Chapter 2 of this dissertation investigates the dynamics of sectoral real exchange rates in the context of the PPP hypothesis. It is widely known that aggregate exchange rates exhibit a considerable degree of persistence, serving as evidence against the PPP hypothesis. Recent studies, however, report that persistence estimates are markedly lower if exchange rate dynamics are examined at the disaggregated level. Given the focus on the dynamics of disaggregated series, a persistence analysis of sectoral exchange rates perfectly fits into the DHP+CSD framework. Consistent with recent studies, our estimation results show that the persistence of sectoral exchange rates is indeed lower than that of aggregate exchange rates. In addition, the persistence estimates from the DHP+CSD model are substantially lower than the estimates from those models that ignored the dynamics of common factors.

This suggests that the estimates of the latter models might be vulnerable to distortions caused by ignoring some key features of the given large panel data set. We also document the difference in responses with respect to common shocks and idiosyncratic shocks. This analysis is possible primarily because the DHP+CSD model can distinguish the two types of shocks. On average, common shocks appear to have approximately 50% more persistent effects on the economy than idiosyncratic shocks.

Chapter 3 aims to assess the persistence of inflation at the disaggregated level. Persistence is widely accepted as one of the key characteristics of inflation. Similar to the recent PPP literature, however, numerous studies have also found considerably lower persistence at the disaggregated level. Since many empirical studies often disregard the possible dynamics of common factors, there is room for refining the existing analysis by adopting the DHP+CSD model. Given the estimated dynamics of sectoral inflation, we also attempt to measure the degree of intrinsic persistence at the disaggregated level. Intrinsic persistence is a useful concept for identifying the structural sources of inflation persistence; a low intrinsic persistence implies that most of the inflation persistence is inherited from the real marginal costs. Because low intrinsic persistence also implies less inertia, it is associated with forward-looking behavior in price-setting. In contrast to the substantial degrees of estimated intrinsic persistence in the literature, we find that price-setting is markedly forward-looking at the disaggregated level; in approximately half of all sectors in the U.S. economy, price-setting is close to purely forward-looking. In measuring intrinsic persistence through the DHP+CSD model, we establish a relationship between the DHP+CSD model and the sectoral New Keynesian Phillips Curves. Recovering the structural parameters of intrinsic persistence from the reduced-form DHP+CSD estimates serves as an alternative framework of structural analysis for inflation dynamics.

In conclusion, this dissertation develops a useful econometric method for analyzing large panel data sets and illustrates its practical value by applying it to two important empirical contexts: the PPP hypothesis and the intrinsic persistence of inflation. With the DHP+CSD model, we can analyze the dynamics of disaggregated series more precisely and shed new light on the discrepancies in persistence between an aggregate series and its underlying disaggregated series. We also illustrate that the developed model has potential as a reduced-form representation of structural models for further structural analysis. All things considered, it is hoped that this dissertation provides a useful econometric framework for large panel data analysis.

The latter two chapters of this dissertation rely on the same econometric method developed in Chapter 1. As a consequence, there are significant overlaps among the chapters. Each chapter contains relevant background information and descriptions on the econometric method such that it is entirely self-contained. A reader interested in only one chapter of this dissertation may focus solely on that chapter.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to sincerely thank my advisor Jushan Bai for his guidance and support.

I am also deeply grateful to John Donaldson, Dennis Kristensen, Ricardo Reis, and Jón Steinsson for their insightful advice and encouragement.

I thank many faculty and colleagues for their thoughtful comments and suggestions: Jisun Baek, Richard Davis, Raffaella Giacomini, Ju Hyun Kim, Ji Hyung Lee, Marcelo Moreira, Serena Ng, Hyun Seung Oh, Alexei Onatski, WooRam Park, Christoph Rothe, Bernard Salanié, and Hyelim Son.

For supporting my research, I thank Columbia University and the Program for Economic Research.

To my beloved wife Hyo Jung

Chapter 1

Asymptotic Theory for Dynamic Heterogeneous Panels with Cross-Sectional Dependence

1.1 Introduction

A growing number of studies in macroeconomics and finance have attempted to utilize large panel data sets. Large panel data sets contain rich information on the dynamics of many cross-sectional units over long time periods. These data sets often consist of numerous series in different categories that reflect the multifaceted aspects of an economy. In other circumstances, data sets are constructed from a large number of series at a highly disaggregated level within the same category so that they can reveal dynamics in greater detail. In either case, many studies have shown the benefits of analyzing large panel data sets; Stock and Watson (2002) significantly improved the forecasting performance, Bernanke et al. (2005) identified the effect of structural shocks, Boivin and Giannoni (2006) estimated a Dynamic Stochastic General Equilibrium (DSGE) model, Boivin et al. (2009) distinguished the effect of common shocks from idiosyncratic shocks, and Pesaran and Chudik (2013) noted the discrepancies between an aggregate series and its underlying disaggregated series.

To gain the most from large panel data sets, econometric models should allow all the key characteristics of these rich data sets without distortion. This chapter considers a new model that is tailored for large panel data sets. There are three key features in the model: ($\mathbb{D}$) dynamics through lagged dependent variables, ($\mathbb{H}$) heterogeneity of the individual coefficients, and ($\mathbb{CSD}$) cross-sectional dependence in the error terms. Put differently, we consider dynamic heterogeneous panels with cross-sectional dependence, hereafter abbreviated as DHP+CSD. Dynamics, heterogeneity and cross-sectional dependence are pervasive and important characteristics of large panel data sets. While there has been a great deal of research on each of these three features, the consequences of incorporating them jointly in a model have not been extensively studied in the existing literature.[1] The

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

primary contribution of this chapter is addressing these three issues ($\mathbb{D}$, $\mathbb{H}$, and $\mathbb{CSD}$) simultaneously. Given the prevalence of these three features in most economic data sets, extending models in this direction is essential for more realistic empirical analysis.

In our model, cross-sectional dependence is modeled through the use of a common factor structure in the error terms.[2] A common factor structure is highly probable, especially in the disaggregated data sets, because each series is likely to be affected by some common shocks. Indeed, researchers frequently find common dynamics among disaggregated series in a variety of empirical contexts. In these cases, cross-sectional dependence is merely a symptom of having a common factor structure in the data generating process; it is these unobservable common factors that induce cross-sectional dependence. In general, unobservable common factors may be correlated with observable regressors in the model in an arbitrary fashion.[3] Such correlations, however, can make the unobservable common factors behave as omitted variables that are difficult to control for. For example, the ordinary least squares (OLS) estimator is not capable of consistently estimating the DHP+CSD model due to the omitted variable bias. To estimate the model consistently, it is crucial to control for the unobservable common factors. In this chapter, we treat the unobservable common factors as parameters to be estimated. By directly estimating the unobservable common factors, we control for the omitted variables and achieve consistency. This approach is primarily possible because we consider a large N and large T setup where the data contain sufficient information for estimating both the common factors and individual coefficients.

The benefit of the general specification of the DHP+CSD model comes at the cost of posing challenges to both the estimation and the asymptotic theory. The difficulty

[2] A spatial approach is an equally popular approach of modeling cross-sectional dependence. See, for example, Anselin (2008) and references therein.

[3] One such circumstance of our particular interest is the dynamic panel model with serially correlated common factors. In this case, the common factors are necessarily correlated with the lagged dependent variables.

of simultaneously addressing the three issues ($\mathbb{D}$, $\mathbb{H}$, and $\mathbb{CSD}$) has been recognized in the existing literature. Phillips and Sul (2003), for instance, note the interdependence of these issues and emphasize the importance of taking a systematic approach. Estimation of the DHP+CSD model involves minimizing a least squares objective function over a large number of parameters that consist of individual coefficients, factors, and loadings. For the convenience of implementation, we adopt the iterative principal component approach proposed in Bai (2009). This approach decomposes the original estimation problem into two steps: the estimation of the individual coefficients given common factors, and the estimation of the common factors given individual coefficients. The final estimates are obtained by iterating these two steps until convergence. Because each step is extremely easy to implement, the iterative approach greatly simplifies the estimation procedure. We prove that the proposed estimator for each individual coefficient is $\sqrt{T}$-consistent under non-stringent conditions of N and T. We also derive the asymptotic distribution of each individual coefficient estimator.

We run extensive Monte Carlo experiments using a simple version of the DHP+CSD model, particularly with serially correlated unobservable common factors. The simulation results show that the DHP+CSD estimator works well in finite samples. The estimator correctly estimates the true individual coefficients, even under heteroskedasticity and weak cross-sectional dependence in the idiosyncratic errors. Although the performance of the DHP+CSD estimator cannot dominate that of an infeasible estimator, the inefficiency gap rapidly narrows down as the sample size increases.[4] By contrast, the OLS estimator suffers from severe omitted variable bias because it neglects controlling for the unobservable common factors. We also examine the performance of the proposed estimator when no common factor structure is present in the data generating process. The simulation results show that the overall performance is satisfactory despite the need for estimating

[4]The infeasible estimator utilizes true common factors as if they are observable, and it is used as a benchmark for assessing performance.

unnecessary parameters. Without the robust properties of the DHP+CSD estimator, the analysis must typically proceed in two stages. In the first stage, one pre-tests the presence of cross-sectional dependence using suitable tests available in the existing literature.[5] In the second stage, the model is estimated using the DHP+CSD estimator if the pre-test detects cross-sectional dependence. Otherwise, the OLS estimator may be used. However, it is well-known that the second stage inference can suffer from size distortion due to pre-testing (see Leeb and Pötscher, 2005 and its subsequent studies). With the DHP+CSD estimator, no pre-testing is necessary due to its robust performance with or without the presence of cross-sectional dependence. Therefore, the proposed estimator facilitates prudent estimation without the known pre-testing problem.

This chapter is closely related to the work of Bai (2009). Since we adopt the proposed iterative estimation procedure to simplify the entire estimation problem, the main frame of the proof strategy in Bai (2009) naturally carries over to our problem as well. The primary difference comes from whether to allow for heterogeneity of the individual coefficients. Because the proof of Bai (2009) is designed for the case of slope homogeneity, modifications are necessary to cope with the heterogeneous coefficients in the DHP+CSD model. The key to the modification is disentangling the complex dependence structure among the individual coefficient estimators. This chapter is also related to the work of Moon and Weidner (2010), who analyze the same estimator using a matrix perturbation argument. Since both Bai (2009) and Moon and Weidner (2010) allow for $\mathbb{D}$ and $\mathbb{CSD}$ (but not $\mathbb{H}$), the extension of this chapter is unavoidable if interest lies in the *heterogeneous* dynamics at the disaggregated level. Another related study is Pesaran (2006), who propose the Common Correlated Effects (CCE) estimator for models with $\mathbb{H}$ and $\mathbb{CSD}$ (but not $\mathbb{D}$). The CCE estimator relies on the novel idea that unobservable common factors can be controlled for by including cross-sectional averages of observables as additional regressors. Chudik and

[5] There exists a vast literature on testing cross-sectional dependence. See Friedman (1937), Breusch and Pagan (1980), Frees (1995), Pesaran (2004), Pesaran et al. (2008), Sarafidis et al. (2009), Baltagi et al. (2011), Chen et al. (2012), among others. Each test is valid under different conditions.

Pesaran (2013) extend the approach to dynamic panels by including additional lags of cross-sectional averages and develop an asymptotic theory under all three features ($\mathbb{D}$, $\mathbb{H}$, and $\mathbb{CSD}$) as in this chapter. The proposed estimator in Chudik and Pesaran (2013) is also easy to implement and shows good finite sample performance. Finally, our work is related to the vast literature on heterogeneous dynamics before cross-sectional dependence gained much attention. Because the literature is mostly focused only on $\mathbb{D}$ and $\mathbb{H}$, there can be an inconsistency issue if unobserved common factors have their own dynamics such that the correlations with lagged dependent variables are unavoidable.

The remainder of the chapter 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 concludes. Finally, Section 7 provides extensive simulation results and all the proofs.

We use the following notations throughout the chapter. 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.

1.2 Model

We consider a panel data model with heterogeneous coefficients:

$$y_{it} = \mathbf{x}_{it}'\beta_{0,i} + u_{it} \tag{1.1}$$

and

$$u_{it} = \lambda_i^{0\prime} F_t^0 + \varepsilon_{it} \tag{1.2}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where $\mathbf{x}_{it}$ is a $k \times 1$ vector of observed individual-specific regressors on the ith cross-section unit at time t, $\beta_{0,i}$ is a $k \times 1$ vector of individual-specific coefficients, and u_{it} is the error term. The main parameter of interest is $\beta_{0,i}$. We allow $\mathbf{x}_{it}$ to include lagged dependent variables so that the current model can cover the case of dynamic panels. The cross-sectional dependence of the error term u_{it} is modeled using a factor structure (1.2) in which F_t^0 is an $r \times 1$ vector of common factors, λ_i^0 is an $r \times 1$ vector of factor loadings, and ε_{it} is the idiosyncratic error. All three components (F_t^0, λ_i^0, and ε_{it}) are unobservable because they are components of the error term u_{it}.

In matrix notations, this model can also be written as

$$Y_i = X_i\beta_{0,i} + F^0\lambda_i^0 + \varepsilon_i \tag{1.3}$$

where

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iT} \end{bmatrix}, \; X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{iT} \end{bmatrix}, \; F^0 = \begin{bmatrix} F_1^0 \\ F_2^0 \\ \vdots \\ F_T^0 \end{bmatrix}, \text{ and } \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}. \tag{1.4}$$

We also define $\Lambda^0 = \left(\lambda_1^0, \lambda_2^0, \cdots, \lambda_N^0\right)'$ in a similar manner.

Note that the current model can be viewed as a generalization of widely used panel data models with individual fixed effects and/or time fixed effects. If there is no common fluctuation in the error term (i.e., F_t^0 is a constant over time) then the model simplifies to a fixed effects model. In that case, the loadings capture unobserved individual heterogeneity. If the common factors have a homogeneous effect on individuals such that λ_i^0 are the same for all i's, then the model reverts to one with time fixed effects. In the special case where $\lambda_i^0 = [1, \alpha_i]'$ and $F_t^0 = [d_t, 1]'$, the current model also covers the two-way fixed effects model.

We allow flexible specifications for λ_i^0 and F_t^0. The common factor F_t^0 can be a general linear covariance stationary process, possibly with a non-zero mean. A similar statement also holds true for λ_i^0; it can be correlated over cross-sections with the possibility of a non-zero mean. Recalling that $\mathbf{x}_{it}$ may include lagged dependent variables, the flexible specification that we allow for λ_i^0 and F_t^0 may lead to a potential correlation between $\mathbf{x}_{it}$ and the common component $\lambda_i^{0\prime}F_t^0$. We do allow $\mathbf{x}_{it}$ to be correlated with F_t^0 and/or λ_i^0 in a general nonlinear fashion. The key idea that enables this flexible specification is to treat both λ_i^0 and F_t^0 as fixed parameters to be estimated. It is possible to estimate F_t^0 and λ_i^0 and thus control for the unobserved heterogeneity because we consider the case where rich information can be obtained from large dimensional data. We postulate the following assumptions to be more specific about the nature of the factors and the loadings.

Assumption A:

A(i-1) $E\left\|F_t^0\right\|^4 \leq M$ for some positive constant $M < \infty$

A(i-2) $\frac{1}{T}\sum_{t=1}^T F_t^0 F_t^{0\prime} \to^p \Sigma_F$ for some $r \times r$ positive definite matrix Σ_F, as $T \to \infty$

A(ii-1) $E\left\|\lambda_i^0\right\|^4 \leq M$ for some positive constant $M < \infty$

A(ii-2) $\frac{1}{N}\Lambda^{0\prime}\Lambda^0 \to^p \Sigma_\Lambda$ for some $r \times r$ positive definite matrix Σ_Λ, as $N \to \infty$

A(iii) There exists a compact set $\mathcal{B}$ such that $\beta_{0,i} \in \mathcal{B}$ for every 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, which means that the common factors affect almost all series in the limit. When we derive the asymptotic results, we will consider the estimation on a compact set around the true value $\beta_{0,i}$ as defined in Assumption A(iii). For example, for a dynamic panel model with a single lag, $\mathcal{B}$ can be an interval between -1 and 1, excluding the boundaries.

Assumption B: Define $M_F = I_T - F\left(F'F\right)^{-1}F'$ for any $T \times r$ matrix F, where I_T denotes a $T \times T$ identity matrix. Also, we define $\mathcal{F}$ as the space of all F that satisfy $\frac{F'F}{T} = I_r$.

B(i) $E\left\|\mathbf{x}_{it}\right\|^4 \leq M$ for some positive constant $M < \infty$

B(ii) $S_{ii} = \frac{X_i' M_{F^0} X_i}{T} \to^p \Sigma_{ii} > 0$ for some $k \times k$ positive definite matrix Σ_{ii}, as $T \to \infty$

B(iii) $\inf_{F\in\mathcal{F}} D(F) > 0$ where $D(F) = \frac{1}{N}\sum_{i=1}^{N} D_i$, $D_i = B_i - C_i' A_i^{-1} C_i$, $A_i = \frac{X_i' M_F X_i}{T}$, $B_i = \left(\lambda_i^0 \lambda_i^{0\prime}\right) \otimes \frac{I_T}{T}$, and $C_i = \lambda_i^{0\prime} \otimes \frac{1}{T}\left(X_i' M_F\right)$

Assumption B contains conditions on $\mathbf{x}_{it}$, including the boundedness of its moments. More importantly, $\mathbf{x}_{it}$ is assumed to exhibit sufficient variation such that the corresponding coefficient $\beta_{0,i}$ is identifiable, which is intuitively appealing. Identifiability of $\beta_{0,i}$ also requires that observed individual regressor $\mathbf{x}_{it}$ does not exhibit multicolinearity with the unobservable true common factors F_t^0. The assumptions above allows common observed regressors in $\mathbf{x}_{it}$ unless the common regressors do not span the same space with the unobservable true common factors. The final assumption guarantees the unique minimizer of the estimation objective function.[6] The notations of $D(F)$ is used to emphasize that the entire term is a function of F.

Assumption C: ε_{it} is independent of $\mathbf{x}_{js}$, $\beta_{0,j}$, F_s^0, and λ_j^0 for all i, j, t, and s.

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 regressors, coefficients, factors, and loadings.

Assumption D:

D(i) $E(\varepsilon_{it}) = 0$ and $E\left|\varepsilon_{it}\right|^8 \leq M$ for some positive constant $M < \infty$

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

D(ii-2) $\forall i, N \quad \sum_{j=1}^{N} \bar{\sigma}_{ij} \leq M$

D(ii-3) $\forall t, s \quad \exists \bar{\tau}_{ts} < \infty \quad s.t. \quad \forall i, j \quad |\sigma_{ij,ts}| = |E(\varepsilon_{it}\varepsilon_{js})| \leq \bar{\tau}_{ts}$

D(ii-4) $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \bar{\tau}_{ts} \leq M$

D(ii-5) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} |E(\varepsilon_{it}\varepsilon_{js})| \leq M$

---

[6]The assumption is a heterogeneous coefficients version of the corresponding Assumption A in Bai (2009).

D(iii) $E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\varepsilon_{is}\varepsilon_{it}-E\left(\varepsilon_{is}\varepsilon_{it}\right)\right]\right|^4 \leq M$

D(iv-1) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|cov(\varepsilon_{iu}\varepsilon_{it},\varepsilon_{ju}\varepsilon_{js})\right| \leq M$

D(iv-2) $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|cov(\varepsilon_{it}\varepsilon_{kt},\varepsilon_{js}\varepsilon_{ks})\right| \leq M$

Assumption D on idiosyncratic errors contains standard assumptions found in the literature on factor models. The conditions allow for "weak" time series and cross-sectional correlations in the sense of the approximate factor model of Chamberlain and Rothschild (1983). This is in contrast to the exact factor model originally developed by Geweke (1977) and Sargent and Sims (1977), which imposes strict uncorrelatedness assumptions. If, however, $\mathbf{x}_{it}$ includes lagged dependent variables, the possibility of serial correlation in the idiosyncratic errors is excluded by Assumption C. Heteroskedasticity is also allowed insofar as the moments are uniformly bounded by the condition D(i). Assumption D(i) to D(iii) are already assumed in Assumption C of Bai (2009). More complicated technical assumptions D(iv-1) and D(iv-2) are weaker than those assumed in Bai (2009) due to the heterogeneous coefficient setup in this chapter. But these assumptions serve the same purpose of limiting the amount of correlation among the idiosyncratic errors. See the discussions therein for the interpretation of the conditions.

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 $\mathbf{x}_{it}$ is absent, various methods have been proposed in the literature to determine the number of factors.[7] 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 chapter, however, it is not obvious whether the existing method remains valid, and a formal analysis goes beyond the scope of this dissertation.

[7] See, for example, Bai and Ng (2002), Hallin and Liska (2007), Amengual and Watson (2007), and Onatski (2010).

1.3 Estimation

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

$$\begin{aligned} SSR\left(\{\beta_i\}_{i=1}^{N}, F, \Lambda\right) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{it}'\beta_i - \lambda_i' F_t\right)^2 \\ &= \sum_{i=1}^{N}\left(Y_i - X_i\beta_i - F\lambda_i\right)'\left(Y_i - X_i\beta_i - F\lambda_i\right) \end{aligned} \tag{1.5}$$

subject to the normalization $\frac{F'F}{T} = I_r$ and $\frac{\Lambda'\Lambda}{N}$ being diagonal. The objective function came originally from the literature on the principal components approach to pure factor models. If β_i are known, the objective function amounts to that of pure factor models, with $Y_i - X_i\beta_i$ being regarded as observed data. In our estimation problem, however, β_i are also unknown along with the factors and the loadings. Therefore, we treat β_i as parameters and minimize the objective function over their space as well.

The objective function (1.5) 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 (1.5) by using the inverse of the variance matrix of the idiosyncratic errors as a weighting matrix. Estimating the variance matrix of the idiosyncratic errors, which we will denote as Σ_ε, is challenging due to its dimensions. If $N > T$, the estimator of Σ_ε 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 Σ_ε. Boivin and Ng (2006) suggested restricting the number of parameters to be estimated by setting the off-diagonal elements of Σ_ε 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 Σ_ε.

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 β_i and F to the minimization problem, each λ_i must satisfy a relationship of the form $\lambda_i = \left(F'F\right)^{-1} F' \left(Y_i - X_i\beta_i\right) = \left(F'F\right)^{-1} F'W_i$, to be a minimizer as well. Defining $W_i = Y_i - X_i\beta_i$ and substituting the expression for the loadings into (1.5), we have the following concentrated objective function:

$$SSR\left(\{\beta_i\}_{i=1}^{N}, F\right) = \sum_{i=1}^{N} \left(Y_i - X_i\beta_i - F\left(F'F\right)^{-1} F'W_i\right)' \left(Y_i - X_i\beta_i - F\left(F'F\right)^{-1} F'W_i\right) \quad (1.6)$$

which, in turn, simplifies to

$$SSR\left(\{\beta_i\}_{i=1}^{N}, F\right) = tr\left[W'M_F W\right] = tr\left[W'W\right] - tr\left[F'\left(WW'\right)F/T\right] \quad (1.7)$$

where $W = (W_1, W_2, \cdots, W_N)$, and $M_F = I_r - F\left(F'F\right)^{-1} F' = I_r - \frac{1}{T}FF'$ denotes an orthogonal projection matrix.

Given all β_i, the minimization problem above is equivalent to the problem of maximizing the last term of (1.7), which, in turn, is equivalent to

$$\max_F tr\left[F'\left(\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\beta_i\right)\left(Y_i - X_i\beta_i\right)'\right)F\right] \qquad subject\ to \quad \frac{F'F}{T} = I_r \quad (1.8)$$

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 $\hat{F}$ equal to the rescaled eigenvectors of $\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\beta_i\right)\left(Y_i - X_i\beta_i\right)'$, corresponding to its r largest eigenvalues. Rescaling the eigenvectors by $\sqrt{T}$ satisfies the normalization restriction $\frac{F'F}{T} = I_r$. Now, given $\hat{F}$, the least squares problem has a well-known solution for $\hat{\beta}_i$, which is $\hat{\beta}_i = \left(X_i'M_{\hat{F}}X_i\right)^{-1} X_i'M_{\hat{F}}Y_i$.

Therefore, the estimator $\left(\left\{\hat{\beta}_i\right\}_{i=1}^{N}, \hat{F}\right)$ should simultaneously solve a system of nonlinear equations

$$\hat{\beta}_i = \left(X_i'M_{\hat{F}}X_i\right)^{-1} X_i'M_{\hat{F}}Y_i \quad (1.9)$$

and

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i - X_i\hat{\beta}_i\right)\left(Y_i - X_i\hat{\beta}_i\right)'\right]\hat{F} = \hat{F}\hat{V}_{NT} \tag{1.10}$$

where $\hat{V}_{NT}$ is a diagonal matrix of r largest eigenvalues corresponding to $\hat{F}$. Actual estimation procedure can be implemented by iterating each of the two steps in (1.9) and (1.10) until convergence. The estimator for λ_i^0 can be obtained by computing $\hat{\lambda}_i = \left(\hat{F}'\hat{F}\right)^{-1}\hat{F}'\left(Y_i - X_i\hat{\beta}_i\right)$. Such iterative estimation procedure has already been adopted by Ahn et al. (2001) and Bai (2009) under a homogeneous coefficient setup, and the convergence of the procedure to the local solution of the original optimization problem has been shown by Sargan (1964). While the iterative procedure substantially reduces the computational burden of finding local optimum, the estimation still requires trying multiple starting points as in the original optimization problem. The procedure, however, is shown to have a good convergence property from our extensive Monte Carlo experiments.

1.4 Asymptotic theory

This section establishes the consistency, the rate of convergence, and the asymptotic normality of the DHP+CSD estimator proposed in the previous section. There are three unique features of the current problem that pose challenges to the econometric theory. First, the proposed estimator does not have a closed-form expression; it consists of a set of equations that should be satisfied simultaneously by $\hat{\beta}_i$ and $\hat{F}_t$. Second, the unobserved common factors are treated as parameters to be estimated, and thus the number of parameters grows with T. Finally, each i has its own parameter of interest $\beta_{0,i}$, and the number of parameters also grows with N.

In general, when estimators have no closed-form expression, the approach in Newey and McFadden (1994) can be applied for a consistency argument. In this chapter, 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.[8] In the appendix, we show that the auxiliary objective function is uniformly close to the original objective function as (N, T) approaches infinity and that the auxiliary objective function is uniquely minimized at the true parameter values. Relying on this framework, we can prove the consistency of the individual coefficient estimator as well as the consistency of the space spanned by the estimated common factors.

Theorem 1. Consistency : Define $P_F = F\left(F'F\right)^{-1}F'$ for any $T \times r$ matrix F. Under Assumption A-D, as $(N,\, T) \to \infty$ jointly, the following statements hold:

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

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

Once consistency is established, we characterize the convergence rate of the individual coefficient estimator $\hat{\beta}_i$. The rate of convergence, however, cannot be characterized in a single step because the estimator does not have a closed-form expression. Instead, we begin with the consistency of $\hat{\beta}_i$ and $\hat{F}$ and then refine their convergence rates in multiple rounds (in an iterative manner) until we obtain the final convergence rate results. This refinement is applied to all terms that consist of $\hat{\beta}_i$ and $\hat{F}$. During this process, extra care must be taken to cope with the heterogeneity of individual coefficients. Each estimator $\hat{\beta}_i$ depends on $\hat{F}$, and $\hat{F}$ in turn depends on all $\hat{\beta}_i$. This circular relationship implies that $\hat{\beta}_i$ is, in effect, dependent upon all other $\hat{\beta}_j$ with $i \neq j$. This complicated dependence structure among the estimators arises from pooling the information over cross-sectional units to control for the unobservable common factors. Pooling cross-sectional information is typical in homogeneous panel models to utilize the rich information in the data, thus boosting the rate of convergence. Nevertheless, no complicated dependence structure arises in the homogeneous panel models because heterogeneous coefficients do not exist. In most models

[8] This proof strategy was first proposed in Bai (1994) and later adopted by Bai (2009) and Bonhomme and Manresa (2012).

with heterogeneous coefficients, consistent estimators typically do not require pooling cross-sections. Consequently, their asymptotic properties can be investigated without considering any complicated dependence structure among the estimators. In contrast, the estimation of the DHP+CSD model involves the consideration of heterogeneous coefficients and the pooling of cross-sections at the same time. This renders the DHP+CSD model unique from the typical setups in the existing literature and also poses challenges to developing an asymptotic theory. We address the issue by exploiting the observation that the model (as well as the estimation procedure) is symmetrical with respect to each cross-section unit. The symmetry enables the collection of particular terms with the same stochastic order, and thus helps in deriving the following interim result.

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 $\frac{\sqrt{T}}{N} \to 0$:

$$\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) = S_{ii}^{-1}\xi_i + \frac{1}{N}\sum_{j=1}^{N} G_{ij} S_{jj}^{-1}\xi_j + o_p(1) \tag{1.11}$$

where $G_{ij} = \left(\frac{X_i' M_{F^0} X_i}{T}\right)^{-1}\left(\frac{X_i' M_{F^0} X_j}{T}\right)\left[\lambda_i^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_j^0\right]$.

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 an infeasible estimator term ($S_{ii}^{-1}\xi_i$) and a weighted average of those terms over all cross-sections. 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; i.e., 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 $S_{jj}^{-1}\xi_j$ 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 ($\mathbf{x}_{it}$ and $\mathbf{x}_{jt}$) are more correlated and/or if their loadings (λ_i^0 and λ_j^0) are more correlated. This relationship highlights the consequences of controlling for cross-sectional dependence by pooling cross-section units in estimation. The rate of convergence can be readily determined from the expression in Proposition 1 when combined with the assumption of limited cross-sectional correlation.

Theorem 2. Rate of Convergence : Under Assumptions A-D, as $(N, T) \to \infty$ jointly, $\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) = O_p(1)$ as long as $\frac{\sqrt{T}}{N} \to 0$.

Pesaran (2006) also establishes the same rate of convergence under the same condition on N and T, although the study does not explicitly cover dynamic panels. Notably, the condition $\frac{\sqrt{T}}{N} \to 0$ is less stringent than the conditions in Bai (2009) or Moon and Weidner (2010), whose estimators for homogeneous coefficients require that $\frac{T}{N} \to 0$. Note that an interesting trade-off exists between the rate of convergence and the stringency of the condition on N and T. If the coefficients are indeed homogeneous, two options are available. Using a homogeneous coefficient estimator facilitates $\sqrt{NT}$-consistency, but it requires a stricter condition on N and T that may or may not be satisfied by the data set at hand. Adopting a heterogeneous coefficient estimator requires less stringent conditions, but it delivers only $\sqrt{T}$-consistency.

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 $\mathbf{X}_i$, λ_i^0, and ε_i jointly over i is sufficient for the weighted average term to vanish. If $\mathbf{x}_{it}$ includes lagged dependent variables, such independence over i is not feasible. Instead, $\widetilde{\mathbf{X}}_i$, λ_i^0, and ε_i are required to be jointly independent over i where $\widetilde{\mathbf{X}}_i = M_{F^0}\mathbf{X}_i$. Here, we present our final asymptotic result for this special case.

Theorem 3. Asymptotic Normality: Suppose that Assumptions A-D hold and that $\widetilde{\mathbf{X}}_i$, λ_i^0, and ε_i are jointly independent over i. As $(N, T) \to \infty$ jointly,

$$\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) \to^d \mathcal{N}(0, \Omega_i) \tag{1.12}$$

a long as $\frac{\sqrt{T}}{N} \to 0$ where $\Omega_i = \Sigma_i^{-1}\Xi_i\Sigma_i^{-1}$, $\Sigma_i = p\lim_{T\to\infty}\frac{X_i' M_{F^0} X_i}{T}$, and $\Xi_i = p\lim_{T\to\infty}\frac{X_i' M_{F^0} E\left(\varepsilon_i\varepsilon_i'\right) M_{F^0} X_i}{T}$.

1.5 Monte Carlo simulation results

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

$$y_{it} = \rho_i y_{i,t-1} + \lambda_i F_t + \varepsilon_{it} \tag{1.13}$$

The DGP is a special case of the general model studied in this chapter. In spite of its simplicity, (1.13) allows for all three features $\mathbb{D}$, $\mathbb{H}$, and $\mathbb{CSD}$. The heterogeneous autoregressive (AR) coefficients ρ_i are drawn from $\mathcal{U}[0, 1)$, and the loadings λ_i are drawn 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 $u_t \sim \mathcal{N}(0, 1)$ being its innovation process:

$$F_t = \rho_F F_{t-1} + u_t \tag{1.14}$$

where the first $1,000$ observations of the generated data are burned-in. We use $\rho_F = 0.7$ to introduce serial correlations in F_t. The serial correlation necessarily leads to the correlations between the regressor $y_{i,t-1}$ and the unobservable common factor F_t. Therefore, it should be noted that any estimator without special treatment of the unobserved factor F_t will not be consistent because of the omitted variable bias. Regarding the idiosyncratic errors ε_{it}, we follow the simulation scheme in Doz et al. (2011) to allow for both weak cross-sectional dependence and heteroskedasticity. The variance σ_i^2 of each idiosyncratic error term ε_{it}

is drawn independently from $\mathcal{U}[0.5, 1.5]$ to introduce heteroskedasticity. In the scheme of Doz et al. (2011) for cross-sectional dependence among the idiosyncratic disturbances, a typical entry of the variance-covariance matrix of the idiosyncratic errors has the form of $\mathcal{T}_{ij} = \tau^{|i-j|}\sqrt{\sigma_i \sigma_j}$ such that the entire variance-covariance matrix has a pattern of Toeplitz matrix. Under this correlation structure, ε_{it} and ε_{jt} become less correlated as the distance $|i - j|$ between the two series increases. The amount of cross-sectional dependence is controlled by a tuning parameter τ . Note that no specific cross-sectional ordering of the series matters; i.e., 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 AR coefficients ρ_i. The procedure is repeated for $B = 1,000$ times to evaluate the performance of five types of estimators: 1) the infeasible estimator, 2) the DHP+CSD estimator, 3) the CCE estimator, 4) the Dynamic CCE estimator, and 5) the OLS estimator. The infeasible estimator utilizes the realized true factor processes as though they were observable, and its performance is used as a benchmark for comparison. The DHP+CSD estimator is the estimator of our primary interest. The CCE estimator was initially proposed by Pesaran (2006) in the context of static panels. We also include the CCE estimator to examine its validity in dynamics panels. Chudik and Pesaran (2011) suggest including more lags of the cross-sectional averages in the case of dynamic panels, and Chudik and Pesaran (2013) develop the econometric theory for such estimator. Following the lag length selection rule in Chudik and Pesaran (2013), we evaluate the performance of the Dynamic CCE estimator by including $3, 3, 4, 4, 5$ lags of cross-sectional averages for $T = 30, 50, 70, 100, 200$, respectively. Finally, the OLS estimator is expected to be inconsistent because it ignores the omitted variable F_t that is correlated with the regressor. We nevertheless include the OLS estimator in the simulation to illustrate the amount of omitted variable bias in our experiment setup.

We evaluate the performance of each estimator in two ways. Following Pesaran

(2006), the first approach compares the performance of the mean group (MG) estimator, $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$. This is equivalent to considering the mean $E(\rho_i)$ of the heterogeneous coefficients as a new parameter of interest and to comparing the performance its consistent estimators. The MG estimator is frequently used in the empirical studies to examine the representative dynamics of heterogeneous individual series. The second approach computes the mean integrated squared error (MISE), $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i - \rho_i|^2\right)$, which is analogous to the corresponding concept in nonparametrics. The MISE better reflects the performance of each individual estimator because the measure is based on the deviation of each estimate from the corresponding true parameter.

Table 1.1 shows the performance of the MG estimator based on the five types of estimators when both weak cross-sectional dependence and heteroskedasticity are present in the idiosyncratic errors. The inconsistency of the OLS estimator stands out in the simulation results. Most of the large 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 highlights 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 even when the sample size is large. The Dynamic CCE estimator performs well, especially when compared to the CCE estimator, suggesting that adding more lags helps to control for the space spanned by the true common factors. The performance of the new estimator proposed in this chapter is reported under the column labeled DHP+CSD. We can see the bias, standard deviation, and 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

Table 1.1: Performance of the MG estimator

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0252	-0.0398	-0.0612	-0.0215	0.1830
200	30	SD	0.0227	0.0238	0.0270	0.0306	0.0703
		RMSE	0.0339	0.0463	0.0669	0.0374	0.1960
		Bias	-0.0148	-0.0231	-0.0173	0.0154	0.1929
200	50	SD	0.0214	0.0220	0.0276	0.0310	0.0535
		RMSE	0.0260	0.0319	0.0326	0.0346	0.2002
		Bias	-0.0104	-0.0162	-0.0079	0.0312	0.1987
200	70	SD	0.0210	0.0213	0.0259	0.0296	0.0427
		RMSE	0.0235	0.0268	0.0270	0.0429	0.2032
		Bias	-0.0071	-0.0110	0.0054	0.0431	0.2018
200	100	SD	0.0205	0.0208	0.0248	0.0278	0.0349
		RMSE	0.0217	0.0235	0.0254	0.0513	0.2048
		Bias	-0.0036	-0.0054	0.0126	0.0542	0.2055
200	200	SD	0.0200	0.0203	0.0230	0.0253	0.0229
		RMSE	0.0203	0.0210	0.0262	0.0599	0.2068
		Bias	-0.0033	-0.0050	0.0144	0.0557	0.2060
100	200	SD	0.0281	0.0283	0.0309	0.0337	0.0274
		RMSE	0.0283	0.0288	0.0341	0.0651	0.2078
		Bias	-0.0039	-0.0055	0.0150	0.0560	0.2056
70	200	SD	0.0338	0.0341	0.0372	0.0405	0.0313
		RMSE	0.0340	0.0346	0.0401	0.0691	0.2079
		Bias	-0.0040	-0.0051	0.0168	0.0573	0.2052
50	200	SD	0.0392	0.0400	0.0430	0.0466	0.0344
		RMSE	0.0395	0.0403	0.0462	0.0738	0.2080
		Bias	-0.0034	-0.0028	0.0220	0.0616	0.2051
30	200	SD	0.0514	0.0522	0.0560	0.0600	0.0429
		RMSE	0.0515	0.0523	0.0602	0.0861	0.2096

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $\rho_F = 0.7$ and $u_t \sim i.i.d.\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\mathcal{N}(1, 0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5, 1.5]$ and $cov\left(\varepsilon_{it}, \varepsilon_{jt}\right) = \tau^{|i-j|}\sqrt{\sigma_i \sigma_j}$ where $\tau = 0.5$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N} \hat{\rho}_i$ with $E\left(\rho_i\right) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

better than the DHP+CSD estimator because the DHP+CSD estimator must estimate unobserved factors in addition to other parameters. In sum, Table 1.1 shows that our new estimator is suitable for analyzing dynamic heterogeneous panels with cross-sectional dependence, whereas the existing methods seem to suffer from the omitted variable bias caused by ignoring the common factor dynamics. The pattern of relative performance of the five types of estimators is also confirmed in Table 1.2, in which performance is measured in terms of the MISE of all individual coefficient estimators.

Table 1.2: MISE performance of the individual coefficient estimators

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0148	0.0162	0.0380	0.0322	0.0691
200	50	0.0080	0.0085	0.0183	0.0202	0.0656
200	70	0.0055	0.0057	0.0125	0.0165	0.0649
200	100	0.0037	0.0039	0.0086	0.0143	0.0643
200	200	0.0018	0.0018	0.0044	0.0121	0.0634

Notes: See also the notes to Table 1.1. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\rho_i|^2\right)$.

An interesting experiment is 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 can still replace the OLS estimator with the DHP+CSD estimator. In this case, it is important for the DHP+CSD estimator to exhibit robust performance even though a factor structure does not exist in the errors. We consider a similar DGP to the one used in the previous experiment, except for one modification; the factors and loadings are not present. Table 1.3 reports the results for all types of estimator except for infeasible estimator; the OLS estimator serves as a benchmark in this experiment. 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,

Table 1.3: Performance of the MG estimator under no common factor structure

N	T		OLS	DHP+CSD	Dynamic CCE	CCE
		Bias	-0.0301	-0.0402	-0.0600	-0.0529
200	30	SD	0.0237	0.0274	0.0270	0.0259
		RMSE	0.0383	0.0486	0.0658	0.0589
		Bias	-0.0191	-0.0214	-0.0351	-0.0325
200	50	SD	0.0224	0.0231	0.0229	0.0226
		RMSE	0.0294	0.0315	0.0419	0.0396
		Bias	-0.0138	-0.0148	-0.0252	-0.0229
200	70	SD	0.0219	0.0221	0.0219	0.0218
		RMSE	0.0259	0.0266	0.0334	0.0316
		Bias	-0.0096	-0.0100	-0.0168	-0.0155
200	100	SD	0.0212	0.0212	0.0211	0.0211
		RMSE	0.0232	0.0234	0.0270	0.0262
		Bias	-0.0049	-0.0050	-0.0078	-0.0069
200	200	SD	0.0205	0.0205	0.0204	0.0204
		RMSE	0.0211	0.0211	0.219	0.0216
		Bias	-0.0039	-0.0041	-0.0058	-0.0043
100	200	SD	0.0289	0.0289	0.0288	0.0288
		RMSE	0.0292	0.0292	0.0294	0.0291
		Bias	-0.0042	-0.0045	-0.0053	-0.0033
70	200	SD	0.0355	0.0356	0.0353	0.0353
		RMSE	0.0358	0.0359	0.0357	0.0355
		Bias	-0.0051	-0.0053	-0.0051	-0.0025
50	200	SD	0.0415	0.0417	0.0412	0.0414
		RMSE	0.0418	0.0420	0.0416	0.0415
		Bias	-0.0071	-0.0071	-0.0047	-0.0004
30	200	SD	0.0529	0.0530	0.0526	0.0528
		RMSE	0.0534	0.0535	0.0528	0.0528

Notes: The dependent variables are generated 1,000 times according to $y_{it} = \rho_i y_{i,t-1} + \varepsilon_{it}$ with $u_t \sim i.i.d.\mathcal{N}(0,1)$. The first 1,000 observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\mathcal{U}[0,1)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\mathcal{U}[0.5,1.5]$ and $cov\left(\varepsilon_{it},\varepsilon_{jt}\right) = \tau^{|i-j|}\sqrt{\sigma_i\sigma_j}$ where $\tau = 0.5$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E\left(\rho_i\right) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

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. The result implies that it might be prudent to always use the DHP+CSD estimator. It is because the DHP+CSD estimator delivers consistent estimates under any circumstances, and the price paid in terms of efficiency is not large even if no factor structure is present. The simulation results show that the performance of the CCE estimator and the Dynamic CCE estimator are also satisfactory even when no common factor structure is present. The pattern of relative performance is also confirmed in Table 1.4, in which the performance is measured in terms of the MISE of all individual coefficient estimators.

Table 1.4: MISE performance of the individual coefficient estimators under no common factor structure

N	T	OLS	DHP+CSD	Dynamic CCE	CCE
200	30	0.0259	0.0289	0.0367	0.0310
200	50	0.0147	0.0152	0.0181	0.0165
200	70	0.0102	0.0104	0.0123	0.0111
200	100	0.0070	0.0071	0.0079	0.0074
200	200	0.0034	0.0034	0.0037	0.0035

Notes: See also the notes to Table 1.3. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i-\rho_i|^2\right)$.

To evaluate the validity of the asymptotic distribution of the DHP+CSD estimator in Theorem 3, we consider a significance test of AR coefficient ρ_i for generic individual i. That is, the null hypothesis becomes $H_0:\ \rho_i = 0$.[9] The data are generated from the same setup of (1.13) and (1.14) except for two modifications. First, data are generated with the AR coefficients set to zero because we are interested in the distribution of the estimator under the null hypothesis. Second, idiosyncratic errors are generated to be independent over cross-sections. This is because the asymptotic distribution in Theorem 3 is derived

[9] It can also be of interest to test a joint null hypothesis, $H_0:\ \rho_i = 0\quad \forall i$. Such test requires consistent estimator for the inverse of a variance-covariance matrix of increasing dimensions. It is beyond the scope of this chapter, and we leave the analysis for future work.

under such special case that idiosyncratic errors are cross-sectionally independent. We repeat data generation and hypothesis testing for $B = 10,000$ times and compute the size of the test at the 5% nominal level. The results are reported in Table 1.5 below for various combinations of N and T. The results show that the test based on the asymptotic distribution has a satisfactory size property without severe distortion.

Table 1.5: Size of the significance test for individual coefficient

$N \setminus T$	30	50	70	100	200
30	0.076	0.071	0.064	0.061	0.060
50	0.070	0.066	0.061	0.057	0.057
70	0.069	0.066	0.059	0.057	0.054
100	0.069	0.064	0.059	0.054	0.054
200	0.067	0.063	0.056	0.054	0.053

Notes: The dependent variables and the common factors are generated $10,000$ times according to (1.13) and (1.14), respectively, with $\rho_F = 0.7$ and $u_t \sim i.i.d.\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The loadings are generated as $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The test statistic is constructed by $\sqrt{T}\hat{\rho}_i\hat{\Omega}_i^{-\frac{1}{2}}$ where $\hat{\Omega}_i$ is the sample analog of the asymptotic variance in Theorem 3. The null hypothesis is $H_0: \ \rho_i = 0$.

The appendix at the end of this chapter contains more extensive Monte Carlo experiments for various common factor persistence as well as for different specifications on the idiosyncratic errors.

1.6 Conclusion

This chapter considers a new econometric model that is tailored for realistic large panel data analysis. Dynamics, heterogeneity, and cross-sectional dependence are simultaneously allowed for in a single model to reflect the pervasive characteristics of large panel data sets. Jointly allowing for these three issues is also important in examining the discrepancies in persistence between an aggregate series and its underlying disaggregated series. We propose an estimator for the DHP+CSD model and develop an asymptotic theory under a large N and large T setup. The estimator is based on an iterative principal component approach to overcome the challenges in estimation arising from the generality of the DHP+CSD model. The proposed estimator is shown to be $\sqrt{T}$-consistent under non-stringent conditions and performs well in finite samples. Furthermore, the overall performance of the estimator is satisfactory even if no factor structure is present. This robust behavior of the DHP+CSD estimator facilitates prudent estimation without requiring the additional procedure of pre-testing cross-sectional dependence.

1.7 Appendix

1.7.1 Additional Monte Carlo simulation results

Table 1.6: Performance of the MG estimator under $\rho_F = 0.8$ and *i.i.d.* idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0308	-0.0462	-0.0699	-0.0308	0.2050
200	30	SD	0.0236	0.0245	0.0276	0.0308	0.0726
		RMSE	0.0388	0.0522	0.0751	0.0436	0.2175
		Bias	-0.0184	-0.0269	-0.0211	0.0115	0.2183
200	50	SD	0.0225	0.0232	0.0282	0.0323	0.0541
		RMSE	0.0291	0.0355	0.0352	0.0343	0.2249
		Bias	-0.0134	-0.0192	-0.0098	0.0298	0.2230
200	70	SD	0.0220	0.0239	0.0269	0.0313	0.0446
		RMSE	0.0258	0.0307	0.0287	0.0432	0.2274
		Bias	-0.0096	-0.0137	0.0056	0.0440	0.2288
200	100	SD	0.0218	0.0220	0.0267	0.0300	0.0348
		RMSE	0.0238	0.0259	0.0273	0.0532	0.2314
		Bias	-0.0052	-0.0072	0.0146	0.0571	0.2330
200	200	SD	0.0214	0.0215	0.0250	0.0274	0.0213
		RMSE	0.0220	0.0227	0.0290	0.0633	0.2339
		Bias	-0.0048	-0.0066	0.0165	0.0585	0.2332
100	200	SD	0.0305	0.0310	0.0340	0.0367	0.0264
		RMSE	0.0309	0.0317	0.0377	0.0691	0.2347
		Bias	-0.0046	-0.0066	0.0177	0.0595	0.2335
70	200	SD	0.0358	0.0360	0.0402	0.0436	0.0298
		RMSE	0.0361	0.0366	0.0439	0.0737	0.2354
		Bias	-0.0036	-0.0053	0.0196	0.0602	0.2338
50	200	SD	0.0419	0.0427	0.0471	0.0513	0.0343
		RMSE	0.0421	0.0430	0.0510	0.0791	0.2363
		Bias	-0.0038	-0.0041	0.0218	0.0610	0.2338
30	200	SD	0.0540	0.0549	0.0604	0.0664	0.0423
		RMSE	0.0542	0.0550	0.0642	0.0901	0.2376

Notes: The dependent variables and the common factors are generated 1,000 times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first 1,000 observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E(\rho_i) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.7: Performance of the MG estimator under $\rho_F = 0.5$ and $i.i.d.$ idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0222	-0.0349	-0.0583	-0.0245	0.1176
200	30	SD	0.0230	0.0242	0.0254	0.0271	0.0667
		RMSE	0.0320	0.0425	0.0635	0.0366	0.1353
		Bias	-0.0138	-0.0210	-0.0216	0.0064	0.1300
200	50	SD	0.0222	0.0228	0.0250	0.0277	0.0512
		RMSE	0.0261	0.0309	0.0330	0.0284	0.1397
		Bias	-0.0102	-0.0153	-0.0136	0.0198	0.1342
200	70	SD	0.0217	0.0220	0.0240	0.0272	0.0430
		RMSE	0.0240	0.0269	0.0276	0.0337	0.1409
		Bias	-0.0075	-0.0111	-0.0024	0.0301	0.1392
200	100	SD	0.0216	0.0219	0.0239	0.0269	0.0358
		RMSE	0.0229	0.0245	0.0240	0.0404	0.1437
		Bias	-0.0042	-0.0060	0.0049	0.0401	0.1427
200	200	SD	0.0214	0.0215	0.0230	0.0256	0.0270
		RMSE	0.0218	0.0223	0.0236	0.0476	0.1453
		Bias	-0.0037	-0.0055	0.0069	0.0418	0.1432
100	200	SD	0.0304	0.0307	0.0319	0.0344	0.0322
		RMSE	0.0306	0.0312	0.0326	0.0542	0.1467
		Bias	-0.0031	-0.0050	0.0085	0.0432	0.1436
70	200	SD	0.0355	0.0358	0.0377	0.0407	0.0356
		RMSE	0.0357	0.0362	0.0386	0.0594	0.1480
		Bias	-0.0023	-0.0041	0.0106	0.0445	0.1442
50	200	SD	0.0417	0.0421	0.0443	0.0480	0.0399
		RMSE	0.0418	0.0423	0.0455	0.0655	0.1496
		Bias	-0.0024	-0.0037	0.0133	0.0461	0.1443
30	200	SD	0.0537	0.0539	0.0570	0.0619	0.0478
		RMSE	0.0538	0.0540	0.0585	0.0772	0.1520

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E(\rho_i) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.8: Performance of the MG estimator under $\rho_F = 0.2$ and $i.i.d.$ idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0186	-0.0280	-0.0512	-0.0241	0.0272
200	30	SD	0.0227	0.0238	0.0245	0.0249	0.0667
		RMSE	0.0294	0.0368	0.0567	0.0347	0.0721
		Bias	-0.0117	-0.0172	-0.0219	0.0007	0.0395
200	50	SD	0.0220	0.0225	0.0235	0.0251	0.0525
		RMSE	0.0249	0.0283	0.0321	0.0251	0.0657
		Bias	-0.0088	-0.0128	-0.0150	0.0115	0.0433
200	70	SD	0.0216	0.0219	0.0227	0.0249	0.0452
		RMSE	0.0233	0.0253	0.0272	0.0274	0.0626
		Bias	-0.0065	-0.0093	-0.0060	0.0199	0.0483
200	100	SD	0.0216	0.0217	0.0227	0.0249	0.0387
		RMSE	0.0226	0.0236	0.0235	0.0319	0.0619
		Bias	-0.0038	-0.0052	0.0007	0.0283	0.0519
200	200	SD	0.0213	0.0214	0.0221	0.0241	0.0311
		RMSE	0.0216	0.0220	0.0221	0.0372	0.0605
		Bias	-0.0032	-0.0047	0.0027	0.0304	0.0525
100	200	SD	0.0304	0.0306	0.0310	0.0331	0.0373
		RMSE	0.0306	0.0309	0.0311	0.0450	0.0644
		Bias	-0.0026	-0.0041	0.0043	0.0320	0.0530
70	200	SD	0.0355	0.0357	0.0365	0.0391	0.0409
		RMSE	0.0356	0.0360	0.0368	0.0505	0.0669
		Bias	-0.0018	-0.0034	0.0065	0.0336	0.0536
50	200	SD	0.0416	0.0418	0.0429	0.0462	0.0455
		RMSE	0.0417	0.0419	0.0434	0.0571	0.0703
		Bias	-0.0021	-0.0035	0.0091	0.0358	0.0536
30	200	SD	0.0535	0.0536	0.0554	0.0594	0.0543
		RMSE	0.0536	0.0537	0.0562	0.0693	0.0763

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E(\rho_i) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.9: Performance of the MG estimator under $\rho_F = 0$ and $i.i.d.$ idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0171	-0.0246	-0.0473	-0.0247	-0.0339
200	30	SD	0.0227	0.0236	0.0245	0.0241	0.0677
		RMSE	0.0284	0.0341	0.0532	0.0345	0.0757
		Bias	-0.0108	-0.0152	-0.0218	-0.0029	-0.0217
200	50	SD	0.0220	0.0224	0.0230	0.0240	0.0534
		RMSE	0.0245	0.0271	0.0317	0.0242	0.0577
		Bias	-0.0081	-0.0113	-0.0153	0.0067	-0.0180
200	70	SD	0.0217	0.0218	0.0223	0.0238	0.0466
		RMSE	0.0232	0.0246	0.0271	0.0247	0.0499
		Bias	-0.0060	-0.0082	-0.0073	0.0143	-0.0128
200	100	SD	0.0217	0.0217	0.0223	0.0239	0.0403
		RMSE	0.0225	0.0232	0.0234	0.0279	0.0423
		Bias	-0.0034	-0.0046	-0.0008	0.0221	-0.0091
200	200	SD	0.0213	0.0214	0.0218	0.0234	0.0329
		RMSE	0.0216	0.0219	0.0218	0.0322	0.0342
		Bias	-0.0029	-0.0042	0.0011	0.0243	-0.0086
100	200	SD	0.0304	0.0305	0.0307	0.0323	0.0400
		RMSE	0.0305	0.0308	0.0307	0.0404	0.0409
		Bias	-0.0026	-0.0040	0.0025	0.0257	-0.0083
70	200	SD	0.0355	0.0356	0.0362	0.0382	0.0441
		RMSE	0.0356	0.0358	0.0363	0.0460	0.0448
		Bias	-0.0019	-0.0033	0.0046	0.0275	-0.0076
50	200	SD	0.0416	0.0417	0.0424	0.0451	0.0490
		RMSE	0.0416	0.0419	0.0427	0.0529	0.0496
		Bias	-0.0022	-0.0036	0.0072	0.0299	-0.0078
30	200	SD	0.0538	0.0539	0.0550	0.0584	0.0591
		RMSE	0.0538	0.0540	0.0554	0.0656	0.0596

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1, 0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E(\rho_i) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.10: Performance of the MG estimator under $\rho_F = 0.8$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0293	-0.0440	-0.0661	-0.0248	0.2112
200	30	SD	0.0230	0.0260	0.0282	0.0326	0.0729
		RMSE	0.0373	0.0511	0.0719	0.0409	0.2234
		Bias	-0.0172	-0.0256	-0.0168	0.0170	0.2218
200	50	SD	0.0215	0.0221	0.0293	0.0334	0.0552
		RMSE	0.0275	0.0338	0.0338	0.0375	0.2286
		Bias	-0.0121	-0.0178	-0.0057	0.0351	0.2282
200	70	SD	0.0211	0.0231	0.0277	0.0317	0.0437
		RMSE	0.0244	0.0292	0.0283	0.0473	0.2323
		Bias	-0.0083	-0.0124	0.0091	0.0486	0.2319
200	100	SD	0.0207	0.0210	0.0264	0.0295	0.0348
		RMSE	0.0223	0.0244	0.0279	0.0568	0.2345
		Bias	-0.0042	-0.0061	0.0170	0.0610	0.2359
200	200	SD	0.0201	0.0204	0.0240	0.0263	0.0210
		RMSE	0.0205	0.0213	0.0294	0.0664	0.2368
		Bias	-0.0040	-0.0055	0.0187	0.0624	0.2364
100	200	SD	0.0282	0.0288	0.0322	0.0350	0.0256
		RMSE	0.0285	0.0293	0.0372	0.0716	0.2378
		Bias	-0.0049	-0.0064	0.0190	0.0624	0.2357
70	200	SD	0.0339	0.0343	0.0386	0.0420	0.0296
		RMSE	0.0343	0.0349	0.0430	0.0752	0.2376
		Bias	-0.0048	-0.0049	0.0209	0.0638	0.2355
50	200	SD	0.0400	0.0419	0.0449	0.0489	0.0329
		RMSE	0.0403	0.0422	0.0495	0.0804	0.2378
		Bias	-0.0042	-0.0024	0.0267	0.0685	0.2355
30	200	SD	0.0519	0.0537	0.0576	0.0622	0.0414
		RMSE	0.0521	0.0537	0.0635	0.0925	0.2391

Notes: The dependent variables and the common factors are generated according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first 1,000 observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5,1.5]$ and $cov\left(\varepsilon_{it},\varepsilon_{jt}\right) = \tau^{|i-j|}\sqrt{\sigma_i\sigma_j}$ where $\tau = 0.5$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E\left(\rho_i\right) = 0.5$ being its true parameter.

Table 1.11: Performance of the MG estimator under $\rho_F = 0.5$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0206	-0.0336	-0.0554	-0.0203	0.1225
200	30	SD	0.0224	0.0235	0.0259	0.0278	0.0687
		RMSE	0.0304	0.0410	0.0612	0.0344	0.1404
		Bias	-0.0124	-0.0198	-0.0189	0.0103	0.1325
200	50	SD	0.0212	0.0220	0.0251	0.0277	0.0528
		RMSE	0.0245	0.0296	0.0314	0.0296	0.1426
		Bias	-0.0087	-0.0140	-0.0110	0.0235	0.1378
200	70	SD	0.0208	0.0213	0.0238	0.0269	0.0433
		RMSE	0.0226	0.0255	0.0263	0.0357	0.1444
		Bias	-0.0060	-0.0096	0.0001	0.0333	0.1413
200	100	SD	0.0203	0.0209	0.0231	0.0259	0.0362
		RMSE	0.0212	0.0230	0.0231	0.0422	0.1458
		Bias	-0.0032	-0.0048	0.0067	0.0430	0.1447
200	200	SD	0.0198	0.0202	0.0218	0.0242	0.0262
		RMSE	0.0201	0.0208	0.0228	0.0494	0.1471
		Bias	-0.0031	-0.0049	0.0082	0.0443	0.1449
100	200	SD	0.0282	0.0284	0.0301	0.0326	0.0307
		RMSE	0.0283	0.0289	0.0312	0.0550	0.1482
		Bias	-0.0040	-0.0053	0.0087	0.0444	0.1443
70	200	SD	0.0337	0.0344	0.0361	0.0391	0.0346
		RMSE	0.0340	0.0348	0.0371	0.0592	0.1484
		Bias	-0.0039	-0.0053	0.0106	0.0460	0.1442
50	200	SD	0.0394	0.0398	0.0419	0.0449	0.0378
		RMSE	0.0396	0.0402	0.0432	0.0643	0.1490
		Bias	-0.0034	-0.0043	0.0151	0.0501	0.1444
30	200	SD	0.0516	0.0521	0.0551	0.0588	0.0467
		RMSE	0.0517	0.0523	0.0572	0.0773	0.1517

Notes: The dependent variables and the common factors are generated according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first 1,000 observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1, 0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5, 1.5]$ and $cov\left(\varepsilon_{it}, \varepsilon_{jt}\right) = \tau^{|i-j|}\sqrt{\sigma_i \sigma_j}$ where $\tau = 0.5$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N} \hat{\rho}_i$ with $E\left(\rho_i\right) = 0.5$ being its true parameter.

Table 1.12: Performance of the MG estimator under $\rho_F = 0.2$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0172	-0.0275	-0.0492	-0.0215	0.0296
200	30	SD	0.0222	0.0234	0.0251	0.0250	0.0688
		RMSE	0.0281	0.0361	0.0552	0.0330	0.0749
		Bias	-0.0104	-0.0165	-0.0201	0.0033	0.0405
200	50	SD	0.0214	0.0220	0.0232	0.0248	0.0537
		RMSE	0.0238	0.0275	0.0307	0.0251	0.0672
		Bias	-0.0075	-0.0117	-0.0133	0.0142	0.0458
200	70	SD	0.0209	0.0213	0.0224	0.0244	0.0455
		RMSE	0.0222	0.0243	0.0261	0.0283	0.0646
		Bias	-0.0051	-0.0081	-0.0041	0.0224	0.0495
200	100	SD	0.0205	0.0209	0.0220	0.0240	0.0391
		RMSE	0.0211	0.0224	0.0223	0.0329	0.0631
		Bias	-0.0027	-0.0041	0.0023	0.0309	0.0535
200	200	SD	0.0201	0.0204	0.0211	0.0230	0.0301
		RMSE	0.0203	0.0208	0.0212	0.0385	0.0614
		Bias	-0.0024	-0.0038	0.0040	0.0326	0.0538
100	200	SD	0.0282	0.0286	0.0293	0.0313	0.0352
		RMSE	0.0283	0.0288	0.0296	0.0452	0.0643
		Bias	-0.0032	-0.0045	0.0046	0.0330	0.0533
70	200	SD	0.0337	0.0344	0.0353	0.0377	0.0392
		RMSE	0.0338	0.0347	0.0356	0.0501	0.0661
		Bias	-0.0032	-0.0045	0.0062	0.0346	0.0532
50	200	SD	0.0394	0.0401	0.0411	0.0434	0.0430
		RMSE	0.0395	0.0403	0.0416	0.0555	0.0684
		Bias	-0.0027	-0.0039	0.0102	0.0391	0.0537
30	200	SD	0.0515	0.0521	0.0539	0.0568	0.0524
		RMSE	0.0516	0.0523	0.0548	0.0690	0.0751

Notes: The dependent variables and the common factors are generated according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5,1.5]$ and $cov\,(\varepsilon_{it},\varepsilon_{jt}) = \tau^{|i-j|}\sqrt{\sigma_i\sigma_j}$ where $\tau = 0.5$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E\,(\rho_i) = 0.5$ being its true parameter.

Table 1.13: Performance of the MG estimator under $\rho_F = 0$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T		Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
		Bias	-0.0161	-0.0244	-0.0459	-0.0226	-0.0341
200	30	SD	0.0223	0.0234	0.0248	0.0241	0.0700
		RMSE	0.0275	0.0338	0.0522	0.0331	0.0779
		Bias	-0.0099	-0.0149	-0.0206	-0.0008	-0.0227
200	50	SD	0.0213	0.0218	0.0223	0.0235	0.0549
		RMSE	0.0235	0.0264	0.0304	0.0235	0.0594
		Bias	-0.0071	-0.0106	-0.0140	0.0090	-0.0172
200	70	SD	0.0209	0.0211	0.0218	0.0234	0.0469
		RMSE	0.0221	0.0237	0.0260	0.0250	0.0500
		Bias	-0.0049	-0.0074	-0.0059	0.0164	-0.0131
200	100	SD	0.0205	0.0207	0.0214	0.0229	0.0405
		RMSE	0.0211	0.0220	0.0222	0.0282	0.0426
		Bias	-0.0026	-0.0039	0.0003	0.0242	-0.0089
200	200	SD	0.0200	0.0202	0.0207	0.0222	0.0320
		RMSE	0.0202	0.0206	0.0207	0.0329	0.0332
		Bias	-0.0027	-0.0040	0.0018	0.0257	-0.0089
100	200	SD	0.0282	0.0285	0.0289	0.0306	0.0376
		RMSE	0.0283	0.0287	0.0290	0.0400	0.0386
		Bias	-0.0036	-0.0049	0.0022	0.0262	-0.0096
70	200	SD	0.0337	0.0342	0.0348	0.0369	0.0419
		RMSE	0.0339	0.0345	0.0349	0.0452	0.0430
		Bias	-0.0035	-0.0049	0.0039	0.0280	-0.0094
50	200	SD	0.0392	0.0398	0.0407	0.0427	0.0461
		RMSE	0.0393	0.0401	0.0409	0.0510	0.0471
		Bias	-0.0029	-0.0046	0.0075	0.0324	-0.0087
30	200	SD	0.0517	0.0523	0.0534	0.0562	0.0567
		RMSE	0.0518	0.0525	0.0540	0.0649	0.0573

Notes: The dependent variables and the common factors are generated according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first 1,000 observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5,1.5]$ and $cov\,(\varepsilon_{it},\varepsilon_{jt}) = \tau^{|i-j|}\sqrt{\sigma_i\sigma_j}$ where $\tau = 0.5$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E\,(\rho_i) = 0.5$ being its true parameter.

Table 1.14: MISE Performance of the individual coefficient estimator under $\rho_F = 0.8$ and *i.i.d.* idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0162	0.0181	0.0395	0.0326	0.0816
200	50	0.0088	0.0095	0.0186	0.0199	0.0796
200	70	0.0060	0.0064	0.0126	0.0162	0.0793
200	100	0.0041	0.0043	0.0088	0.0143	0.0796
200	200	0.0020	0.0020	0.0047	0.0123	0.0792

Table 1.15: MISE Performance of the individual coefficient estimator under $\rho_F = 0.5$ and *i.i.d.* idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0140	0.0151	0.0368	0.0309	0.0422
200	50	0.0078	0.0082	0.0179	0.0186	0.0369
200	70	0.0053	0.0055	0.0122	0.0146	0.0350
200	100	0.0036	0.0037	0.0081	0.0121	0.0341
200	200	0.0018	0.0018	0.0040	0.0094	0.0324

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i - \rho_i|^2\right)$. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.16: MISE Performance of the individual coefficient estimator under $\rho_F = 0.2$ and *i.i.d.* idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0139	0.0146	0.0352	0.0296	0.0245
200	50	0.0078	0.0081	0.0175	0.0175	0.0158
200	70	0.0054	0.0055	0.0119	0.0132	0.0124
200	100	0.0037	0.0038	0.0079	0.0104	0.0102
200	200	0.0018	0.0018	0.0037	0.0073	0.0074

Table 1.17: MISE Performance of the individual coefficient estimator under $\rho_F = 0$ and *i.i.d.* idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0144	0.0150	0.0345	0.0290	0.0257
200	50	0.0081	0.0083	0.0173	0.0169	0.0147
200	70	0.0056	0.0057	0.0118	0.0125	0.0102
200	100	0.0038	0.0039	0.0078	0.0095	0.0072
200	200	0.0019	0.0019	0.0037	0.0062	0.0035

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i - \rho_i|^2\right)$. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.18: MISE Performance of the individual coefficient estimator under $\rho_F = 0.8$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0161	0.0179	0.0395	0.0331	0.0843
200	50	0.0087	0.0093	0.0188	0.0207	0.0818
200	70	0.0060	0.0063	0.0129	0.0171	0.0817
200	100	0.0040	0.0042	0.0090	0.0152	0.0814
200	200	0.0020	0.0020	0.0048	0.0132	0.0818

Table 1.19: MISE Performance of the individual coefficient estimator under $\rho_F = 0.5$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0139	0.0149	0.0369	0.0313	0.0433
200	50	0.0076	0.0080	0.0178	0.0191	0.0381
200	70	0.0052	0.0054	0.0122	0.0152	0.0363
200	100	0.0036	0.0036	0.0082	0.0126	0.0348
200	200	0.0017	0.0017	0.0040	0.0100	0.0337

Notes: The dependent variables and the common factors are generated 1,000 times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first 1,000 observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5,1.5]$ and $cov\,(\varepsilon_{it},\varepsilon_{jt}) = \tau^{|i-j|}\sqrt{\sigma_i\sigma_j}$ where $\tau = 0.5$. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i - \rho_i|^2\right)$. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.20: MISE Performance of the individual coefficient estimator under $\rho_F = 0.2$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0139	0.0145	0.0355	0.0301	0.0249
200	50	0.0077	0.0079	0.0174	0.0179	0.0163
200	70	0.0053	0.0054	0.0120	0.0136	0.0128
200	100	0.0036	0.0037	0.0079	0.0108	0.0102
200	200	0.0018	0.0018	0.0038	0.0078	0.0077

Table 1.21: MISE Performance of the individual coefficient estimator under $\rho_F = 0$ and heteroskedastic and cross-sectionally correlated idiosyncratic errors

N	T	Infeasible	DHP+CSD	Dynamic CCE	CCE	OLS
200	30	0.0144	0.0148	0.0348	0.0294	0.0260
200	50	0.0080	0.0082	0.0173	0.0172	0.0149
200	70	0.0055	0.0056	0.0118	0.0128	0.0103
200	100	0.0038	0.0038	0.0078	0.0098	0.0071
200	200	0.0018	0.0018	0.0037	0.0066	0.0035

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$, and the loadings follow $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_i^2)$ with $\sigma_i^2 \sim i.i.d.\,\mathcal{U}[0.5,1.5]$ and $cov\,(\varepsilon_{it},\varepsilon_{jt}) = \tau^{|i-j|}\sqrt{\sigma_i\sigma_j}$ where $\tau = 0.5$. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i - \rho_i|^2\right)$. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.22: Performance of the MG estimator under no common factor structure and *i.i.d.* idiosyncratic shocks

N	T		OLS	DHP+CSD	Dynamic CCE	CCE
200	30	Bias	-0.0310	-0.0555	-0.0630	-0.0561
		SD	0.0222	0.0287	0.0251	0.0240
		RMSE	0.0381	0.0625	0.0678	0.0610
200	50	Bias	-0.0196	-0.0233	-0.0371	-0.0349
		SD	0.0210	0.0218	0.0212	0.0211
		RMSE	0.0288	0.0319	0.0428	0.0408
200	70	Bias	-0.0145	-0.0162	-0.0271	-0.0253
		SD	0.0205	0.0208	0.0204	0.0204
		RMSE	0.0251	0.0264	0.0339	0.0325
200	100	Bias	-0.0104	-0.0113	-0.0188	-0.0180
		SD	0.0205	0.0206	0.0203	0.0203
		RMSE	0.0230	0.0234	0.0277	0.0271
200	200	Bias	-0.0056	-0.0058	-0.0093	-0.0090
		SD	0.0201	0.0202	0.0200	0.0200
		RMSE	0.0209	0.0210	0.0220	0.0219
100	200	Bias	-0.0061	-0.0063	-0.0089	-0.0084
		SD	0.0286	0.0286	0.0284	0.0284
		RMSE	0.0292	0.0293	0.0298	0.0296
70	200	Bias	-0.0057	-0.0061	-0.0079	-0.0072
		SD	0.0352	0.0354	0.0348	0.0349
		RMSE	0.0357	0.0359	0.0357	0.0356
50	200	Bias	-0.0057	-0.0061	-0.0070	-0.0060
		SD	0.0422	0.0423	0.0419	0.0419
		RMSE	0.0426	0.0428	0.0425	0.0423
30	200	Bias	-0.0060	-0.0066	-0.0055	-0.0039
		SD	0.0538	0.0541	0.0535	0.0535
		RMSE	0.0541	0.0546	0.0538	0.0536

Notes: The dependent variables are generated $1,000$ times according to $y_{it} = \rho_i y_{i,t-1} + \varepsilon_{it}$ with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The bias, standard deviation, and RMSE are those of the MG estimator $\hat{\rho} = \frac{1}{N}\sum_{i=1}^{N}\hat{\rho}_i$ with $E(\rho_i) = 0.5$ being its true parameter. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.23: MISE Performance of the individual coefficient estimator under no common factor structure and *i.i.d.* idiosyncratic shocks

N	T	OLS	DHP+CSD	Dynamic CCE	CCE
200	30	0.0257	0.0317	0.0370	0.0314
200	50	0.0147	0.0157	0.0183	0.0167
200	70	0.0102	0.0107	0.0125	0.0113
200	100	0.0070	0.0072	0.0080	0.0075
200	200	0.0034	0.0035	0.0037	0.0035

Notes: The dependent variables are generated $1,000$ times according to $y_{it} = \rho_i y_{i,t-1} + \varepsilon_{it}$ with $u_t \sim i.i.d.\,\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The AR coefficients are generated as $\rho_i \sim i.i.d.\,\mathcal{U}[0,1)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The MISE performance is computed by $\frac{1}{B}\sum_{k=1}^{B}\left(\frac{1}{N}\sum_{i=1}^{N}|\hat{\rho}_i - \rho_i|^2\right)$. The results for the CCE and the Dynamic CCE estimators are based on the methods proposed in Pesaran (2006) and Chudik and Pesaran (2013), respectively.

Table 1.24: Size of the significance test with $\rho_F = 0.8$

$N \setminus T$	30	50	70	100	200
30	0.094	0.066	0.057	0.062	0.060
50	0.085	0.065	0.056	0.058	0.057
70	0.081	0.062	0.050	0.057	0.052
100	0.080	0.063	0.053	0.055	0.060
200	0.085	0.062	0.054	0.051	0.058

Table 1.25: Size of the significance test with $\rho_F = 0.5$

$N \setminus T$	30	50	70	100	200
30	0.081	0.072	0.068	0.073	0.061
50	0.083	0.074	0.067	0.070	0.059
70	0.079	0.073	0.067	0.065	0.058
100	0.069	0.066	0.064	0.064	0.061
200	0.070	0.063	0.060	0.062	0.057

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The loadings are generated as $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The test statistic is constructed by $\sqrt{T}\hat{\rho}_i\hat{\Omega}_i^{-\frac{1}{2}}$ where $\hat{\Omega}_i$ is the sample analog of the asymptotic variance in Theorem 3. The null hypothesis is $H_0:\ \rho_i = 0$.

Table 1.26: Size of the significance test with $\rho_F = 0.2$

$N \setminus T$	30	50	70	100	200
30	0.079	0.075	0.068	0.061	0.065
50	0.083	0.073	0.068	0.063	0.066
70	0.076	0.064	0.065	0.062	0.061
100	0.072	0.066	0.061	0.062	0.062
200	0.072	0.063	0.059	0.064	0.057

Table 1.27: Size of the significance test with $\rho_F = 0$

$N \setminus T$	30	50	70	100	200
30	0.077	0.076	0.071	0.062	0.067
50	0.077	0.070	0.067	0.062	0.059
70	0.073	0.059	0.062	0.062	0.058
100	0.074	0.057	0.061	0.059	0.059
200	0.067	0.059	0.060	0.061	0.058

Notes: The dependent variables and the common factors are generated $1,000$ times according to (1.13) and (1.14), respectively, with $u_t \sim i.i.d.\mathcal{N}(0,1)$. The first $1,000$ observations of the generated data are burned-in. The loadings are generated as $\lambda_i \sim i.i.d.\,\mathcal{N}(1,0.5)$. The idiosyncratic errors are generated from $\varepsilon_{it} \sim i.i.d.\,\mathcal{N}(0,1)$. The test statistic is constructed by $\sqrt{T}\hat{\rho}_i\hat{\Omega}_i^{-\frac{1}{2}}$ where $\hat{\Omega}_i$ is the sample analog of the asymptotic variance in Theorem 3. The null hypothesis is $H_0:\ \rho_i = 0$.

1.7.2 Proofs

Lemma A.1: Under Assumptions A-D,

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

(ii) $\frac{1}{\sqrt{T}}\left\|X_i' M_{\hat{F}}\right\| = O_p(1)$

(iii) $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{\tau}_{ts}^2 = O(1)$ where $\bar{\tau}_{ts}$ 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,

$$
\begin{aligned}
E\left|\frac{1}{NT^2}\sum_{i=1}^{N}\|X_i\|^4\right| &= E\left[\frac{1}{NT^2}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\|\mathbf{x}_{it}\|^2\right)^2\right] \\
&= \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\|\mathbf{x}_{it}\|^2\|\mathbf{x}_{is}\|^2\right) \\
&\le \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(E\|\mathbf{x}_{it}\|^4 E\|\mathbf{x}_{is}\|^4\right)^{\frac{1}{2}}
\end{aligned}
$$

The entire expression above is bounded because $\mathbf{x}_{it}$ has a bounded 4th moment for all i and t by Assumption B(i).

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

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\left\|X_i' M_{\hat{F}}\right\| &= \frac{1}{\sqrt{T}}\left\|X_i'\left(I_T - \frac{1}{T}\hat{F}\hat{F}'\right)\right\| \\
&\le \frac{1}{\sqrt{T}}\left(\|X_i\| + \frac{1}{T}\left\|X_i\hat{F}\hat{F}'\right\|\right) \\
&\le \frac{1}{\sqrt{T}}\|X_i\| + \left(\frac{1}{\sqrt{T}}\|X_i\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)
\end{aligned}
$$

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 $\frac{1}{\sqrt{T}}\left\|\hat{F}\right\| = \sqrt{r}$ by the normalizing assumption. Combining these results, we obtain the desired result.

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

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{\tau}_{ts}^{2} &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\bar{\tau}_{ts}\bar{\tau}_{ts}\right) \\ &\leq M\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\bar{\tau}_{ts}\right)\end{aligned}$$

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

Lemma A.2: Define $C_{NT} = \min\left\{\sqrt{N}, \sqrt{T}\right\}$. Let $\mathcal{F}$ be the space of all F that satisfy $\frac{F'F}{T} = I_r$. Under Assumptions A-D,

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

(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) $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t\varepsilon_{it}\right\|^2 = O_p\left(\frac{1}{C_{NT}}\right) \qquad \forall F \in \mathcal{F}$

Proof:

(i) From the definition of norms,

$$\begin{aligned}E\left|\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_{it}\varepsilon_{it}\right\|^2\right| &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\mathbf{x}_{it}'\mathbf{x}_{is}\varepsilon_{it}\varepsilon_{is}\right) \\ &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\|\mathbf{x}_{it}\|\,\|\mathbf{x}_{is}\|\right)\right| \times \left|E\left(\varepsilon_{it}\varepsilon_{is}\right)\right| \\ &\leq M\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{is}\right)\right|\right\} \qquad (1.15)\end{aligned}$$

where the second statement utilized the independence assumption between the regressors and the idiosyncratic errors. The last statement follows from the Cauchy-Schwarz inequality, $|E(\|\mathbf{x}_{it}\| \|\mathbf{x}_{is}\|)| \leq \left(E\|\mathbf{x}_{it}\|^2 E\|\mathbf{x}_{is}\|^2\right)^{\frac{1}{2}}$, and the boundedness assumption in B(i). That is, the boundedness of (1.15) is determined by the following term:

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

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 (1.15) is bounded.

(ii) Using the properties of norms,

$$\begin{aligned}E\left|\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\varepsilon_{it}\right\|^2\right| &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(F_t^{0\prime}F_s^0\varepsilon_{it}\varepsilon_{is}\right) \qquad (1.16)\\ &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\left\|F_t^0\right\|\left\|F_s^0\right\|\right)\right| \times |E(\varepsilon_{it}\varepsilon_{is})|\\ &\leq M\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E(\varepsilon_{it}\varepsilon_{is})|\right\}\end{aligned}$$

where we used the assumption of independence between the true factors and idiosyncratic errors to separate out the expectation. Note that $\left|E\left(\left\|F_t^0\right\|\left\|F_s^0\right\|\right)\right| \leq \left(E\left\|F_t^0\right\|^2 E\left\|F_s^0\right\|^2\right)^{\frac{1}{2}}$ by the Cauchy-Schwarz inequality, where the right hand side is bounded by Assumption A(i-1). Combined with the result in part (i), this leads to the conclusion that (1.16) is bounded.

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

Lemma A.3: Under Assumptions A-D,

(i) $\sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left(\frac{X_i' M_F \varepsilon_i}{T} \right) \right\| = o_p(1)$

(ii) $\sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \lambda_i^{0\prime} F^{0\prime} M_F \varepsilon_i \right\| = o_p(1)$

(iii) $\sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' P_F \varepsilon_i \right\| = o_p(1)$

(iv) $\sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i' P_{F^0} \varepsilon_i \right\| = o_p(1)$

Proof:

Recall that $\mathcal{B}$ is a compact set around the true parameter $\beta_{0,i}$ defined in Assumption A(iii) and that $\mathcal{F}$ is defined in Lemma A.2.

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

$$\sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left(\frac{X_i' M_F \varepsilon_i}{T} \right) \right\| \leq \sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left(\frac{X_i' \varepsilon_i}{T} \right) \right\| + \sup_{\beta_i \in \mathcal{B}, F \in \mathcal{F}} \left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left(\frac{X_i' P_F \varepsilon_i}{T} \right) \right\| \tag{1.17}$$

Regarding the first term of (1.17), by the Cauchy-Schwarz inequality, we have

$$\left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left(\frac{X_i' \varepsilon_i}{T} \right) \right\| \leq \frac{1}{\sqrt{T}} \left(\frac{1}{N} \sum_{i=1}^{N} \|\beta_i - \beta_{0,i}\|^2 \right)^{\frac{1}{2}} \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \mathbf{x}_{it} \varepsilon_{it} \right\|^2 \right)^{\frac{1}{2}} \tag{1.18}$$

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

Next, by applying the Cauchy-Schwarz inequality twice to the second term of (1.17),

$$\left\| \frac{1}{N} \sum_{i=1}^{N} (\beta_i - \beta_{0,i})' \left(\frac{X_i' P_F \varepsilon_i}{T} \right) \right\| \leq \left(\frac{1}{N} \sum_{i=1}^{N} \|\beta_i - \beta_{0,i}\|^4 \right)^{\frac{1}{4}} \left(\frac{1}{NT^2} \sum_{i=1}^{N} \|X_i\|^4 \right)^{\frac{1}{4}} \times \left(\frac{1}{\sqrt{T}} \|F\| \right) \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \frac{1}{T} \sum_{t=1}^{T} F_t \varepsilon_{it} \right\|^2 \right)^{\frac{1}{2}} \tag{1.19}$$

Note that the first three components of the product are all bounded. The first term $\frac{1}{N}\sum_{i=1}^{N}\|\beta_i-\beta_{0,i}\|^4$ is bounded because we consider β_i on a bounded set $\mathcal{B}$ around the true $\beta_{0,i}$, and the second term $\frac{1}{NT^2}\sum_{i=1}^{N}\|X_i\|^4$ is bounded by Lemma A.1(i). Finally, $\frac{1}{\sqrt{T}}\|F\|=\sqrt{r}$ for all $F\in\mathcal{F}$ by the definition of the set $\mathcal{F}$. The last component of (1.19), $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}F_t\varepsilon_{it}\right\|^2$, is of order $O_p\left(\frac{1}{C_{NT}}\right)$ by Lemma A.2(iii). From these results, $\sup_{\beta_i\in\mathcal{B},F\in\mathcal{F}}\left\|\frac{1}{N}\sum_{i=1}^{N}(\beta_i-\beta_{0,i})'\left(\frac{X_i'P_F\varepsilon_i}{T}\right)\right\|=o_p(1)$ as well.

Combining above results that the two terms in (1.17) 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),

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

by the Cauchy-Schwarz inequality. But $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t^0\varepsilon_{it}\right\|^2$ is bounded by Lemma A.2(ii), and $\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\|$ is also bounded by Assumption A(i-2). Moreover, this result holds uniformly as the expression involves neither β_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

$$S_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right) \equiv \frac{1}{NT}\sum_{i=1}^{N}(Y_i - X_i\beta_i)' M_F (Y_i - X_i\beta_i) - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' M_{F^0}\varepsilon_i \quad (1.20)$$

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

$$\begin{aligned} S_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right) = &\frac{1}{N}\sum_{i=1}^{N}(\beta_i - \beta_{0,i})'\left(\frac{X_i' M_F X_i}{T}\right)(\beta_i - \beta_{0,i}) \\ &+ \frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime} F^{0\prime} M_F F^0 \lambda_i^0 \quad (1.21) \\ &+ 2\frac{1}{NT}\sum_{i=1}^{N}(\beta_i - \beta_{0,i})' X_i' M_F F^0\lambda_i^0 + 2\frac{1}{N}\sum_{i=1}^{N}(\beta_i - \beta_{0,i})'\left(\frac{X_i' M_F \varepsilon_i}{T}\right) \\ &+ 2\frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime} F^{0\prime} M_F\varepsilon_i + \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' P_F\varepsilon_i - \frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i' P_{F^0}\varepsilon_i, \quad (1.22) \end{aligned}$$

Note that Lemma A.3 implies that the last four terms in (1.22) are $o_p(1)$ uniformly over the entire space of β_i and F. By defining the sum of the first three terms as an auxiliary objective function

$$\begin{aligned} \tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right) &= \frac{1}{N}\sum_{i=1}^{N}(\beta_i - \beta_{0,i})'\left(\frac{X_i' M_F X_i}{T}\right)(\beta_i - \beta_{0,i}) \\ &\quad + \frac{1}{NT}\sum_{i=1}^{N}\lambda_i^{0\prime} F^{0\prime} M_F F^0\lambda_i^0 + 2\frac{1}{NT}\sum_{i=1}^{N}(\beta_i - \beta_{0,i})' X_i' M_F F^0\lambda_i^0 \\ &= \frac{1}{N}\sum_{i=1}^{N}\left\{(\beta_i - \beta_{0,i})'\left(\frac{X_i' M_F X_i}{T}\right)(\beta_i - \beta_{0,i})\right. \\ &\quad \left. + \lambda_i^{0\prime}\left(\frac{F^{0\prime} M_F F^0}{T}\right)\lambda_i^0 + 2(\beta_i - \beta_{0,i})'\left(\frac{X_i' M_F F^0}{T}\right)\lambda_i^0\right\} \quad (1.23) \end{aligned}$$

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, $\tilde{S}_{NT}\left(\{\beta_{0,i}\}_{i=1}^N, F^0H\right)$ 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 (1.23) as

$$\begin{aligned}\lambda_i^{0\prime}\left(\frac{F^{0\prime}M_F F^0}{T}\right)\lambda_i^0 &= \frac{1}{T}\times tr\left[F^{0\prime}M_F F^0\lambda_i^0\lambda_i^{0\prime}\right]\\ &= tr\left[\left(F^{0\prime}M_F\right)\left(\frac{I_T}{T}\right)\left(M_F F^0\right)\left(\lambda_i^0\lambda_i^{0\prime}\right)\right]\\ &= vec\left(M_F F^0\right)'\left[\left(\lambda_i^0\lambda_i^{0\prime}\right)\otimes\frac{I_T}{T}\right]vec\left(M_F F^0\right)\\ &= \eta' B_i\eta\end{aligned}$$

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

By the same argument, the third term inside the curly braces of (1.23) becomes

$$\begin{aligned}\left(\beta_i-\beta_{0,i}\right)'\left(\frac{X_i'M_F F^0}{T}\right)\lambda_i^0 &= \frac{1}{T}\times tr\left[\left(\beta_i-\beta_{0,i}\right)'X_i'M_F F^0\lambda_i^0\right]\\ &= \frac{1}{T}\times tr\left[\left(\beta_i-\beta_{0,i}\right)'\left(X_i'M_F\right)\left(M_F F^0\right)\left(\lambda_i^0\right)\right]\\ &= \left(\beta_i-\beta_{0,i}\right)'\left[\lambda_i^{0\prime}\otimes\frac{1}{T}\left(X_i'M_F\right)\right]vec\left(M_F F^0\right)\\ &= \left(\beta_i-\beta_{0,i}\right)'C_i\eta\end{aligned}$$

where we define $C_i = \lambda_i^{0\prime}\otimes\frac{1}{T}\left(X_i'M_F\right)$.

Using the two expressions derived above,

$$\tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right) = \frac{1}{N}\sum_{i=1}^{N}\left\{(\beta_i - \beta_{0,i})' A_i (\beta_i - \beta_{0,i}) + \eta' B_i \eta + 2(\beta_i - \beta_{0,i})' C_i \eta\right\}$$

where we define $A_i = \frac{X_i' M_F X_i}{T}$. Now, by completing the squares, we obtain

$$\tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right) = \eta' D \eta + \frac{1}{N}\sum_{i=1}^{N}\theta_i' A_i \theta_i \tag{1.24}$$

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

Note that the weighting matrix D is positive definite by Assumption B(iii) and that A_i are at least positive semi-definite by construction. Therefore, $\tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right) \geq 0$ due to the quadratic form of the auxiliary objective function (1.24). It is easy to see that $\tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right)$ attains a minimum of zero at $\left(\{\beta_{0,i}\}_{i=1}^{N}, F^o H\right)$. We now show that, for any $\left(\{\beta_i\}_{i=1}^{N}, F\right)$ to be a minimizer, it should be equal to $\left(\{\beta_{0,i}\}_{i=1}^{N}, F^o H\right)$. If $F \neq F^o H$, it follows that $\eta \neq 0$ and that $\tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, F\right)$ cannot achieve its minimum due to the quadratic form with positive definite D. That is, $F = F^o H$ for any minimizer. Given $F = F^o H$, we have $\theta_i = \beta_i - \beta_{0,i}$ and $A_i = \frac{X_i' M_{F^0} X_i}{T}$ where A_i is now positive definite by Assumption B(ii). Therefore, it should be $\theta_i = 0$, or equivalently, $\beta_i = \beta_{0,i}$ for $\tilde{S}_{NT}$ to achieve its minimum of zero. In other words, $\left(\{\beta_{0,i}\}_{i=1}^{N}, F^o H\right)$ is the unique minimizer of the auxiliary objective function. Thus, it follows that individual coefficient estimators are consistent for their corresponding true parameters. 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 $\tilde{S}_{NT}\left(\{\beta_i\}_{i=1}^{N}, \hat{F}\right) = o_p(1)$ as well as $\frac{1}{T} F^{0\prime} M_{\hat{F}} F^0 = o_p(1)$. Then, it is easy to see that

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

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

Proposition A.1: Under Assumptions A-D,

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

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

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

$$\left[\frac{1}{NT}\sum_{i=1}^{N}\left(Y_i-X_i\hat{\beta}_i\right)\left(Y_i-X_i\hat{\beta}_i\right)'\right]\hat{F}=\hat{F}V_{NT} \tag{1.25}$$

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:

$$\begin{aligned}\hat{F}V_{NT}=&\frac{1}{NT}\sum_{i=1}^{N}X_i\left(\beta_{0,i}-\hat{\beta}_i\right)\left(\beta_{0,i}-\hat{\beta}_i\right)'X_i'\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}X_i\left(\beta_{0,i}-\hat{\beta}_i\right)\lambda_i^{0\prime}F^{0\prime}\hat{F}\\&+\frac{1}{NT}\sum_{i=1}^{N}X_i\left(\beta_{0,i}-\hat{\beta}_i\right)\varepsilon_i'\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\left(\beta_{0,i}-\hat{\beta}_i\right)'X_i'\hat{F}\\&+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\left(\beta_{0,i}-\hat{\beta}_i\right)'X_i'\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\varepsilon_i'\hat{F}\\&+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\lambda_i^{0\prime}F^{0\prime}\hat{F}+\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\hat{F}\\&+\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\lambda_i^{0\prime}F^{0\prime}\hat{F}\\=&\,I1+I2+I3+I4+I5+I6+I7+I8+F^o\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right)\end{aligned} \tag{1.26}$$

where we labeled the first eight terms of (1.26) as $I1$ to $I8$, respectively. By rearranging terms and rescaling both sides by $\frac{1}{\sqrt{T}}$, we have

$$\begin{aligned}\frac{1}{\sqrt{T}}\left\|\hat{F}V_{NT}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{N}\right)^{-1}-F^{0}\right\| &= \frac{1}{\sqrt{T}}\left\|(I1+I2+\cdots+I8)\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{N}\right)^{-1}\right\| \\ &\leq \left(\frac{1}{\sqrt{T}}\|I1\|+\frac{1}{\sqrt{T}}\|I2\|+\cdots+\frac{1}{\sqrt{T}}\|I8\|\right) \\ &\quad \times \left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{N}\right)^{-1}\right\| \end{aligned} \tag{1.27}$$

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

$$\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|^{2} = tr\left[\left(\frac{F^{0\prime}F^{0}}{T}-\frac{F^{0\prime}M_{\hat{F}}F^{0}}{T}\right)^{-1}\right] \tag{1.28}$$

Note that $\frac{F^{0\prime}M_{\hat{F}}F^{0}}{T}=o_p(1)$ as already discussed above and that the existence and boundedness of $\left(\frac{F^{0\prime}F^{0}}{T}\right)^{-1}$ are guaranteed by Assumption A(i-2). Thus, we conclude that $\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|=O_p(1)$. As both $\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|$ and $\left\|\left(\frac{\Lambda^{0\prime}\Lambda^{0}}{N}\right)^{-1}\right\|$ are bounded, the order of (1.27) is determined by the sum of the eight terms.

Now we examine the stochastic order of the eight terms from $\frac{1}{\sqrt{T}}\|I1\|$ to $\frac{1}{\sqrt{T}}\|I8\|$. For the first term, using Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{\sqrt{T}}\|I1\| &= \frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\sum_{i=1}^{N}X_i\left(\beta_{0,i}-\hat{\beta}_i\right)\left(\beta_{0,i}-\hat{\beta}_i\right)'X_i'\hat{F}\right\| \\ &\leq \left(\frac{1}{NT^2}\sum_{i=1}^{N}\|X_i\|^4\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\beta}_i-\beta_{0,i}\right\|^4\right)^{\frac{1}{2}}\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\end{aligned}$$

Note that the model considered in this chapter is symmetric with respect to every cross-sectional unit. Furthermore, the proposed estimator processes all individual data in a symmetric manner as well. That is, each $\left\|\hat{\beta}_i-\beta_{0,i}\right\|$ 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 $\left\|\hat{\beta}_i - \beta_{0,i}\right\|$ by $O_p\left(B_{NT}\right)$, the order of $\left(\frac{1}{N}\sum_{i=1}^N \left\|\hat{\beta}_i - \beta_{0,i}\right\|^4\right)^{\frac{1}{2}}$ 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 $\frac{1}{\sqrt{T}}\left\|I1\right\| = O_p\left(B_{NT}^2\right)$.

Next, for $\frac{1}{\sqrt{T}}\left\|I2\right\|$

$$\begin{aligned}
\frac{1}{\sqrt{T}}\left\|I2\right\| &= \frac{1}{\sqrt{T}}\left\|\frac{1}{NT}\sum_{i=1}^N X_i\left(\beta_{0,i} - \hat{\beta}_i\right)\lambda_i^{0\prime}F^{0\prime}\hat{F}\right\| \\
&\leq \left(\frac{1}{NT}\sum_{i=1}^N \left\|X_i\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^N \left\|\hat{\beta}_i - \beta_{0,i}\right\|^4\right)^{\frac{1}{4}}\left(\frac{1}{N}\sum_{i=1}^N \left\|\lambda_i^0\right\|^4\right)^{\frac{1}{4}} \\
&\qquad \times\left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)
\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality. As in the case of $\frac{1}{\sqrt{T}}\left\|I1\right\|$, it follows that $\left(\frac{1}{N}\sum_{i=1}^N \left\|\hat{\beta}_i - \beta_{0,i}\right\|^4\right)^{\frac{1}{4}} = O_p\left(B_{NT}\right)$ while all other terms are simply $O_p(1)$. Therefore, we obtain $\frac{1}{\sqrt{T}}\left\|I2\right\| = O_p\left(B_{NT}\right)$. By the same argument, we can show that $\frac{1}{\sqrt{T}}\left\|I3\right\|$, $\frac{1}{\sqrt{T}}\left\|I4\right\|$ and $\frac{1}{\sqrt{T}}\left\|I5\right\|$ are also $O_p\left(B_{NT}\right)$. The results for $\frac{1}{\sqrt{T}}\left\|I6\right\|$, $\frac{1}{\sqrt{T}}\left\|I7\right\|$ and $\frac{1}{\sqrt{T}}\left\|I8\right\|$ are already provided in Theorem 1 of Bai and Ng (2002). All three terms are shown to be of order $O_p\left(\frac{1}{C_{NT}}\right)$. These terms do not involve any individual estimator $\left\|\hat{\beta}_i - \beta_{0,i}\right\|$ that the results still remain true in this setup. Summing over the eight terms in (1.27), we can show that $\frac{1}{\sqrt{T}}\left\|\hat{F}V_{NT}\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} - F^0\right\| = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}}\right)$.

Lemma A.4: Under Assumptions A-D,

(i) $\frac{1}{T}F^{0\prime}\left(I1 + \cdots + I8\right) = o_p(1)$

(ii) $\left\|V_{NT}\right\| = O_p(1)$

(iii) V_{NT} is invertible and $\left\|V_{NT}^{-1}\right\| = O_p(1)$

(iv) $\|H\| = O_p(1)$ where $H = \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)\left(\frac{F^{0\prime}\hat{F}}{T}\right)V_{NT}^{-1}$

(v) H is invertible and $\left\|H^{-1}\right\| = O_p(1)$

Proof:

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

$$\begin{aligned}\left\|\frac{1}{T}F^{0\prime}\left(I1+\cdots+I8\right)\right\| &\leq \left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|(I1+\cdots+I8)\right\|\right)\\ &\leq \left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|I1\right\|+\cdots+\frac{1}{\sqrt{T}}\left\|I8\right\|\right)\end{aligned}$$

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 $O_p\left(B_{NT}\right)+O_p\left(\frac{1}{C_{NT}}\right)$ in Proposition A.1. Therefore, we conclude that $\left\|\frac{1}{T}F^{0\prime}\left(I1+\cdots+I8\right)\right\| = o_p(1)$.

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

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

Proof:

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

$$\begin{aligned}\frac{1}{\sqrt{T}}\left\|\hat{F}-F^0H\right\| &= \frac{1}{\sqrt{T}}\left\|\left(\hat{F}H^{-1}-F^0\right)H\right\|\\ &\leq \frac{1}{\sqrt{T}}\left\|\hat{F}H^{-1}-F^0\right\|\left\|H\right\|\end{aligned}$$

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 $O_p(B_{NT}) + O_p\left(\frac{1}{C_{NT}}\right)$. The last term, $\|H\|$, 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) $\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}F_t^{0\prime}\right\| = O_p(1)$

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

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

Proof:

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

$$E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}F^{0\prime}\varepsilon_i\lambda_i^{0\prime}\right\|^2 = \frac{1}{NT}\times E\left(tr\left[\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\lambda_i^{0}\varepsilon_{it}F_t^{0\prime}F_s^{0}\varepsilon_{js}\lambda_j^{0\prime}\right]\right) \quad (1.29)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\lambda_j^{0\prime}\lambda_i^{0}F_t^{0\prime}F_s^{0}\varepsilon_{it}\varepsilon_{js}\right)$$

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

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\lambda_j^{0\prime}\lambda_i^{0}F_t^{0\prime}F_s^{0}\right)\right|\times\left|E\left(\varepsilon_{it}\varepsilon_{js}\right)\right| \quad (1.30)$$

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

$$\left|E\left(\lambda_j^{0\prime}\lambda_i^{0}F_t^{0\prime}F_s^{0}\right)\right| \le \left(E\left\|\lambda_j^{0}\right\|^4 E\left\|\lambda_i^{0}\right\|^4 E\left\|F_t^{0}\right\|^4 E\left\|F_s^{0}\right\|^4\right)^{\frac{1}{4}} \quad (1.31)$$

by repetitive application of the Cauchy-Schwarz inequality. All four terms on the right hand side of (1.31) are bounded by Assumption A(i-1) and A(ii-1). Thus, the order of (1.30) is determined by $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\varepsilon_{it}\varepsilon_{js}\right)\right|$, 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) $\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\lambda_i^{0\prime}\right\| = O_p\left(\frac{1}{C_{NT}\sqrt{N}}\right)$

(iv) $\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i\varepsilon_i'F^0\right\| = O_p\left(\frac{1}{C_{NT}\sqrt{T}}\right)$

(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),

$$\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}E\left(\varepsilon_{kt}\varepsilon_{it}\right)\right|^2 \leq \sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\bar{\sigma}_{ki}\right|^2 \tag{1.32}$$

$$= \sum_{i=1}^{N}\bar{\sigma}_{ki}^2 \tag{1.33}$$

where the last equality holds because $\bar{\sigma}_{ki}$ in (1.32) does not depend on the subscript t. Note that (1.33) 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

$$E\left|\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\varepsilon_{kt}\varepsilon_{it}-E\left(\varepsilon_{kt}\varepsilon_{it}\right)\right]\right|^2\right| = E\left(\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\zeta_{it}\right|^2\right) \tag{1.34}$$

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

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

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\zeta_{it}\zeta_{js}\right)\right|$$

which takes additional summation over subscript j. But, from the definition of ζ_{it},

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\zeta_{it}\zeta_{js}\right)\right| = \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|cov\left(\varepsilon_{kt}\varepsilon_{it},\varepsilon_{ks}\varepsilon_{js}\right)\right|$$

which is bounded by Assumption D(iv-2). Thus, (1.34) is bounded.

(iii) By adding and subtracting $E\left(\varepsilon_{it}\varepsilon_{kt}\right)$,

$$\begin{aligned}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}\varepsilon_{kt}\lambda_i^{0\prime}\right\| &= \frac{1}{NT}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{it}\varepsilon_{kt}-E\left(\varepsilon_{it}\varepsilon_{kt}\right)+E\left(\varepsilon_{it}\varepsilon_{kt}\right)\right]\lambda_i^{0\prime}\right\| \\ &\le \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{it}\varepsilon_{kt}-E\left(\varepsilon_{it}\varepsilon_{kt}\right)\right]\lambda_i^{0\prime}\right\| \\ &\quad + \frac{1}{N}\left\|\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left(\varepsilon_{it}\varepsilon_{kt}\right)\lambda_i^{0\prime}\right\| \qquad (1.35)\end{aligned}$$

where the last inequality follows from the triangle inequality. We examine the first term of (1.35):

$$\begin{aligned}E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{it}\varepsilon_{kt}-E\left(\varepsilon_{it}\varepsilon_{kt}\right)\right]\lambda_i^{0\prime}\right\|^2 &= E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\zeta_{it}\lambda_i^{0\prime}\right\|^2 \qquad (1.36) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left(\lambda_j^{0\prime}\lambda_i^{0}\zeta_{it}\zeta_{js}\right) \\ &\le \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\lambda_j^{0\prime}\lambda_i^{0}\right)\right|\times\left|E\left(\zeta_{it}\zeta_{js}\right)\right|\end{aligned}$$

where the first statement follows from the use of ζ_{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. Note that $\left|E\left(\lambda_j^{0\prime}\lambda_i^{0}\right)\right| \le \left(E\left\|\lambda_j^{0}\right\|^2 E\left\|\lambda_i^{0}\right\|^2\right)^{\frac{1}{2}}$ 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 (1.36) is bounded by $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left|E\left(\zeta_{it}\zeta_{js}\right)\right|$, which is, in turn, bounded by Assumption D(iv-2). In sum, (1.36) is bounded.

Now, we examine the second term of (1.35)

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} E\left(\varepsilon_{it}\varepsilon_{kt}\right) \lambda_i^{0\prime} \right\| \leq \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} \left| E\left(\varepsilon_{it}\varepsilon_{kt}\right) \right| \right) \left\| \lambda_i^0 \right\| \qquad (1.37)$$

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

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

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 (1.37) 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 $\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}\varepsilon_{kt}\lambda_i^{0\prime} \right\| = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N}\right) = O_p\left(\frac{1}{C_{NT}\sqrt{N}}\right)$.

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

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

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

$$+ \frac{1}{T} \left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_k' E\left(\varepsilon_i\varepsilon_i'\right) F^0 \right\| \qquad (1.38)$$

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

First, using the Cauchy-Schwarz inequality,

$$\begin{aligned}\left\| \frac{1}{\sqrt{NT^3}} \sum_{i=1}^{N} \varepsilon_k' \left[\varepsilon_i \varepsilon_i' - E\left(\varepsilon_i \varepsilon_i'\right)\right] F^0 \right\| &= \frac{1}{\sqrt{NT^3}} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{kt} \left[\varepsilon_{it}\varepsilon_{is} - E\left(\varepsilon_{it}\varepsilon_{is}\right)\right] F_s^{0\prime} \right\| \\ &\leq \left(\frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{s=1}^{T} \left[\varepsilon_{it}\varepsilon_{is} - E\left(\varepsilon_{it}\varepsilon_{is}\right)\right] F_s^{0\prime} \right\|^2 \right)^{\frac{1}{2}} \\ &\quad \times \left(\frac{1}{T} \sum_{t=1}^{T} \left\|\varepsilon_{kt}\right\|^2 \right)^{\frac{1}{2}} \qquad (1.39)\end{aligned}$$

Note that the last term of (1.39) is bounded by Assumption D(i). To examine the order of the first term of (1.39), consider

$$\begin{aligned}E\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{s=1}^{T} \left[\varepsilon_{iu}\varepsilon_{is} - E\left(\varepsilon_{iu}\varepsilon_{is}\right)\right] F_s^{0\prime} \right\|^2 &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} E\left(F_t^{0\prime} F_s^0 \zeta_{is}\zeta_{jt}\right) \\ &\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{s=1}^{T} \sum_{t=1}^{T} \left|E\left(F_t^{0\prime} F_s^0\right)\right| \times \left|E\left(\zeta_{is}\zeta_{jt}\right)\right| \\ &\qquad (1.40)\end{aligned}$$

where we used new symbol $\zeta_{is} = \varepsilon_{iu}\varepsilon_{is} - E\left(\varepsilon_{iu}\varepsilon_{is}\right)$ for notational simplicity, and the last statement follows from the independence between idiosyncratic errors and factors. Note that $\left|E\left(F_t^{0\prime} F_s^0\right)\right| \leq \left(E\left\|F_t^0\right\|^2 E\left\|F_s^0\right\|^2\right)^{\frac{1}{2}}$ by the Cauchy-Schwarz inequality, where the right hand side of inequality is bounded by Assumption A(i-1). Therefore, overall bound of (1.40) is determined by $\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}\left|E\left(\zeta_{is}\zeta_{jt}\right)\right|$, which is assumed to be bounded by Assumption D(iv-1). In sum, we conclude that the first term of (1.38) is $\frac{1}{\sqrt{NT}} \times O_p(1)$.

Next, related to the second term of (1.38),

$$\begin{aligned}E\left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_k' E\left(\varepsilon_i\varepsilon_i'\right) F^0 \right\| &= E\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \varepsilon_{kt} E\left(\varepsilon_{it}\varepsilon_{is}\right) F_s^{0\prime} \right\| \\ &\leq \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left\|\varepsilon_{kt} F_s^{0\prime}\right\| \times \left|E\left(\varepsilon_{it}\varepsilon_{is}\right)\right| \qquad (1.41)\end{aligned}$$

where the last statement follows from the properties of norms and the fact that $E\left(\varepsilon_{it}\varepsilon_{is}\right)$ is a non-random object. Note that $E\left\|\varepsilon_{kt}F_s^{0\prime}\right\| \leq \left(E\left\|\varepsilon_{kt}\right\|^2 E\left\|F_s^0\right\|^2\right)^{\frac{1}{2}}$ by the Cauchy-Schwarz inequality, where the right hand side is bounded by Assumption A(i-1) and D(i). Now the order of (1.41) is determined by $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T\left|E\left(\varepsilon_{it}\varepsilon_{is}\right)\right|$. 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 (1.38): $\left\|\frac{1}{NT^2}\sum_{i=1}^N\varepsilon_k'\varepsilon_i\varepsilon_i'F^0\right\| = O_p\left(\frac{1}{C_{NT}\sqrt{T}}\right)$.

(v) By adding and subtracting $E\left(\varepsilon_i\varepsilon_i'\right)$, we have

$$\begin{aligned}\left\|\frac{1}{NT}\sum_{i=1}^N\varepsilon_i\varepsilon_i'\right\| &= \frac{1}{NT}\left\|\sum_{i=1}^N\left[\varepsilon_i\varepsilon_i' - E\left(\varepsilon_i\varepsilon_i'\right) + E\left(\varepsilon_i\varepsilon_i'\right)\right]\right\| \\ &\leq \frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{NT^2}}\sum_{i=1}^N\left[\varepsilon_i\varepsilon_i' - E\left(\varepsilon_i\varepsilon_i'\right)\right]\right\| + \frac{1}{\sqrt{T}}\left\|\frac{1}{\sqrt{N^2T}}\sum_{i=1}^N E\left(\varepsilon_i\varepsilon_i'\right)\right\| \end{aligned} \tag{1.42}$$

where the last inequality follows from the triangle inequality. For the first term of (1.42),

$$\begin{aligned} E\left\|\frac{1}{\sqrt{NT^2}}\sum_{i=1}^N\left[\varepsilon_i\varepsilon_i' - E\left(\varepsilon_i\varepsilon_i'\right)\right]\right\| &= \frac{1}{\sqrt{NT^2}} \times E\left[\left(\sum_{t=1}^T\sum_{s=1}^T\left|\sum_{i=1}^N\left[\varepsilon_{it}\varepsilon_{is} - E\left(\varepsilon_{it}\varepsilon_{is}\right)\right]\right|^2\right)^{\frac{1}{2}}\right] \\ &\leq \left(\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\varepsilon_{it}\varepsilon_{is} - E\left(\varepsilon_{it}\varepsilon_{is}\right)\right]\right|^2\right)^{\frac{1}{2}} \end{aligned}$$

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 (1.42),

$$\begin{aligned}\left\|\frac{1}{\sqrt{N^2T}}\sum_{i=1}^N E\left(\varepsilon_i\varepsilon_i'\right)\right\| &\leq \frac{1}{\sqrt{N^2T}}\sum_{i=1}^N\left\|E\left(\varepsilon_i\varepsilon_i'\right)\right\| \\ &= \frac{1}{N}\sum_{i=1}^N\left(\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\left|E\left(\varepsilon_{it}\varepsilon_{is}\right)\right|^2\right)^{\frac{1}{2}} \end{aligned} \tag{1.43}$$

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 (1.43) is bounded.

From the results above, we can calculate the stochastic order of (1.42): $\left\| \frac{1}{NT} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i' \right\| = O_p\left(\frac{1}{C_{NT}}\right)$.

Lemma A.7: Under Assumptions A-D,

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

(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 $F^0 H$,

$$\frac{1}{NT^2} \sum_{k=1}^{N} \left\| \varepsilon_k' \hat{F} \right\|^2 = \frac{1}{NT^2} \sum_{k=1}^{N} \left\| \varepsilon_k' \left(\hat{F} - F^0 H + F^0 H \right) \right\|^2$$

$$\leq \left(\frac{1}{NT} \sum_{k=1}^{N} \|\varepsilon_k\|^2 \right) \left(\frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 \right) + \frac{1}{T} \left(\frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{kt} F_t^{0\prime} \right\|^2 \right) \|H\|^2$$

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. By Corollary A.1(i), $\frac{1}{T} \left\| \hat{F} - F^0 H \right\|^2 = O_p\left(B_{NT}^2\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$. Both $\frac{1}{N} \sum_{k=1}^{N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_{kt} F_t^{0\prime} \right\|^2$ and $\|H\|$ are bounded by Lemma A.5(i) and Lemma A.4(iv), respectively. Combining these results, we have $\frac{1}{NT^2} \sum_{k=1}^{N} \left\| \varepsilon_k' \hat{F} \right\|^2 = O_p\left(B_{NT}^2\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$.

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

$$\left\| \frac{1}{\sqrt{NT^2}} \sum \lambda_i^0 \varepsilon_i' \hat{F} \right\| = \frac{1}{\sqrt{NT^2}} \left\| \sum_{k=1}^{N} \left(\hat{F} - F^0 H + F^0 H \right)' \varepsilon_k \lambda_k^{0\prime} \right\|$$

$$\leq \left(\frac{1}{\sqrt{T}} \left\| \hat{F} - F^0 H \right\| \right) \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} \varepsilon_k \lambda_k^{0\prime} \right\| + \frac{1}{\sqrt{T}} \times \|H\| \left\| \frac{1}{\sqrt{NT}} \sum_{k=1}^{N} F^{0\prime} \varepsilon_k \lambda_k^{0\prime} \right\|$$

where the last inequality follows from the properties of norms. By Lemma A.5 (ii) and (iii), both $\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}\varepsilon_k\lambda_k^{0\prime}\right\|$ and $\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N}F^{0\prime}\varepsilon_k\lambda_k^{0\prime}\right\|$ are bounded. Lemma A.4(iv) proved that $\|H\|$ is bounded. Lastly, $\frac{1}{\sqrt{T}}\left\|\hat{F}-F^0H\right\| = O_p(B_{NT})+O_p\left(\frac{1}{C_{NT}}\right)$ by Corollary A.1(i). Thus, we conclude that $\left\|\frac{1}{\sqrt{NT^2}}\sum\lambda_i^0\varepsilon_i'\hat{F}\right\| = O_p(B_{NT})+O_p\left(\frac{1}{C_{NT}}\right)$.

Lemma A.8: Under Assumptions A-D,

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

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

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

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

Proof:

(i) From (1.26), we have

$$\begin{aligned}\frac{1}{T}\left\|F^{0\prime}\left(\hat{F}-F^0H\right)\right\| &= \frac{1}{T}\left\|F^{0\prime}\left(I1+\cdots+I8\right)V_{NT}^{-1}\right\| \\ &\le \frac{1}{T}\left\|F^{0\prime}I1V_{NT}^{-1}\right\|+\cdots+\frac{1}{T}\left\|F^{0\prime}I8V_{NT}^{-1}\right\| \qquad (1.44)\end{aligned}$$

where the last statement follows from the triangle inequality. We shall characterize the stochastic orders of the eight terms in (1.44). For the first five terms, we have

$$\frac{1}{T}\left\|F^{0\prime}\,IX\,V_{NT}^{-1}\right\| \le \left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|IX\right\|\right)\left\|V_{NT}^{-1}\right\|$$

where a generic notation IX is used to denote each of $I1$, $I2$, $I3$, $I4$ and $I5$. We have already shown that $\frac{1}{\sqrt{T}}\|IX\| = O_p(B_{NT})$ in Proposition A.1.

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

$$\begin{aligned}\frac{1}{T}\left\|F^{0\prime}I6V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|F^{0\prime}\left(\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\varepsilon_i'\hat{F}\right)V_{NT}^{-1}\right\| \\ &\le \frac{1}{\sqrt{N}}\left\|\frac{F^{0\prime}F^0}{T}\right\|\left\|\frac{1}{\sqrt{NT^2}}\sum\lambda_i^0\varepsilon_i'\hat{F}\right\|\left\|V_{NT}^{-1}\right\|\end{aligned}$$

where the last inequality follows from the properties of norms. Each of terms in the above expression is investigated already: $\left\|\frac{F^{0\prime}F^0}{T}\right\|$ is bounded by Assumption A(i-2), $\left\|\frac{1}{\sqrt{NT^2}}\sum\lambda_i^0\varepsilon_i'\hat{F}\right\| = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}}\right)$ 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 $\frac{1}{T}\left\|F^{0\prime}I6V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{\sqrt{N}}\right) + O_p\left(\frac{1}{C_{NT}\sqrt{N}}\right)$.

Next term $\frac{1}{T}\left\|F^{0\prime}I7V_{NT}^{-1}\right\| = \frac{1}{T}\left\|F^{0\prime}\left(\frac{1}{NT}\sum_{i=1}^N\varepsilon_i\lambda_i^{0\prime}F^{0\prime}\hat{F}\right)V_{NT}^{-1}\right\|$ does not involve any β_i. It is already shown to be of order $O_p\left(\frac{1}{C_{NT}^2}\right)$ in Bai (2003).

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

$$\begin{aligned}\frac{1}{T}\left\|F^{0\prime}I8V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|F^{0\prime}\left(\frac{1}{NT}\sum_{i=1}^N\varepsilon_i\varepsilon_i'\hat{F}\right)V_{NT}^{-1}\right\| \\ &\leq \frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^0\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT^2}\sum_{i=1}^N\left\|\varepsilon_i'\hat{F}\right\|^2\right)^{\frac{1}{2}}\left\|V_{NT}^{-1}\right\|\end{aligned}$$

We have already shown that both $\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T F_t^0\varepsilon_{it}\right\|^2$ and $\left\|V_{NT}^{-1}\right\|$ are bounded by Lemma A.2(ii) and by Lemma A.4(iii), respectively. Lemma A.7(i) showed that $\frac{1}{NT^2}\sum_{i=1}^N\left\|\varepsilon_i'\hat{F}\right\|^2 = O_p\left(B_{NT}^2\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$. Combining all the results, we can show that $\frac{1}{T}\left\|F^{0\prime}I8V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{\sqrt{T}}\right) + O_p\left(\frac{1}{C_{NT}\sqrt{T}}\right)$.

Finally, summing over the eight terms in (1.44), we obtain the result of part (i).

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

$$\begin{aligned}\left\|\frac{1}{T}\hat{F}'\left(\hat{F}-F^0H\right)\right\| &= \left\|\frac{1}{T}\left(\hat{F}-F^0H+F^0H\right)'\left(\hat{F}-F^0H\right)\right\| \\ &\leq \frac{1}{T}\left\|\hat{F}-F^0H\right\|^2 + \|H\|\left\|\frac{1}{T}F^{0\prime}\left(\hat{F}-F^0H\right)\right\|\end{aligned}$$

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

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

$$\frac{1}{T}\left\|X_k'\left(\hat{F}-F^0H\right)\right\| = \frac{1}{T}\left\|X_k'\left(I1+\cdots+I8\right)V_{NT}^{-1}\right\|$$
$$\leq \frac{1}{T}\left\|X_k'I1V_{NT}^{-1}\right\|+\cdots+\frac{1}{T}\left\|X_k'I8V_{NT}^{-1}\right\| \quad (1.45)$$

Again, the first five terms can be shown to be of order $O_p\left(B_{NT}\right)$ because

$$\begin{aligned}\frac{1}{T}\left\|X_k'\,IX\,V_{NT}^{-1}\right\| &\leq \left(\frac{1}{\sqrt{T}}\left\|X_k\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|IX\right\|\right)\left\|V_{NT}^{-1}\right\| \\ &= O_p(1)\,O_p\left(B_{NT}\right)\,O_p(1) \\ &= O_p\left(B_{NT}\right)\end{aligned}$$

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

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

$$\begin{aligned}\frac{1}{T}\left\|X_k'I6V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|X_k'\left(\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\varepsilon_i'\hat{F}\right)V_{NT}^{-1}\right\| \\ &\leq \frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{T}}\left\|X_k\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left\|\frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_i'\hat{F}\right\|\left\|V_{NT}^{-1}\right\|\end{aligned}$$

Note that $\left\|\frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_i'\hat{F}\right\| = O_p\left(B_{NT}\right)+O_p\left(\frac{1}{C_{NT}}\right)$ as shown in Lemma A.7(ii) while all other terms are bounded. Therefore, it follows that $\frac{1}{T}\left\|X_k'I6V_{NT}^{-1}\right\| = O_p\left(B_{NT}\right)+O_p\left(\frac{1}{C_{NT}}\right)$.

Next term $\frac{1}{T}\left\|X_k'I7V_{NT}^{-1}\right\| = \frac{1}{T}\left\|X_k'\left(\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\lambda_i^{0\prime}F^{0\prime}\hat{F}\right)V_{NT}^{-1}\right\|$, again, does not involve any β_i. In the exactly same manner with the part (i), we can show that $\frac{1}{T}\left\|X_k'I7V_{NT}^{-1}\right\| = O_p\left(\frac{1}{C_{NT}^2}\right)$

Lastly, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\frac{1}{T}\left\|X_k' I8 V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|X_k'\left(\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\hat{F}\right)V_{NT}^{-1}\right\| \\ &\leq \frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_{kt}\varepsilon_{it}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT^2}\sum_{i=1}^{N}\left\|\varepsilon_i'\hat{F}\right\|^2\right)^{\frac{1}{2}}\left\|V_{NT}^{-1}\right\|\end{aligned}$$

Note that $\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_{kt}\varepsilon_{it}\right\|^2$ and $\left\|V_{NT}^{-1}\right\|$ are bounded by Lemma A.2(i) and Lemma A.4(iii), respectively. From Lemma A.7(i) , we have $\left(\frac{1}{NT^2}\sum_{i=1}^{N}\left\|\varepsilon_i'\hat{F}\right\|^2\right)^{\frac{1}{2}} = O_p\left(B_{NT}\right)+O_p\left(\frac{1}{C_{NT}}\right)$. Thus, we conclude that $\frac{1}{T}\left\|X_k' I8 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{\sqrt{T}}\right)+O_p\left(\frac{1}{C_{NT}\sqrt{T}}\right)$.

Finally, summing over the eight terms in (1.45), we obtain the desired result.

(iv) Using (1.26) and the triangle inequality,

$$\begin{aligned}\left\|\frac{1}{T}\varepsilon_k'\left(\hat{F}-F^0H\right)\right\| &= \frac{1}{T}\left\|\varepsilon_k'\left(I1+\cdots+I8\right)V_{NT}^{-1}\right\| \\ &\leq \frac{1}{T}\left\|\varepsilon_k' I1 V_{NT}^{-1}\right\|+\cdots+\frac{1}{T}\left\|\varepsilon_k' I8 V_{NT}^{-1}\right\| \qquad (1.46)\end{aligned}$$

From the definition of $I1$ and the Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{T}\left\|\varepsilon_k' I1 V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|\varepsilon_k'\left[\frac{1}{NT}\sum_{i=1}^{N}X_i\left(\beta_{0,i}-\hat{\beta}_i\right)\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\hat{F}\right]V_{NT}^{-1}\right\| \\ &\leq \frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}\mathbf{x}_{it}'\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\beta_{0,i}-\hat{\beta}_i\right\|^8\right)^{\frac{1}{4}} \\ &\quad\times\left(\frac{1}{NT^2}\sum_{i=1}^{N}\left\|X_i\right\|^4\right)^{\frac{1}{4}}\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|V_{NT}^{-1}\right\|\end{aligned}$$

Note that $\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\beta_{0,i}-\hat{\beta}_i\right\|^8\right)^{\frac{1}{4}} = O_p\left(B_{NT}^2\right)$ 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 $\frac{1}{T}\left\|\varepsilon_k' I1 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}^2}{\sqrt{T}}\right)$.

The stochastic order of next three terms in (1.46) can be shown in the same manner. For $\frac{1}{T}\left\|\varepsilon_k' I2 V_{NT}^{-1}\right\|$,

$$\begin{aligned}
\frac{1}{T}\left\|\varepsilon_k' I2 V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|\varepsilon_k'\left[\frac{1}{NT}\sum_{i=1}^{N} X_i\left(\beta_{0,i}-\hat{\beta}_i\right)\lambda_i^{0\prime}F^{0\prime}\hat{F}\right]V_{NT}^{-1}\right\| \\
&\leq \frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}\mathbf{x}_{it}'\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\beta_{0,i}-\hat{\beta}_i\right\|^4\right)^{\frac{1}{4}} \\
&\qquad \times\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\lambda_i^0\right\|^4\right)^{\frac{1}{4}}\left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|V_{NT}^{-1}\right\| \\
&= \frac{1}{\sqrt{T}}\times O_p(1)\,O_p\left(B_{NT}\right)\,O_p(1)\,O_p(1)\,O_p(1)\,O_p(1) \\
&= O_p\left(\frac{B_{NT}}{\sqrt{T}}\right)
\end{aligned}$$

where the second statement follows from applying the Cauchy-Schwarz inequality. The proof for $\frac{1}{T}\left\|\varepsilon_k' I3 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{\sqrt{T}}\right)$ and $\frac{1}{T}\left\|\varepsilon_k' I4 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{\sqrt{T}}\right)$ are omitted.

Next,

$$\begin{aligned}
\left\|\frac{1}{T}\varepsilon_k' I5 V_{NT}^{-1}\right\| &= \left\|\frac{1}{T}\varepsilon_k'\left[\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\hat{F}\right]V_{NT}^{-1}\right\| \\
&\leq \frac{1}{NT\sqrt{T}}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}\varepsilon_{it}\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\right\|\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|V_{NT}^{-1}\right\| \qquad (1.47)
\end{aligned}$$

Note that $\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|$ and $\left\|V_{NT}^{-1}\right\|$ are bounded by the normalizing assumption and Lemma A.4(iii), respectively. Therefore, $\frac{1}{NT\sqrt{T}}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}\varepsilon_{it}\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\right\|$ determines the order of (1.47). Let $\Upsilon = \frac{1}{NT\sqrt{T}}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}\varepsilon_{it}\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\right\|$ only temporarily for notational convenience. Then, by adding and subtracting $E(\varepsilon_{kt}\varepsilon_{it})$,

$$\begin{aligned}
\Upsilon &= \frac{1}{NT\sqrt{T}}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it}) + E(\varepsilon_{kt}\varepsilon_{it})\right]\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\right\| \\
&\leq \frac{1}{NT\sqrt{T}}\left\{\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it})\right]\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\right\| + \left\|\sum_{i=1}^{N}\sum_{t=1}^{T}E(\varepsilon_{kt}\varepsilon_{it})\left(\beta_{0,i}-\hat{\beta}_i\right)' X_i'\right\|\right\} \\
&\leq \left\{\frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\varepsilon_{kt}\varepsilon_{it} - E(\varepsilon_{kt}\varepsilon_{it})\right]\right|^2\right)^{\frac{1}{2}} + \frac{1}{\sqrt{N}}\left(\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}E(\varepsilon_{kt}\varepsilon_{it})\right|^2\right)^{\frac{1}{2}}\right\} \\
&\qquad \times\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\beta_{0,i}-\hat{\beta}_i\right\|^4\right)^{\frac{1}{4}}\left(\frac{1}{NT^2}\sum_{i=1}^{N}\|X_i\|^4\right)^{\frac{1}{4}} \qquad (1.48)
\end{aligned}$$

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 (1.48) is of order $\left\{\frac{1}{\sqrt{T}}\times O_p(1) + \frac{1}{\sqrt{N}}\times O_p(1)\right\} = O_p\left(\frac{1}{C_{NT}}\right)$. The other terms can be shown to be $\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\beta_{0,i}-\hat{\beta}_i\right\|^4\right)^{\frac{1}{4}} = O_p(B_{NT})$ and $\left(\frac{1}{NT^2}\sum_{i=1}^{N}\|X_i\|^4\right)^{\frac{1}{4}} = O_p(1)$, respectively. In sum, $\left\|\frac{1}{T}\varepsilon_k' I5 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{C_{NT}}\right)$.

The proof for $\left\|\frac{1}{T}\varepsilon_k' I6 V_{NT}^{-1}\right\|$ is much simpler. Using the Cauchy-Schwarz inequality,

$$\begin{aligned}
\frac{1}{T}\left\|\varepsilon_k' I6 V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|\varepsilon_k'\left[\frac{1}{NT}\sum_{i=1}^{N}F^0\lambda_i^0\varepsilon_i'\hat{F}\right]V_{NT}^{-1}\right\| \\
&\leq \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}F_t^{0\prime}\right\|\left\|\frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_i'\hat{F}\right\|\left\|V_{NT}^{-1}\right\|
\end{aligned}$$

We have already shown that $\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{kt}F_t^{0\prime}\right\|$ and $\left\|V_{NT}^{-1}\right\|$ are bounded by Lemma A.5(i) and Lemma A.4(iii), respectively. Finally, $\left\|\frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\lambda_i^0\varepsilon_i'\hat{F}\right\| = O_p(B_{NT}) + O_p\left(\frac{1}{C_{NT}}\right)$ from Lemma A.7(ii). Therefore, it follows that $\frac{1}{T}\left\|\varepsilon_k' I6 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{\sqrt{NT}}\right) + O_p\left(\frac{1}{C_{NT}\sqrt{NT}}\right)$.

Next,

$$\begin{aligned}
\frac{1}{T}\left\|\varepsilon_k' I7 V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|\varepsilon_k'\left[\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\lambda_i^{0\prime}F^{0\prime}\hat{F}\right]V_{NT}^{-1}\right\| \\
&\leq \left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}\varepsilon_{it}\lambda_i^{0\prime}\right\|\left(\frac{1}{\sqrt{T}}\left\|F^0\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|V_{NT}^{-1}\right\| \\
&= O_p\left(\frac{1}{C_{NT}\sqrt{N}}\right)O_p(1)\,O_p(1)\,O_p(1)
\end{aligned}$$

where the inequality follows from the properties of norms, and the rate result for the first term, $\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{kt}\varepsilon_{it}\lambda_i^{0\prime}\right\|$, comes from Lemma A.6(iii). Thus, we have $\frac{1}{T}\left\|\varepsilon_k' I7 V_{NT}^{-1}\right\| = O_p\left(\frac{1}{C_{NT}\sqrt{N}}\right)$.

Lastly,

$$\begin{aligned}
\frac{1}{T}\left\|\varepsilon_k' I8 V_{NT}^{-1}\right\| &= \frac{1}{T}\left\|\varepsilon_k'\left[\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\hat{F}\right]V_{NT}^{-1}\right\| \\
&= \frac{1}{NT^2}\left\|\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i\varepsilon_i'\left(\hat{F}-F^0H+F^0H\right)V_{NT}^{-1}\right\| \\
&\leq \left\{\left(\frac{1}{\sqrt{T}}\left\|\varepsilon_k\right\|\right)\left\|\frac{1}{NT}\sum_{i=1}^{N}\varepsilon_i\varepsilon_i'\right\|\left(\frac{1}{\sqrt{T}}\left\|\left(\hat{F}-F^0H\right)\right\|\right)\right. \\
&\qquad \left. +\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i\varepsilon_i'F^0\right\|\left\|H\right\|\right\}\left\|V_{NT}^{-1}\right\|
\end{aligned}$$

where the second statement comes from adding and subtracting F^0H, 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), $\frac{1}{\sqrt{T}}\left\|\left(\hat{F}-F^0H\right)\right\| = O_p\left(B_{NT}\right)+O_p\left(\frac{1}{C_{NT}}\right)$ from Corollary A.1(i), $\left\|\frac{1}{NT^2}\sum_{i=1}^{N}\varepsilon_k'\varepsilon_i\varepsilon_i'F^0\right\| = O_p\left(\frac{1}{C_{NT}\sqrt{T}}\right)$ from Lemma A.6(iv) and the fact that other terms are bounded, we can show that $\frac{1}{T}\left\|\varepsilon_k' I8 V_{NT}^{-1}\right\| = O_p\left(\frac{B_{NT}}{C_{NT}}\right)+O_p\left(\frac{1}{C_{NT}^2}\right)$.

Summing over the eight terms in (1.46), we have the desired result $\frac{1}{T}\varepsilon_k'\left(\hat{F}-F^0H\right) = O_p\left(\frac{B_{NT}}{C_{NT}}\right)+O_p\left(\frac{1}{C_{NT}^2}\right)$.

Corollary A.2: Under Assumptions A-D, $\frac{1}{T}\left\|X_i' M_{\hat{F}} F^0 H\right\| = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$

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

$$\begin{aligned}
\frac{1}{T}\left\|X_i' M_{\hat{F}} F^0 H\right\| &= \left\|\frac{1}{T}X_i' M_{\hat{F}}\left(\hat{F} - F^0 H\right)\right\| \\
&= \left\|\frac{1}{T}X_i'\left(I - \frac{1}{T}\hat{F}\hat{F}'\right)\left(\hat{F} - F^0 H\right)\right\| \\
&\leq \frac{1}{T}\left\|X_i'\left(\hat{F} - F^0 H\right)\right\| + \left(\frac{1}{\sqrt{T}}\left\|X_i\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|\frac{1}{T}\hat{F}'\left(\hat{F} - F^0 H\right)\right\|
\end{aligned}$$

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

Lemma A.9: Under Assumptions A-D,

(i) $HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$

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

(iii) $\frac{1}{\sqrt{T}}X_i' M_{\hat{F}}\varepsilon_i = \frac{1}{\sqrt{T}}X_i' M_{F^0}\varepsilon_i + O_p\left(B_{NT}\right) + O_p\left(\frac{B_{NT}\sqrt{T}}{C_{NT}}\right) + O_p\left(\frac{\sqrt{T}}{C_{NT}^2}\right)$

Proof:

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

$$\begin{aligned}
\left\|HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\| &= \left\|\left[HH'\left(\frac{F^{0\prime}F^0}{T}\right) - I\right]\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\| \\
&\leq \left\|HH'\left(\frac{F^{0\prime}F^0}{T}\right) - I\right\|\left\|\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right\| \qquad (1.49)
\end{aligned}$$

where the last inequality follows from the properties of norms. Note that $\left\| \left(\frac{F^{0\prime}F^0}{T} \right)^{-1} \right\|$ is bounded that the order of (1.49) depends on the order of $\left\| HH' \left(\frac{F^{0\prime}F^0}{T} \right) - I \right\|$. By pre-multiplying and post-multiplying $HH^{-1} = I$, we have

$$\begin{aligned} \left\| HH' \left(\frac{F^{0\prime}F^0}{T} \right) - I \right\| &= \left\| HH^{-1} \left[HH' \left(\frac{F^{0\prime}F^0}{T} \right) - I \right] HH^{-1} \right\| \\ &\leq \|H\| \left\| H' \left(\frac{F^{0\prime}F^0}{T} \right) H - I \right\| \left\| H^{-1} \right\| \end{aligned} \tag{1.50}$$

Again, the order of the above object in (1.50) is determined by $\left\| H' \left(\frac{F^{0\prime}F^0}{T} \right) H - I \right\|$ because both $\|H\|$ and $\left\| H^{-1} \right\|$ are shown to be bounded in Lemma A.4. By adding and subtracting $H' \left(\frac{F^{0\prime}\hat{F}}{T} \right)$ and using the fact that $I = \frac{\hat{F}'\hat{F}}{T}$ by the normalization assumption, we have

$$\begin{aligned} \left\| H' \left(\frac{F^{0\prime}F^0}{T} \right) H - I \right\| &= \left\| H' \left(\frac{F^{0\prime}F^0}{T} \right) H - H' \left(\frac{F^{0\prime}\hat{F}}{T} \right) + H' \left(\frac{F^{0\prime}\hat{F}}{T} \right) - \frac{\hat{F}'\hat{F}}{T} \right\| \\ &= \left\| H' \left[\frac{1}{T} F^{0\prime} \left(F^0 H - \hat{F} \right) \right] + \left[\frac{1}{T} \left(H' F^{0\prime} - \hat{F}' \right) \hat{F} \right] \right\| \\ &\leq \|H\| \left\| \frac{1}{T} F^{0\prime} \left(\hat{F} - F^0 H \right) \right\| + \left\| \frac{1}{T} \hat{F}' \left(\hat{F} - F^0 H \right) \right\| \end{aligned}$$

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), $\left\| \frac{1}{T} F^{0\prime} \left(\hat{F} - F^0 H \right) \right\| = O_p \left(B_{NT} \right) + O_p \left(\frac{1}{C_{NT}^2} \right)$ by Lemma A.8(i), and $\left\| \frac{1}{T} \hat{F}' \left(\hat{F} - F^0 H \right) \right\| = O_p \left(B_{NT} \right) + O_p \left(\frac{1}{C_{NT}^2} \right)$ by Lemma A.8(ii). Combining all the results above, we obtain the desired lemma: $HH' - \left(\frac{F^{0\prime}F^0}{T} \right)^{-1} = O_p \left(B_{NT} \right) + O_p \left(\frac{1}{C_{NT}^2} \right)$.

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

$$\begin{aligned}\frac{1}{T}X_i'\left(M_{F^0} - M_{\hat{F}}\right)X_j &= \frac{1}{T^2}X_i'\hat{F}\hat{F}'X_j - \frac{1}{T^2}X_i'F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}F^{0\prime}X_j \\ &= \frac{1}{T^2}X_i'\left(\hat{F} - F^0H + F^0H\right)\left(\hat{F} - F^0H + F^0H\right)'X_j \\ &\quad - \frac{1}{T^2}X_i'F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}F^{0\prime}X_j \\ &= \frac{1}{T}X_i'\left(\hat{F} - F^0H\right)H'\left(\frac{1}{T}F^{0\prime}X_j\right) \\ &\quad + \left[\frac{1}{T}X_i'\left(\hat{F} - F^0H\right)\right]\left[\frac{1}{T}\left(\hat{F} - F^0H\right)'X_j\right] \\ &\quad + \left(\frac{1}{T}X_i'F^0\right)H\left[\frac{1}{T}\left(\hat{F} - F^0H\right)'X_j\right] \\ &\quad + \left(\frac{1}{T}X_i'F^0\right)\left[HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\left(\frac{1}{T}F^{0\prime}X_j\right) \qquad (1.51)\end{aligned}$$

where the second statement follows from adding and subtracting F^0H, and the last equality comes form expanding all terms. We can easily examine the order of the four terms in (1.51) because we have already derived the order of each subcomponent. For the first term in (1.51), we know $\frac{1}{T}X_i'\left(\hat{F} - F^0H\right) = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$ while other subcomponents are all bounded. The order of the second term $\left[\frac{1}{T}X_i'\left(\hat{F} - F^0H\right)\right]\left[\frac{1}{T}\left(\hat{F} - F^0H\right)'X_j\right]$ is determined by the product of $\frac{1}{T}X_i'\left(\hat{F} - F^0H\right) = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$ and $\frac{1}{T}\left(\hat{F} - F^0H\right) = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$, hence of order $O_p\left(B_{NT}^2\right) + O_p\left(\frac{B_{NT}}{C_{NT}^2}\right) + O_p\left(\frac{1}{C_{NT}^4}\right)$. The third term simply has the same rate with $\frac{1}{T}\left(\hat{F} - F^0H\right) = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$. The order of the last term is determined by $HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$ form the part (i) of this lemma. Summing over the four terms in (1.51), we obtain $\frac{1}{T}X_i'\left(M_{F^0} - M_{\hat{F}}\right)X_j = o_p(1)$. That is, we have the desired lemma: $\frac{1}{T}X_i'M_{\hat{F}}X_j = \frac{1}{T}X_i'M_{F^0}X_j + o_p(1)$.

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

$$\begin{aligned}
\frac{1}{\sqrt{T}}X_i'\left(M_{F^0} - M_{\hat{F}}\right)\varepsilon_i &= \frac{1}{\sqrt{T}} \times \frac{1}{T}X_i'\hat{F}\hat{F}'\varepsilon_i - \frac{1}{\sqrt{T}} \times \frac{1}{T}X_i'F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}F^{0\prime}\varepsilon_i \\
&= \frac{1}{\sqrt{T}} \times \frac{1}{T}X_i'\left(\hat{F} - F^0H + F^0H\right)\left(\hat{F} - F^0H + F^0H\right)'\varepsilon_i \\
&\quad - \frac{1}{\sqrt{T}} \times \frac{1}{T}X_i'F^0\left(\frac{F^{0\prime}F^0}{T}\right)^{-1}F^{0\prime}\varepsilon_i \\
&= \frac{1}{T}X_i'\left(\hat{F} - F^0H\right)H'\left(\frac{1}{\sqrt{T}}F^{0\prime}\varepsilon_i\right) \\
&\quad + \sqrt{T}\left[\frac{1}{T}X_i'\left(\hat{F} - F^0H\right)\right]\left[\frac{1}{T}\left(\hat{F} - F^0H\right)'\varepsilon_i\right] \\
&\quad + \sqrt{T}\left(\frac{1}{T}X_i'F^0\right)H\left[\frac{1}{T}\left(\hat{F} - F^0H\right)'\varepsilon_i\right] \\
&\quad + \left(\frac{1}{T}X_i'F^0\right)\left[HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1}\right]\left(\frac{1}{\sqrt{T}}F^{0\prime}\varepsilon_i\right) \qquad (1.52)
\end{aligned}$$

where, again, the second statement follows from adding and subtracting F^0H, and the last equality comes form expanding all terms. The order of the four terms in (1.52) can be determined by the existing results: $\frac{1}{T}X_i'\left(\hat{F} - F^0H\right) = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$, $\frac{1}{T}\varepsilon_k'\left(\hat{F} - F^0H\right) = O_p\left(\frac{B_{NT}}{C_{NT}}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$ and $HH' - \left(\frac{F^{0\prime}F^0}{T}\right)^{-1} = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$. Combining all those results, we can easily show that $\frac{1}{\sqrt{T}}X_i'\left(M_{F^0} - M_{\hat{F}}\right)\varepsilon_i = O_p\left(B_{NT}\right) + O_p\left(\frac{B_{NT}\sqrt{T}}{C_{NT}}\right) + O_p\left(\frac{\sqrt{T}}{C_{NT}^2}\right)$, thus proving the lemma. The exact order of $O_p\left(B_{NT}\right) + O_p\left(\frac{B_{NT}\sqrt{T}}{C_{NT}}\right) + O_p\left(\frac{\sqrt{T}}{C_{NT}^2}\right)$ will be needed later in the proof of Proposition 1.

Proof of Proposition 1:

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

$$\begin{aligned}
\left(\frac{1}{T} X_i' M_{\hat{F}} X_i\right)\left(\hat{\beta}_i - \beta_{0,i}\right) &= \frac{1}{T} X_i' M_{\hat{F}} \varepsilon_i + \frac{1}{T} X_i' M_{\hat{F}} F^0 \lambda_i^0 \\
&= \frac{1}{T} X_i' M_{\hat{F}} \varepsilon_i \\
&\quad + \frac{1}{T} X_i' M_{\hat{F}} \left[\hat{F} H^{-1} - (I1 + \cdots + I8)\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right]\lambda_i^0 \\
&= \frac{1}{T} X_i' M_{\hat{F}} \varepsilon_i + J1 + J2 + J3 + J4 + J5 + J6 + J7 + J8 \qquad (1.53)
\end{aligned}$$

where the second statement follows from (1.26), and the last statement is just relabeling of the eight terms after expansion.

We examine the stochastic order of the eight terms in (1.53) one at a time. First, by the definition of $J1$,

$$\begin{aligned}
\|J1\| &= \left\|(-1)\frac{1}{T} X_i' M_{\hat{F}} I1 \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&\leq \left(\frac{1}{\sqrt{T}}\left\|X_i' M_{\hat{F}}\right\|\right)\left(\frac{1}{\sqrt{T}}\|I1\|\right)\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\|\left\|\lambda_i^0\right\| \qquad (1.54)
\end{aligned}$$

where the inequality comes from the properties of norms. We have already derived the order of each term in (1.54): $\frac{1}{\sqrt{T}}\left\|X_i' M_{\hat{F}}\right\| = O_p(1)$ by Lemma A.1, $\frac{1}{\sqrt{T}}\|I1\| = O_p\left(B_{NT}^2\right)$ by Proposition A.1(ii), $\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\| = O_p(1)$ by (1.28) in Proposition A.1, $\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\| = O_p(1)$ by Assumption A(ii-2), and, lastly, $\left\|\lambda_i^0\right\| = O_p(1)$ by Assumption A(ii-1). Combining all the results, we conclude that $\|J1\| = o_p\left(B_{NT}\right)$.

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

$$\begin{aligned}
J2 &= -\frac{1}{T}X_i'M_{\hat{F}}I2\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0 \\
&= -\frac{1}{T}X_i'M_{\hat{F}}\left[\frac{1}{NT}\sum_{j=1}^{N}X_j\left(\beta_{0,j}-\hat{\beta}_j\right)\lambda_j^{0\prime}F^{0\prime}\hat{F}\right]\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0 \\
&= -\frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i'M_{\hat{F}}X_j}{T}\right)\left(\beta_{0,j}-\hat{\beta}_j\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right] \\
&= \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i'M_{\hat{F}}X_j}{T}\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\left(\hat{\beta}_j-\beta_{0,j}\right)
\end{aligned}$$

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

For $J3$,

$$\begin{aligned}
\|J3\| &= \left\|(-1)\frac{1}{T}X_i'M_{\hat{F}}I3\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&= \left\|(-1)\frac{1}{T}X_i'M_{\hat{F}}\left[\frac{1}{NT}\sum_{k=1}^{N}X_k\left(\beta_{0,k}-\hat{\beta}_k\right)\varepsilon_k'\hat{F}\right]\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&\leq \left(\frac{1}{\sqrt{T}}\left\|X_i'M_{\hat{F}}\right\|\right)\left(\frac{1}{NT^2}\sum_{k=1}^{N}\|X_k\|^4\right)^{\frac{1}{4}}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\beta_{0,k}-\hat{\beta}_k\right\|^4\right)^{\frac{1}{4}} \\
&\quad\times\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|\varepsilon_k'\hat{F}\right\|^2\right)^{\frac{1}{2}}\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\|\left\|\lambda_i^0\right\|
\end{aligned}$$

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 $\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\beta_{0,k}-\hat{\beta}_k\right\|^4\right)^{\frac{1}{4}} =$

$O_p(B_{NT})$ and that $\left(\frac{1}{NT^2}\sum_{k=1}^N \left\|\varepsilon_k'\hat{F}\right\|^2\right)^{\frac{1}{2}} = O_p(B_{NT}) + O_p\left(\frac{1}{C_{NT}}\right)$ while other terms are bounded. Therefore, we conclude that $J3 = o_p(B_{NT})$.

Next, consider $J4$

$$\begin{aligned}
\|J4\| &= \left\|(-1)\frac{1}{T}X_i' M_{\hat{F}} I4 \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&= \left\|\frac{1}{T}X_i' M_{\hat{F}}\left[\frac{1}{NT}\sum_{k=1}^N F^0\lambda_k^0\left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\hat{F}\right]\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&\leq \left(\frac{1}{T}\left\|X_i' M_{\hat{F}} F^0 H\right\|\right)\left\|H^{-1}\right\|\left(\frac{1}{N}\sum_{k=1}^N \left\|\lambda_k^0\right\|^4\right)^{\frac{1}{4}}\left(\frac{1}{N}\sum_{k=1}^N\left\|\beta_{0,k}-\hat{\beta}_k\right\|^4\right)^{\frac{1}{4}} \\
&\quad \times \left(\frac{1}{NT}\sum_{k=1}^N \|X_k\|^2\right)^{\frac{1}{2}}\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\| \\
&\quad \times \left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\|\left\|\lambda_i^0\right\|
\end{aligned}$$

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 $\frac{1}{T}\left\|X_i' M_{\hat{F}} F^0 H\right\| = O_p(B_{NT}) + O_p\left(\frac{1}{C_{NT}^2}\right)$ by Corollary A.2 and that $\left(\frac{1}{N}\sum_{k=1}^N \left\|\beta_{0,k}-\hat{\beta}_k\right\|^4\right)^{\frac{1}{4}} = O_p(B_{NT})$. Other terms are already shown to be bounded. Combining all these results, we can easily see that $J4 = o_p(B_{NT})$.

Now

$$\begin{aligned}
\|J5\| &= \left\|(-1)\frac{1}{T}X_i' M_{\hat{F}} I5 \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&= \frac{1}{T}\left\|X_i' M_{\hat{F}}\left[\frac{1}{NT}\sum_{k=1}^N \varepsilon_k\left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\hat{F}\right]\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&\leq \frac{1}{NT\sqrt{T}}\left\|\sum_{k=1}^N X_i' M_{\hat{F}}\varepsilon_k\left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\|\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\|\left\|\lambda_i^0\right\|
\end{aligned} \tag{1.55}$$

where the last inequality hold true by the Cauchy-Schwarz inequality. Since the last four terms of (1.55) are bounded, the first term $\frac{1}{NT\sqrt{T}}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\|$ solely determines the order of $J5$. Let $\Upsilon = \frac{1}{NT\sqrt{T}}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\|$ only temporarily for notational convenience. Then, by substituting the definition of $M_{\hat{F}} = I - \frac{1}{T}\hat{F}\hat{F}'$, we have

$$\begin{aligned}
\Upsilon &= \frac{1}{NT\sqrt{T}}\left\|\sum_{k=1}^{N} X_i'\left[I_T - \frac{1}{T}\hat{F}\hat{F}'\right]\varepsilon_k\left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\| \\
&\leq \frac{1}{NT\sqrt{T}}\left\{\left\|\sum_{k=1}^{N} X_i'\varepsilon_k\left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\| + \frac{1}{T}\left\|X_i'\hat{F}\sum_{k=1}^{N}\hat{F}'\varepsilon_k\left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\|\right\} \\
&\leq \left\{\frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_{it}\varepsilon_{kt}\right\|^2\right)^{\frac{1}{2}} + \left(\frac{1}{\sqrt{T}}\left\|X_i\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|\hat{F}'\varepsilon_k\right\|^2\right)^{\frac{1}{2}}\right\} \\
&\qquad \times\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\beta_{0,k}-\hat{\beta}_k\right\|^4\right)^{\frac{1}{4}}\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|X_k\right\|^4\right)^{\frac{1}{4}} \qquad (1.56)
\end{aligned}$$

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 (1.56) sum up to yield $O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}}\right)$. Note that $\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\beta_{0,k}-\hat{\beta}_k\right\|^4\right)^{\frac{1}{4}} = O_p\left(B_{NT}\right)$ and that $\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|X_k\right\|^4\right)^{\frac{1}{4}}$ is simply bounded. Therefore, we conclude that $\frac{1}{NT\sqrt{T}}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \left(\beta_{0,k}-\hat{\beta}_k\right)' X_k'\right\| = o_p\left(B_{NT}\right)$, which in turn implies that $J5 = o_p\left(B_{NT}\right)$.

Next,

$$
\begin{aligned}
\|J6\| &\equiv \left\| (-1)\frac{1}{T}X_i' M_{\hat{F}} I6 \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \lambda_i^0 \right\| \\
&= \left\| -\frac{1}{T}X_i' M_{\hat{F}} \left[\frac{1}{NT}\sum_{k=1}^{N} F^0 \lambda_k^0 \varepsilon_k' \hat{F}\right] \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \lambda_i^0 \right\| \\
&\leq \frac{1}{\sqrt{N}} \left(\frac{1}{T}\left\|X_i' M_{\hat{F}} F^0 H\right\|\right) \left\|H^{-1}\right\| \left\| \frac{1}{\sqrt{NT^2}} \sum_{k=1}^{N} \lambda_k^0 \varepsilon_k' \hat{F} \right\| \left\| \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} \right\| \\
&\qquad \times \left\| \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \right\| \left\|\lambda_i^0\right\|
\end{aligned}
$$

where the last inequality follows from the Cauchy-Schwarz inequality. From Corollary A.2, we have $\frac{1}{T}\left\|X_i' M_{\hat{F}} F^0 H\right\| = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$. Note also that $\left\|\frac{1}{\sqrt{NT^2}}\sum_{k=1}^{N} \lambda_k^0 \varepsilon_k' \hat{F}\right\| = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}}\right)$ by Lemma A.7(ii) while remaining three terms are bounded. Combining the results, we can easily show that $J6 = o_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^3\sqrt{N}}\right)$.

Consider $J7$:

$$
\begin{aligned}
\|J7\| &\equiv \left\| (-1)\frac{1}{T}X_i' M_{\hat{F}} I7 \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \lambda_i^0 \right\| \\
&= \left\| -\frac{1}{T}X_i' M_{\hat{F}} \left[\frac{1}{NT}\sum_{k=1}^{N} \varepsilon_k \lambda_k^{0\prime} F^{0\prime} \hat{F}\right] \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1} \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \lambda_i^0 \right\| \\
&= \frac{1}{NT} \left\| \sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \lambda_k^{0\prime} \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \lambda_i^0 \right\| \\
&\leq \frac{1}{NT} \left\| \sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \lambda_k^{0\prime} \right\| \left\| \left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1} \right\| \left\|\lambda_i^0\right\| \qquad (1.57)
\end{aligned}
$$

where the third statement follows from the fact that $\frac{F^{0\prime}\hat{F}}{T}$ cancels out with $\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}$, and the last inequality comes from the properties of norms. Given the boundedness of $\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\|$ and $\left\|\lambda_i^0\right\|$, it is $\frac{1}{NT}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \lambda_k^{0\prime}\right\|$ that determines the order of (1.57).

Using the definition of $M_{\hat{F}} = I - \frac{1}{T}\hat{F}\hat{F}'$,

$$\begin{aligned}
\frac{1}{NT}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \lambda_k^{0\prime}\right\| &= \frac{1}{NT}\left\|\sum_{k=1}^{N} X_i' \left(I - \frac{1}{T}\hat{F}\hat{F}'\right)\varepsilon_k \lambda_k^{0\prime}\right\| \\
&\leq \frac{1}{\sqrt{NT}}\left\|\frac{1}{\sqrt{NT}}\sum_{k=1}^{N} X_i'\varepsilon_k \lambda_k^{0\prime}\right\| \\
&\quad + \frac{1}{\sqrt{T}}\left(\frac{1}{\sqrt{T}}\left\|X_i\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left\|\frac{1}{\sqrt{NT^2}}\sum_{k=1}^{N}\hat{F}'\varepsilon_k \lambda_k^{0\prime}\right\| \\
&= \frac{1}{\sqrt{NT}} \times O_p(1) + \frac{1}{\sqrt{T}} \times O_p(1)\, O_p(1)\, \left[O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}}\right)\right] \\
&= o_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}\sqrt{T}}\right)
\end{aligned}$$

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

Finally, consider $J8$:

$$\begin{aligned}
\left\|J8\right\| &= \left\|-\frac{1}{T}X_i' M_{\hat{F}} I8 \left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&= \left\|-\frac{1}{T}X_i' M_{\hat{F}}\left[\frac{1}{NT}\sum_{k=1}^{N}\varepsilon_k\varepsilon_k'\hat{F}\right]\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right\| \\
&\leq \frac{1}{NT^2}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}}\varepsilon_k\varepsilon_k'\hat{F}\right\|\left\|\left(\frac{F^{0\prime}\hat{F}}{T}\right)^{-1}\right\|\left\|\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\right\|\left\|\lambda_i^0\right\| \qquad (1.58)
\end{aligned}$$

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 (1.58) are already shown to be bounded. Thus, we examine $\frac{1}{NT^2}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}}\varepsilon_k\varepsilon_k'\hat{F}\right\|$ which determines the order of $J8$. Using the definition of $M_{\hat{F}} = I - \frac{1}{T}\hat{F}\hat{F}'$,

$$\begin{aligned}\frac{1}{NT^2}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \varepsilon_k' \hat{F}\right\| &= \frac{1}{NT^2}\left\|\sum_{k=1}^{N} X_i'\left(I_T - \frac{1}{T}\hat{F}\hat{F}'\right)\varepsilon_k \varepsilon_k' \hat{F}\right\| \\ &\leq \frac{1}{\sqrt{T}}\left(\frac{1}{N}\sum_{k=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{x}_{it}\varepsilon_{kt}\right\|^2\right)^{\frac{1}{2}}\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|\varepsilon_k'\hat{F}\right\|^2\right)^{\frac{1}{2}} \\ &\quad + \left(\frac{1}{\sqrt{T}}\left\|X_i\right\|\right)\left(\frac{1}{\sqrt{T}}\left\|\hat{F}\right\|\right)\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|\varepsilon_k'\hat{F}\right\|^2\right)\end{aligned}$$

where the last inequality follows from the combination of the triangle inequality and the Cauchy-Schwarz inequality. Note that $\left(\frac{1}{NT^2}\sum_{k=1}^{N}\left\|\varepsilon_k'\hat{F}\right\|^2\right)^{\frac{1}{2}} = O_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}}\right)$ from Lemma A.7(i) while other terms are bounded. Thus, $\frac{1}{NT^2}\left\|\sum_{k=1}^{N} X_i' M_{\hat{F}} \varepsilon_k \varepsilon_k' \hat{F}\right\| = o_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$, which in turn leads to $J8 = o_p\left(B_{NT}\right) + O_p\left(\frac{1}{C_{NT}^2}\right)$.

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

$$\begin{aligned}\left(\frac{1}{T}X_i' M_{\hat{F}} X_i\right)\left(\hat{\beta}_i - \beta_{0,i}\right) &= \frac{1}{T}X_i' M_{\hat{F}}\varepsilon_i + J1 + J2 + J3 + J4 + J5 + J6 + J7 + J8 \\ &= \frac{1}{T}X_i' M_{\hat{F}}\varepsilon_i + \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i' M_{\hat{F}} X_j}{T}\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\left(\hat{\beta}_j - \beta_{0,j}\right) \\ &\quad + O_p\left(\frac{1}{C_{NT}^2}\right) + o_p\left(\frac{1}{\sqrt{T}}\right) + o_p\left(B_{NT}\right) \qquad (1.59)\end{aligned}$$

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

$$\begin{aligned}\left(\frac{1}{T}X_i' M_{\hat{F}} X_i\right)\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) &= \frac{1}{\sqrt{T}}X_i' M_{\hat{F}}\varepsilon_i \\ &\quad + \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i' M_{\hat{F}} X_j}{T}\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\sqrt{T}\left(\hat{\beta}_j - \beta_{0,j}\right) \\ &\quad + O_p\left(\frac{\sqrt{T}}{C_{NT}^2}\right) + o_p(1)\end{aligned}$$

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 $M_{\hat{F}}$ in the above expression by M_{F^0} using Lemma A.9 (ii) and (iii).

$$
\begin{aligned}
&\left[\left(\frac{1}{T}X_i'M_{F^0}X_i + o_p(1)\right) + o_p(1)\right]\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) \\
&\qquad = \left[\frac{1}{\sqrt{T}}X_i'M_{F^0}\varepsilon_i + O_p\left(B_{NT}\right) + O_p\left(\frac{B_{NT}\sqrt{T}}{C_{NT}}\right) + O_p\left(\frac{\sqrt{T}}{C_{NT}^2}\right)\right] \\
&\qquad\qquad + \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i'M_{F^0}X_j}{T} + o_p(1)\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\sqrt{T}\left(\hat{\beta}_j - \beta_{0,j}\right) \\
&\qquad\qquad + O_p\left(\frac{\sqrt{T}}{C_{NT}^2}\right) + o_p(1) \\
&\qquad = \left(\frac{1}{\sqrt{T}}X_i'M_{F^0}\varepsilon_i\right) \\
&\qquad\qquad + \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i'M_{F^0}X_j}{T}\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\sqrt{T}\left(\hat{\beta}_j - \beta_{0,j}\right) + o_p(1)
\end{aligned}
$$

where the last equality holds under an additional condition on N and T, that is $\frac{\sqrt{T}}{N} \to 0$. 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:

$$
\begin{aligned}
\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) &= \left[\left(\frac{1}{T}X_i'M_{F^0}X_i\right) + o_p(1)\right]^{-1}\left\{\frac{1}{\sqrt{T}}X_i'M_{F^0}\varepsilon_i\right. \\
&\qquad \left. + \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i'M_{F^0}X_j}{T}\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\sqrt{T}\left(\hat{\beta}_j - \beta_{0,j}\right) + o_p(1)\right\} \\
&= \left(\frac{X_i'M_{F^0}X_i}{T}\right)^{-1}\left(\frac{1}{\sqrt{T}}X_i'M_{F^0}\varepsilon_i\right) \\
&\quad + \frac{1}{N}\sum_{j=1}^{N}\left(\frac{X_i'M_{F^0}X_i}{T}\right)^{-1}\left(\frac{X_i'M_{F^0}X_j}{T}\right)\left[\lambda_j^{0\prime}\left(\frac{\Lambda^{0\prime}\Lambda^0}{N}\right)^{-1}\lambda_i^0\right]\sqrt{T}\left(\hat{\beta}_j - \beta_{0,j}\right) \\
&\quad + o_p(1)
\end{aligned} \tag{1.60}
$$

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

$$G_{ij} \equiv \left(\frac{X_i' M_{F^0} X_i}{T} \right)^{-1} \left(\frac{X_i' M_{F^0} X_j}{T} \right) \left[\lambda_j^{0\prime} \left(\frac{\Lambda^{0\prime} \Lambda^0}{N} \right)^{-1} \lambda_i^0 \right].$$

Then the expression (1.60) above can be rewritten as

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

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

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

Defining

$$\sqrt{T}\left(\hat{\beta} - \beta_0\right) = \begin{bmatrix} \sqrt{T}\left(\hat{\beta}_1 - \beta_{0,1}\right) \\ \sqrt{T}\left(\hat{\beta}_2 - \beta_{0,2}\right) \\ \vdots \\ \sqrt{T}\left(\hat{\beta}_N - \beta_{0,N}\right) \end{bmatrix}, \qquad S = \begin{bmatrix} S_{11}^{-1} & 0 & \cdots & 0 \\ 0 & S_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & S_{NN}^{-1} \end{bmatrix}$$

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

we have $\sqrt{T}\left(\hat{\beta} - \beta_0\right) = S^{-1}\xi + \left(\frac{1}{N}G\right)\sqrt{T}\left(\hat{\beta} - \beta_0\right) + o_p(1)$. Solving for $\sqrt{T}\left(\hat{\beta} - \beta_0\right)$ yields

$$\begin{aligned} \sqrt{T}\left(\hat{\beta} - \beta_0\right) &= \left(I - \frac{1}{N}G\right)^{-1} S^{-1}\xi + o_p(1) \\ &= \left(I + \frac{1}{N}G + \frac{1}{N^2}G^2 + \frac{1}{N^3}G^3 + \cdots\right) S^{-1}\xi + o_p(1). \end{aligned}$$

Then, a generic individual i's estimator has an expression

$$\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right) = S_{ii}^{-1}\xi_i + \frac{1}{N}\sum_{j=1}^{N} G_{ij}S_{jj}^{-1}\xi_j + o_p(1). \tag{1.61}$$

Note that higher order terms related to $\frac{1}{N^k}G^k S^{-1}\xi$ can be ignored as $o_p(1)$ due 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 (1.61) 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:

Since an average of independent terms $\frac{1}{N}\sum_{j=1}^{N} G_{ij}S_{jj}^{-1}\xi_j$ in (1.61) converges faster under cross-sectional independence, $S_{ii}^{-1}\xi_i$ becomes the only leading term of $\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right)$. By cross-sectional independence, we mean that $\widetilde{\mathbf{X}}_i$, λ_i^0, and ε_i are jointly independent over i, where we define $\widetilde{\mathbf{X}}_i = M_{F^o}\mathbf{X}_i$. If no lagged dependent variables are included as regressors, replacing $\widetilde{\mathbf{X}}_i$ by $\mathbf{X}_i$ provides a sufficient condition. Note that $S_{ii}^{-1}\xi_i = \left(\frac{X_i' M_{F^0} X_i}{T}\right)^{-1} \frac{1}{\sqrt{T}} X_i' M_{F^0}\varepsilon_i$ corresponds to the infeasible estimator where the latent common factor F^0 is treated as if observable. That is, under cross-sectional independence, the individual estimator $\sqrt{T}\left(\hat{\beta}_i - \beta_{0,i}\right)$ asymptotically behaves like the infeasible estimator, and its asymptotic distribution can be characterized as a normal distribution with mean zero and variance Ω_i. Here, the asymptotic variance Ω_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 $\Sigma_i = p\lim_{T\to\infty}\frac{X_i' M_{F^0} X_i}{T}$ and $\Xi_i = p\lim_{T\to\infty}\frac{X_i' M_{F^0} E\left(\varepsilon_i\varepsilon_i'\right) M_{F^0} X_i}{T}$.

Chapter 2

Aggregation and Purchasing Power Parity

2.1 Introduction

The theory of purchasing power parity (PPP) predicts that the real exchange rate between two economies, defined as the relative price levels denominated in the same currency, should be constant. In a stricter version of the PPP theory, two price levels in the same currency denomination should be equal. In contrast to these theoretical predictions, numerous empirical studies have found large and persistent deviations from the PPP in real exchange rate data. In a survey, Rogoff (1996) presented a range of three to five years as the consensus half-life estimates in the empirical literature.[1] This failure to reconcile the theoretical predictions with the empirical findings is referred to as the PPP puzzle. Obstfeld and Rogoff (2001) identified the PPP puzzle as one of the six major puzzles in international economics.

Among the various attempts to resolve the PPP puzzle, Imbs et al. (2005) noted a potential upward bias in the half-life estimates based on aggregate exchange rate data. This argument is closely related to the observation of Granger (1980): the persistence of an aggregate series can be considerably different from the persistence of its underlying disaggregated series because of the aggregation procedure. In the context of the PPP hypothesis, this observation implies that the deviation of the aggregate real exchange rate from the PPP may be highly persistent even though the disaggregated real exchange rate dynamics quickly revert to the PPP. Motivated by the idea, Imbs et al. (2005) directly analyzed sectoral real exchange rate dynamics and compared the results with the dynamics of the aggregate exchange rates. The authors estimated the mean group (MG) dynamics of the sectoral real exchange rates and found half-life estimates in the range of 11 to 26 months, depending on the specifications.[2] These estimates are clearly much shorter than

[1] Half-life is widely used in the PPP literature to quantify the persistence of deviation from the PPP. It is defined as the time required for half of the initial impact of a given shock to dissipate.

[2] The MG model was proposed by Pesaran and Smith (1995). It is estimated by averaging the least squares estimates of the autoregressive coefficients across individuals.

the consensus estimates and thus favor the implications of the PPP theory. It is argued that this finding confirms the role of aggregation bias in resolving the PPP puzzle.

Carvalho and Nechio (2011) (hereafter CN) extended the analysis of Imbs et al. (2005) and developed an elaborate framework to assess the aggregation bias. In their framework, the difference between the persistence of aggregate exchange rates and that of the MG dynamics of sectoral exchange rates is called the *total heterogeneity effect.* That is, the aforementioned finding of Imbs et al. (2005) is equivalent to discovering a large total heterogeneity effect. As another useful concept, the *aggregation effect* is defined as the difference between the persistence of aggregate exchange rates and the average persistence of sectoral exchange rates. Using Eurostat data, CN empirically found that the total heterogeneity effect is large whereas the aggregation effect is small.[3]

This chapter aims to revisit the aforementioned empirical findings of CN by relying on their insightful framework. The estimation results of CN crucially depend on the validity of the ordinary least squares (OLS) estimation used to analyze the sectoral exchange rate dynamics. We reexamine the sectoral exchange rate dynamics by replacing the OLS approach with the dynamic heterogeneous panels with cross-sectional dependence (DHP+CSD) approach proposed in Chapter 1. There are several reasons why it is more beneficial to analyze the sectoral exchange rate dynamics using the DHP+CSD model. First of all, the DHP+CSD approach facilitates a more general specification and robust estimation of the disaggregated dynamics than the OLS estimator. While the OLS approach only allows for heterogeneous dynamics, the DHP+CSD model additionally considers the possible cross-sectional dependence that arises from a common factor structure. The OLS estimator suffers from inconsistency in the general case of serially correlated common factors, but the DHP+CSD estimator properly controls for the unobservable common factors and thus remain consistent. Furthermore, the DHP+CSD estimator performs in

[3]The authors argue that the debate on aggregation bias between Imbs et al. (2005) and Chen and Engel (2005) can be reconciled by carefully distinguishing the two different effects.

a robust manner, even if no cross-sectional dependence is present in the data. Second, the specification of DHP+CSD model provides an appropriate empirical framework for analyzing aggregation bias. Since the pioneering work of Granger (1980), subsequent studies have focused on common factors as well as heterogeneous dynamics; Zaffaroni (2004) extended the analysis of Granger (1980) by also including common factors, and Pesaran and Chudik (2013) found that common factor dynamics also contribute to the exaggeration of aggregate persistence. Therefore, the difference in persistence between aggregate exchange rates and sectoral exchange rates can be measured more precisely through the use of the DHP+CSD model because it allows for common factor dynamics.[4] Finally, the DHP+CSD model enables the distinguishing of impulse response functions with respect to common shocks and idiosyncratic shocks. It is natural to allow for the transmission mechanisms to be different depending on the nature of shocks. In a similar vein, but in a different context of price dynamics, Boivin et al. (2009) uncovered several stylized facts on the impulse responses with respect to common shocks and idiosyncratic shocks, respectively. These empirical findings became a motivation for subsequent studies (e.g. Carvalho and Lee, 2011) that develop a more refined structural model. Similarly, documenting the difference in responses with respect to common shocks and idiosyncratic shocks in the PPP context can be beneficial for future structural analysis.

It should be noted that the goal of this study is not to provide a conclusive answer to the debate on the persistence of exchange rates. Rather, it aims to illustrate the difference in the results one can obtain by simply adopting a more general specification for the disaggregated exchange rate dynamics. In this respect, we embrace the disclaimer of CN that a number of important but unaddressed methodological issues exist. One of these issues is obtaining correct confidence intervals for the estimated persistence measures (see Killian and Zha, 1999; and Rossi, 2005). It is frequently found that computed confidence

[4] Imbs et al. (2005), as well as Murray and Papell (2005) and Choi et al. (2006), explicitly addressed the cross-sectional dependence (that may arise from common factors), but maintained the assumption of no serial correlation.

intervals can be uninformative, having extremely large upper bounds. In our analysis, we focus on how much the point estimate of the persistence measure changes when a more general econometric model is adopted.

Bearing this caveat in mind, we find a large total heterogeneity effect that is consistent with the finding of CN. Our estimation results, however, show that the aggregation effect is also substantial, in contrast to the finding of CN. While CN measured the aggregation effect at approximately 14% of the total heterogeneity effect, we find that the aggregation effect is approximately 79% of the total heterogeneity effect. This difference arises from adopting a more general econometric model that also allows for the dynamics of common factors. Our estimation results also show that common shocks exert approximately 50% more persistent effect on the economy compared to idiosyncratic shocks.

The remainder of the chapter is organized as follows. Section 2 describes a widely used data set in the PPP literature. Section 3 explains the framework of CN to distinguish the aggregation effect from the total heterogeneity effect. Section 4 presents the empirical strategy of this chapter to estimate the sectoral exchange rate dynamics. Corresponding empirical results are provided in Section 5. Finally, Section 6 concludes.

2.2 Data

We utilize a widely used data set in the PPP literature, including Carvalho and Nechio (2011), Imbs et al. (2005), and Chen and Engel (2005), because our aim is to revisit an existing issue by using a new model. The data set is constructed by the statistical agency of European Union, the Eurostat. The Eurostat data provide the price indices for the consumption of goods and services in 11 European countries: Belgium, Denmark, Finland, France, Germany, Greece, Italy, Netherlands, Portugal, Spain, and the U.K. Detailed information on price indices are available at two-digit levels for 19 categories of goods and services: alcohol, books, bread, clothing, communications, diary, domestic appliances, footwear, fruits, furniture, hotels, leisure, meat, public transportation, rents, sound, tobacco, and vehicles. As seen from the list, the characteristics of the categories are widely different from one another (*e.g.,* tradable versus non-tradable). Therefore, we expect that the dynamics in each category are heterogeneous, which would be relevant for the issue of aggregation as discussed above. Based on these sectoral price indices, the CPI-based real exchange rates are constructed against the U.S. dollar. For a given country, the real exchange rates are computed as:

$$RER_{it} = E_t \times \frac{P_{it}^*}{P_{it}^{US}} \tag{2.1}$$

where P_{it}^* and P_{it}^{US} denote the price indices of category i at time t for the given country and the U.S., respectively, and E_t denotes the nominal bilateral exchange rate between the two countries at time t. Finally, the sample used in the analysis runs from 1981 through 1995 at monthly frequency. See Section IV and Appendix 3 of Imbs et al. (2005) for a more detailed description of the data.

2.3 Total heterogeneity effect versus aggregation effect

This section explains the framework of CN and provides the precise definitions of the total heterogeneity effect and the aggregation effect. Understanding the difference between the two conceptual objects is essential for the subsequent sections of this chapter. CN model the dynamics of sectoral real exchange rates using autoregressive (AR) processes:

$$y_{it} = c_i + a_{i1}y_{i,t-1} + a_{i2}y_{i,t-2} + \cdots + a_{ip}y_{i,t-p} + e_{it} \tag{2.2}$$

where $y_{it} = ln\ (RER_{it})$. Assuming that e_{it} are *i.i.d.* shocks, CN use the OLS estimator to estimate the autoregressive coefficients $\hat{A}_i(L) = 1 - \hat{a}_{i1}L - \hat{a}_{i2}L^2 - \cdots - \hat{a}_{ip}L^p$ for each sectoral exchange rate series. The half-life of each sectoral exchange rate is computed from the estimated dynamics, and denoted by $\mathcal{HL}_i$.

Given the estimated dynamics $\hat{A}_i(L) = 1 - \hat{a}_{i1}L - \hat{a}_{i2}L^2 - \cdots - \hat{a}_{ip}L^p$ for all sectoral exchange rates, CN also estimate the MG dynamics of sectoral exchange rates as:

$$\hat{A}_{MG}(L) = 1 - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{a}_{i1}\right)L - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{a}_{i2}\right)L^2 - \cdots - \left(\frac{1}{N}\sum_{i=1}^{N}\hat{a}_{ip}\right)L^p \tag{2.3}$$

as per Pesaran and Smith (1995). Then, $\mathcal{HL}_{MG}$ is defined as the half-life that corresponds to the MG dynamics and interpreted as a representative half-life measure of the sectoral real exchange rates. Proposition 5 of CN provides a theoretical interpretation of $\mathcal{HL}_{MG}$ as the half-life of a counterfactual one-sector economy. The counterfactual one-sector economy is identical to a heterogeneous multi-sector economy except for one feature: there is only one sector, of which the frequency of price changes is equal to the average frequency of price changes of the heterogeneous multi-sector economy. It is argued that the one-sector counterfactual economy allows a thought experiment for isolating the role of heterogeneity from other parts of the model.

Defining $\mathcal{HL}_{Agg}$ as the half-life of aggregate real exchange rates, the two key concepts of CN are defined as:

$$\text{aggregation effect} = \mathcal{HL}_{Agg} - \frac{1}{N}\sum_{i=1}^{N}\mathcal{HL}_i \tag{2.4}$$

$$\text{total heterogeneity effect} = \mathcal{HL}_{Agg} - \mathcal{HL}_{MG}. \tag{2.5}$$

2.4 Empirical strategy

One of the key empirical findings of CN is that the aggregation effect is small whereas the total heterogeneity effect is large. From the definitions (2.4) and (2.5) of the two effects, this statement is equivalent to finding that $\frac{1}{N}\sum_{i=1}^{N}\mathcal{HL}_i$ is measured to be large whereas $\mathcal{HL}_{MG}$ is measured to be small. These estimates crucially depend on the validity of the OLS estimator that the authors used to estimate the dynamics of disaggregated exchange rates. This section reexamines these two key estimates by using the DHP+CSD model in Chapter 1. The purpose of this study is to investigate the robustness of the existing empirical findings when a more general specification is allowed.

We model the dynamics of the sectoral exchange rates by using a DHP+CSD model of the following form:

$$y_{it} = c_i + a_{i1}y_{i,t-1} + a_{i2}y_{i,t-2} + \cdots + a_{ip}y_{i,t-p} + \lambda_i F_t + \varepsilon_{it} \tag{2.6}$$

where F_t denotes an unobservable common factor, λ_i denotes the factor loading that captures the heterogeneous effects of the common factor, and ε_{it} are *i.i.d.* idiosyncratic shocks. The dynamics of the common factor are allowed through an autoregression $B(L)F_t = u_t$, where u_t denotes an *i.i.d.* common shock. We assume that the sectoral real exchange rates are stationary processes relying on the existing panel unit-root test results available in the existing literature. Using the same Eurostat data set, Imbs et al. (2005) reported that four out of five variants of unit-root tests based on Levin and Lin (2002) and Im et al. (2003) could reject the null hypothesis of non-stationarity.[5] Their conclusion is also consistent with the existing literature (see Frankel and Rose, 1996; Oh, 1996; Wu, 1996; Lothian, 1997).

Allowing for common dynamics is appealing if we recall how sectoral real exchange rate series are constructed; all the relative prices are computed against the corresponding U.S. price indices. Therefore, shocks to the U.S. economy are likely to generate common dynamics among the sectoral exchange rates. In the DHP+CSD model, such common fluctuations in sectoral real exchange rate dynamics can be captured by the common factor F_t. Regarding the specification, one may proceed with the analysis in two stages as is typical in the literature. In the first stage, one pre-tests the presence of cross-sectional dependence from the common factor structure using available tests.[6] In the second stage, the model is estimated using the DHP+CSD estimator if the pre-test detects cross-sectional dependence.

[5] In the test of Levin and Lin (2002), the null hypothesis is that all series are non-stationary against that all series are stationary. Im et al. (2003) generalized the test to heterogeneous panel models.

[6] Among many available tests of cross-sectional dependence are Friedman (1937), Breusch and Pagan (1980), Frees (1995), Pesaran (2004), Pesaran et al. (2008), Sarafidis et al. (2009), Baltagi et al. (2011), Chen et al. (2012). Each test is valid under different conditions.

However, it is well-known that pre-testing can affect the second stage inference (see Leeb and Pötscher (2005), for example). This section directly utilizes the robust properties of the DHP+CSD estimator, instead of proceeding in two stages after a pre-test. In the simulation results in Chapter 1, the DHP+CSD estimator is shown to perform in a robust manner with or without the presence of cross-sectional dependence.

It should be noted that the specification of (2.6) is more general than that of (2.2). That is, (2.6) nests the model (2.2) as a special case under an additional restriction that no common factor structure exists. Consequently, the DHP+CSD estimator will deliver similar estimation results if the restriction imposed by CN is indeed valid. Otherwise, the difference in empirical results can be interpreted as possible distortion introduced by not allowing a common factor structure and its own dynamics. It is also worth mentioning that the specification (2.6) also nests that of other studies in the literature. For instance, Imbs et al. (2005) performed an analysis with possible cross-sectional dependence that may arise from common factor structure, but did not allow for the dynamics of common factors.[7] Such specification is again a special case of (2.6) with an additional restriction that the common factor itself is the common shock. All in all, (2.6) is sufficiently general to cover a number of specifications used in the existing literature.

The key to our general specification is that the common factor is allowed to have its own dynamics. Pesaran and Chudik (2013) share the same motivation: allowing common factor dynamics is important in assessing the role of aggregation. In the context of sectoral inflation dynamics, these authors considered a more general specification than Altissimo et al. (2009), who only allowed serially uncorrelated factors. By doing so, Pesaran and Chudik (2013) found that the discrepancies in persistence between the aggregate inflation series and its underlying disaggregated inflation series could be explained by serially correlated common factors combined with the heterogeneity of the AR coefficients.

[7] Murray and Papell (2005) and Choi et al. (2006) also addressed cross-sectional dependence explicitly, but maintained the assumption that the error terms are serially uncorrelated.

Although the context is different, their main idea also carries over to our empirical context; the persistence of sectoral real exchange rates can be precisely measured when common factor dynamics are also taken into account in the model.

Our general specification, however, cannot be consistently estimated by the OLS approach of CN. From the perspective of the OLS estimation, the error term is the sum of two unobservable components $(\lambda_i F_t + \varepsilon_{it})$. Since the common factor has its own serial correlation, F_t is necessarily correlated with the lagged dependent variables. That is, the common factor plays the role of an omitted variable that is correlated with regressors in the equation. Consequently, the OLS is no longer consistent due to the omitted variable bias. For consistent estimation, it is crucial to control for the unobserved common factors. In contrast to the OLS estimator, the DHP+CSD estimator treats the unobservable common factors as parameters and directly estimate them from the rich information in the large cross-sectional units (large N). By doing so, the DHP+CSD estimator delivers consistent estimates for the AR coefficients while allowing any arbitrary dynamics of the unobservable common factor; the common factor can be a general linear covariance stationary process, possibly with a non-zero mean.

The model (2.6) can be estimated by minimizing the sum of squared residuals over all parameters: heterogeneous AR coefficients, common factors, and loadings. For the ease of implementation, however, we estimate the model in an iterative manner. We first group the parameters into two subgroups: 1) heterogeneous AR coefficients, and 2) factors and loadings. The intuition behind the iterative estimation procedure is that, given a subgroup of parameters, it is extremely easy to estimate the other subgroup of parameters, and *vice versa*. For example, we can simply run regressions equation-by-equation to estimate the heterogeneous AR coefficients if the common factors are known. Conversely, if all heterogeneous AR coefficients are known, the model simplifies to a pure factor model that can be readily estimated by existing methods such as the principal component estimation. By iterating the two steps of estimation until convergence, we obtain the estimates for both groups of parameters. It is shown by Sargan (1964) that the outcome of the iteration converges to the solution of the original problem.

Except for allowing common factors with their own dynamics, we closely follow CN in modeling the dynamics of sectoral real exchange rates. We chose a lag order of 19 as our benchmark, following the choice of both CN and Imbs et al. (2005), for the dynamics of sectoral real exchange rates and the common factor. We also tried lag orders of 2 and 12, following the additional specifications of CN and Imbs et al. (2005), respectively, for a robustness check.

2.5 Estimation results

We report the estimated persistence of the MG dynamics of sectoral real exchange rates in Table 2.1. The persistence measures include the half-life, the sum of autoregressive coefficients ($\mathcal{SAC}$), the largest autoregressive root ($\mathcal{LAR}$), and the cumulative impulse response ($\mathcal{CIR}$). Since the debate on sectoral exchange rate persistence is centered around the half-life measures, we also focus on comparing the half-life estimates in the following discussion. The first row of Table 2.1 reports the persistence of the MG dynamics constructed from the OLS estimates. That is, each sectoral exchange rate dynamics is estimated by the OLS estimator without considering the possibility of unobserved common factor or its own dynamics. The corresponding half-life of 26 months is considerably lower than the consensus half-life estimates of three to five years by Rogoff (1996) or the estimate of 46 months by Imbs et al. (2005) based on aggregate exchange rates. Note that the difference is defined as the total heterogeneity effect (2.5) in the framework of CN. Therefore, CN confirms the argument of Imbs et al. (2005) that the total heterogeneity is large. The bottom row of Table 2.1 reports the estimated persistence of the MG dynamics based on the DHP+CSD estimator. Note that the persistence measures are separately reported with respect to common shocks and idiosyncratic shocks. While the OLS estimator cannot distinguish common shocks from idiosyncratic shocks, the general specification of the DHP+CSD facilitates this detailed analysis. Consistent with the finding of CN, our results also show that the persistence of the MG dynamics is much lower than that of aggregate exchange rates. The quantitative difference is even larger if we consider the response with respect to idiosyncratic shocks. If we apply the structural interpretation of the MG dynamics as in CN, the estimation results imply that the one-sector counterfactual economy exhibits a different pattern of dynamics from that of aggregate exchange rates.

Table 2.1: Estimated persistence of the MG dynamics of sectoral real exchange rates

		$\mathcal{HL}$	$\mathcal{SAC}$	$\mathcal{LAR}$	$\mathcal{CIR}$
OLS		26 months	0.97	0.95	33.2
DHP+CSD	common shock	27 months	0.97	0.94	34.9
	idiosyncratic shock	18 months	0.95	0.94	20.9

Notes: $\mathcal{HL}$, $\mathcal{SAC}$, $\mathcal{LAR}$, and $\mathcal{CIR}$ denote half-life, the sum of autoregressive coefficients, the largest autoregressive root, and the cumulative impulse response, respectively.

Now we turn into the dynamics of each sectoral exchange rate series instead of the MG dynamics. Figure 2.1 plots the histogram of the half-life estimates from the analysis of sectoral exchange rate dynamics by using the DHP+CSD estimator. Recall again that the general specification of the DHP+CSD model allows for discriminating the nature of shocks. The top panel of Figure 2.1 shows the histogram of half-life estimates with respect to idiosyncratic shocks, whereas the bottom panel presents an analogous histogram with respect to common shocks. The two histograms make plain that common shocks have more persistent effect on the economy compared to idiosyncratic shocks; the estimated half-lives with respect to common shocks are, on average, 50% more persistent than those with respect to idiosyncratic shocks. The difference suggests that the transmission mechanisms are substantially different depending on the nature of shocks. This new stylized fact should be a consideration for future structural modeling. Similar attempts have been made in a different context of price dynamics; Carvalho and Lee (2011) refined the existing structural models based on the findings of Boivin et al. (2009) that impulse responses differ substantially depending on the nature of shocks.

Substantial heterogeneity in the half-life estimates stands out in Figure 2.1; while

Figure 2.1: Histograms of half-life estimates of sectoral real exchange rates

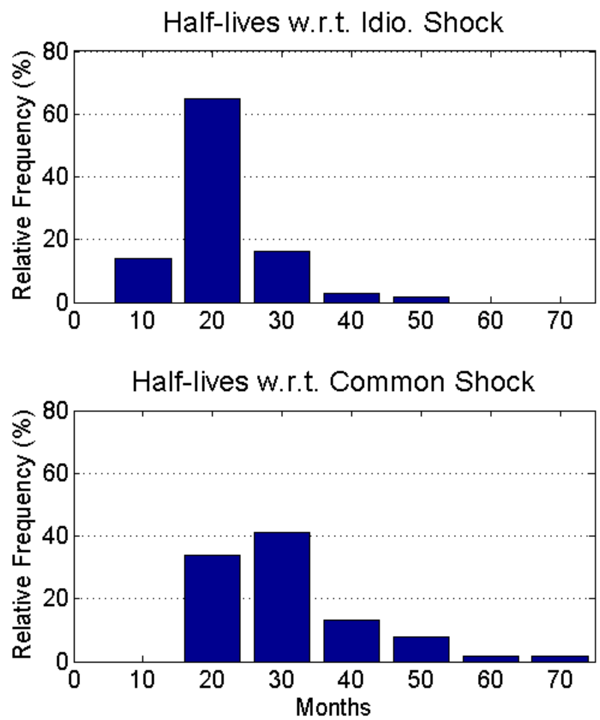


Table 2.2: Average of the estimated persistence of sectoral real exchange rates

		$\mathcal{HL}$	$\mathcal{SAC}$	$\mathcal{LAR}$	$\mathcal{CIR}$
OLS		43.2 months	0.97	0.94	59.5
DHP+CSD	common shock	31.0 months	0.97	0.97	41.2
	idiosyncratic shock	20.6 months	0.95	0.96	24.7

Notes: $\mathcal{HL}$, $\mathcal{SAC}$, $\mathcal{LAR}$, and $\mathcal{CIR}$ denote half-life, the sum of autoregressive coefficients, the largest autoregressive root, and the cumulative impulse response, respectively.

some sectors quickly reverts to the PPP, shocks to other sectors last long for several years. Such heterogeneity in persistence is important in examining the role of aggregation bias as already observed by Granger (1980). This issue is closely related to measuring the magnitude of the aggregation effect in CN. As discussed in previous sections, CN confirms the argument of Chen and Engel (2005) that the aggregation effect is small. By the definition of the aggregation effect in (2.4), small aggregation effect is equivalent to low average half-life of sectoral exchange rates. Table 2.2 presents the average of the estimated half-lives at the disaggregated level. The persistence estimates in the first row comes from the estimation results based on the OLS estimator. Note that the estimated half-life of 43.2 months is within the consensus range of three to five years of aggregate exchange rate half-life by Rogoff (1996). It is also close to the half-life estimate of 46 months by Imbs et al. (2005) for aggregate exchange rates. The estimation results, however, change substantially if common factor dynamics are allowed for through the use of the DHP+CSD model. The bottom row of Table 2.2 reports the persistence estimates with respect to common shocks and idiosyncratic shocks from the DHP+CSD estimates. Note that the half-life estimate is only 31 months for common shocks, which is considerably lower than the half-life of aggregate exchange rates. The quantitative difference is even larger if we consider the case for idiosyncratic shocks. Since the DHP+CSD estimator is known to perform robustly even if no common factor structure is present, the difference in estimation results between the OLS estimator and the DHP+CSD estimator implies that omitted variables are likely to be present. In such cases, the half-life estimates from the OLS can be biased, whereas the estimates from our general specification remain correct.

In sum, the general specification of the DHP+CSD model only partially confirms the claims in CN; the total heterogeneity effect is estimated to be large, but the aggregation effect also appears to be substantial. Large aggregation effect can be explained by the considerable heterogeneity in dynamics in Figure 2.1; by the mechanism of Granger (1980), the persistence of aggregate exchange rates appears to be exaggerated due to the aggregation of heterogeneous sectoral exchange rate dynamics.

2.6 Conclusion

This chapter investigates the dynamics of sectoral real exchange rates in the context of the PPP hypothesis. Relying on the framework of Carvalho and Nechio (2011), we revisit the issue of aggregation bias in explaining the discrepancies in persistence between the aggregate exchange rates and the sectoral exchange rates. We utilize the general specification of the DHP+CSD model developed in Chapter 1 to model the dynamics of disaggregated exchange rates. Dynamic, heterogeneity, and common factor dynamics in the DHP+CSD model is known to be important in explaining the discrepancies in persistence between an aggregate series and its underlying disaggregated series. Our estimation results show that the persistence of sectoral exchange rates is indeed lower than that of aggregate exchange rates. In addition, the persistence estimates from the DHP+CSD model are substantially lower than the estimates from the existing models that ignored the dynamics of common factors. This implies that the persistence estimates of the latter models might have been distorted by ignoring the common factor dynamics. We also find that the impulse responses with respect to common shocks are different from those with respect to idiosyncratic shocks. On average, the effect of common shocks on the economy is approximately 50% more persistent than that of idiosyncratic shocks. Such difference in transmission mechanism deserves further research. All in all, the analysis illustrates the importance of precisely modeling the disaggregated dynamics by allowing for all the key features of large panel data sets.

Chapter 3

Aggregation and Inflation Persistence

3.1 Introduction

Many researchers assume that inflation is a highly persistent series. In the long-run neutrality literature, for example, numerous authors assumed that inflation is a unit-root process. (see Fisher and Seater, 1993; King and Watson, 1997; Fair, 2000, among others). There are other papers that also confirmed the persistent behavior of the inflation series. (see, for example, Crowder and Hoffman, 1996; Evans and Lewis, 1995; Ng and Perron, 2001). Some authors, such as Cogley and Sargent (2001) and Cogley and Sargent (2005), argued that the persistence of inflation has declined considerably, but numerous studies again confirmed the persistent behavior of inflation (see Stock, 2001, Pivetta and Reis, 2007, among others).[1] All in all, persistence is the key characteristic of inflation that has received a great attention in the literature.

In contrast to the aforementioned studies on the persistence of aggregate inflation, some authors directly analyzed disaggregated inflation data and found a considerably lower degree of persistence (see, for example, Bils and Klenow, 2004; Clark, 2006; Altissimo et al., 2009; Pesaran and Chudik, 2013). Some argue that direct investigation of disaggregated inflation can be a more appealing approach because disaggregated data sets are closer to the object of our interest:

[W]e have focused on macroeconomic, aggregate evidence bearing on inflation persistence. Yet the dominant models in the literature aim to provide microeconomic foundations for inflation, based on the price-setting decisions of individual firms. In this regards, it is striking that the now large literature that examines micro-price data has emerged relatively recently, ... (Fuhrer, 2010, p.478)

The aim of this chapter is to reassess the persistence of inflation at the sectoral level by using the dynamic heterogeneous panels with cross-sectional dependence (DHP+CSD) model developed in Chapter 1. Using the DHP+CSD model can be beneficial in this

[1] See also Garcia and Perron (1996), Evans and Wachtel (1993), Kim (1993), among others, for the results of unit root tests with possible regime changes.

context due to its general empirical specification. The model simultaneously allows for dynamics, heterogeneity, and cross-sectional dependence, which are the key characteristics of large panels of disaggregated data sets. Among these three characteristics, dynamics and heterogeneity are already well-incorporated by the aforementioned empirical studies. The empirical frameworks of both Bils and Klenow (2004) and Clark (2006) are based on the equation-by-equation ordinary least squares (OLS) estimation, where all the autoregressive (AR) coefficients are allowed to be different across sectors to allow for heterogeneous dynamics. These authors, however, maintained the independence assumption of the innovation terms, which is necessary for the OLS estimation. Nevertheless, a common factor structure is highly likely in the disaggregated data because the inflation dynamics in all sectors are likely to be affected by economy-wide fluctuations. Indeed, the presence of a common factor structure is confirmed by Clark (2006), Boivin et al. (2009), and Reis and Watson (2010) for the U.S. economy.[2] In general, common factors can have their own dynamics that reflect economy-wide fluctuations. In such cases, the unobservable common factors are necessarily correlated with the lagged dependent variables. Therefore, the unobservable common factors become the omitted variables that are correlated with the regressors in the equation. This correlation makes the OLS estimation no longer consistent due to the omitted variable bias, and it can permit a distortion in the persistence estimates of the sectoral inflation series. In contrast, the DHP+CSD estimator treats the unobservable common factors as parameters to be estimated, and directly estimates them from the rich information of the large panels. By controlling for the space spanned by the unobserved factors using the direct estimates, the DHP+CSD estimator restores the consistency for the AR coefficients. Therefore, it is particularly relevant to use the DHP+CSD estimator in revisiting the analysis of the sectoral inflation persistence.

Using the DHP+CSD model is also important in a different perspective for comparing

[2]See Altissimo et al. (2009), Mumtaz et al. (2009), and Pesaran and Chudik (2013) for the common factor structures of sectoral inflation series in European countries.

the persistence between an aggregate series and its underlying disaggregated series. It is well-known that the aggregation procedure plays an important role in causing discrepancies in persistence. The theoretical analysis of Granger (1980) showed that the persistence property can be drastically different between an aggregate series and its underlying disaggregated series.[3] Zaffaroni (2004) extended the theoretical analysis by also considering the role of (serially uncorrelated) common factors, and Altissimo et al. (2009) applied the framework to explain the difference in persistence between aggregate inflation and sectoral inflation in France, Germany, and Italy. Pesaran and Chudik (2013) further allowed for serial correlation in common factors and applied the theoretical results to the same inflation data set of Altissimo et al. (2009). Based on a direct comparison of the empirical results, Pesaran and Chudik (2013) concluded that common factor dynamics also play an important role in explaining the discrepancy in persistence when combined with the heterogeneity in sectoral inflation dynamics. We share the same perspective with Pesaran and Chudik (2013) in terms of the role of serially correlated common factors, and it is the prime motivation for revisiting the existing analysis using the DHP+CSD model that allows for common factor dynamics. This chapter, however, differs from Pesaran and Chudik (2013) in several aspects. The theoretical approach of the DHP+CSD estimator is based on the iterative principal component analysis as described in Chapter 1. In contrast, Pesaran and Chudik (2013) relied on the Common Correlated Effects (CCE) approach. The basic idea of their approach is to include the cross-sectional averages of observables and their lags as additional regressors to control for the space spanned by the unobservable common factors. It was initially proposed by Pesaran (2006) in the context of static panels, but Pesaran and Chudik (2013) suggest including more lags of cross-sectional averages in the case of dynamics panels. In terms of the data used, Pesaran and Chudik (2013) focused on the sectoral inflation data in European countries to directly compare with the empirical results in Altissimo et al. (2009). This chapter focuses on the dynamics of sectoral inflation in the

[3] Robinson (1978) also pointed out a similar observation earlier.

U.S. to compare the results with the existing literature on the U.S. economy.

The specification of the DHP+CSD model is sufficiently general to encompass the disaggregated versions of potentially many different models of inflation dynamics. For example, a sectoral version of the accelerationist Phillips curve becomes a DHP+CSD model if the deviation from the nonaccelerating inflation rate of unemployment (NAIRU) has heterogeneous effects and/or if the supply-shifter at the sectoral level has a common factor structure. As another example, a sectoral version of the New Keynesian Phillips Curves (NKPC) 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 contrast, series-by-series AR dynamics driven by independent innovation terms are not suitable as a reduced-form representation of these models of inflation dynamics. This illustrates the usefulness of the DHP+CSD model that can better serve as a reduced-form model of inflation dynamics.

In this chapter, we also aim to perform a structural analysis on the dynamics of inflation using the sectoral NKPC derived in Imbs et al. (2011) (hereafter abbreviated as IJP). We first show that the DHP+CSD model can be a reduced-form representation of the sectoral NKPC as briefly described above. However, the structural parameters in the sectoral NKPC are not readily identifiable from the reduced-form estimates of the DHP+CSD because, in general, the mapping is not one-to-one. To complete the mapping, we rely on the structural assumptions imposed by IJP. These assumptions correspond to the partial indexation scheme proposed by Smets and Wouters (2003), which is widely used in the NKPC literature to derive the *hybrid* form of NKPC where both lagged inflation and expected inflation are present. This scheme essentially imposes further structure between the backward-looking and forward-looking parameters in the NKPC, and this tighter parameterization enables the identification of both backward- and forward-looking parameters from the reduced-form estimates of the DHP+CSD model.

Given the recovered backward-looking parameters, we investigate the degree of intrinsic persistence of inflation at the sectoral level. Intrinsic persistence is a useful concept for identifying the structural source of inflation persistence: the persistence of inflation is either inherited from the driving process (such as the real marginal cost process) or intrinsic to the inflation series itself through the lagged inflation. The knowledge of the degree of intrinsic persistence is important in determining the optimal path of the output-gap-adjusted price level, and thus crucial for monetary policy making.[4] In the literature based on aggregate inflation, a wide variety of models allow for intrinsic persistence in their specifications.[5] Therefore, it is an interesting question whether the inflation dynamics at the sectoral level also exhibit a substantial degree of intrinsic persistence. We attempt to answer this question through the aforementioned framework, which consists of the reduced-form estimates of the DHP+CSD and their mapping to the structural parameters in the sectoral NKPC model.

This chapter shares the basic motivation of IJP, who argue that heterogeneity is the key in explaining the discrepancy between the NKPC at both the aggregate level and the sectoral level. Indeed, we rely on their version of the sectoral NKPC model as our structural model of inflation dynamics. Our analysis, however, differs from that of IJP in several aspects. We focus on the intrinsic persistence of sectoral NKPC, whereas IJP also covered the frequency of price changes and the effect of real marginal costs. In terms of estimation, IJP used the labor market data at the sectoral level as a proxy for the sectoral real marginal costs and performed maximum likelihood estimations. In contrast, we use the DHP+CSD estimator to obtain the reduced-form estimates of the dynamics and then recover the intrinsic persistence parameters by utilizing structural assumptions. In doing so, we treat the sectoral real marginal costs as latent processes that are controlled for by the factor structure of the DHP+CSD estimator. This still serves our analysis purposes because we are interested in the intrinsic persistence parameters. The benefit of treating the real marginal costs as latent is that we can consider a large number of sectors at a highly disaggregated level; the proxy measures from labor market information are typically

[4]See Woodford (2010) and reference therein.

[5]See, for example, Woodford (2007) and references therein for a survey of structural models for explaining the microfoundations of intrinsic persistence.

limited in their availability and their levels of disaggregation. Finally, IJP used French data to utilize the labor market data at a disaggregated level, whereas we use U.S. personal consumption expenditure price inflation at a highly disaggregated level.

Our estimation results show that the intrinsic persistence at the sectoral level is considerably lower on average than the intrinsic persistence of aggregate inflation. That is, the price-setting at the sectoral level appears to be more forward-looking than suggested in the literature. In approximately half of all sectors in the U.S. economy, the implied price-setting behavior is close to purely forward-looking. Moreover, the estimated intrinsic persistence is found to be highly heterogeneous across sectors. The pattern is in line with the findings of IJP, who argue that the aggregation of sectoral NKPC with heterogeneous persistence exaggerates the persistence of inflation dynamics at the aggregate level.

The remainder of the chapter is organized as follows. Section 2 establishes the mapping between the DHP+CSD model and the sectoral NKPC model. Section 3 describes the sectoral inflation data set used in the analysis. Section 4 explains our empirical strategy for estimating the sectoral inflation dynamics, and Section 5 presents the empirical results. Finally, Section 6 concludes.

3.2 Mapping between DHP+CSD and sectoral NKPC

In this section, we show that the DHP+CSD model developed in Chapter 1 is a reduced-form representation of the sectoral NKPC. A reader who is only interested in the reduced-form analysis may skip this section.

We start from the sectoral NKPC derived in IJP:

$$\pi_{it} = \gamma_i^b \pi_{i,t-1} + \gamma_i^f E_t \pi_{i,t+1} + \kappa_i \widehat{mc}_{it} + \xi_{it} \tag{3.1}$$

where π_{it} is sector i's inflation at time t, $\widehat{mc}_{it}$ is the log-deviation of the real marginal cost, and ξ_{it} is typically interpreted as markup shocks. The NKPC model basically describes the dynamics of inflation in each sector in terms of lagged inflation, the expectation of future inflation, and real marginal cost. The parameters γ_i^b and γ_i^f capture the degree to which the pricing decisions are backward-looking and forward-looking, respectively. The effect of the real marginal costs is captured by the parameter κ_i. As this chapter focuses on intrinsic persistence, γ_i^b is the primary parameter of interest.

Each sector's NKPC equation is a rational expectations model that can be solved as

$$\pi_{it} = \delta_{1i}\pi_{i,t-1} + \left(\frac{1}{\delta_{2i}\gamma_i^f}\right)\sum_{k=0}^{\infty}\left(\delta_{2i}^{-1}\right)^k E_t\left[\kappa_i \widehat{mc}_{i,t+k} + \xi_{i,t+k}\right] \tag{3.2}$$

where $\delta_{1i} = \frac{1-\sqrt{1-4\gamma_i^b\gamma_i^f}}{2\gamma_i^f} < 1$ and $\delta_{2i} = \frac{1+\sqrt{1-4\gamma_i^b\gamma_i^f}}{2\gamma_i^f} > 1$.

Regarding the sectoral real marginal costs, we assume a factor structure

$$\widehat{mc}_{it} = a_i \widehat{\mathbf{mc}}_t + v_{it} \tag{3.3}$$

where $\widehat{\mathbf{mc}}_t$ 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. Moreover, sector-specific shocks are allowed to be correlated across sectors with similar structure or characteristics. 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. Given the specification of sectoral NKPC, the factor structure of the real marginal costs is consistent with numerous existing empirical studies that have found a factor structure in the sectoral inflation data (see, for example, Clark, 2006; Reis and Watson, 2010). Assuming a factor structure in $\widehat{mc}_{it}$ has an additional benefit of avoiding the measurement issue of the proxies for the real marginal costs.[6] Instead of using 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

[6]See, for example, Rudd and Whelan (2005); Mazumder (2010) for discussions on the proxy issue.

persistence due to the aggregation procedure. However, the number of sectors has been limited in 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 thus free from this constraint. Notably, it is also possible to 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

$$\begin{aligned}\pi_{it} &= \delta_{1i}\pi_{i,t-1} + \left(\frac{1}{\delta_{2i}\gamma_i^f}\right)\sum_{k=0}^{\infty}\left(\delta_{2i}^{-1}\right)^k E_t\left[\kappa_i a_i \widehat{\mathbf{mc}}_{t+k} + \kappa_i v_{i,t+k} + \xi_{i,t+k}\right] && (3.4)\\ &= \delta_{1i}\pi_{i,t-1} + \left(\frac{\kappa_i a_i}{\delta_{2i}\gamma_i^f}\right)\sum_{k=0}^{\infty}\left(\delta_{2i}^{-1}\right)^k E_t\widehat{\mathbf{mc}}_{t+k} + \left(\frac{1}{\delta_{2i}\gamma_i^f}\right)\sum_{k=0}^{\infty}\left(\delta_{2i}^{-1}\right)^k E_t\left[\kappa_i v_{i,t+k} + \xi_{i,t+k}\right] && (3.5)\end{aligned}$$

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. In the literature, the markup shock is typically assumed to be independent over time such that $E_t\xi_{i,t+k} = 0$ for all future terms.[7] Consequently, only the current term ξ_{it} survives. 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 $\widehat{\mathbf{mc}}_t$. For illustration purposes, suppose that

[7]Schorfheide (2008) discusses the difficulty of identifying the intrinsic persistence γ_i^b when the markup shock is allowed to be serially correlated.

$\widehat{\mathbf{mc}}_t$ follows a simple AR(1) dynamics of $\widehat{\mathbf{mc}}_t = \phi \widehat{\mathbf{mc}}_{t-1} + u_t$ where the innovation process u_t is orthogonal to the sector-specific shocks and the markup shocks. This specification leads to a representation of sectoral inflation without the expectation of the future terms as follows:

$$\pi_{it} = \delta_{1i}\pi_{i,t-1} + \left(\frac{\kappa_i a_i}{\delta_{2i}\gamma_i^f}\right)\left\{\sum_{k=0}^{\infty}\left(\frac{\phi}{\delta_{2i}}\right)^k\right\}\widehat{\mathbf{mc}}_t + \left(\frac{\kappa_i v_{it} + \xi_{it}}{\delta_{2i}\gamma_i^f}\right). \tag{3.6}$$

As $|\phi| < 1$ and $\delta_{2i} > 1$, the infinite sum converges to $(1 - \phi/\delta_{2i})^{-1}$. Finally, we have a DHP+CSD model

$$\pi_{it} = \delta_{1i}\pi_{i,t-1} + \lambda_i F_t + \varepsilon_{it} \tag{3.7}$$

where $F_t = \widehat{\mathbf{mc}}_t$, $\lambda_i = \frac{\kappa_i a_i}{\delta_{2i}\gamma_i^f(1-\phi/\delta_{2i})}$ and $\varepsilon_{it} = \frac{\kappa_i v_{it} + \xi_{it}}{\delta_{2i}\gamma_i^f}$. 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 $\widehat{\mathbf{mc}}_t$, it is easy to show that we still have a static factor representation where the static factor F_t spans the space spanned by $\widehat{\mathbf{mc}}_t$ and its lags. In such cases, the static factor F_t is no longer a scalar but rather is an $r \times 1$ vector.

Given the mapping $\delta_{1i} = \frac{1-\sqrt{1-4\gamma_i^b\gamma_i^f}}{2\gamma_i^f}$ from the rational expectations solution, we cannot separately identify γ_i^b and γ_i^f from the reduced-form coefficient δ_{1i}. To complete the mapping, further structural assumptions are required. In this chapter, we rely on the assumptions of the underlying structural model behind the sectoral NKPC. 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 is 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 the backward-looking parameter $\gamma_i^b = \frac{\omega_i}{1+\beta\omega_i}$, where the structural parameter ω_i is the fraction of firms that index to past inflation. The household discount factor β can be calibrated following the widely used values in the literature. Then, there exists a one-to-one mapping between γ_i^b and ω_i. The coefficient γ_i^f of the forward-looking term has an expression of $\gamma_i^f = \frac{\beta}{1+\beta\omega_i}$.[8] In essence, the partial indexation scheme imposes tight parameterization between γ_i^b and γ_i^f such that they can be separately identified from the reduced-form parameter estimate δ_{1i}. The specific expressions for γ_i^b and γ_i^f imply that the reduced-form persistence δ_{1i} is equivalent to the degree of partial indexation, ω_i. Thus, it is straightforward to see that if no firm follows the indexation scheme, the dynamics of inflation are solely driven by the real marginal costs with lagged inflation, and thus all persistence is inherited from the real marginal costs.

3.3 Data

The data set used in the analysis is a large panel of inflation series at a highly disaggregated level. We use the same data set as the one used by Reis and Watson (2010) so that the analysis can be insulated from possible quality issues of the disaggregated data set.[9] The original data set comes from the underlying detail tables of Personal Consumption Expenditures by type of product in the National Income and Product Accounts. The observations are available at monthly frequency, but we use quarterly data for comparability with most other studies in the literature. The annualized quarterly inflation is calculated as $\pi_{it} = 4 \times 100 \times ln\left(\frac{P_{it}}{P_{i,t-1}}\right)$, where P_{it} correspond to the observations in March, June, September, and December. Several filtering criteria are applied to the data in selecting which inflation series to include in the analysis. Any series with no price change for more than five years are excluded from the analysis. Among any series that have correlations higher than 0.99 at the level or in the first difference, only one series are included for the analysis. Obvious outliers are replaced by the centered seven-quarter local median values. This filtering process results in 187 inflation series at a highly disaggregated level, covering various categories of durables, non-durables, and services. The sample runs from the first quarter of 1959 to the second quarter of 2006.

[8] See IJP for further details of the model.

[9] We are grateful to Reis and Watson for making their data set publicly available on the journal website.

3.4 Empirical strategy

We estimate the dynamics of inflation at the sectoral level using the DHP+CSD estimator developed in Chapter 1:

$$\pi_{it} = \delta_{1i}\pi_{i,t-1} + \lambda_i' F_t + \varepsilon_{it} \tag{3.8}$$

where F_t denotes an $r \times 1$ vector of unobservable common factors, λ_i denotes the corresponding vector of factor loadings that capture the heterogeneous effects of the common factors, and ε_{it} are *i.i.d.* idiosyncratic shocks. As described in the previous section, π_{it} are the annualized quarterly inflation rates for various categories of durables, non-durables, and services. The model (3.8) extends the specification of Bils and Klenow (2004), who modeled the dynamics of sectoral inflation series as a collection of independent AR(1) processes. That is, the specification (3.8) nests the specification of Bils and Klenow (2004) as a special case under an additional assumption that no factor structure is present. This special case can still be consistently estimated by the DHP+CSD model as long as the AR coefficients are the objects of interest. This is because the DHP+CSD estimator performs in a robust manner even though cross-sectional dependence is not present, as shown in the Monte Carlo simulation study of Chapter 1. Allowing for common factors also makes the specification (3.8) consistent with existing empirical studies where analyses are based on factor models (see, for example, Clark, 2006; Reis and Watson, 2010).

The estimation is performed by iterating the two steps of estimation: regression and factor extraction. In the regression step, we simply run regressions to estimate the AR coefficients δ_{1i} given factors as known. In doing so, we impose a restriction of $\delta_{1i} \geq 0$, which is implied by a structural restriction that the partial indexation parameter should be $0 \leq \omega_i \leq 1$. In the factor extraction step, we use a principal component method to estimate the common factors F_t given AR coefficients δ_{1i}.[10] By iterating the two steps until

[10] Note that, given AR coefficients, $\pi_{it} - \delta_{1i}\pi_{i,t-1}$ can be treated as a panel of observables that are generated from a pure factor model.

convergence, we can obtain the estimates for both the AR coefficients and the common factors. The estimates for the loadings are obtained as a by-product of the factor extraction step. The iteration may start from any initial values. One possibility is to begin with running regressions with lagged inflation only, as if the factors are not present. Although the initial estimates are biased due to the omitted variable, the solution after the convergence is valid. We can also begin by extracting factors first, as if no lagged inflation is relevant in the equation. The initial factor estimates do not span the correct space because the sectoral inflation series are not from a pure factor model, but the final estimates at convergence will still be valid. However, the number of iterations may substantially differ depending on how we begin the iteration.

In the existing literature, the number of static factors (r) is assumed to be known.[11] For the actual estimation, however, choice of the dimension of the static factor F_t is required. Numerous methods are proposed in the literature for determining the number of static factors in the pure factor model.[12] The statistical method for determining the number of factors has not yet been studied in the context of the DHP+CSD model. In the analysis, we choose $r = 4$ as our benchmark to allow for a reasonable degree of dynamics in the economy-wide real marginal cost process. In addition, we perform a robustness check for different numbers of common factors up to seven.

Given the estimates of δ_{1i} from the DHP+CSD estimation, we recover the backward-looking parameter γ_i^b by using the mapping to the sectoral NKPC as described in Section 2. The household discount factor is calibrated as $\beta = 0.99$ following the existing literature. The intrinsic persistence parameters γ_i^b are calculated from the rational expectations solution $\delta_{1i} = \frac{1-\sqrt{1-4\gamma_i^b\gamma_i^f}}{2\gamma_i^f}$, where $\gamma_i^b = \frac{\omega_i}{1+\beta\omega_i}$ and $\gamma_i^f = \frac{\beta}{1+\beta\omega_i}$ follows from the parameterization of

[11]The approach in Pesaran and Chudik (2013) does not require the knowledge of the number of factors. Instead, the method requires choosing the number of lags to include for the cross-sectional averages of observables in the auxiliary regression.

[12]See Bai and Ng (2002), Hallin and Liska (2007), Amengual and Watson (2007), and Onatski (2010), among others.

the partial indexation scheme.

3.5 Estimation results

Figure 3.1 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. The estimated $\hat{\delta}_{1i}$ values are scattered widely from low to high values reaching nearly up to 0.9. Under such pattern of large heterogeneity, the persistence of the aggregate inflation may not represent the persistence of its underlying sectoral inflation, as observed by Granger (1980) and Zaffaroni (2004). Our estimation results lead to the same conclusion as in Bils and Klenow (2004) and Clark (2006); inflation exhibits low persistence at the sectoral level, and the persistence of the aggregate inflation appears to be exaggerated by the aggregation procedure. Altissimo et al. (2009) and Pesaran and Chudik (2013) draw a similar conclusion by analyzing disaggregated inflation data for European countries.

Figure 3.1: Histogram of the estimated reduced-form persistence of sectoral inflation

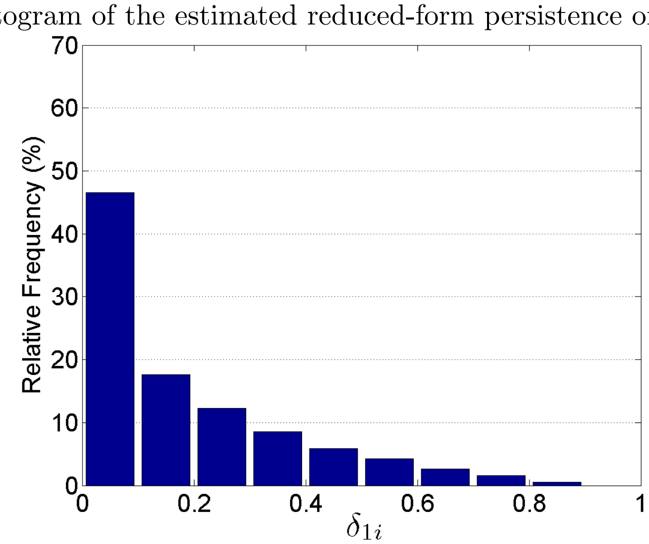
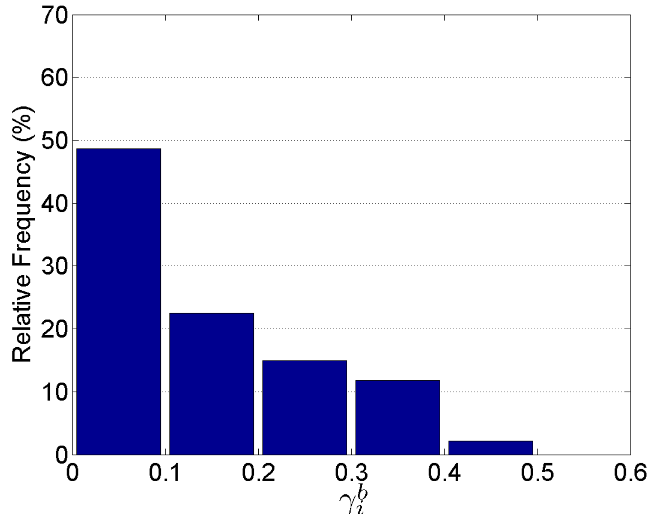


Figure 3.2 presents the estimated intrinsic persistence parameter $\hat{\gamma}_i^b$ for all sectors. The intrinsic persistence parameter estimates are calculated from the reduced-form estimates $\hat{\delta}_{1i}$ by using the mapping described the previous sections. The histogram of the $\hat{\gamma}_i^b$ estimates

Figure 3.2: Histogram of the estimated intrinsic persistence of sectoral inflation



reveals that there is substantial heterogeneity in intrinsic persistence as well; the sectors with high intrinsic persistence coexist in the sectoral inflation data with those sectors that "inherit" most of their persistence from the real marginal costs. It is worth noting that the price-setting behavior is close to purely forward-looking in approximately half of all sectors in the U.S. economy. The average intrinsic persistence at the sectoral level is only around the value of 0.13. Low average intrinsic persistence combined with high cross-sectional heterogeneity is in line with the findings of other empirical studies on the sectoral NKPC. IJP find that the average intrinsic persistence is around 0.3, and that there exists large heterogeneity across sectors. Their analysis is based on the estimated NKPC in 16 sectors of the French economy. Leith and Malley (2007) focus on the NKPC of 18 manufacturing sectors in the U.S. economy. The average of their estimated intrinsic persistence is also around 0.3, and substantial heterogeneity across sectors is again confirmed. IJP argue that the intrinsic persistence of aggregate inflation is measured to be higher than the actual intrinsic persistence at the sectoral level possibly due to the aggregation of heterogeneous processes. Our estimation results are consistent with their argument.

To compare with the intrinsic persistence of aggregate inflation, we conduct a meta-

Table 3.1: Meta-analysis of (mean) intrinsic persistence in the NKPC literature

$N = 1$		$N \simeq 20$		$N \simeq 200$	
Galí and Rabanal (2004)	0.02				
Andrés et al. (2005)	0.50				
Rabanal and Rubio-Ramirez (2005)	0.43	Leith and Malley (2007)	0.30		
Lindé (2005)	0.54				
Bouakez et al. (2005)	0.50				
Salemi (2006)	0.43			DHP+CSD	0.13
Cho and Moreno (2006)	0.44				
Boivin and Giannoni (2006)	0.50				
Fernández-Villaverde and Rubio-Ramírez (2007)	0.13	Imbs et al. (2011)	0.30		
Laforte (2007)	0.40				
Rabanal (2007)	0.50				

Notes: Imbs et al. (2011) is based on the French inflation data, whereas other studies are based on U.S. inflation data.

analysis of the literature on the aggregate NKPC. Table 3.1 reports the intrinsic persistence estimates from a number of selected empirical studies listed in a survey by Schorfheide (2008). These empirical studies rely on the assumption of treating the real marginal costs as latent processes, so does our empirical framework. The intrinsic persistence estimates in a majority of these studies are around 0.4 or higher. IJP also used the value 0.4 as a typical estimate of intrinsic persistence in aggregate NKPC literature to compare their estimates at the sectoral level. Table 3.1 also summarizes other empirical studies at different levels of aggregation. Analyzing the sectoral NKPC that consists of less than 20 sectors, both IJP and Leith and Malley (2007) discover that the estimated intrinsic persistence parameters are on average lower than the estimates from the aggregate NKPC literature. Our analysis involves a much more disaggregated data set of inflation covering 187 categories of goods and services. Our estimate of the average intrinsic persistence at value 0.13 turns out to be far lower than those reported in the aforementioned empirical studies. Therefore, there exists an interesting pattern that the estimated intrinsic persistence is inversely associated with the degree of disaggregation. Such an intriguing pattern is supportive of the claim in

IJP that the price-setting at the disaggregated level tends to be much more forward-looking than suggested by the aggregate NKPC analysis in the literature.

Although it is conventional in the NKPC literature to compare the estimated intrinsic persistence parameter values as in the meta-analysis above, there are pitfalls with the direct comparisons. Most notably, the microfoundations and structural assumptions differ from one model to another. In addition, the inflation data sets used in the analysis may also differ in terms of the time span. Acknowledging such limitations, we adopt an alternative approach to assess the role of aggregation in examining the intrinsic persistence. The approach analyzes the aggregate inflation dynamics within the same empirical framework used for the disaggregated inflation analysis. The basic idea is to utilize the extracted common factors in the disaggregated analysis to control for the economy-wide real marginal cost process in the aggregate NKPC. To illustrate the intuition, suppose we estimate a typical specification of the aggregate NKPC:

$$\pi_t = \gamma_b \pi_{t-1} + \gamma_f E_t \pi_{t+1} + \kappa \widehat{\mathbf{mc}}_t + \xi_t. \tag{3.9}$$

The rational expectations solution of (3.9) expresses the inflation dynamics in terms of the lagged inflation and the discounted sum of future real marginal costs at the economy level as follows:

$$\pi_t = \delta_1 \pi_{t-1} + \underbrace{\left(\frac{\kappa}{\delta_2 \gamma_f}\right) \sum_{k=0}^{\infty} \left(\delta_2^{-1}\right)^k E_t \widehat{\mathbf{mc}}_{t+k}}_{\Gamma' F_t} + \underbrace{\left(\frac{1}{\delta_2 \gamma_i^f}\right) \xi_t}_{e_t}. \tag{3.10}$$

Recall that the discounted sum of future economy-wide marginal costs also appears in our empirical framework for the sector NKPC in (3.5). Under our maintained assumption of a common factor structure in the real marginal costs, the entire sum is controlled for by the unobserved static factors F_t. Consequently, the same infinite sum in (3.10) can be controlled for by the common factor estimates $\hat{F}_t$ that are extracted from the panel of sectoral inflation series. This approach analyzes both the aggregate NKPC and the sectoral NKPC within the same empirical framework using the same factor estimates. Consequently, the consequences of considering the NKPC at different levels of disaggregation can be isolated from other specifications of the analysis.

We use the aggregate PCE inflation data from the NIPA accounts, and estimate the reduced-form parameter δ_1 in (3.10). Then, we recover the parameter of interest γ_b using the same relationship among the parameters previously described. This procedure yields $\hat{\gamma}_b = 0.48$ as the intrinsic persistence estimate of aggregate inflation. The estimate is comparable to the estimates from the aggregate NKPC literature listed in Table 3.1. Note that $\hat{\gamma}_b = 0.48$ is much higher than the average intrinsic persistence of value 0.13 in our sectoral NKPC analysis. Since both estimates are based on the same empirical framework, it can be argued that the difference between the two estimates comes solely from analyzing the aggregate NKPC instead of sectoral NKPC.

3.6 Conclusion

This chapter assesses the persistence of inflation at the sectoral level by using the DHP+CSD model developed in Chapter 1. We show that the general specification of the DHP+CSD serves as a reduced-form representation of the sectoral NKPC model proposed by IJP. We establish a mapping between the reduced-form parameters of the DHP+CSD model and the structural parameters of the sectoral NKPC model by adopting the widely used partial indexation scheme. Through the mapping, we recover the intrinsic persistence parameters from the reduced-form estimates. We find that intrinsic persistence of sectoral inflation is much lower than suggested in the aggregate NKPC literature. The results reveal that the price-setting in approximately half of all sectors in the U.S. economy is close to purely forward-looking. Based on a meta-analysis of the parameter estimates in the literature, we find that the estimated intrinsic persistence is inversely associated with the level of disaggregation. This pattern is in line with the existing conjecture that the high degree of estimated intrinsic persistence of aggregate inflation might be due to the aggregation of heterogeneous sectoral inflation series.

Bibliography

Ahn, S. C., Hoon Lee, Y., and Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101(2):219–255.

Altissimo, F., Mojon, B., and Zaffaroni, P. (2009). Can aggregation explain the persistence of inflation? *Journal of Monetary Economics*, 56(2):231–241.

Amengual, D. and Watson, M. W. (2007). Consistent estimation of the number of dynamic factors in a large N and T panel. *Journal of Business and Economic Statistics*, 25(1):91–96.

Andrés, J., López-Salido, J. D., and Nelson, E. (2005). Sticky-price models and the natural rate hypothesis. *Journal of Monetary Economics*, 52(5):1025–1053.

Anselin, L. (2008). *Spatial econometrics: methods and models*, volume 4. Kluwer Academic Publishers.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, 15(5):453–472.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Baltagi, B. H., Feng, Q., and Kao, C. (2011). Testing for sphericity in a fixed effects panel data model. *The Econometrics Journal*, 14(1):25–47.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly Journal of Economics*, 120(1):387–422.

Bils, M. and Klenow, P. J. (2004). Some evidence on the importance of sticky prices. *Journal of Political Economy*, 112(5):947–985.

Boivin, J. and Giannoni, M. (2006). DSGE models in a data-rich environment. Technical report, National Bureau of Economic Research.

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky prices and monetary policy: Evidence from disaggregated US data. *The American Economic Review*, 92(1):350–384.

Boivin, J. and Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, 132(1):169–194.

Bonhomme, S. and Manresa, E. (2012). Grouped patterns of heterogeneity in panel data. Working papers, CEMFI.

Bouakez, H., Cardia, E., and Ruge-Murcia, F. J. (2005). Habit formation and the persistence of monetary shocks. *Journal of Monetary Economics*, 52(6):1073–1088.

Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1):239–253.

Calvo, G. A. (1983). Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398.

Carvalho, C. and Lee, J. W. (2011). Sectoral price facts in a sticky-price model. *FRB of New York Staff Report*, (495).

Carvalho, C. and Nechio, F. (2011). Aggregation and the PPP puzzle in a sticky-price model. *The American Economic Review*, 101(6):2391–2424.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51(5):1281–1304.

Chen, J., Gao, J., and Li, D. (2012). A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models. *Econometric Theory*, 28(5):1114–1163.

Chen, S. S. and Engel, C. (2005). Does aggregation bias explain the PPP puzzle? *Pacific Economic Review*, 10(1):49–72.

Cho, S. and Moreno, A. (2006). A small-sample study of the New-Keynesian macro model. *Journal of Money, Credit and Banking*, 38(6):1461–1481.

Choi, C. Y., Mark, N. C., and Sul, D. (2006). Unbiased estimation of the half-life to PPP convergence in panel data. *Journal of Money, Credit, and Banking*, 38(4):921–938.

Chudik, A. and Pesaran, M. H. (2011). Infinite-dimensional VARs and factor models. *Journal of Econometrics*, 163(1):4–22.

Chudik, A. and Pesaran, M. H. (2013). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Technical report, Mimeo, March.

Clark, T. E. (2006). Disaggregate evidence on the persistence of consumer price inflation. *Journal of Applied Econometrics*, 21(5):563–587.

Cogley, T. and Sargent, T. J. (2001). Evolving post-World War II US inflation dynamics. In *NBER Macroeconomics Annual 2001, Volume 16*, pages 331–388. MIT Press.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics*, 8(2):262–302.

Crowder, W. J. and Hoffman, D. L. (1996). The long-run relationship between nominal interest rates and inflation: the Fisher equation revisited. *Journal of Money, Credit and Banking*, 28(1):102–118.

Doz, C., Giannone, D., and Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1):188–205.

Evans, M. and Wachtel, P. (1993). Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit and Banking*, 25(3):475–511.

Evans, M. D. and Lewis, K. K. (1995). Do expected shifts in inflation affect estimates of the long-run Fisher relation? *The Journal of Finance*, 50(1):225–253.

Fair, R. C. (2000). Testing the NAIRU model for the United States. *Review of Economics and Statistics*, 82(1):64–71.

Fan, J., Liao, Y., and Mincheva, M. (2011a). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320–3356.

Fan, J., Liao, Y., and Mincheva, M. (2011b). Large covariance estimation by thresholding principal orthogonal complements. *Available at SSRN 1977673*.

Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2007). How structural are structural parameters? In *NBER Macroeconomics Annual 2007, Volume 22*, pages 83–137. University of Chicago Press.

Fisher, M. E. and Seater, J. J. (1993). Long-run neutrality and superneutrality in an ARIMA framework. *The American Economic Review*, 83(3):402–415.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005). The generalized dynamic factor model. *Journal of the American Statistical Association*, 100(471):830–840.

Frankel, J. A. and Rose, A. K. (1996). A panel project on purchasing power parity: mean reversion within and between countries. *Journal of International Economics*, 40(1):209–224.

Frees, E. W. (1995). Assessing cross-sectional correlation in panel data. *Journal of econometrics*, 69(2):393–414.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701.

Fuhrer, J. C. (2010). Inflation persistence. volume 3 of *Handbook of Monetary Economics*, pages 423 – 486. Elsevier.

Galí, J. and Rabanal, P. (2004). Technology shocks and aggregate fluctuations: How well does the real business cycle model fit postwar US data? In *NBER Macroeconomics Annual 2004, Volume 19*, pages 225–318. MIT Press.

Garcia, R. and Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78(1):111–125.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *D.J. Aigner, A.S. Goldberger (Eds.), Latent variables in socio-economic models, North-Holland, Amsterdam.*

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14(2):227–238.

Hallin, M. and Liska, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, 102(478):603–617.

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115(1):53–74.

Imbs, J., Jondeau, E., and Pelgrin, F. (2011). Sectoral Phillips curves and the aggregate Phillips curve. *Journal of Monetary Economics*, 58(4):328–344.

Imbs, J., Mumtaz, H., Ravn, M. O., and Rey, H. (2005). PPP strikes back: Aggregation and the real exchange rate. *The Quarterly Journal of Economics*, 120(1):1–43.

Killian, L. and Zha, T. (1999). Quantifying the half-life of deviations from PPP. *Journal of Applied Econometrics*, 17(2):107–25.

Kim, C.-J. (1993). Unobserved-component time series models with Markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty. *Journal of Business & Economic Statistics*, 11(3):341–349.

King, R. and Watson, M. (1997). Testing long-run neutrality. *FRB Richmond Economic Quarterly*, 83(3):69–101.

Laforte, J.-P. (2007). Pricing models: a Bayesian DSGE approach for the US economy. *Journal of Money, Credit and Banking*, 39(s1):127–154.

Leeb, H. and Pötscher, B. M. (2005). Model selection and inference: Facts and fiction. *Econometric Theory*, 21(1):21–59.

Leith, C. and Malley, J. (2007). A sectoral analysis of price-setting behavior in US manufacturing industries. *The Review of Economics and Statistics*, 89(2):335–342.

Levin, A. T. and Lin, C.-F. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1–24.

Lindé, J. (2005). Estimating New-Keynesian Phillips curves: A full information maximum likelihood approach. *Journal of Monetary Economics*, 52(6):1135–1149.

Lothian, J. R. (1997). Multi-country evidence on the behavior of purchasing power parity under the current float. *Journal of International Money and Finance*, 16(1):19–35.

Mazumder, S. (2010). The New Keynesian Phillips curve and the cyclicality of marginal cost. *Journal of Macroeconomics*, 32(3):747–765.

Moon, H. and Weidner, M. (2010). Dynamic linear panel regression models with interactive fixed effects. *Unpublished manuscript, University of Southern California*.

Mumtaz, H., Zabczyk, P., and Ellis, C. (2009). What lies beneath: what can disaggregated data tell us about the behaviour of prices? Bank of England working papers 364, Bank of England.

Murray, C. J. and Papell, D. H. (2005). Do panels help solve the purchasing power parity puzzle? *Journal of Business & Economic Statistics*, 23(4):410–415.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111–2245. Elsevier.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6):1519–1554.

Obstfeld, M. and Rogoff, K. (2001). The six major puzzles in international macroeconomics: Is there a common cause? In *NBER Macroeconomics Annual 2000, Volume 15*, pages 339–412. MIT press.

Oh, K.-Y. (1996). Purchasing power parity and unit root tests using panel data. *Journal of International Money and Finance*, 15(3):405–418.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016.

Pesaran, M. (2004). General diagnostic tests for cross section dependence in panels. Technical report, Faculty of Economics, University of Cambridge.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

Pesaran, M. H. and Chudik, A. (2013). Aggregation in large dynamic panels. *Forthcoming in Journal of Econometrics*.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of econometrics*, 68(1):79–113.

Pesaran, M. H., Ullah, A., and Yamagata, T. (2008). A bias-adjusted LM test of error cross-section independence. *The Econometrics Journal*, 11(1):105–127.

Phillips, P. C. B. and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross-section dependence. *Econometrics Journal*, 6(1):217–259.

Pivetta, F. and Reis, R. (2007). The persistence of inflation in the united states. *Journal of economic Dynamics and Control*, 31(4):1326–1358.

Rabanal, P. (2007). Does inflation increase after a monetary policy tightening? answers based on an estimated DSGE model. *Journal of Economic Dynamics and Control*, 31(3):906–937.

Rabanal, P. and Rubio-Ramirez, J. F. (2005). Comparing New Keynesian models of the business cycle: A Bayesian approach. *Journal of Monetary Economics*, 52(6):1151 – 1166.

Reis, R. and Watson, M. W. (2010). Relative goods' prices, pure inflation, and the Phillips correlation. *American Economic Journal: Macroeconomics*, 2(3):128–57.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, 5(3):163–168.

Rogoff, K. (1996). The purchasing power parity puzzle. *Journal of Economic Literature*, 34(2):647–668.

Rossi, B. (2005). Confidence intervals for half-life deviations from purchasing power parity. *Journal of Business & Economic Statistics*, 23(4):432–442.

Rudd, J. and Whelan, K. (2005). Does labor's share drive inflation? *Journal of Money, Credit and Banking*, 37(2):297–312.

Salemi, M. K. (2006). Econometric policy evaluation and inverse control. *Journal of Money, Credit, and Banking*, 38(7):1737–1764.

Sarafidis, V., Yamagata, T., and Robertson, D. (2009). A test of cross-section dependence for a linear dynamic panel model with regressors. *Journal of Econometrics*, 148(2):149–161.

Sargan, J. D. (1964). Wages and prices in the united kingdom: a study in econometric methodology. *Econometric analysis for national economic planning*, 16:25–54.

Sargent, T. J. and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. Technical report.

Schorfheide, F. (2008). DSGE model-based estimation of the new Keynesian Phillips curve. *Economic Quarterly*, 94(4):397–433.

Smets, F. and Wouters, R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123–1175.

Stock, J. H. (2001). Evolving post-World War II US inflation dynamics: Comment. pages 379–387.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for VAR analysis. Technical report, National Bureau of Economic Research.

Woodford, M. (2007). Interpreting inflation persistence: Comments on the conference on "quantitative evidence on price determination". *Journal of Money, Credit and Banking*, 39:203–210.

Woodford, M. (2010). Optimal monetary stabilization policy. volume 3 of *Handbook of Monetary Economics*, pages 723–828. Elsevier.

Wu, Y. (1996). Are real exchange rates nonstationary? evidence from a panel-data test. *Journal of Money, Credit and Banking*, 28(1):54–63.

Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics*, 120(1):75–102.