Efficient estimation of heterogenous coefficients in panel data models with common shocks*

Kunpeng Li† and Lina Lu‡

October 11, 2014

Abstract

This paper investigates efficient estimation of heterogeneous coefficients in panel data models with common shocks, which have been a particular focus of recent theoretical and empirical literature. We propose a new two-step method to estimate the heterogeneous coefficients. In the first step, the maximum likelihood (ML) method is first conducted to estimate the loadings and idiosyncratic variances. The second step estimates the heterogeneous coefficients by using the structural relations implied by the model and replacing the unknown parameters with their ML estimates. We establish the asymptotic theory of our estimator, including consistency, asymptotic representation, and limiting distribution. The two-step estimator is asymptotically efficient in the sense that it has the same limiting distribution as the infeasible generalized least squares (GLS) estimator. Intensive Monte Carlo simulations show that the proposed estimator performs robustly in a variety of data setups.

Key Words: Factor analysis; Block diagonal covariance; Panel data models; Common shocks; Maximum likelihood estimation, heterogeneous coefficients; Inferential theory

*We thank Jushan Bai, Bernard Salanié, and participants in seminars at Columbia University, Huazhong University of Science and Technology, Renmin University of China, and University of International Business and Economics for their helpful comments.

†International School of Economics and Management, Capital University of Economics and Business, Beijing, China. Email: likp.07@sem.tsinghua.edu.cn.

‡Department of Economics, Columbia University, New York, NY, 10027.

1 Introduction

It has been long recognized and well documented in the literature that a small number of factors can explain a large fraction of the comovement of financial, macroeconomic and sectorial variables, for example, Ross (1976), Sargent and Sims (1977), Geweke (1977) and Stock and Watson (1998). Based on this fact, recent econometric literature places particular focus on panel data models with common shocks. These models specify that the dependent variable and explanatory variables both have a factor structure. A typical example can be written as

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \lambda_i' f_t + \epsilon_{it}, \\ x_{it} &= \nu_i + \gamma_i' f_t + v_{it}, \qquad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T. \end{aligned} \tag{1.1}$$

where y_{it} denotes the dependent variable; x_{it} denotes a $k \times 1$ vector of explanatory variables; and f_t is an $r \times 1$ vector of unknown factors, which represents the unobserved economic shocks. The factor loadings γ_i and λ_i capture the heterogeneous responses to the shocks. A salient feature of this paper is that the coefficients of x_{it} are assumed to be individual-dependent.

Due to the presence of factor f_t, the error term of the y equation (i.e., $\lambda_i' f_t + \epsilon_{it}$) is correlated with the explanatory variables. The usual estimation methods, such as ordinary least squares method, are not applicable. The instrumental variables (IV) method appears to be an intuitive way to address this issue, but the validity of IV is difficult to justify in practice. A remarkable result from recent studies is that, even without IV, model (1.1) can still be consistently estimated. The related literature includes Pesaran (2006), Bai (2009), Moon and Weidner (2012), Song (2013) and Bai and Li (2014), among others.

Bai (2009) proposes the iterated principal components (PC) method to estimate a model with homogeneous coefficients. His analysis has been reexamined and extended by perturbation theory in Moon and Weidner (2012). Both studies find that a bias arises from cross-sectional heteroscedasticity. Bai and Li (2014) therefore consider the quasi maximum likelihood method to eliminate this bias from the estimator. All these studies focus on the case of homogeneous coefficient. If the underlying coefficients are heterogeneous, misspecification of homogeneity would lead to inconsistent estimation (see the simulation of Kapetanios et al. (2011)).

There are several studies on the estimation of heterogeneous coefficients. Pesaran (2006) proposes the common correlated effect (CCE) estimation method to estimate the heterogeneous coefficients (1.1). The intuition of his method is approximating the unknown projection space of the factors f_t by the space spanned by the cross-sectional average of the observations $(y_{it}, x_{it}')'$. To this end, some rank condition is needed. Song (2013) alternatively considers the iterated principal components method, which extends the analysis of Bai (2009) to the case of heterogeneous coefficients. In this paper, we propose a new method to estimate (1.1). Our estimation method is motivated by both Pesaran's and Song's methods having their limitations in estimating the heterogeneous coefficients for some particular data setups. The CCE estimator has a reputation for computational simplicity and excellent finite sample properties. However, we note that in some cases rank condition alone is not enough for a good approximation. When good approximation breaks down, the CCE estimator would perform poorly. With Song's method, although his theory is beautiful, the minimizer of the objective function is not easily obtained, especially for the data with heavy cross-sectional heteroscedasticity. As far as we know, there is no good way to address this issue. The limitations of the CCE method and the iterated principal components method are manifested by simulations in Section 6.

Our estimation method is a two-step method. In the first step, we use the maximum likelihood (ML) method to estimate a pure factor model. Next, the heterogeneous coefficients are estimated by using relations implied by the model and replacing the parameters with their ML estimates. The proposed estimation method aims to strike a balance between efficiency and computational economy. We note that in model (1.1) the computational burden cannot be ignored due to a huge number of βs being estimated, especially when N is large. This problem is made worse because we can only compute β_i $(i = 1, 2, \dots, N)$ sequentially, instead of all β_i simultaneously by matrix algebra. As a result, the iterated computation method, which requires updating β_i one by one in each iteration, may not be attractive because of the heavy computational burden. Our estimation method overcomes this problem by using the iterated computation method to estimate a pure factor model, delaying the estimation of β_i to the second step. Nevertheless, as we will show, the two-step estimators are asymptotically efficient.

The rest of the paper is organized as follows. Section 2 illustrates the idea of our estimation. Section 3 presents some theoretical results of the factor models, in which the covariance matrix of idiosyncratic errors are block-diagonal. These results are very useful for the subsequent analysis. Section 4 presents the asymptotic properties of the proposed estimator. Section 5 extends our method to the case with zero restrictions on the loadings in the y equation. We show that when zero restrictions are present, the loadings contain information for β. We propose a minimum distance estimator to achieve the efficiency. Section 6 extends the model to nonzero restrictions. Section 7 conducts extensive simulations to investigate the finite sample properties of the proposed estimator and provides some comparisons with the competitors. Section 8 concludes. Throughout the paper, the norm of a vector or matrix is that of Frobenius; that is, $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$ for matrix A. In addition, we use $\dot{v}_t$ to denote $v_t - \frac{1}{T}\sum_{s=1}^{T} v_s$ for any column vector v_t and M_{wv} to denote $\frac{1}{T}\sum_{t=1}^{T} \dot{w}_t \dot{v}_t'$ for any vectors w_t and v_t.

2 Key idea of the estimation

To illustrate the idea of our estimation, first substitute the second equation of model (1.1) into the first one. Then

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \nu_i \end{bmatrix} + \begin{bmatrix} \beta_i'\gamma_i' + \lambda_i' \\ \gamma_i' \end{bmatrix} f_t + \begin{bmatrix} \beta_i' v_{it} + \epsilon_{it} \\ v_{it} \end{bmatrix}.$$

Let $z_{it} = (y_{it}, x_{it}')'$, $\mu_i = (\alpha_i, \nu_i')'$, $u_{it} = (\beta_i' v_{it} + \epsilon_{it}, v_{it}')'$ and Λ_i' be the factor loadings matrix before f_t in the above equation. Now we have

$$z_{it} = \mu_i + \Lambda_i' f_t + u_{it}. \tag{2.1}$$

Let Ω_i be the covariance matrix of v_{it} and $\sigma_{\epsilon i}^2$ the variance of ϵ_{it}. Throughout the paper, we assume that ϵ_{it} is independent of v_{js} for all i, j, t, s. This assumption is crucial to the models with common shocks and is maintained by all the related studies; for example, Bai (2009), Bai and Li (2014), Pesaran (2006), and Moon and Weidner (2012). The covariance of u_{it}, denoted by Σ_{ii}, now is

$$\Sigma_{ii} = \begin{bmatrix} \Sigma_{i,11} & \Sigma_{i,12} \\ \Sigma_{i,21} & \Sigma_{i,22} \end{bmatrix} = \begin{bmatrix} \beta_i'\Omega_i\beta_i + \sigma_{\epsilon i}^2 & \beta_i'\Omega_i \\ \Omega_i\beta_i & \Omega_i \end{bmatrix}. \tag{2.2}$$

This leads to

$$\Sigma_{i,22}\beta_i = \Sigma_{i,21}. \tag{2.3}$$

Suppose that we have obtained a consistent estimator of Σ_{ii}, β_i is then estimated by

$$\hat{\beta}_i = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21} \tag{2.4}$$

We call the above estimator CoVariance estimator, denoted by $\hat{\beta}_i^{CV}$ since the estimation for β_i only involves the covariance of u_{it}.

The remaining problem is to consistently estimate Σ_{ii}. A striking feature of the model (2.1) is that the variance matrix of its idiosyncratic errors is block-diagonal. So we need to extend the usual factor analysis to accommodate this feature.

3 Factor models

Let $i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$. Consider the following factor models

$$z_{it} = \mu_i + \Lambda_i' f_t + u_{it}, \tag{3.1}$$

where z_{it} is a $\bar{K}\times 1$ vector of observations with $\bar{K} = k+1$; u_{it} is a $\bar{K}\times 1$ vector of error terms; Λ_i is an $r\times\bar{K}$ loading matrix; and f_t is an $r\times 1$ vector of factors. Let $z_t = (z_{1t}', z_{2t}', \ldots, z_{Nt}')'$, $\mu = (\mu_1', \mu_2', \ldots, \mu_N')'$, $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_N)'$ and $u_t = (u_{1t}', u_{2t}', \ldots, u_{Nt}')'$, then we can rewrite (3.1) as

$$z_t = \mu + \Lambda f_t + u_t. \tag{3.2}$$

Without loss of generality, we assume that $\bar{f} = T^{-1}\sum_{t=1}^T f_t = 0$ throughout the paper since the model can be rewritten as $z_t = \mu + \Lambda\bar{f} + \Lambda(f_t - \bar{f}) + u_t = \mu^* + \Lambda f_t^* + u_t$ with $\mu^* = \mu + \Lambda\bar{f}$ and $f_t^* = f_t - \bar{f}$. To analyze (3.2), we make the following assumptions:

Assumption A: The factor f_t is a sequence of constant. Let $M_{ff} = T^{-1}\sum_{t=1}^T \dot{f}_t\dot{f}_t'$ with $\dot{f}_t = f_t - T^{-1}\sum_{t=1}^T f_t$. We assume that $\overline{M}_{ff} = \lim_{T\to\infty} M_{ff}$ is a strictly positive definite matrix.

Assumption B: The idiosyncratic error term u_{it} is assumed such that

B.1 u_{it} is independent and identically distributed ($i.i.d$) over t and uncorrelated over i with $E(u_{it}) = 0$ and $E(\|u_{it}^4\|) \leq \infty$ for all $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Let Σ_{ii} be the variance of u_{it} and $\Psi = \text{diag}(\Sigma_{11}, \Sigma_{22}, \ldots, \Sigma_{NN})$ be the variance of u_t.

B.2 f_t is independent of u_{js} for all (j, t, s).

Assumption C: There exists a positive constant C sufficiently large such that

C.1 $\|\Lambda_i\| \le C$ for all $i = 1, \cdots, N$.

C.2 $C^{-1} \le \tau_{min}(\Sigma_{ii}) \le \tau_{max}(\Sigma_{ii}) \le C$ for all $i = 1, \cdots, N$, where $\tau_{min}(\cdot)$ and $\tau_{max}(\cdot)$ denote the smallest and largest eigenvalues of its argument, respectively.

C.3 There exists an $r \times r$ positive matrix Q such that $Q = \lim_{N\to\infty} N^{-1}\Lambda'\Psi^{-1}\Lambda$, where $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_N)'$ and Ψ is the variance of $u_t = (u'_{1t}, u'_{2t}, \ldots, u'_{Nt})'$.

Assumption D: The variances Σ_{ii} for all i are estimated in a compact set; that is, all the eigenvalues of $\hat{\Sigma}_{ii}$ are in an interval $[C^{-1}, C]$ for sufficiently large constant C.

Assumptions A-D are usually made in the context of factor analysis; for example, Bai and Li (2012a, 2014). Readers are referred to Bai and Li (2012a) for the related discussions on these assumptions.

3.1 Estimation

The objective function used to estimate (3.2) is

$$\ln \mathscr{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}[M_{zz}\Sigma_{zz}^{-1}] \tag{3.3}$$

where $\theta = (\Lambda, \Psi, M_{ff})$ and $\Sigma_{zz} = \Lambda M_{ff}\Lambda' + \Psi$; $M_{zz} = \frac{1}{T}\sum_{t=1}^{T} \dot{z}_t\dot{z}_t'$ is the data matrix where $\dot{z}_t = z_t - \frac{1}{T}\sum_{s=1}^{T} z_s$. Suppose that f_t is random and follows $N(0, M_{ff})$, the above objective function is the corresponding likelihood function after concentrating out the intercept μ. Although the factors f_t are assumed to be fixed constants, we still use the above objective function and call the maximizer $\hat{\theta} = (\hat{\Lambda}, \hat{\Psi}, \hat{M}_{ff})$, defined by

$$\hat{\theta} = \underset{\theta\in\Theta}{\mathrm{argmax}}\ \ln \mathscr{L}(\theta),$$

the quasi maximum likelihood estimator, or the MLE, where Θ is the parameter space specified by Assumption D.

It is known in factor analysis that the loadings and factors can only be identified up to a rotation. To see this, let $\hat{\theta} = (\hat{\Lambda}, \hat{\Psi}, \hat{M}_{ff})$ be the maximizer of (3.3), then $\hat{\theta}^\dagger = (\hat{\Lambda}\hat{M}_{ff}^{1/2}, \hat{\Psi}, I_r)$ is also a qualified maximizer. From this perspective, it is no loss of generality to normalize that

$$M_{ff} = \frac{1}{T}\sum_{t=1}^{T} f_t f_t' = I_r.$$

Under this normalization, Σ_{zz} is simplified as $\Sigma_{zz} = \Lambda\Lambda' + \Psi$.

Maximizing the objective function (3.3) with respect to Λ and Ψ gives the following two first order conditions.

$$\hat{\Lambda}'\hat{\Psi}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = 0 \tag{3.4}$$

$$\mathrm{Bdiag}(M_{zz} - \hat{\Sigma}_{zz}) = 0 \tag{3.5}$$

where $\mathrm{Bdiag}(\cdot)$ is the block-diagonal operator, which puts the element of its argument to zero if the counterpart of Ψ is nonzero, otherwise unspecified. $\hat{\Lambda}$ and $\hat{\Psi}$ denote the MLE and $\hat{\Sigma}_{zz} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

3.2 Asymptotic properties of the MLE

This section presents the asymptotic results of the MLE for (3.3). Since we only impose $M_{ff} = I_r$ in (3.2), the loadings and factors still cannot be fully identified. We adopt the treatment of Bai (2003), in which the rotational matrix appears in the asymptotic representation. This treatment has two advantages in the present context. First, it simplifies our analysis. Second, it clarifies that the estimation and inferential theory of β is invariant to the rotational matrix. Alternatively, we can impose some additional restrictions to uniquely fix the rotational matrix; see Bai and Li (2012a) for full identification strategies. The following theorem, which serves as the base for the subsequent analysis, gives the asymptotic representations of the MLE.

Theorem 3.1 *Under Assumptions A-D, as $N, T \to \infty$, we have*

$$\hat{\Lambda}_i - R'\Lambda_i = R'\frac{1}{T}\sum_{t=1}^{T} f_t u'_{jt} + o_p(T^{-1/2})$$

$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u'_{it} - \Sigma_{ii}) + o_p(T^{-1/2})$$

where $R = \Lambda'\hat{\Psi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}$.

Remark 3.1 Notice that the rotational matrix R only enters in the asymptotic representation of $\hat{\Lambda}_i$. This is consistent with only loadings and factors having rotational indeterminacy and idiosyncratic errors not having such a problem.

Remark 3.2 By the above theorem, we immediately have $\hat{\Lambda}_i - R'\Lambda_i = O_p(T^{-1/2})$ and $\hat{\Sigma}_{ii} - \Sigma_{ii} = O_p(T^{-1/2})$. These two results continue to hold when N is fixed since the model falls within the scope of traditional factor analysis. But the asymptotic representations will be more complicated when N is finite. An implication of this result is that the covariance estimator $\hat{\beta}_i^{CV}$ is consistent even when N is finite.

4 Asymptotic results for the covariance estimator

Now we use the results in Theorem 3.1 to derive the asymptotic representation of $\hat{\beta}_i^{CV}$. Notice $\hat{\beta}_i^{CV} = (\hat{\Sigma}_{i,22})^{-1}\hat{\Sigma}_{i,21}$ and $\beta_i = (\Sigma_{i,22})^{-1}\Sigma_{i,21}$. Given $\hat{\Sigma}_{ii} = \Sigma_{ii} + o_p(1)$ by Theorem 3.1, the consistency of $\hat{\beta}_i$ is immediately obtained by the continuous mapping theorem. Furthermore, by Theorem 3.1,

$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u'_{it} - \Sigma_{ii}) + O_p(T^{-1}).$$

Then it follows

$$\hat{\Sigma}_{i,21} - \Sigma_{i,21} = \frac{1}{T}\sum_{t=1}^{T}[v_{it}(\epsilon_{it} + v'_{it}\beta_i) - \Omega_i\beta_i] + O_p(T^{-1}); \tag{4.1}$$

$$\hat{\Sigma}_{i,22} - \Sigma_{i,22} = \frac{1}{T}\sum_{t=1}^{T}[v_{it}v_{it}' - \Omega_i] + O_p(T^{-1}). \tag{4.2}$$

Notice that

$$\begin{aligned}\hat{\beta}_i - \beta_i &= (\hat{\Sigma}_{i,22})^{-1}\hat{\Sigma}_{i,21} - \Sigma_{i,22}^{-1}\Sigma_{i,21} \\ &= (\hat{\Sigma}_{i,22})^{-1}\Big[(\hat{\Sigma}_{i,21} - \Sigma_{i,21}) - (\hat{\Sigma}_{i,22} - \Sigma_{i,22})\Sigma_{i,22}^{-1}\Sigma_{i,21}\Big]\end{aligned} \tag{4.3}$$

Substituting (4.1) and (4.2) into (4.3) and noting that $\hat{\Sigma}_{i,22} \xrightarrow{p} \Omega_i$ and $\beta_i = \Sigma_{i,22}^{-1}\Sigma_{i,21}$, we have the following theorem on $\hat{\beta}_i^{CV}$.

Theorem 4.1 *Under Assumptions A-D, when $N, T \to \infty$, we have*

$$\sqrt{T}(\hat{\beta}_i^{CV} - \beta_i) = \Omega_i^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}\epsilon_{it}\Big) + o_p(1) \tag{4.4}$$

Remark 4.1 The above asymptotic result implies that our estimator is asymptotically efficient. To see this, suppose that the factors f_t are observed, then the GLS estimator has the asymptotic representation:

$$\sqrt{T}(\hat{\beta}_i^{GLS} - \beta_i) = \Omega_i^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}\epsilon_{it}\Big) + o_p(1), \tag{4.5}$$

which is the same as that of Theorem 4.1, implying the asymptotic efficiency of the CV estimator.

Remark 4.2 Although the asymptotic result of $\hat{\beta}_i^{CV}$ is derived under Assumption B, we point out that the proposed method works in a very general setup given the results of Bai and Li (2012b), which show that the quasi maximum likelihood method can be used to estimate approximate factor models (Chamberlain and Rothschild, 1983). More specifically, let $\Sigma_{ii,t}$ be the variance of u_{it}, where the covariance matrix has an additional superscript t to indicate that it is time-varying. Partition $\Sigma_{ii,t}$ as

$$\Sigma_{ii,t} = \begin{bmatrix} \Sigma_{ii,t,11} & \Sigma_{ii,t,12} \\ \Sigma_{ii,t,21} & \Sigma_{ii,t,22} \end{bmatrix}.$$

Under the assumption that ϵ_{it} is independent of v_{it}, we have $\Sigma_{ii,t,22}\beta_i = \Sigma_{ii,t,21}$ for all t, which implies that

$$\Big(\frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t,22}\Big)\beta_i = \frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t,21}.$$

To consistently estimate β_i, it suffices to consistently estimate $\frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t}$. As shown in Bai and Li (2012b), if the underlying covariance is time-varying but misspecified to be time-invariant in the estimation, the resulting estimator of the covariance is a consistent estimator for the average underlying covariance over time, that is, $\frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t}$ happens to be estimated by the MLE.

Remark 4.3 For the basic model, the CCE estimator of Pesaran (2006) and the iterated PC estimator of Song (2013) have the same asymptotic representations as in Theorem 4.1 and hence are asymptotically efficient. However, different methods require different conditions for the asymptotic theory. Except for the rank condition, the CCE estimator potentially requires N be large, otherwise the average error over the cross section cannot be negligible. The PC estimator is derived under the cross-sectional homoscedasticity. If heteroscedasticity is present, a large N is needed to ensure the consistency. For the CV estimator, the consistency can be maintained for a fixed N even in the presence of the cross-sectional heteroscedasticity. So the CV estimator requires the least restrictive condition for the consistency.

Remark 4.4 With slight modification, our method can be used to estimate the homogeneous coefficient. Suppose $\beta_i \equiv \beta$ for all i. Now we have $\Sigma_{i,22}\beta = \Sigma_{i,21}$ for all i, which leads to

$$\Big(\sum_{i=1}^{N} \Sigma_{i,22}\Big)\beta = \sum_{i=1}^{N} \Sigma_{i,21}.$$

So a consistent estimator for β is

$$\hat{\beta} = \big(\sum_{i=1}^{N} \hat{\Sigma}_{i,22}\big)^{-1}\big(\sum_{i=1}^{N} \hat{\Sigma}_{i,21}\big). \tag{4.6}$$

The asymptotic properties of $\hat{\beta}$ will not be pursued in this paper. In section 6, we conduct a small simulation to examine its finite sample performance.

Corollary 4.1 *Under the assumptions of Theorem 4.1, we have*

$$\sqrt{T}(\hat{\beta}_i^{CV} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2 \Omega_i^{-1}\big),$$

where $\sigma_{\epsilon i}^2$ is the variance of ϵ_{it} and Ω_i is the variance of v_{it}. The variance $\sigma_{\epsilon i}^2\Omega_i^{-1}$ can be consistently estimated by $\hat{\sigma}_{\epsilon i}^2 \hat{\Sigma}_{i,22}^{-1}$, where $\hat{\sigma}_{\epsilon i}^2 = \hat{\Sigma}_{i,11} - \hat{\beta}_i^{CV\prime}\hat{\Sigma}_{i,22}\hat{\beta}_i^{CV}$.

5 Models with zero restrictions

In this section, we consider the following restricted model:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \epsilon_{it} \\ x_{it} &= \nu_i + \gamma_i^{g\prime} g_t + \gamma_i^{h\prime} h_t + v_{it} \end{aligned} \tag{5.1}$$

where the dimensions of g_t and h_t are $r_1 \times 1$ and $r_2 \times 1$, respectively. A salient feature of model (5.1) is that the explanatory variables include more factors than the error of the y equation. This specification aims to accommodate that both endogenous and exogenous shocks exist in the economic system. Endogenous shocks such as unexpected monetary supply would directly affect all economic variables. Exogenous shocks such as oil prices would first affect the energy-related industries and then gradually affect other economic variables. In model (5.1), g_t denotes the endogenous shocks that directly affect y and x, and h_t denotes the exogenous shocks that affect first x then y[①].

The y equation of (5.1) can be written as

$$y_{it} = \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \phi_i' h_t + \epsilon_{it}$$

with $\phi_i = 0$ for all i. Let $f_t = (g_t', h_t')'$, $\lambda_i = (\psi_i', \phi_i')'$ and $\gamma_i = (\gamma_i^{g\prime}, \gamma_i^{h\prime})'$, we have the same representation as (1.1). From this perspective, model (5.1) can be viewed as a restricted version of model (1.1). This implies that the two-step method proposed in Section 4 is applicable to (5.1). However, this estimation method is not efficient. Consider the ideal case that g_t is observable. To eliminate the endogenous ingredient $\psi_i' g_t$, we post-multiply $M_G = I - G(G'G)^{-1}G'$ on both sides of the y equation. The remaining part of x_{it} includes v_{it} and $\gamma_i^{h\prime}(h_t - H'G(G'G)^{-1}g_t)$, which both provide the information for β. However, as shown in Theorem 4.1, only the variations of v_{it} are used to signal β_i in $\hat{\beta}_i^{CV}$. Therefore, partial information is discarded and the two-step method in Section 4 is inefficient.

The preceding discussion provides some insights on the improvement of efficiency. To efficiently estimate model (5.1), we need to use information contained in the common components of x_{it}. Rewrite model (5.1) as

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \nu_i \end{bmatrix} + \begin{bmatrix} \beta_i'\gamma_i^{g\prime} + \psi_i' & \beta_i'\gamma_i^{h\prime} \\ \gamma_i^{g\prime} & \gamma_i^{h\prime} \end{bmatrix} \begin{bmatrix} g_t \\ h_t \end{bmatrix} + \begin{bmatrix} \beta_i' v_{it} + \epsilon_{it} \\ v_{it} \end{bmatrix} \tag{5.2}$$

We use Λ_i' to denote the loadings matrix before $f_t = (g_t', h_t')'$. The symbols μ_i, z_{it} and u_{it} are defined the same as in the previous section. We then have the same equation as (2.1). Further partition the loadings matrix Λ_i into four blocks,

$$\Lambda_i = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,12} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} = \begin{bmatrix} \psi_i + \gamma_i^g \beta_i & \gamma_i^g \\ \gamma_i^h \beta_i & \gamma_i^h \end{bmatrix}. \tag{5.3}$$

So we have $\Lambda_{i,22}\beta_i = \Lambda_{i,21}$. This result together with (2.3) leads to

$$\begin{bmatrix} \Lambda_{i,22} \\ \Sigma_{i,22} \end{bmatrix} \beta_i = \begin{bmatrix} \Lambda_{i,21} \\ \Sigma_{i,21} \end{bmatrix} \tag{5.4}$$

Given the above structural relationship, a routine to estimate β_i is replacing $\Lambda_{i,22}, \Lambda_{i,21}, \Sigma_{i,22}$ and $\Sigma_{i,21}$ with their MLE and minimizing the distance on the both sides of the equation with some weighting matrix. While this method is intuitive, it is not correct since $\hat{\Lambda}_{i,22}$ and $\hat{\Lambda}_{i,21}$ are not consistent estimators of $\Lambda_{i,22}$ and $\Lambda_{i,21}$, as shown in Theorem 3.1. Let $\Lambda_i^* = R'\Lambda_i$ represent the underlying parameters that the MLE corresponds to, where R is the rotation matrix defined in Theorem 3.1. Then

$$\Lambda_i^{*\prime} = \begin{bmatrix} \Lambda_{i,11}^{*\prime} & \Lambda_{i,21}^{*\prime} \\ \Lambda_{i,12}^{*\prime} & \Lambda_{i,22}^{*\prime} \end{bmatrix} = \Lambda_i' R = \begin{bmatrix} \Lambda_{i,11}' & \Lambda_{i,21}' \\ \Lambda_{i,12}' & \Lambda_{i,22}' \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} \beta_i'\gamma_i^{g\prime} + \psi_i' & \beta_i'\gamma_i^{h\prime} \\ \gamma_i^{g\prime} & \gamma_i^{h\prime} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

[①] Another way to see this point is as follows. Notice that the x equation can always be written as

$$x_{it} = \nu_i + (\gamma_i^{g\prime} + \gamma_i^{h\prime} H'G(G'G)^{-1})g_t + \gamma_i^{h\prime}(h_t - H'G(G'G)^{-1}g_t) + v_{it} = \nu_i + \gamma_i^{*g\prime} g_t + \gamma_i^{h\prime} h_t^* + v_{it}.$$

In the last equation, g_t is uncorrelated with h_t^*. Given this expression, it is no loss of generality to assume that h_t is uncorrelated with g_t. Now we see that g_t causes the endogeneity problem but h_t does not. So we say that g_t represents endogenous shocks and h_t represents exogenous shocks.

implying

$$\Lambda^*_{i,21} = (R'_{12}\gamma_i^g + R'_{22}\gamma_i^h)\beta_i + R'_{12}\psi_i \tag{5.5}$$

$$\Lambda^*_{i,22} = R'_{12}\gamma_i^g + R'_{22}\gamma_i^h \tag{5.6}$$

From (5.5) and (5.6), we see that unless $\psi_i = 0$, $\Lambda^*_{i,22}\beta_i = \Lambda^*_{i,21}$ does not hold. But when $\psi_i = 0$, we see from (5.1) that the model is free of the endogeneity problem and the ordinary least squares method is applicable. The preceding analysis indicates that the existence of the rotational indeterminacy for loadings impedes the use of the underlying relation $\Lambda_{i,22}\beta_i = \Lambda_{i,21}$ in the estimation of β_i.

Although this result is a little disappointing, we now show that with some transformation, $\Lambda_{i,22}\beta_i = \Lambda_{i,21}$ can still be used to estimate β_i. First by $\Lambda_i^{*\prime} = \Lambda_i' R$,

$$\Lambda^*_{i,11} = (R'_{11}\gamma_i^g + R'_{21}\gamma_i^h)\beta_i + R'_{11}\psi_i \tag{5.7}$$

$$\Lambda^*_{i,12} = R'_{11}\gamma_i^g + R'_{21}\gamma_i^h \tag{5.8}$$

By the expressions (5.5)-(5.8), we have the following equation:

$$(\Lambda^*_{i,21} - \Lambda^*_{i,22}\beta_i) = R'_{12}R'^{-1}_{11}(\Lambda^*_{i,11} - \Lambda^*_{i,12}\beta_i) = V(\Lambda^*_{i,11} - \Lambda^*_{i,12}\beta_i) \tag{5.9}$$

where $V = R'_{12}R'^{-1}_{11}$, an $r_2 \times r_1$ rotational matrix. The preceding equation can be written as

$$(\Lambda^*_{i,22} - V\Lambda^*_{i,12})\beta_i = \Lambda^*_{i,21} - V\Lambda^*_{i,11} \tag{5.10}$$

Given the above result, together with (2.3), we have

$$\begin{bmatrix} \Lambda^*_{i,22} - V\Lambda^*_{i,12} \\ \Sigma_{i,22} \end{bmatrix} \beta_i = \begin{bmatrix} \Lambda^*_{i,21} - V\Lambda^*_{i,11} \\ \Sigma_{i,21} \end{bmatrix} \tag{5.11}$$

If V is known, then we can replace $\Lambda^*_{i,11}, \Lambda^*_{i,12}, \Lambda^*_{i,21}, \Lambda^*_{i,22}$ with the corresponding estimates, and β_i is efficiently estimated. Although V is unknown, it can be consistently estimated by (5.9) since β_i can be consistently (albeit not efficiently) estimated by $\hat{\beta}_i^{CV} = \hat{\Sigma}^{-1}_{i,22}\hat{\Sigma}_{i,21}$. Given the above analysis, we propose the following estimation procedure:

1. *Use the maximum likelihood method to obtain the estimates $\hat{\Sigma}_{ii}, \hat{\Lambda}_i, \hat{f}_t$ for all i and t.*

2. *Calculate $\hat{\beta}_i^{CV} = \hat{\Sigma}^{-1}_{i,22}\hat{\Sigma}_{i,21}$ and*

$$\hat{V} = \Big[\sum_{i=1}^N (\hat{\Lambda}_{i,21} - \hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]\Big[\sum_{i=1}^N (\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]^{-1}.$$

3. *Calculate $\hat{\beta}_i = (\hat{\Delta}_i' W_i^{-1}\hat{\Delta}_i)^{-1}\hat{\Delta}_i' W_i^{-1}\hat{\delta}_i$, where W_i is a predetermined weighting matrix that is specified below, and*

$$\hat{\Delta}_i = \begin{bmatrix} \hat{\Lambda}_{i,22} - \hat{V}\hat{\Lambda}_{i,12} \\ \hat{\Sigma}_{i,22} \end{bmatrix}, \qquad \hat{\delta}_i = \begin{bmatrix} \hat{\Lambda}_{i,21} - \hat{V}\hat{\Lambda}_{i,11} \\ \hat{\Sigma}_{i,21} \end{bmatrix} \tag{5.12}$$

where we call the resulting estimator the Loading-coVariance estimators, denoted by $\hat{\beta}_i^{LV}$.

Remark 5.1 We can iterate the second and third steps by using the updated estimator of β_i to calculate $\hat{V}$. We call the estimator resulting from this iterating procedure the *Iterated-LV* estimator, denoted by $\hat{\beta}_i^{ILV}$. The iterated estimator has the same asymptotic representation as the LV estimator, but better finite sample performance; see the simulation results in Section 6.

5.1 The optimal weighting matrix

To carry out the estimation procedure, we need to specify the weighting matrix W_i. It can be shown that the theoretically optimal weighting matrix is

$$W_i^{opt} = \begin{bmatrix} R_{22\cdot1}' M_{hh\cdot g}^{-1} R_{22\cdot1} & 0_{r_2\times k} \\ 0_{k\times r_2} & \Sigma_{i,22} \end{bmatrix},$$

where $R_{22\cdot1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$ and $M_{hh\cdot g} = M_{hh} - M_{hg}M_{gg}^{-1}M_{gh}$. This weighting matrix can be consistently estimated by

$$\hat{W}_i = \begin{bmatrix} \Big[\big(\frac{1}{T}\sum_{t=1}^T \hat{h}_t\hat{h}_t'\big) - \big(\frac{1}{T}\sum_{t=1}^T \hat{h}_t\hat{\eta}_t'\big)\big(\frac{1}{T}\sum_{t=1}^T \hat{\eta}_t\hat{\eta}_t'\big)^{-1}\big(\frac{1}{T}\sum_{t=1}^T \hat{\eta}_t\hat{h}_t'\big)\Big]^{-1} & 0_{r_2\times k} \\ 0_{k\times r_2} & \hat{\Sigma}_{i,22} \end{bmatrix} \quad (5.13)$$

with $\hat{\eta}_t = \hat{g}_t + \hat{V}'\hat{h}_t$, where $\hat{g}_t$ and $\hat{h}_t$ are given by

$$\begin{bmatrix} \hat{g}_t \\ \hat{h}_t \end{bmatrix} = \Big(\sum_{i=1}^N \hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\Big)^{-1}\Big(\sum_{i=1}^N \hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}z_{it}\Big).$$

5.2 The asymptotic result

The following theorem gives the asymptotic representation of the LV estimator with some remarks following.

Theorem 5.1 *Under Assumptions A-D, when $N, T \to \infty$, we have*

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) = \big(\gamma_i^{h\prime}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})\gamma_i^h + \Omega_i\big)^{-1} \times \frac{1}{\sqrt{T}}\sum_{t=1}^T \Big[\gamma_i^{h\prime}(\dot{h}_t - M_{hg}M_{gg}^{-1}\dot{g}_t) + v_{it}\Big]\epsilon_{it} + o_p(1)$$

Given Theorem 5.1, we have the following corollary:

Corollary 5.1 *Under the assumptions of Theorem 5.1, we have*

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2(\gamma_i^{h\prime}\overline{M}_{hh\cdot g}\gamma_i^h + \Omega_i)^{-1}\big).$$

where $\overline{M}_{hh\cdot g} = \underset{T\to\infty}{\text{plim}}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})$. The above asymptotic result can be presented alternatively as

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2\big[\underset{T\to\infty}{\text{plim}}\frac{1}{T}X_i'M_{\overline{G}}X_i\big]^{-1}\big).$$

with $\overline{G} = (\mathbf{1}_T, G)$, where $\mathbf{1}_T$ is a T-dimensional vector with all the elements equal to 1.

Remark 5.2 Consider the "y" equation, which can be written as

$$Y_i = \alpha_i \mathbf{1}_T + X_i\beta_i + G\psi_i + E_i \tag{5.14}$$

where $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $X_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$, and E_i is defined similarly as Y_i. If the factors g_t are observable, the infeasible GLS estimator for β_i is

$$\hat{\beta}_i^{GLS} = (X_i' M_{\overline{G}} X_i)^{-1}(X_i' M_{\overline{G}} Y_i).$$

By (5.14), we have

$$\hat{\beta}_i^{GLS} - \beta_i = (X_i' M_{\overline{G}} X_i)^{-1}(X_i' M_{\overline{G}} E_i).$$

Notice $\text{var}(E_i) = \sigma_{\epsilon i}^2 I_T$. Thus the limiting distribution of $\hat{\beta}_i^{GLS} - \beta_i$ conditional on X_i is

$$\sqrt{T}(\hat{\beta}_i^{GLS} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2 \big[\operatorname*{plim}_{T\to\infty} \frac{1}{T} X_i' M_{\overline{G}} X_i\big]^{-1}\big).$$

the same as that of Corollary (5.1). This means that the LV estimator $\hat{\beta}_i^{LV}$ is asymptotically efficient.

Remark 5.3 Consider the following model, in which zero restrictions exist in both the x equation and the y equation:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \epsilon_{it} \\ x_{it} &= \nu_i + \gamma_i^{h\prime} h_t + v_{it} \end{aligned} \tag{5.15}$$

where g_t and h_t are assumed to be correlated. Model (5.15) is a special case of (5.1) in view that γ_i^g is restricted to zero. So the loading-covariance two-step method can be directly applied to (5.15). We note that the LV estimator is efficient even in the presence of additional zero restrictions $\gamma_i^g = 0$. To see this point, notice that Λ_i in model (5.15) is

$$\Lambda_i = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,12} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} = \begin{bmatrix} \psi_i & 0 \\ \gamma_i^h \beta_i & \gamma_i^h \end{bmatrix}.$$

The coefficient β_i can only be estimated by the relations of $\Lambda_{i,21}$ and $\Lambda_{i,22}$, which is the same as Model (5.1). By the same arguments, we conclude that the model

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \phi_i' h_t + \epsilon_{it}, \\ x_{it} &= \nu_i + \gamma_i^{h\prime} h_t + v_{it}. \end{aligned}$$

is efficiently estimated by the CV method.

Remark 5.4 If the underlying coefficients are identical, we can also use the information contained in the loadings to improve the efficiency. Let

$$\hat{g}_i(V, \beta) = \begin{bmatrix} \hat{\Lambda}_{i,22} - V\hat{\Lambda}_{i,12} \\ \hat{\Sigma}_{i,22} \end{bmatrix} \beta - \begin{bmatrix} \hat{\Lambda}_{i,21} - V\hat{\Lambda}_{i,11} \\ \hat{\Sigma}_{i,21} \end{bmatrix}.$$

Given equation (5.11) (notice that now $\beta_i \equiv \beta$ for all i) we can consistently estimate β by

$$(\hat{\beta}^{LV}, \hat{V}) = \operatorname*{argmin}_{\beta, V} \sum_{i=1}^{N} \hat{g}_i(V, \beta)' \hat{W}_i^{-1} \hat{g}_i(V, \beta). \tag{5.16}$$

where $\hat{W}_i$ is defined in (5.13). Notice that if Λ is identified, we can estimate β by (5.4), replacing the unknown parameters with their estimates. So the additional estimation of V can be regarded as the cost we pay for the rotational indeterminacy. The finite sample properties of the above LV estimator will be investigated in Section 7.

6 Discussions on models with time-invariant regressors

In some applications, it is of interest to include some time-invariant variables, such as gender, race, education, and so forth. In this section, we address this concern. Consider the following model with time-invariant variables:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \phi_i' h_t + \epsilon_{it} \\ x_{it} &= \nu_i + \gamma_i^{g\prime} g_t + \gamma_i^{h\prime} h_t + v_{it} \end{aligned} \tag{6.1}$$

where ϕ_i are observable and represent the time-invariant regressors. Model (6.1) specifies that the coefficients of ϕ_i are time-varying. We believe that this is a sensible way to make the model flexible enough. Now we show that our estimation idea can be used to estimate (6.1). As in the previous section, rewrite model (6.1) as

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \nu_i \end{bmatrix} + \begin{bmatrix} \beta_i'\gamma_i^{g\prime} + \psi_i' & \beta_i'\gamma_i^{h\prime} + \phi_i' \\ \gamma_i^{g\prime} & \gamma_i^{h\prime} \end{bmatrix} \begin{bmatrix} g_t \\ h_t \end{bmatrix} + \begin{bmatrix} \beta_i' v_{it} + \epsilon_{it} \\ v_{it} \end{bmatrix} \tag{6.2}$$

Let Λ_i' be the loadings matrix before $f_t = (g_t', h_t')'$ and partition it into four blocks, we have

$$\Lambda_i = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,12} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} = \begin{bmatrix} \psi_i + \gamma_i^g \beta_i & \gamma_i^g \\ \phi_i + \gamma_i^h \beta_i & \gamma_i^h \end{bmatrix} \tag{6.3}$$

Let $\Lambda_i^* = R'\Lambda_i$ be the underlying parameters that the estimators correspond to. So we have

$$\Lambda_i^{*\prime} = \begin{bmatrix} \Lambda_{i,11}^{*\prime} & \Lambda_{i,21}^{*\prime} \\ \Lambda_{i,12}^{*\prime} & \Lambda_{i,22}^{*\prime} \end{bmatrix} = \Lambda_i' R = \begin{bmatrix} \Lambda_{i,11}' & \Lambda_{i,21}' \\ \Lambda_{i,12}' & \Lambda_{i,22}' \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

This leads to

$$\begin{aligned} \Lambda_{i,11}^* &= (R_{11}'\gamma_i^g + R_{21}'\gamma_i^h)\beta_i + R_{11}'\psi_i + R_{21}'\phi_i, \qquad & \Lambda_{i,12}^* &= R_{11}'\gamma_i^g + R_{21}'\gamma_i^h \qquad (6.4) \\ \Lambda_{i,21}^* &= (R_{12}'\gamma_i^g + R_{22}'\gamma_i^h)\beta_i + R_{12}'\psi_i + R_{22}'\phi_i, \qquad & \Lambda_{i,22}^* &= R_{12}'\gamma_i^g + R_{22}'\gamma_i^h \qquad (6.5) \end{aligned}$$

From (6.4)$-$(6.5), we have

$$R_{12}'R_{11}'^{-1}(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i) + R_{22\cdot 1}'\phi_i = (\Lambda_{i,21}^* - \Lambda_{i,22}^*\beta_i) \tag{6.6}$$

where $R_{22\cdot 1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$. Given (6.6) together with $\Sigma_{i,22}\beta_i = \Sigma_{i,21}$, we have

$$\begin{bmatrix} \Lambda_{i,22}^* - V\Lambda_{i,12}^* \\ \Sigma_{i,22} \end{bmatrix} \beta_i = \begin{bmatrix} \Lambda_{i,21}^* - V\Lambda_{i,11}^* - R_{22\cdot 1}'\phi_i \\ \Sigma_{i,21} \end{bmatrix} \tag{6.7}$$

where $V = R_{12}'R_{11}'^{-1}$. If V and $R_{22\cdot 1}$ are known, we can use (6.7) to efficiently estimate β_i. Similarly as in the previous section, we can use $\hat{\beta}_i^{CV}$ to get a preliminary estimators for V and $R_{22\cdot 1}$. This leads to the following estimation procedures:

1. *Use the maximum likelihood method to obtain the estimates* $\hat{\Sigma}_{ii}, \hat{\Lambda}_i$ *and* $\hat{f}_t$ *for all* i *and* t.

2. *Calculate* $\hat{\beta}_i^{CV} = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$ *and* $\hat{V}$ *and* $\hat{R}_{22\cdot 1}$ *by*

$$[\hat{V}, \hat{R}'_{22\cdot 1}] = \Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,21} - \hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})\Xi_i\Big]\Big[\sum_{i=1}^{N}\Xi_i\Xi_i'\Big]^{-1}$$

   where $\Xi_i = [(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})', \phi_i']'$.

3. *Calculate* $\hat{\beta}_i^{LV} = (\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i)^{-1}\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\gamma}_i$, *where*

$$\hat{\Delta}_i = \begin{bmatrix} \hat{\Lambda}_{i,22} - \hat{V}\hat{\Lambda}_{i,12} \\ \hat{\Sigma}_{i,22} \end{bmatrix}, \qquad \hat{\gamma}_i = \begin{bmatrix} \hat{\Lambda}_{i,21} - \hat{V}\hat{\Lambda}_{i,11} - \hat{R}'_{22\cdot 1}\phi_i \\ \hat{\Sigma}_{i,21} \end{bmatrix}$$

   and $\hat{W}_i$ *is the predetermined weighting matrix, which is the same as (5.13).*

Similarly we can iterate Steps 2 and 3 by replacing $\hat{\beta}_i^{CV}$ with the updated LV estimator. This leads to the iterated LV estimator. Under the same conditions of Theorem (5.1), we can show

$$\begin{aligned}\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) = {}& \big(\gamma_i^{h\prime}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})\gamma_i^h + \Omega_i\big)^{-1} \\ & \times \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\gamma_i^{h\prime}(\dot{h}_t - M_{hg}M_{gg}^{-1}\dot{g}_t) + v_{it}\Big]\epsilon_{it} + o_p(1)\end{aligned}$$

The above asymptotic result can be interpreted in a similar way as in Remark 5.2. So the LV estimator is asymptotically efficient.

7 Finite sample properties

In this section, we run Monte Carlo simulations to investigate the finite sample properties of the proposed estimators. The model considered in the simulation consists of one explanatory variable ($K = 1$) and two factors ($r = 2$), which can be presented as

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}\beta_i + \psi_i g_t + \phi_i h_t + \epsilon_{it}, \\ x_{it} &= \mu_i + \gamma_i^g g_t + \gamma_i^h h_t + v_{it}, \end{aligned} \tag{7.1}$$

where g_t and h_t are both scalars. We consider the following different specifications on the models (M), loadings (L), errors (E) and coefficients(C):

M1: ψ_i and ϕ_i are random variables for all i;

M2: ϕ_i is zero for all i and ψ_i is random variable.

L1: ψ_i and ϕ_i (if not zero) are generated according to $\psi_i = 2 + N(0,1)$ and $\phi_i = 1 + N(0,1)$; similarly γ_i^g and γ_i^h are generated by $\gamma_i^g = 1 + N(0,1)$ and $\gamma_i^h = 2 + N(0,1)$.

L2: ψ_i and ϕ_i (if not zero) are generated from $N(0,1)$; γ_i^g and γ_i^h are generated according to $\gamma_i^g = \psi_i + N(0,1)$ and $\gamma_i^h = \phi_i + N(0,1)$.

E1: Let Ξ be a $N(K+1)$ dimensional vector with all its elements being 1. Let $\Upsilon = \text{diag}(\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_N)$ be an $N(K+1) \times N(K+1)$ block diagonal matrix, where

$\Upsilon_i = \text{diag}(1, (M_i' M_i)^{-1/2} M_i)$ with M_i being a $K \times K$ standard normal random matrix. Then u_t is generated according to $u_t = \sqrt{\text{diag}(\Xi)} \Upsilon \varepsilon_t$, where ε_t is an $N(K+1) \times 1$ vector with all its elements being i.i.d from $N(0, 1)$.

E2: u_t is generated as in **E1** except that

$$\Xi_i = 0.1 + \frac{\eta_i}{1 - \eta_i} \iota_i' \iota_i, \qquad i = 1, 2, \cdots, N(K+1)$$

where ι_i' is the ith row of Λ, and η_i is drawn independently from $U[u, 1-u]$ with $u = 0.1$.

C1: $\beta_i = 1 + N(0, 0.04)$ for all i.

C2: $\beta_i = 1$ for all i.

Remark 7.1 Two specifications in M denotes the two models considered in the paper. M1 corresponds to the basic model, and M2 corresponds to the model with zero restrictions. We consider two different sets of loadings. In L1 all the loadings have the same mean, but in L2 only the loadings corresponding to the same individual share the same component. Both specifications lead to the correlated loadings, but as will be seen below, the CCE estimator performs quite differently in the two setups. We also consider the cross-sectional heteroscedasticity and homoscedasticity in the simulation, which correspond to E1 and E2, respectively. When generating heteroscedasticity, we add 0.1 to the expression, avoiding the variance being too close to zero. Our approach to generating the idiosyncratic errors is similar to Doz et al. (2012) and Bai and Li (2014). We also consider two specifications for the coefficients. While we mainly focus on the performance of the estimation of heterogeneous coefficients, we also use simulations to examine the finite sample properties of the two estimators proposed in Remarks 4.4 and 5.4.

The other parameters including $g_t, h_t, \alpha_i, \nu_i$ are all generated independently from $N(0, 1)$. To evaluate the performance of estimators, we use the average of the root mean square error (RMSE) to measure the goodness-of-fit, which is calculated by

$$\sqrt{\frac{1}{N\mathcal{S}} \sum_{s=1}^{\mathcal{S}} \sum_{i=1}^{N} (\hat{\beta}_i^{(s)} - \beta_i)^2},$$

where $\hat{\beta}_i^{(s)}$ is the estimator of the ith unit in the sth experiment, and β_i is the underlying true value. $\mathcal{S}$ is the number of repetitions, which is set to 1000 in the simulation.

7.1 Determining the number of factors

We now discuss the determination of the number of factors, which is a relevant issue in the factor-analysis-based method. In the basic model, determining the number of factors is relatively easier. In the first step, we estimate a pure factor model. So the existing determination methods, such as Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013), can be used. Although these methods do work well in the present setup, to be consistent with the theory established in Section 3, we instead consider the following MLE-based information criterion in the simulation

$$\hat{r} = \underset{0 \le m \le r_{\max}}{\text{argmin}} \; IC(m) \tag{7.2}$$

where

$$IC(m) = \frac{1}{N\bar{K}} \ln |\hat{\Lambda}^m \hat{\Lambda}^{m\prime} + \hat{\Psi}^m| + m\frac{N\bar{K}+T}{NT\bar{K}} \ln \min(N\bar{K}, T).$$

where $\hat{\Lambda}^m$ and $\hat{\Psi}^m$ are the respective estimator of Λ and Ψ when the number of factors is set to m and $\bar{K} = K + 1$. For the model with zero restrictions, we need to determine the factor numbers in the y equation and the x equation, respectively. Following Bai and Li (2014), we consider a two-step method to determine them. First, we use (7.2) to obtain the total number $r = r_1 + r_2$, denoted by $\hat{r}$, and the associated CV estimator $\hat{\beta}_i^{\hat{r}}$; we then use (7.2) again to determine the factor number of the residual matrix $\mathscr{R} = (\mathscr{R}_{it})$ with $\mathscr{R}_{it} = \dot{y}_{it} - \dot{x}_{it}'\hat{\beta}_i^{\hat{r}}$, which we use $\hat{r}_1$ to denote. Then $\hat{r}_2 = \hat{r} - \hat{r}_1$. In the simulation, we set $r_{\max} = 3$.

In practice, the basic model and the model with zero restrictions cannot be differentiated. We therefore suggest estimating the two models in a unified way. More specifically, for a given data set, we calculate r and r_1. If $\hat{r} = \hat{r}_1$, we turn to the basic model; if $\hat{r} > \hat{r}_1$, we turn to the model with zero restrictions.

Table 1 reports the percentages that the number of factors is correctly estimated by (7.2) based on 1000 repetitions. From the table, we see that the number of factors can be correctly estimated with very high probability. This result is robust to all combinations of listed specifications on loadings, errors and models.

Table 1: The percentage of correctly estimating the number of factors

		M1				M2			
	T	50	100	150	200	50	100	150	200
		L1+E1				L1+E1			
	50	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		L1+E2				L1+E2			
	50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		L2+E1				L2+E1			
	50	99.8	100.0	100.0	100.0	99.9	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		L2+E2				L2+E2			
	50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

7.2 Finite sample properties of several estimators

In this section, we examine the performance of the CV and LV estimators. For the purpose of comparison, we also calculate Pesaran's CCE estimator, Song's PC estimator, and the infeasible GLS estimator. The infeasible GLS estimator, which is calculated by assuming that the factors are observed, serves as the benchmark for comparison. Since the previous subsection has confirmed that the number of factors can be correctly estimated with high probability, we assume that the number of factors is known in this subsection.

Tables 2-3 report the performance of the CCE, PC, CV and infeasible GLS (denoted by INF) estimators under different loading and error choices in the basic model. In summary, we see that the CCE estimator performs well under L1, but poorly under L2; the PC estimator performs well under E1, but poorly under E2; the CV estimator performs well under all setups.

First consider the different loading choices. Under L1, the performance of the CCE estimator is considerably good and very close to that of the CV estimator. The performance of these two estimators is only slightly inferior to the infeasible GLS estimator regardless of homoscedasticity or heteroscedasticity. However, under L2 the performance of the CCE estimator is poor. Not only does it have a large average RMSE, but it also exhibits a slowly decreasing rate for the average RMSE. In contrast, the CV estimator performs closely with the infeasible GLS estimator. The average RMSE of the CV estimator decreases almost at the same speed with that of the infeasible estimator.

The reason for the different performance of the CCE estimator under different loading sets is that the space spanned by $\tilde{z}_t = \frac{1}{N}\sum_{i=1}^{N} \dot{z}_{it}$ with $\dot{z}_{it} = (\dot{y}_{it}, \dot{x}'_{it})'$ provides a good approximation to the space spanned by f_t under L1, but a poor approximation under L2. To see this point more clearly, consider (2.1), which can be written as $\dot{z}_{it} = \Lambda'_i f_t + \dot{u}_{it}$. Taking the average over i, we have $\tilde{z}_t = \tilde{\Lambda}' f_t + \tilde{u}_t$, where $\tilde{\Lambda}$ and $\tilde{u}_t$ are defined similarly to $\tilde{z}_t$. With some transformation, we have $f_t = (\tilde{\Lambda}\tilde{\Lambda}')^{-1}\tilde{\Lambda}(\tilde{z}_t - \tilde{u}_t)$. So a good approximation requires two conditions. First, $\tilde{z}_t$ dominates $\tilde{u}_t$ so that $\tilde{u}_t$ is negligible. Second, $\tilde{\Lambda}\tilde{\Lambda}'$ is invertible when N goes to infinity. The loadings in L1 satisfy these two conditions, but the loadings in L2 violate the first one. In fact, the terms $\tilde{\Lambda}' f_t$ and $\tilde{u}_t$ are of the same magnitude under L2. So a good approximation fails. There are cases in which the second condition breaks down. For example, if all rows of Λ share the same mean, then $\tilde{\Lambda}$ is of rank one asymptotically, which in turn leads to $\tilde{\Lambda}'\tilde{\Lambda}$ being singular asymptotically. The simulation results confirm that the CCE estimator performs poorly in this case.

Table 2: The performance of the four estimators in the basic model

		L1+E1				L2+E1			
N	T	CCE	PC	CV	INF	CCE	PC	CV	INF
50	50	0.1517	0.1596	0.1537	0.1501	0.3980	0.1603	0.1533	0.1492
100	50	0.1499	0.1538	0.1512	0.1494	0.3985	0.1543	0.1508	0.1489
150	50	0.1491	0.1519	0.1500	0.1489	0.3961	0.1526	0.1503	0.1492
50	100	0.1052	0.1087	0.1049	0.1024	0.3868	0.1095	0.1051	0.1026
100	100	0.1034	0.1058	0.1040	0.1029	0.3855	0.1060	0.1037	0.1025
150	100	0.1029	0.1046	0.1033	0.1025	0.3863	0.1049	0.1034	0.1026
50	150	0.0857	0.0878	0.0848	0.0830	0.3819	0.0883	0.0847	0.0828
100	150	0.0839	0.0855	0.0841	0.0832	0.3826	0.0858	0.0841	0.0832
150	150	0.0834	0.0846	0.0836	0.0831	0.3819	0.0848	0.0836	0.0830
50	200	0.0749	0.0760	0.0733	0.0717	0.3832	0.0763	0.0732	0.0716
100	200	0.0723	0.0737	0.0723	0.0715	0.3815	0.0741	0.0726	0.0718
150	200	0.0719	0.0729	0.0720	0.0716	0.3813	0.0731	0.0722	0.0717

Consider then the different choices of the errors. Table 3 shows that the PC estimator performs poorly in the presence of cross-sectional heteroscedasticity (E2). In addition, we find that the performance of the PC estimator is improved marginally under E1, but significantly under E2, when N becomes larger. According to the theory of Song (2013), the PC estimate is $\sqrt{T}$-consistent, implying that the performance of the PC estimator should be closely related to T and loosely related to N. This theoretical result is supported by Table 2 but contradicted in Table 3. We think that the underlying reason is due to the computation problem of the minimizer of the objective function in the iterated PC method, as mentioned in Section 1. The extent of this problem depends on the strength of heteroscedasticity. In our simulation, we generate heavy heteroscedasticity, which magnifies the computational problem of the iterated PC method. [2]

Table 3: The performance of the four estimators in the basic model

[2] In the case of a homogeneous coefficient, this computational problem does not exist. First, as shown in the next subsection, the PC estimator generally has a better convergence under a homogeneous coefficient. Second, as pointed out in Moon and Weidner (2012), the objective function of the PC method can be written into a trace form, which only depends on β. So we can first use the method suggested by Bai (2009) to obtain a preliminary estimator, and then turn to the Newton-Raphson algorithm to get a better estimator.

		L1+E2				L2+E2			
N	T	CCE	PC	CV	INF	CCE	PC	CV	INF
50	50	0.3505	3.4677	0.3667	0.3581	0.4079	2.2194	0.2456	0.2377
100	50	0.3426	2.7550	0.3592	0.3545	0.4084	1.6894	0.2390	0.2362
150	50	0.3470	2.6504	0.3569	0.3543	0.4128	1.2141	0.2382	0.2363
50	100	0.2515	2.8863	0.2494	0.2427	0.3870	2.0866	0.1672	0.1630
100	100	0.2380	2.5816	0.2430	0.2399	0.3856	1.5579	0.1630	0.1616
150	100	0.2417	2.6489	0.2447	0.2430	0.3864	0.9734	0.1644	0.1630
50	150	0.2141	2.9851	0.2008	0.1956	0.3773	1.9264	0.1333	0.1302
100	150	0.2029	2.7919	0.1996	0.1977	0.3804	1.4195	0.1340	0.1326
150	150	0.1973	2.4904	0.1988	0.1973	0.3791	1.0475	0.1319	0.1310
50	200	0.1944	3.5289	0.1763	0.1718	0.3769	1.8067	0.1168	0.1141
100	200	0.1781	3.0194	0.1715	0.1694	0.3787	1.1939	0.1142	0.1131
150	200	0.1726	2.4151	0.1717	0.1705	0.3771	0.8777	0.1128	0.1122

Tables 4-7 report the simulation results for the models with zero restrictions and heterogeneous coefficients. Overall, these tables reaffirm the result that the CCE estimator performs poorly under L2, and the PC estimator performs poorly under E2. Besides this result, there are several additional points worth noting. First, the CCE and CV estimators are inefficient. Under the L1+E1 setup, even when N and T are large, say $N = 150, T = 200$, the average RMSEs of these two estimators are considerably larger than the remaining four estimators. This is not surprising since the two estimation methods do not use the information contained in the zero restrictions; see the discussion in Section 5. Second, several iterations over the LV estimator indeed improve the finite sample performance, especially when N and T are small or moderate. In all combinations of N and T, the ILV estimator outperforms the LV one. Third, under homoscedasticity, the PC, LV and ILV estimators are seen to be efficient since their performance is very close to that of the infeasible GLS estimator, especially when N and T are large.

Table 4: The performance of the six estimators under M2+L1+E1

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.1486	0.0811	0.1527	0.0891	0.0822	0.0790
100	50	0.1483	0.0797	0.1503	0.0868	0.0808	0.0787
150	50	0.1488	0.0792	0.1501	0.0862	0.0803	0.0785
50	100	0.1023	0.0560	0.1046	0.0588	0.0564	0.0546
100	100	0.1026	0.0552	0.1039	0.0575	0.0555	0.0545
150	100	0.1024	0.0549	0.1032	0.0571	0.0552	0.0545
50	150	0.0831	0.0454	0.0849	0.0470	0.0456	0.0443
100	150	0.0831	0.0449	0.0840	0.0463	0.0450	0.0443
150	150	0.0828	0.0445	0.0834	0.0457	0.0447	0.0442
50	200	0.0718	0.0391	0.0732	0.0404	0.0392	0.0382
100	200	0.0717	0.0387	0.0725	0.0396	0.0388	0.0382
150	200	0.0715	0.0384	0.0720	0.0392	0.0385	0.0381

Table 5: The performance of the six estimators under M2+L2+E1

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.2716	0.1231	0.1533	0.1215	0.1210	0.1193
100	50	0.2673	0.1218	0.1512	0.1210	0.1209	0.1200
150	50	0.2674	0.1205	0.1504	0.1201	0.1200	0.1194
50	100	0.2532	0.0849	0.1047	0.0838	0.0836	0.0825
100	100	0.2563	0.0835	0.1034	0.0830	0.0829	0.0823
150	100	0.2562	0.0833	0.1033	0.0829	0.0829	0.0825
50	150	0.2469	0.0691	0.0849	0.0681	0.0680	0.0672
100	150	0.2500	0.0683	0.0845	0.0679	0.0678	0.0674
150	150	0.2476	0.0676	0.0836	0.0673	0.0673	0.0670
50	200	0.2475	0.0595	0.0732	0.0588	0.0587	0.0580
100	200	0.2474	0.0586	0.0725	0.0582	0.0582	0.0578
150	200	0.2476	0.0584	0.0720	0.0581	0.0581	0.0579

Table 6: The performance of the six estimators under under M2+L1+E2

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.2794	0.7402	0.3002	0.2293	0.2172	0.2103
100	50	0.2905	0.2507	0.3020	0.2223	0.2130	0.2081
150	50	0.2980	0.3511	0.3053	0.2282	0.2201	0.2159
50	100	0.2017	0.5204	0.2100	0.1531	0.1495	0.1462
100	100	0.1993	0.1610	0.2081	0.1517	0.1487	0.1468
150	100	0.2057	0.1871	0.2112	0.1524	0.1496	0.1481
50	150	0.1665	0.4558	0.1727	0.1220	0.1198	0.1170
100	150	0.1645	0.3249	0.1675	0.1196	0.1180	0.1166
150	150	0.1641	0.1282	0.1669	0.1202	0.1184	0.1174
50	200	0.1463	0.3222	0.1461	0.1064	0.1048	0.1027
100	200	0.1462	0.1510	0.1484	0.1050	0.1039	0.1027
150	200	0.1447	0.1128	0.1472	0.1043	0.1032	0.1023

Table 7: The performance of the six estimators under under M2+L2+E2

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.2891	1.2307	0.1940	0.1606	0.1600	0.1554
100	50	0.2913	0.7183	0.1910	0.1570	0.1567	0.1545
150	50	0.2894	0.4762	0.1879	0.1567	0.1567	0.1557
50	100	0.2710	0.9264	0.1310	0.1091	0.1080	0.1062
100	100	0.2748	0.6029	0.1306	0.1097	0.1097	0.1086
150	100	0.2720	0.4254	0.1297	0.1079	0.1078	0.1070
50	150	0.2567	0.7998	0.1057	0.0895	0.0882	0.0865
100	150	0.2615	0.5410	0.1061	0.0894	0.0890	0.0880
150	150	0.2654	0.3370	0.1057	0.0887	0.0887	0.0881
50	200	0.2593	0.7218	0.0900	0.0754	0.0748	0.0734
100	200	0.2603	0.5082	0.0901	0.0766	0.0766	0.0759
150	200	0.2566	0.3009	0.0898	0.0749	0.0749	0.0742

7.3 Homogeneous coefficient

In this subsection, we investigate the finite sample properties of the CV and LV estimators suggested in (4.6) and (5.16). We also compute the iterated PC estimator of Bai (2009) and the ML estimator of Bai and Li (2014) for comparison. For simplicity, only the setup "L2+E2" is considered. Table 8 presents the simulation results. Overall, we see that the CV (LV) estimation method gives a consistent estimation for the homogeneous coefficient. Additionally, we see that the performance of the CV(LV) estimator is superior to that of the iterated PC estimator, but inferior to that of the ML estimator. This result is consistent with the two-step method partially taking the cross-sectional heteroscedasticity into account, the iterated PC method not accounting for the cross-sectional heteroscedasticity, and the ML method fully taking the cross-sectional heteroscedasticity into account.

Table 8: The performance of the CV(LV), PC and ML estimators under L2+E2+C2

		CV(LV)		PC		ML	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE
		M1					
50	50	0.0004	0.0216	-0.0002	0.0259	-0.0004	0.0102
100	50	0.0004	0.0151	0.0000	0.0159	0.0002	0.0066
150	50	0.0007	0.0118	0.0008	0.0121	0.0004	0.0052
50	100	0.0006	0.0146	0.0005	0.0194	-0.0000	0.0071
100	100	0.0000	0.0108	-0.0000	0.0117	0.0002	0.0047
150	100	-0.0003	0.0081	-0.0003	0.0086	-0.0002	0.0036
50	150	-0.0000	0.0122	0.0005	0.0181	-0.0000	0.0052
100	150	0.0004	0.0084	0.0002	0.0101	0.0001	0.0037
150	150	-0.0000	0.0067	-0.0002	0.0072	0.0000	0.0031
50	200	0.0008	0.0105	0.0006	0.0173	-0.0001	0.0047
100	200	0.0001	0.0073	0.0002	0.0089	0.0000	0.0033
150	200	0.0000	0.0060	0.0001	0.0065	0.0000	0.0025
		M2					
50	50	0.0003	0.0140	0.0088	0.0224	-0.0001	0.0053
100	50	-0.0002	0.0097	0.0023	0.0111	0.0000	0.0037
150	50	-0.0000	0.0080	0.0012	0.0085	0.0000	0.0030
50	100	0.0003	0.0098	0.0077	0.0185	-0.0001	0.0043
100	100	-0.0001	0.0068	0.0022	0.0086	-0.0001	0.0026
150	100	-0.0001	0.0057	0.0008	0.0063	0.0000	0.0022
50	150	0.0001	0.0075	0.0071	0.0172	0.0002	0.0029
100	150	0.0002	0.0053	0.0025	0.0079	0.0000	0.0021
150	150	0.0000	0.0044	0.0010	0.0052	-0.0001	0.0017
50	200	0.0001	0.0066	0.0075	0.0166	0.0000	0.0026
100	200	0.0001	0.0047	0.0023	0.0071	-0.0000	0.0018
150	200	0.0001	0.0039	0.0010	0.0046	0.0000	0.0015

8 Conclusion

This paper considers the estimation of heterogeneous coefficients in panel data models with common shocks. We propose a two-step method to estimate heterogeneous coefficients, in which the ML method is first used to estimate the loadings and variances of the idiosyncratic errors in a pure factor model, and heterogeneous coefficients are then estimated based on the estimates and structural relations implied by the model. Asymptotic properties of the proposed estimators including the asymptotic representations and limiting distributions are investigated and provided.

In addition, we extend our method to the models with zero restrictions on the partial loadings in the y equation. We point out that efficiency can be gained by using the information contained in the loadings. The asymptotic representation and limiting distribution of the new two-step estimator are studied. We also consider the model with time-invariant regressors.

The proposed estimators are asymptotically efficient in the sense that they have the same limiting distributions as the infeasible GLS estimators. Monte Carlo simulations confirm our theoretical results and show encouraging evidence that the two-step estimators perform robustly in all data setups.

References

Ahn, S. C., and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3), 1203-1227.

Ahn, S. G., Lee, Y. H. and Schmidt, P. (2013). Panel data models with multiple time-varying effects, *Journal of Econometrics*, 174, 1-14.

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 Li, K. (2012a). Statistical analysis of factor models of high dimension, *The Annals of Statistics*, 40(1), 436–465.

Bai, J. and Li, K. (2012b). Maximum likelihood estimation and inference of approximate factor mdoels of high dimension, *manuscipt*.

Bai, J. and Li, K. (2014). Theory and methods of panel data models with interactive effects, *The Annals of Statistics*, 42(1), 142–170.

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

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

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. *Review of Economics and Statistics*, 94, 1014-1024.

Geweke, John. (1977). The dynamic factor analysis of economic time series, in Dennis J. Aigner and Arthur S. Goldberger (eds.) Latent Variables in Socio-Economic Models (Amsterdam: North-Holland).

Kapetanios, G., Pesaran, M. H., and Yamagata, T. (2011). Panels with non-stationary multifactor error structures. Journal of Econometrics, 160(2), 326-348.

Moon, H. and M. Weidner (2012) Linear regression for panel with unknown number of factors as interactive fixed effect. *Manuscipt*, USC

Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, 77(5), 1447-1479.

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

Ross, S. A. (1976). The arbitrage theory of capital asset pricing, *Journal of Economic Theory*,**13(3)**, 341–360.

Sargent, T. J., and Sims, C. A. (1977). Business cycle modeling without pretending to have too much a priori economic theory. New methods in business cycle research, 1, 145-168.

Stock, J. H., and Watson, M. W. (1998). Diffusion indexes (No. w6702). *National Bureau of Economic Research.*

Song, M. (2013). Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications. *Manuscipt*, Columbia University.

Appendix A: Proof of Theorem 3.1

Throughout the appendix, we use C to denote a generic finite constant large enough, which need not to be the same at each appearance. In addition, we introduce following notations for ease of exposition.

$$H = (\Lambda'\Psi^{-1}\Lambda)^{-1}; \quad \hat{H} = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}; \quad R = M_{ff}\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}.$$

We first show that $R = O_p(1)$. The following lemma is useful.

Lemma A.1 *Under Assumptions A-D,*

$$(a) \quad R = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(1)$$

$$(b) \quad R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$$

$$(c) \quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi)\Big]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$$

$$(d) \quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}(\hat{\Psi} - \Psi)\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot o_p\Big(\big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big]^{1/2}\Big)$$

Proof of Lemma A.1: Consider (a). By the definition of R and $\hat{H}$, we have

$$R = M_{ff}\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1} = M_{ff}(\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}^{1/2})\hat{H}^{1/2}$$

By the Cauchy-Schwarz inequality,

$$\|\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}^{1/2}\| = \|\sum_{i=1}^{N}\Lambda_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}^{1/2}\| \le \Big(\sum_{i=1}^{N}\|\Lambda_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2\Big)^{1/2}$$

However,

$$\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2 = \mathrm{tr}\big[\sum_{i=1}^{N}\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}^{1/2}\big] = \mathrm{tr}\big[\hat{H}^{1/2}\hat{H}^{-1}\hat{H}^{1/2}\big] = r \tag{A.1}$$

Given (A.1), together with the boundedness of $\hat{\Sigma}_{ii}^{-1/2}$ and Λ_i, we have

$$\|\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}^{1/2}\| = O_p(N^{1/2})$$

Then (a) follows.

Consider (b). We first show

$$\frac{1}{T}\sum_{t=1}^{T} u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t u_{it}'\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H} = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1/2}) \tag{A.2}$$

By the Cauchy-Schwarz inequality,

$$\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t u_{it}'\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H} \le C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T} f_t u_{it}'\|^2\Big)^{1/2}$$
$$\times\Big(\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2\Big)^{1/2}\|N^{1/2}\hat{H}^{1/2}\|$$

So (A.2) follows by (A.1). Given (A.2) together with result (a), we have (b).

Consider (c), which is equal to

$$\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')])\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}.$$

The above expression is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|^2(\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\|^2\Big)^{1/2}$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(T^{-1/2})$ by (A.1). Then (c) follows.

Consider (d), which is equal to

$$\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}.$$

The above epression is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\|$$

By (A.1), we have $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2=r$, which means $\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|\le\sqrt{r}$ for all i. Given this result, together with the boundedness of $\hat{\Sigma}_{ii}^{-1}$, we have that the above expression is bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|\cdot\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|$$

which is further bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2)^{1/2}(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2)^{1/2},$$

implying (d). □

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

$$\|N^{1/2}\hat{H}^{1/2}\|=O_p(1),\qquad\qquad R=O_p(1).$$

Proof of Proposition A.1: By (3.4), we have $\hat{\Lambda}'\hat{\Psi}^{-1}(M_{zz}-\hat{\Sigma}_{zz})\hat{\Psi}^{-1}\hat{\Lambda}=0$. By

$$M_{zz}=\Lambda M_{ff}\Lambda'+\Lambda\frac{1}{T}\sum_{t=1}^{T}f_tu_t'+\frac{1}{T}\sum_{t=1}^{T}u_tf_t'\Lambda'+\frac{1}{T}\sum_{t=1}^{T}(u_tu_t'-\Psi)+\Psi$$

and $\hat{\Sigma}_{zz}=\hat{\Lambda}\hat{\Lambda}'+\hat{\Psi}$, we have

$$I_r=R'M_{ff}^{-1}R+R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T}f_tu_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}+\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_tf_t'M_{ff}^{-1}R$$

$$+\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi)\Big]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} - \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}(\hat{\Psi}-\Psi)\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} \tag{A.3}$$

Consider the right hand side of (A.3). By Lemma A.1, the first term is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(1)$ and the 2nd-4th terms are all $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$. The last term is equivalent to

$$\hat{H}^{1/2}\Big(\sum_{i=1}^{N}\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2})\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i\hat{H}^{1/2}\Big)\hat{H}^{1/2}$$

which is bounded by

$$\frac{1}{N}\|N^{1/2}\hat{H}^{1/2}\|^2(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\|)$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(N^{-1})$ by $\|\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\| = O_p(1)$ and (A.1). So the last term is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(N^{-1})$. However, by the equation (A.10) of Bai and Li (2012a), we have

$$\|N^{1/2}\hat{H}^{1/2}\|^2 = \mathrm{tr}(N\hat{H}) = \mathrm{tr}\Big[R'M_{ff}^{-1}\Big(\frac{1}{N}\Lambda'\Psi^{-1}\Lambda\Big)^{-1}M_{ff}^{-1}R\Big] + o_p(1).$$

Given these results, we have that the first term dominates the remaining four terms. If R is stochastically unbounded, the right hand side of (A.3) will also be unbounded. However, the left hand side is an identity matrix. A contradiction is obtained. So $R = O_p(1)$, which means $\|N^{1/2}\hat{H}^{1/2} = O_p(1)\|$ by Lemma A.1(a). This completes the proof. $\square$

Lemma A.2 *Under Assumptions A-D with $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$, we have*

$$(a)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t u_{jt}'\Big\|^2 = O_p(T^{-1})$$

$$(b)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Lambda_j\Big\|^2 = O_p(T^{-1})$$

$$(c)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{jt}' - E(u_t u_{jt}')]\Big\|^2 = O_p(T^{-1})$$

$$(d)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})\Big\|^2 = o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\Big)$$

Proof of Lemma A.2. Consider (a), which is bounded by

$$\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda\|\cdot\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t u_{jt}'\Big\|^2$$

By Lemma A.1(a) and Proposition A.1, we have $\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda\| = O_p(1)$. So we have (a).

Consider (b), which is bounded by

$$\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\|^2(\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_j\|^2)$$

which is $O_p(T^{-1})$ by (A.2). Then (b) follows.

Consider (c). The left hand side of (c) is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\Big\|^2\Big)$$

which is $O_p(T^{-1})$ by Proposition A.1. Then (c) follows.

Consider (d). The left hand side of (d) is bounded by

$$C\|\hat{H}^{1/2}\|^2\frac{1}{N}\sum_{j=1}^{N}\Big(\|\hat{H}^{1/2}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{1/2}\|^2\cdot\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\Big)$$

Since $\sum_{j=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1/2}\|^2=r$, the above expression is bounded by

$$Cr\|N^{1/2}\hat{H}^{1/2}\|^2\frac{1}{N^2}\sum_{j=1}^{N}\Big(\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\Big)$$

which is $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\cdot O_p(N^{-1})$ by Proposition A.1. Thus we have (d). □

Proposition A.2 *Under Assumptions A-D, we have*

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j-R'\Lambda_j\|^2=O_p(T^{-1}),\qquad \frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2=O_p(T^{-1}).$$

Proof of Proposition A.2: Consider (3.4), which is equivalent to

$$(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})\hat{\Lambda}_j=(\hat{\Lambda}\hat{\Psi}^{-1}\Lambda)M_{ff}\Lambda_j+(\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_tu_{jt}' \tag{A.4}$$

$$+\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_tf_t'\Lambda_j+\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_tu_{jt}'-E(u_tu_{jt}')]-\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})$$

So we have

$$\hat{\Lambda}_j-R'\Lambda_j=R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T}f_tu_{jt}'+\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_tf_t'\Lambda_j \tag{A.5}$$

$$+\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_tu_{jt}'-E(u_tu_{jt}')]-\hat{H}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})$$

where $R'=(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}(\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda)M_{ff}$ and $\hat{H}=(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}$. We use a_{j1},a_{j2},a_{j3} and a_{j4} to denote the right hand side of (A.5). By triangular inequality,

$$\|\hat{\Lambda}_j-R'\Lambda_j\|\le\|a_{j1}\|+\|a_{j2}\|+\|a_{j3}\|+\|a_{j4}\|$$

Then we have

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j-R'\Lambda_j\|^2\le4\frac{1}{N}\sum_{j=1}^{N}\Big(\|a_{j1}\|^2+\cdots+\|a_{j4}\|^2\Big)$$

Using the results in Lemma A.2, we have

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2 = O_p(T^{-1}) + o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2\Big) \tag{A.6}$$

Consider (3.5), which can be written as

$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) + \Lambda_i'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t u_{it}'\Big) + \Big(\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'\Big)\Lambda_i \tag{A.7}$$
$$-\Lambda_i'R(\hat{\Lambda}_i - R'\Lambda_i) - (\hat{\Lambda}_i - R'\Lambda_i)'R'\Lambda_i - (\hat{\Lambda}_i - R'\Lambda_i)'(\hat{\Lambda}_i - R'\Lambda_i) - \Lambda_i'(RR' - M_{ff})\Lambda_i$$

We use $b_{i1}, b_{i2}, \dots, b_{i7}$ to denote the seven terms on the right hand side. By the Cauchy-Schwarz inequality, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \le 7\frac{1}{N}\sum_{i=1}^{N}\Big(\|b_{i1}\|^2 + \cdots + \|b_{i7}\|^2\Big) \tag{A.8}$$

The first three terms are all $O_p(T^{-1})$. Consider the fourth term, which is bounded in norm by

$$C\|R\|^2\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2 = O_p(T^{-1}) + o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2\Big)$$

by $R = O_p(1)$ and (A.6). The fifth is just the transpose of the fourth. The sixth is $O_p(T^{-2}) + o_p(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2)$, which can be verified by substituting (A.5) in it. Consider the last term. Since the last term is bounded in norm by $\|RR' - M_{ff}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\Lambda_i\|^4$, it suffices to consider the term $RR' - M_{ff}$, which we will show to be $O_p(T^{-1/2}) + o_p([\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2})$. For ease of exposition, we use S to denote the last fourth terms of (A.3). By Lemma A.1 together with $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$, we have

$$S = O_p(T^{-1/2}) + o_p\big([\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2}\big).$$

Now equation (A.3) can be written as $I_r = R'M_{ff}^{-1}R + S$, which is equivalent to $RR' - M_{ff} = -RSR^{-1}M_{ff}$. Since $R = O_p(1)$, if $R \neq o_p(1)$, then $R^{-1} = O_p(1)$. However, R is impossible to be $o_p(1)$ since $I_r = R'M_{ff}^{-1}R + o_p(1)$. So we have

$$RR' - M_{ff} = O_p(T^{-1/2}) + o_p([\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2}), \tag{A.9}$$

implying that the last term is $O_p(T^{-1}) + o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2)$. Given the above results, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1}) + o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2)$$

which implies $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1})$. Substituting this result into (A.6), we have the remaining result of the proposition. This completes the proof of this proposition. □

To prove Theorem 3.1, we further need the following two lemmas.

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

$$(a)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(b)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{jt}' - E(u_t u_{jt}')] = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(c)\quad \hat{H}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj} - \Sigma_{jj}) = O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) + \|\hat{\Lambda}_j - R'\Lambda_j\| \cdot o_p(1)$$

$$(d)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi)\Big]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1}T^{-1/2}) + O_p(T^{-1})$$

$$(e)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}(\hat{\Psi} - \Psi)\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1/2}T^{-1/2})$$

Proof of Lemma A.3: Consider (a). The left hand side of (a) is equal to

$$\hat{H}\sum_{i=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t' + \hat{H}R'\sum_{i=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'$$
$$+ \hat{H}R'\sum_{i=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'$$

The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\big)^{1/2}\big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'\|^2\big)^{1/2}$$

which is $O_p(T^{-1})$ by Proposition A.1 and A.2. The second term is bounded in norm by

$$C^3\|N^{1/2}\hat{H}^{1/2}\|^2\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big)^{1/2}\big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'\|^2\big)^{1/2}$$

which is also $O_p(T^{-1})$ by Proposition A.1 and A.2. The third term is $O_p(N^{-1/2}T^{-1/2})$ by Proposition A.1. So we have (a).

Consider (b). The left hand side (b) is equal to

$$\hat{H}\sum_{i=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]$$
$$+\hat{H}R'\sum_{i=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]$$
$$+\hat{H}R'\sum_{i=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')].$$

The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\big)^{1/2}\big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\|^2\big)^{1/2}$$

which is $O_p(T^{-1})$ by Proposition A.1 and A.2. The second term is bounded in norm by

$$C^3\|N^{1/2}\hat{H}^{1/2}\|^2(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2)^{1/2}(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\|^2)^{1/2}$$

which is also $O_p(T^{-1})$ by Proposition A.1 and A.2. The third term is $O_p(T^{-1/2}T^{-1/2})$ by Proposition A.1. Given these results, we have (b).

Consider (c). The left hand side of (a) is equal to

$$\hat{H}(\hat{\Lambda}_j-R'\Lambda_j)\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})+\hat{H}R'\Lambda_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj}) \tag{A.10}$$

By the boundedness of $\hat{\Sigma}_{jj},\Sigma_{jj}$ and $\hat{H}=O_p(N^{-1})$ (since $\|N^{1/2}\hat{H}^{1/2}\|=O_p(1)$), we have the first term is $\|\hat{\Lambda}_j-R'\Lambda_j\|\cdot o_p(1)$. Consider the second term of (A.10). Substituting (A.7) into the second term, we obtain an expression consisting of 7 terms. The first three terms are all $O_p(N^{-1}T^{-1/2})$ by the boundedness of $\hat{\Sigma}_{ii}$ and $\hat{H}=O_p(N^{-1})$. The fourth and fifth terms are both $\|\hat{\Lambda}_j-R\Lambda_j\|\cdot o_p(1)$. The sixth term is equal to

$$\hat{H}R'\Lambda_j\hat{\Sigma}_{jj}^{-1}(\hat{\Lambda}_j-R'\Lambda_j)'(\hat{\Lambda}_j-R'\Lambda_j)$$

which is bounded by

$$\|N^{1/2}\hat{H}^{1/2}\|\cdot\|R\|\cdot\|\hat{\Sigma}_{jj}^{-1}\|\cdot\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j-R'\Lambda_j\|^2$$

By Propositions A.1 and A.2 and the boundedness of $\hat{\Sigma}_{ii}$, the above expression is $O_p(T^{-1})$. The seventh term is $O_p(N^{-1}T^{-1/2})$ since $S=O_p(T^{-1/2})$. Given these results, we have the second term of (A.10) is $O_p(N^{-1}T^{-1/2})+O_p(T^{-1})+\|\hat{\Lambda}_j-R'\Lambda_j\|\cdot o_p(1)$. Then (c) follows.

Consider (d). The left hand side of (d) is equal to

$$\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}-E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}$$

which is equivalent to

$$\begin{aligned}
&\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Lambda}_i-R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}\\
&\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}(\hat{\Lambda}_j-R'\Lambda_j)'\hat{H}\\
&\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\Lambda_j'R\hat{H}\\
&\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')](\hat{\Sigma}_{jj}^{-1}-\Sigma_{jj}^{-1})\Lambda_j'R\hat{H}\\
&\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\Sigma_{jj}^{-1}\Lambda_j'R\hat{H}
\end{aligned} \tag{A.11}$$

The first term is bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}\|^3 \cdot \Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\Big(\sum_{j=1}^N \|\hat{\Sigma}_{jj}^{-1/2}\hat{\Lambda}_j\hat{H}^{1/2}\|^2\Big)^{1/2}$$
$$\times \Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T [u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)^{1/2} = O_p(T^{-1})$$

by Proposition A.1 and (A.1). The second term is bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}\|^4 \cdot \|R\| \cdot \Big(\frac{1}{N}\sum_{i=1}^N \|\Lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^N \|\hat{\Lambda}_j - R'\Lambda_j\|^2\Big)^{1/2}$$
$$\times \Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T [u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)^{1/2} = O_p(T^{-1})$$

by Propositions A.1 and A.2. The third and fourth terms are both bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}|^4 \cdot \|R\|^2 \cdot \Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^N \|\Lambda_j\|^2\Big)^{1/2}$$
$$\times \Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T [u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)^{1/2} = O_p(T^{-1})$$

by Propositions A.1 and A.2. The last term is $O_p(N^{-1}T^{-1/2})$. Given these results, we have (d).

Consider (e). The left hand side of (c) is equal to

$$\hat{H}^{1/2}\Big(\sum_{i=1}^N \hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}^{1/2}\Big)\hat{H}^{1/2}.$$

The above expression is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|\Big).$$

Since $\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$, then $\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\| \leq \sqrt{r}$ uniformly in i. So the above expression is further bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|\Big).$$

By the Cauchy-Schwarz inequality, the preceding expression is bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2},$$

which is $O_p(N^{-1/2}T^{-1/2})$ by Propositions A.1 and A.2 and (A.1). Then (e) follows. □

Lemma A.4 *Under Assumptions A-D, we have*

$$RR' - M_{ff} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Proof of Lemma A.4. Consider (A.3). Given the results in Lemma A.3, we have

$$R'M_{ff}^{-1}R = I_r + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Taking inverse on the both sides yields

$$R^{-1}M_{ff}R^{-1\prime} = I_r + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Pre-multiplying R and post-multiplying R', together with $R = O_p(1)$, we have Lemma A.4. $\square$

Proof of Theorem 3.1: Consider (A.5). The last three terms of the right hand side of (A.5) are summarized in Lemma A.3(a)-(c). So we have

$$\hat{\Lambda}_j - R'\Lambda_j = R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t u_{jt}' + \|\hat{\Lambda}_j - R'\Lambda_j\| \cdot o_p(1) + o_p(T^{-1/2}). \tag{A.12}$$

The first term of the right hand side is $O_p(T^{-1/2})$. The second term is of smaller order term than the left hand side and hence negligible. Given this result, we have

$$\hat{\Lambda}_j - R'\Lambda_j = O_p(T^{-1/2}). \tag{A.13}$$

Substituting (A.13) into (A.12), we have

$$\hat{\Lambda}_j - R'\Lambda_j = R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t u_{jt}' + o_p(T^{-1/2}).$$

Now consider (A.7). Substituting (A.5) into (A.7), we have

$$\begin{aligned}
\hat{\Sigma}_{ii} - \Sigma_{ii} = {} & \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) - (\hat{\Lambda}_i - R'\Lambda_i)'(\hat{\Lambda}_i - R'\Lambda_i) - \Lambda_i'(RR' - M_{ff})\Lambda_i \\
& - \frac{1}{T}\sum_{t=1}^{T} u_{it}f_t'M_{ff}^{-1}(RR' - M_{ff})\Lambda_i - \Lambda_i'(RR' - M_{ff})M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t u_{it}' \\
& - \Lambda_i'R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T} u_t f_t'\Lambda_i - \Lambda_i'R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{it}' - E(u_t u_{it}')] \\
& - \Lambda_i'\frac{1}{T}\sum_{t=1}^{T} f_t u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}R'\Lambda_i - \frac{1}{T}\sum_{t=1}^{T}[u_{it}u_t' - E(u_{it}u_t')]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}R'\Lambda_i \\
& + \Lambda_i'R\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii}) + (\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}R'\Lambda_i
\end{aligned} \tag{A.14}$$

The second term is $O_p(T^{-1})$ by (A.13). The third term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma A.4. The fourth and fifth terms are both $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ by Lemma A.4. The sixth and eighth terms are both $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma A.3(a).

The seventh and ninth terms are also $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma A.3(b). The last two terms are of smaller order terms than the left hand side and hence negligible. Given these results, we have

$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) + o_p(T^{-1/2})$$

This completes the proof of Theorem 3.1. $\square$

Appendix B: Proofs of the results in Section 4

Lemma B.1 *Under Assumptions A-D, we have*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV} - \beta_i\| = O_p(T^{-1}).$$

Proof of Lemma B.1. Notice

$$\begin{aligned}\hat{\beta}_i^{CV} - \beta_i &= \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21} - \Sigma_{i,22}^{-1}\Sigma_{i,21} \\ &= \hat{\Sigma}_{i,22}^{-1}\big[(\hat{\Sigma}_{i,21} - \Sigma_{i,21}) - (\hat{\Sigma}_{i,22} - \Sigma_{i,22})\Sigma_{i,22}^{-1}\Sigma_{i,21}\big]\end{aligned}$$

By the boundedness of $\hat{\Sigma}_{ii}, \Sigma_{ii}$, we have $\|\hat{\Sigma}_{i,22}^{-1}\| < C, \|\Sigma_{i,22}^{-1}\Sigma_{i,21}\| < C$. Then

$$\|\hat{\beta}_i^{CV} - \beta_i\|^2 \le C\|\hat{\Sigma}_{i,21} - \Sigma_{i,21}\|^2 + C\|\hat{\Sigma}_{i,22} - \Sigma_{i,22}\|^2 \le 2C\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2$$

So $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV} - \beta_i\| = O_p(T^{-1})$ by $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1})$. $\square$

Lemma B.2 *Under Assumptions A-D, we have*

$$(a)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\Lambda}_{i,11}' - \Lambda_{i,21}^{*}\Lambda_{i,11}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\Lambda}_{i,11}' - \Lambda_{i,22}^{*}\beta_i\Lambda_{i,11}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,21}^{*}\beta_i'\Lambda_{i,12}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,22}^{*}\beta_i\beta_i'\Lambda_{i,12}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

where $\Lambda_j^* = R'\Lambda_j$ and $\Lambda_{i,pq}^*$ is defined similarly as $\Lambda_{i,pq}$.

Proof of Lemma B.2. By (A.5), we have

$$\begin{aligned}\hat{\Lambda}_{i,11} - \Lambda_{i,11}^{*} &= \boldsymbol{v}_1 R' M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + \boldsymbol{v}_1\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} u_t f_t'\Lambda_i\boldsymbol{w}_1\Big) \\ &\quad + \boldsymbol{v}_1\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t e_{it} - E(u_t e_{it})] - \boldsymbol{v}_1\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_1\end{aligned} \tag{B.1}$$

$$\hat{\Lambda}_{i,21} - \Lambda^*_{i,21} = \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} u_t f_t' \Lambda_i \boldsymbol{w}_1\Big) + \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t e_{it} - E(u_t e_{it})] - \boldsymbol{v}_2 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_1 \tag{B.2}$$

$$\hat{\Lambda}_{i,12} - \Lambda^*_{i,12} = \boldsymbol{v}_1 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^{T} f_t v_{it}' + \boldsymbol{v}_1 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} u_t f_t' \Lambda_i \boldsymbol{w}_2\Big) + \boldsymbol{v}_1 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t v_{it}' - E(u_t v_{it}')] - \boldsymbol{v}_1 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_2 \tag{B.3}$$

$$\hat{\Lambda}_{i,22} - \Lambda^*_{i,22} = \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^{T} f_t v_{it}' + \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} u_t f_t' \Lambda_i \boldsymbol{w}_2\Big) + \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t v_{it}' - E(u_t v_{it}')] - \boldsymbol{v}_2 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_2 \tag{B.4}$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are defined as $I_r = [\boldsymbol{v}_1, \boldsymbol{v}_2]$ with $\boldsymbol{v}_1$ an $r \times r_1$ matrix and $\boldsymbol{v}_2$ an $r \times r_2$ matrix, respectively. $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are defined as $I_{K+1} = [\boldsymbol{w}_1, \boldsymbol{w}_2]$ with $\boldsymbol{w}_1$ an $(K+1)\times 1$ vector and $\boldsymbol{w}_2$ an $(K+1)\times K$ matrix. In addition, $e_{it} = \epsilon_{it} + v_{it}'\beta_i$.

Consider (a). The left hand side of (a) is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21} - \Lambda^*_{i,21})\Lambda^{*\prime}_{i,11} + \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,21}(\hat{\Lambda}_{i,11} - \Lambda^*_{i,11})'$$
$$+\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21} - \Lambda^*_{i,21})(\hat{\Lambda}_{i,11} - \Lambda^*_{i,11})' = ii_1 + ii_2 + ii_3 \quad \text{say}$$

Consider ii_1. By (B.2), we have

$$\begin{aligned} ii_1 &= \boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t e_{it}\Lambda^{*\prime}_{i,11} + \boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_t e_{it} - E(u_t e_{it})]\Lambda^{*\prime}_{i,11} \\ &\quad + \boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T} u_t f_t'\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda^{*\prime}_{i,11}\Big) - \boldsymbol{v}_2\frac{1}{N}\sum_{i=1}^{N}\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_1\Lambda^{*\prime}_{i,11} \\ &= iii_1 + iii_2 + iii_3 - iii_4 \end{aligned}$$

Consider iii_1. By $\Lambda^{*\prime}_{i,11} = \Lambda'_{i,11}R_{11} + \Lambda'_{i,21}R_{21}$, we have

$$iii_1 = \boldsymbol{v}_2 R' M_{ff}^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t e_{it}\Lambda'_{i,11}\Big)R_{11} + \boldsymbol{v}_2 R' M_{ff}^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t e_{it}\Lambda'_{i,21}\Big)R_{21}$$

which is $O_p(N^{-1/2}T^{-1/2})$ by $R = O_p(1)$. Consider iii_2, which is equivalent to

$$\boldsymbol{v}_2\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}e_{jt}' - E(u_{it}e_{jt})]\Lambda^{*\prime}_{j,11}$$

which is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\| \cdot \Big[\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big]^{-1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}[u_{it}e_{jt}' - E(u_{it}e_{jt}')]\Lambda^{*\prime}_{j,11}\Big\|^2\Big]^{1/2}$$

By $\Lambda^{*\prime}_{j,11} = \Lambda'_{j,11}R_{11} + \Lambda'_{j,21}R_{21}$, we have

$$\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}[u_{it}e_{jt} - E(u_{it}e_{jt})]\Lambda^{*\prime}_{j,11} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}[u_{it}e_{jt} - E(u_{it}e_{jt})]\Lambda'_{j,11}R_{11}$$
$$+\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}[u_{it}e_{jt} - E(u_{it}e_{jt})]\Lambda'_{j,21}R_{21} = O_p(N^{-1/2}T^{-1/2}).$$

Given this result, together with $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$ and (A.1), we have $iii_2 = O_p(N^{-1/2}T^{-1/2})$.

Consider iii_3. Notice that

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda^{*\prime}_{i,11} = \Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda'_{i,11}\Big)R_{11} + \Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda'_{i,21}\Big)R_{21} = O_p(1).$$

Given the above result, together with Lemma A.3(a), we obtain $iii_3 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider iii_4, which is equal to

$$\Big[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_2\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1\Lambda'_{i,11}\Big]R_{11}+\Big[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_2\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1\Lambda'_{i,21}\Big]R_{21}. \quad \text{(B.5)}$$

Consider the first term of the above expression. Ignore R_{11} and $\boldsymbol{v}_2$, the expression in the bracket is bounded in norm by

$$\frac{1}{N}\|\hat{H}^{1/2}\|\sum_{i=1}^{N}\big(\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|\cdot\|\hat{\Sigma}_{ii}^{-1/2}\|\cdot\|\boldsymbol{w}_1\Lambda'_{i,11}\|\cdot\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|\big)$$

which is further bounded by

$$\frac{1}{N}\|N^{1/2}\hat{H}^{1/2}\|\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\boldsymbol{w}_1\Lambda'_{i,11}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2}$$

The above expression is $O_p(N^{-1}T^{-1/2})$ by $\|\hat{\Sigma}_{ii}^{-1/2}\| < C, \|\boldsymbol{w}_1\Lambda'_{i,11}\| < C$ and Propositions A.1 and A.2 as well as (A.1). Given this result, together with $R = O_p(1)$, we have the first term of (B.5) is $O_p(N^{-1}T^{-1/2})$. The second term can be proved to be $O_p(N^{-1}T^{-1/2})$ similarly as the first term. So we have $iii_4 = O_p(N^{-1}T^{-1/2})$. Summarizing all the results, we have $ii_1 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Term ii_2 can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ similarly as ii_1 and the details are omitted. For term ii_3, notice that it is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_{i,21}-\Lambda^*_{i,21}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_{i,11}-\Lambda^*_{i,11}\|^2\Big)^{1/2}$$

However, we have $\|\hat{\Lambda}_{i,21}-\Lambda^*_{i,21}\| \le \|\hat{\Lambda}_i-\Lambda^*_i\| = \|\hat{\Lambda}_i-R'\Lambda_i\|$ and $\|\hat{\Lambda}_{i,11}-\Lambda^*_{i,11}\| \le \|\hat{\Lambda}_i-\Lambda^*_i\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$. Given this result, the preceding expression is bounded by

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2 = O_p(T^{-1})$$

by Proposition A.2. Summarizing the results on ii_1, ii_2 and ii_3, we obtain (a).

Consider (b). The left hand side of (b) can be written as

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\Lambda}'_{i,11} - \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11}\Big)$$
$$+\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} - \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\beta_i\Lambda^{*\prime}_{i,11}\Big) = ii_4 + ii_5, \quad \text{say}$$

Consider ii_4, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}-\Lambda^*_{i,22})\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} + \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\hat{\beta}_i^{CV}(\hat{\Lambda}_{i,11}-\Lambda^*_{i,11})'$$
$$+\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}-\Lambda^*_{i,22})\hat{\beta}_i^{CV}(\hat{\Lambda}_{i,11}-\Lambda^*_{i,11})' = iii_5 + iii_6 + iii_7, \quad \text{say}$$

Consider iii_5. By (B.4), we have

$$\begin{aligned} iii_5 = \boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v'_{it}\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} + \boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f'_t\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\boldsymbol{w}_2\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11}\Big) \\ +\boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_t v'_{it} - E(u_t v'_{it})]\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} \\ -\boldsymbol{v}_2\hat{H}\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_2\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} \end{aligned} \tag{B.6}$$

Consider the first term of (B.6), which can be written as

$$\boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v'_{it}\beta_i\Lambda^{*\prime}_{i,11} + \boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v'_{it}(\hat{\beta}_i^{CV}-\beta_i)\Lambda^{*\prime}_{i,11}. \tag{B.7}$$

Treating $v'_{it}\beta_i$ as a new e_{it}, the first term of the above expression can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the iii_1. By $\Lambda^{*\prime}_{i,11} = \Lambda'_{i,11}R_{11} + \Lambda'_{i,21}R_{21}$, the second term is equal to

$$\boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v'_{it}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,11}R_{11} + \boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v'_{it}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,21}R_{21}.$$

The first term of the above expression is bounded in norm by

$$\|\boldsymbol{v}_2 R'\|\cdot\|M_{ff}^{-1}\|\cdot\|\Lambda_{i,11}\|\cdot\|R_{11}\|\cdot\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t v'_{it}\Big\|^2\Big)^{1/2},$$

which is $O_p(T^{-1})$ by Lemma B.1 and $R = O_p(1)$. The second term can be proved similarly as the first term. Given these results, we have the expression of (B.7) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. So the first term of (B.6) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider the second term. First note that

$$\|\hat{\beta}_i^{CV}\| < C, \qquad \forall\, i \le N \tag{B.8}$$

The above result is due to the boundedness of $\hat{\Sigma}_{ii}$ and $\hat{\beta}_i^{CV} = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$. Given this result, we have

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_2 \hat{\beta}_i^{CV}\Lambda_{i,11}^{*\prime} = O_p(1).$$

Given the above result, together with Lemma A.3(a), we have the second term of (B.6) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider the third term, which is equal to

$$\boldsymbol{v}_2 \frac{1}{NT}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime} \tag{B.9}$$

Ignore $\boldsymbol{v}_2$, The above expression can be rewritten as

$$\frac{1}{NT}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}$$
$$\frac{1}{NT}\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}$$
$$\frac{1}{NT}\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}$$

The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2\Big)^{1/2}$$
$$\times\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\Big\|^2\Big)^{1/2},$$

By the boundedness of $\hat{\beta}_i^{CV}$ and $\Lambda_{i,11}^{*\prime} = \Lambda_{i,11}'R_{11} + \Lambda_{i,21}'R_{21}$,

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2 &\le C\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_{j,11}^{*\prime}\|^2 \\ &\le 2C\big(\|R_{11}\|^2\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_{j,11}\|^2 + \|R_{21}\|^2\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_{j,21}\|^2\big)\end{aligned}$$

Given this result, we have that the first term is $O_p(T^{-1})$ by Propositions A.1 and (A.2). The second term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \|R\| \cdot \Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2\Big)^{1/2}$$
$$\times\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\Big\|^2\Big)^{1/2},$$

which is $O_p(T^{-1})$ by the same arguments. The last term is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \|R\| \cdot (\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2)^{1/2}$$
$$\times(\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_i\Sigma_{ii}^{-1}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\|^2)^{1/2}$$

which is $O_p(N^{-1/2}T^{-1/2})$. Given these results, we have the third term of (B.6) is $O_p(N^{-1/2}T^{-1/2})+ O_p(T^{-1})$.

Consider the fourth term. Ignore $\boldsymbol{v}_2$, this term is bounded in norm by

$$C\|\hat{H}^{1/2}\|\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}\Lambda_{i,11}^{*\prime}\|^2\Big)^{1/2}$$

Notice $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$. So $\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 \le r$ for all i. This leads to

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \le r\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2.$$

Given the above result, together with $\hat{H} = O_p(N^{-1})$, we have the fourth term is $O_p(N^{-1/2}T^{-1/2})$. Summarizing all the results, we have $iii_5 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Term iii_6 can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as iii_5 and the details are omitted. Consider iii_7, which is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_{i,22} - \Lambda_{i,22}^*\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}\|^2 \cdot \|\hat{\Lambda}_{i,11} - \Lambda_{i,11}^*\|^2\Big)^{1/2}$$

By the boundedness of $\hat{\beta}_i^{CV}$, together with $\|\hat{\Lambda}_{i,22} - \Lambda_{i,22}^*\| \le \|\hat{\Lambda}_i - \Lambda_i^*\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$ and $\|\hat{\Lambda}_{i,11} - \Lambda_{i,11}^*\| \le \|\hat{\Lambda}_i - \Lambda_i^*\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$, we have that the preceding expression is bounded by

$$C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2 = O_p(T^{-1})$$

by Proposition A.2. Summarizing the results on iii_5, iii_6 and iii_7, we have

$$ii_4 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}).$$

We proceed to consider ii_5, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^*(\hat{\beta}_i^{CV} - \beta_i)\Lambda_{i,11}^{*\prime}.$$

By $\Lambda_{i,22}^* = R_{12}'\Lambda_{i,12} + R_{22}'\Lambda_{i,22}$ and $\Lambda_{i,11}^{*\prime} = \Lambda_{i,11}'R_{11} + \Lambda_{i,21}'R_{21}$, the above expression can be written as

$$R_{12}'\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}(\hat{\beta}_i^{CV} - \beta_i)\Lambda_{i,11}'\Big)R_{11} + R_{12}'\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}(\hat{\beta}_i^{CV} - \beta_i)\Lambda_{i,21}'\Big)R_{21}$$

$$R'_{22}\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,11}\Big)R_{11}++R'_{22}\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,21}\Big)R_{21}$$

The derivations on the above four terms are almost the same. So we only choose the first one to illustrate. Ignore R'_{12} and R_{11}. By

$$\hat{\beta}_i^{CV}-\beta_i=\hat{\Sigma}_{i,22}^{-1}[(\hat{\Sigma}_{i,21}-\Sigma_{i,21})-(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\beta_i],$$

we can rewrite the expression in the bracket as

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\hat{\Sigma}_{i,22}^{-1}(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11}-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\hat{\Sigma}_{i,22}^{-1}(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\beta_i\Lambda'_{i,11}.$$

Again, the derivations on the above two terms are almost the same. So we only choose the first term to illustrate. This term is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}(\hat{\Sigma}_{i,22}^{-1}-\Sigma_{i,22}^{-1})(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11}+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11}. \quad \text{(B.10)}$$

The first term of the above expression is

$$-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\hat{\Sigma}_{i,22}^{-1}(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\Sigma_{i,22}^{-1}(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11},$$

which is bounded in norm by $C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2=O_p(T^{-1})$ by $\|\Lambda_{i,21}\|<C, \|\Lambda_{i,11}\|<C, \|\hat{\Sigma}_{i,22}^{-1}\|<C, \|\Sigma_{i,22}^{-1}\|<C$ as well as $\|\hat{\Sigma}_{i,22}-\Sigma_{i,22}\|\le\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|$ and $\|\hat{\Sigma}_{i,21}-\Sigma_{i,21}\|\le\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|$. So the first term of (B.10) is $O_p(T^{-1})$. Consider the second term, which, by (A.14), be be written as

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{i,12}\Sigma_{i,22}^{-1}[v_{it}e_{it}-E(v_{it}e_{it})]\Lambda'_{i,11}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}'_2(\hat{\Lambda}_i-R'\Lambda_i)'(\hat{\Lambda}_i-R'\Lambda_i)\boldsymbol{w}_1\Lambda'_{i,11}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}'_2\Lambda'_i(RR'-M_{ff})\Lambda_i\boldsymbol{w}_1\Lambda'_{i,11}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}v_{it}f'_t\Big)M_{ff}^{-1}(RR'-M_{ff})\Lambda_i\boldsymbol{w}_1\Lambda'_{i,11}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}'_2\Lambda'_i(RR'-M_{ff})M_{ff}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big)\Lambda'_{i,11}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}'_2\Lambda'_iR\Big(\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_tf'_t\Big)\Lambda_i\boldsymbol{w}_1\Lambda'_{i,11} \qquad \text{(B.11)}\\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}'_2\Lambda'_iR\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_te_{it}-E(u_te_{it})]\Lambda'_{i,11}
\end{aligned}$$

$$
-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}\Big)R'\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}'
$$

$$
-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{it}u_t'-E(v_{it}u_t')]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}R'\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}'
$$

$$
+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'R\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1\Lambda_{i,11}'
$$

$$
+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}R'\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}'.
$$

where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are defined as $I_{K+1}=[\boldsymbol{w}_1,\boldsymbol{w}_2]$ where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are $(K+1)\times 1$ and $(K+1)\times K$, respectively. The first term is $O_p(N^{-1/2}T^{-1/2})$. The second term is bounded in norm by $C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i-R'\Lambda_i\|^2=O_p(T^{-1})$ by Proposition A.2. The third term is $C\|RR'-M_{ff}\|$ which is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma A.4. The fourth term is bounded in norm by

$$
\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Lambda_{i,12}\Sigma_{i,22}^{-1}\|\cdot\|\frac{1}{T}\sum_{t=1}^{T}v_{it}f_t'\|\cdot\|\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}'\|\Big)\cdot\|M_{ff}^{-1}\|\cdot\|RR'-M_{ff}\|,
$$

which is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$ by Lemma A.4. The fifth term is also $O_p(N^{-1/2}T^{-1})$ $+O_p(T^{-3/2})$ by the similar arguments in the fourth. The sixth and eighth terms are both bounded in norm by

$$
C\|R\|\cdot\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Big\|
$$

which is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma A.3(a). For the seventh term, we temporarily use L_i to denote $\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'$. Notice the left hand side of the seventh term is an $r_1\times r_1$ matrix. So it suffices to show its the (p,q)th element $(p,q=1,2\dots,r_1)$, which is equal to

$$
\frac{1}{N}\sum_{i=1}^{N}L_{i,p}R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t e_{it}-E(u_t e_{it})]\Lambda_{i,11,q}
$$

where $L_{i,p}$ is the pth row of L_i and $\Lambda_{i,11,q}$ is the qth element of $\Lambda_{i,11}$. The above expression can be rewritten as

$$
tr\Big[R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_t e_{it}-E(u_t e_{it})]\Lambda_{i,11,q}L_{i,p}\Big)\Big].
$$

The expression in the trace operator is bounded in norm by

$$
C\cdot\|N^{1/2}\hat{H}^{1/2}\|\cdot\|R\|\cdot\Big(\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}^{-1/2}\hat{\Lambda}_j\hat{H}^{1/2}\|^2\Big)^{1/2}
$$

$$
\times\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_{jt}e_{it}-E(u_{jt}e_{it})]\Lambda_{i,11,q}L_{i,p}\Big\|^2\Big)^{1/2}
$$

which is $O_p(N^{-1/2}T^{-1/2})$. So the seventh term is $O_p(N^{-1/2}T^{-1/2})$. The ninth term can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the seventh. Consider the tenth term, which is bounded in norm by

$$CN^{-1/2}\|\hat{H}^{1/2}\| \cdot \|R\| \cdot \Big(\sum_{i=1}^{N}\|\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2},$$

which is $O_p(N^{-1}T^{-1/2})$. So the tenth term is $O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})$. The eleventh term can be proved to be $O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})$ similarly as the tenth. Summarizing all the results, we have the second term of (B.10) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. This leads to $ii_5 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Given the results on ii_4 and ii_5, we have (b).

Result (c) can be proved similarly as (b) and the details are omitted.$\boldsymbol{\beta}, \beta$

Consider (d). The left hand side of (d) can be written as

$$\begin{aligned}&\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^*\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\Big)\\ +&\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^*\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} - \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^*\beta_i\beta_i'\Lambda_{i,12}^{*\prime}\Big) = ii_6 + ii_7 \quad \text{say}\end{aligned}$$

We first consider ii_6, which is equal to

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}-\Lambda_{i,22}^*)\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} + \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^*\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}(\hat{\Lambda}_{i,12}-\Lambda_{i,12}^*)'\\ +&\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}-\Lambda_{i,22}^*)\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}(\hat{\Lambda}_{i,12}-\Lambda_{i,12}^*)' = iii_8 + iii_9 + iii_{10} \quad \text{say}\end{aligned}$$

Consider iii_8, which is equal to

$$\begin{aligned}iii_8 =\ & \boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{it}'\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\\ &+\boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\boldsymbol{w}_2\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\Big)\\ &+\boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_t v_{it}' - E(u_t v_{it}')]\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\\ &-\boldsymbol{v}_2\hat{H}\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1^{-}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\end{aligned}\tag{B.12}$$

Consider the first term. Ignore $\boldsymbol{v}_2 R' M_{ff}^{-1}$, the remaining expression can be written as

$$\begin{aligned}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{it}'(\hat{\beta}_i^{CV}-\beta_i)(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{it}'\beta_i\beta_i'\Lambda_{i,12}^{*\prime}\\ +&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{it}'(\hat{\beta}_i^{CV}-\beta_i)\beta_i'\Lambda_{i,12}^{*\prime} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{it}'\beta_i(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime}.\end{aligned}$$

The first term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t v_{it}'\Big\|^2\Big)^{1/2}$$

However, by the boundedness of $\hat{\beta}_i^{CV}$ and β_i ($\hat{\beta}_i^{CV}$ is bounded due to the boundedness of $\hat{\Sigma}_{ii}$ and $\hat{\beta}_i^{CV}=\hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$), we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^4\le C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2$$

Given the above result, we have the first term is $O_p(T^{-1})$. The second term is $O_p(N^{-1/2}T^{-1/2})$. The third and fourth terms are both bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t v_{it}'\Big\|^2\Big)^{1/2}\|R\|=O_p(T^{-1}).$$

Given these results, we have the first term of (B.12) is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. The second term is also $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma A.3(a) and the fact

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\boldsymbol{w}_2\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}=O_p(1).$$

The third term can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ similarly as proving (B.9) by replacing $\hat{\beta}_i^{CV}\Lambda_{i,11}^{*\prime}$ with $\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}$. The last term of (B.12) can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the last one of (B.6). Given these results, we have $iii_8=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. Term iii_9 is also $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$, which can be proved by the same arguments in deriving iii_8. Term iii_{10} can be shown to be $O_p(T^{-1})$ similarly as iii_7. Summarizing the results on iii_8, iii_9 and iii_{10}, we have $ii_6=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$.

Consider ii_7, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}(\hat{\beta}_i^{CV}-\beta_i)\beta_i'\Lambda_{i,12}^{*\prime}+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}\beta_i(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}(\hat{\beta}_i^{CV}-\beta_i)(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime}$$

Treating $\beta_i'\Lambda_{i,12}^{*\prime}$ as a new $\Lambda_{i,11}^{*\prime}$, the first term can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ similarly as ii_5. The second term is also $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by the same arguments. The third term is bounded in norm by

$$C\|R\|^2\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2=O_p(T^{-1}).$$

Given the above results, we have $ii_7=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. Summarizing the results on ii_6 and ii_7, we have (d).

This completes the proof of Lemma B.2. $\square$

Lemma B.3 *Under Assumptions A-D, we have*

$$(a)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\Lambda}'_{i,11}-\Lambda^*_{i,11}\Lambda^{*\prime}_{i,11}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\Lambda}'_{i,11}-\Lambda^*_{i,12}\beta_i\Lambda^{*\prime}_{i,11}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\beta}_i^{CV\prime}\hat{\Lambda}'_{i,12}-\Lambda^*_{i,11}\beta_i'\Lambda^{*\prime}_{i,12}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}'_{i,12}-\Lambda^*_{i,12}\beta_i\beta_i'\Lambda^{*\prime}_{i,12}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

Proof of Lemma B.3. The proof of Lemma B.3 is quite similar as the one of Lemma B.2. So we omit it. □

Lemma B.4 *Under Assumptions A-D, we have*

$$\hat{V}-V = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

Proof of Lemma B.4. By the definitions of $\hat{V}$ and V, i.e.,

$$\hat{V} = \Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}-\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]\Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]^{-1}$$

and

$$V = \Big[\sum_{i=1}^{N}(\Lambda^*_{i,21}-\Lambda^*_{i,22}\beta_i)(\Lambda^*_{i,11}-\Lambda^*_{i,12}\beta_i)'\Big]\Big[\sum_{i=1}^{N}(\Lambda^*_{i,11}-\Lambda^*_{i,12}\beta_i)(\Lambda^*_{i,11}-\Lambda^*_{i,12}\beta_i)'\Big]^{-1},$$

together with the fact that $\hat{A}\hat{B}^{-1}-AB^{-1} = ((\hat{A}-A)-AB^{-1}(\hat{B}-B))\hat{B}^{-1}$, we have

$$\hat{V}-V = (J_1-VJ_2)\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]^{-1},$$

where

$$J_1 = \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}-\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})' - \frac{1}{N}\sum_{i=1}^{N}(\Lambda^*_{i,21}-\Lambda^*_{i,22}\beta_i)(\Lambda^*_{i,11}-\Lambda^*_{i,12}\beta_i)'$$

$$J_2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})' - \frac{1}{N}\sum_{i=1}^{N}(\Lambda^*_{i,11}-\Lambda^*_{i,12}\beta_i)(\Lambda^*_{i,11}-\Lambda^*_{i,12}\beta_i)'$$

Consider J_1, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\Lambda}'_{i,11}-\Lambda^*_{i,21}\Lambda^{*\prime}_{i,11}) - \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\Lambda}'_{i,11}-\Lambda^*_{i,22}\beta_i\Lambda^{*\prime}_{i,11})$$

$$-\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,21}^*\beta_i'\Lambda_{i,12}^{*\prime}) + \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,22}^*\beta_i\beta_i'\Lambda_{i,12}^{*\prime}).$$

By Lemma B.2, we have $J_1 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. Consider J_2, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\Lambda}_{i,11}' - \Lambda_{i,11}^*\Lambda_{i,11}^{*\prime}) - \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\Lambda}_{i,11}' - \Lambda_{i,12}^*\beta_i\Lambda_{i,11}^{*\prime})$$
$$-\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,11}^*\beta_i'\Lambda_{i,12}^{*\prime}) + \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,12}^*\beta_i\beta_i'\Lambda_{i,12}^{*\prime}).$$

By Lemma B.3, we have $J_2 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. Given $J_1 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ and $J_2 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$, together with $V = O_p(1)$, we have $\hat{V} - V = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. □

Lemma B.5 *Under Assumptions A-D, we have*

$$(a)\quad \hat{\Lambda}_{i,22} - \hat{V}\hat{\Lambda}_{i,12} - (\Lambda_{i,22}^* - V\Lambda_{i,12}^*) = R_{22\cdot1}'\frac{1}{T}\sum_{t=1}^{T}h_t^\star v_{it}' + o_p(T^{-1/2})$$
$$(b)\quad \hat{\Lambda}_{i,21} - \hat{V}\hat{\Lambda}_{i,11} - (\Lambda_{i,21}^* - V\Lambda_{i,11}^*) = R_{22\cdot1}'\frac{1}{T}\sum_{t=1}^{T}h_t^\star e_{it} + o_p(T^{-1/2})$$

where $R_{22\cdot1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$, $f_t^\star = M_{ff}^{-1}f_t \equiv [g_t^{\star\prime}, h_t^{\star\prime}]'$ *and* $e_{it} = \epsilon_{it} + \beta_i'v_{it}$.

Proof of Lemma B.5. The left hand side of (a) is equal to

$$(\hat{\Lambda}_{i,22} - \Lambda_{i,22}^*) - V(\hat{\Lambda}_{i,12} - \Lambda_{i,12}^*) - (\hat{V} - V)\hat{\Lambda}_{i,12} \tag{B.13}$$

The last term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by $\hat{\Lambda}_{i,12} = \Lambda_{i,12}^* + o_p(1)$ and Lemma B.4. Substituting (B.3) and (B.4) into (B.13), we can rewrite the first two term of (B.13) (denoted by i_1) as

$$i_1 = (\boldsymbol{v}_2 - V\boldsymbol{v}_1)R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T}f_tv_{it}' + (\boldsymbol{v}_2 - V\boldsymbol{v}_1)\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}u_tf_t'\Lambda_i\boldsymbol{w}_2\Big)$$
$$+ (\boldsymbol{v}_2 - V\boldsymbol{v}_1)\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_tv_{it}' - E(u_tv_{it}')] - (\boldsymbol{v}_2 - V\boldsymbol{v}_1)\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_2$$

By Lemma A.3, the last three terms are $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. However,

$$(\boldsymbol{v}_2 - V\boldsymbol{v}_1)R' = [0_{r_2\times r_1}, R_{22\cdot1}'].$$

So we have

$$i_1 = [0_{r_2\times r_1}, R_{22\cdot1}']M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T}f_tv_{it}' + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

implies (a).

Result (b) can be proved similarly as (a). The details are omitted. □

Lemma B.6 *Under Assumptions A-D, we have, for all i,*

$$\hat{W}_i = W_i + o_p(1).$$

Proof of Lemma B.6. Let

$$\hat{W}_{i,11} = \Big\{ \Big[\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t'\Big] - \Big[\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}_t'\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t'\Big]^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{h}_t'\Big]\Big\}^{-1}$$

We first show that

$$\hat{W}_{i,11} = W_{i,11} + o_p(1) \tag{B.14}$$

where

$$W_{i,11} = R_{22\cdot 1}'M_{hh}^{\star}R_{22\cdot 1} = R_{22\cdot 1}'(T^{-1}\sum_{t=1}^{T}h_t^{\star}h_t^{\star\prime})R_{22\cdot 1} = R_{22\cdot 1}'(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})^{-1}R_{22\cdot 1}$$

with $R_{22\cdot 1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$. The last equation of the above expression is due to the definition of $f_t^{\star}$, i.e., $f_t^{\star} \equiv (g_t^{\star\prime}, h_t^{\star\prime})' = M_{ff}^{-1}f_t$.

Let $f_t^* = [g_t^{*\prime}, h_t^{*\prime}]' = R^{-1}f_t$ and $\eta_t^* = g_t^* + V'h_t^*$. By $f_t^* = R^{-1}f_t$, we have

$$g_t^* = (R_{11}^{-1} + R_{11}^{-1}R_{12}R_{22\cdot 1}^{-1}R_{21}R_{11}^{-1})g_t - R_{11}^{-1}R_{12}R_{22\cdot 1}^{-1}h_t \tag{B.15}$$

$$h_t^* = -R_{22\cdot 1}^{-1}R_{21}R_{11}^{-1}g_t + R_{22\cdot 1}^{-1}h_t \tag{B.16}$$

From (B.15) and (B.16), together with $V = R_{11}^{-1}R_{12}$, we have

$$\eta_t^* = g_t^* + V'h_t^* = R_{11}^{-1}g_t. \tag{B.17}$$

Thus,

$$\Big\{ \Big[\frac{1}{T}\sum_{t=1}^{T}h_t^*h_t^{*\prime}\Big] - \Big[\frac{1}{T}\sum_{t=1}^{T}h_t^*\eta_t^{*\prime}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime}\Big]^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}\eta_t^*h_t^{*\prime}\Big]\Big\}^{-1} \tag{B.18}$$

$$= \Big\{R_{22\cdot 1}^{-1}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})R_{22\cdot 1}^{-1\prime}\Big\}^{-1} = R_{22\cdot 1}'(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})^{-1}R_{22\cdot 1} = W_{i,11}$$

So to prove (B.14), it suffices to prove

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t' - \frac{1}{T}\sum_{t=1}^{T}h_t^*h_t^{*\prime} = o_p(1), \tag{B.19}$$

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}_t' - \frac{1}{T}\sum_{t=1}^{T}h_t^*\eta_t^{*\prime} = o_p(1), \tag{B.20}$$

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}_t' - \frac{1}{T}\sum_{t=1}^{T}\eta_t^*\eta_t^{*\prime} = o_p(1). \tag{B.21}$$

Notice that

$$\hat{f}_t = (\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i')^{-1}(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}z_{it}).$$

Then we have

$$\hat{f}_t - f_t^* = -\Big(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\Big)^{-1}\Big(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\Big)f_t^* + \Big(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\Big)^{-1}\Big(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}u_{it}\Big). \tag{B.22}$$

where $f_t^* = R^{-1}f_t$. Equation (B.22) leads to

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - f_t^*)f_t^{*\prime} = o_p(1), \tag{B.23}$$

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - f_t^*)(\hat{f}_t - f_t^*)' = o_p(1). \tag{B.24}$$

To see this, notice the left hand side of (B.23) is equal to

$$-\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t^*f_t^{*\prime}\Big) + \hat{H}\Big(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}u_{it}f_t^{*\prime}\Big),$$

where $\hat{H} = (\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i')^{-1}$. The second term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma A.3(a). The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t^*f_t^{*\prime}\Big\|$$

which is $O_p(T^{-1/2})$ by Proposition A.2. So we obtain (B.23).

Proceed to consider (B.24). The left hand side of (B.24) is bounded in norm by

$$2\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\Big\|^2\Big(\frac{1}{T}\sum_{t=1}^{T}\|f_t^*\|^2\Big) + 2\frac{1}{T}\sum_{T=1}^{T}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}u_{it}\Big\|^2.$$

The first term is $O_p(T^{-1/2})$ since

$$\begin{aligned}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\Big\| \le & C\|N^{1/2}\hat{H}^{1/2}\|\cdot\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2} \\ & \times\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\end{aligned}$$

which is $O_p(T^{-1/2})$. Ignore 2, the second term is equal to

$$tr\Big[\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}\Big] + tr\Big[\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i\hat{H}\Big]$$

which is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma A.3(d) and (e). So we have (B.24).

Given (B.23) and (B.24), we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t\hat{f}_t' - \frac{1}{T}\sum_{t=1}^{T}f_t^*f_t^{*\prime} = o_p(1) \tag{B.25}$$

From (B.25), we immediately obtain (B.19). Now consider (B.20). By the definition of $\hat{\eta}_t$,

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}_t' = \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(\hat{g}_t + \hat{V}'\hat{h}_t)' = \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{g}_t' + (\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t')\hat{V}$$
$$= \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{g}_t' + (\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t')V + (\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t')(\hat{V}-V)$$

From (B.25), we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{g}_t' = \frac{1}{T}\sum_{t=1}^{T}h_t^*g_t^{*\prime} + o_p(1)$$

Given the above result, together with (B.19) and Lemma B.4, we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}_t' = \frac{1}{T}\sum_{t=1}^{T}h_t^*g_t^{*\prime} + (\frac{1}{T}\sum_{t=1}^{T}h_t^*h_t^{*\prime})V + o_p(1) = \frac{1}{T}\sum_{t=1}^{T}h_t^*\eta_t^{*\prime} + o_p(1).$$

Equation (B.21) can be proved similarly as (B.20) and the proof is omitted. Given (B.19), (B.20) and (B.21), we have (B.14).

Given (B.14), in combination with $\hat{\Sigma}_{i,22} = \Sigma_{i,22} + o_p(1)$, we have $\hat{W}_i = W_i + o_p(1)$. This completes the proof. □

Proof of Theorem 5.1. The consistency of $\hat{\beta}_i^{LV}$ is implied by the asymptotic expression. So we only focus on the derivation of the asymptotic expression. Notice that $\hat{\beta}_i^{LV}$ is defined by

$$\hat{\beta}_i^{LV} = (\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i)^{-1}(\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\delta}_i).$$

By $\Delta_i\beta_i = \delta_i$, we also have

$$\beta_i = (\Delta_i'\hat{W}_i^{-1}\Delta_i)^{-1}(\Delta_i'\hat{W}_i^{-1}\delta_i),$$

From the two preceding equations, we have

$$\begin{aligned}\hat{\beta}_i^{LV} - \beta_i &= (\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i)^{-1}[(\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\delta}_i - \Delta_i'\hat{W}_i^{-1}\delta_i) - (\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i - \Delta_i'\hat{W}_i^{-1}\Delta_i)\beta_i] \\ &= \Big\{(\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i)^{-1}[\Delta_i'\hat{W}_i^{-1}(\hat{\delta}_i-\delta_i) - \Delta_i'\hat{W}_i^{-1}(\hat{\Delta}_i-\Delta_i)\beta_i]\Big\} \qquad \text{(B.26)} \\ &+ \Big\{(\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i)^{-1}[(\hat{\Delta}_i-\Delta_i)\hat{W}_i^{-1}(\hat{\delta}_i-\delta_i) - (\hat{\Delta}_i-\Delta_i)\hat{W}_i^{-1}(\hat{\Delta}_i-\Delta_i)\beta_i]\Big\}\end{aligned}$$

By Lemma B.5 and Theorem 3.1, we have

$$\hat{\Delta}_i - \Delta_i = \begin{bmatrix} R_{22\cdot1}' & 0 \\ 0 & I_K \end{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} h_t^{\star}v_{it}' \\ v_{it}v_{it}' - E(v_{it}v_{it}') \end{bmatrix} + o_p(T^{-1/2}) \qquad \text{(B.27)}$$

and

$$\hat{\delta}_i - \delta_i = \begin{bmatrix} R_{22\cdot1}' & 0 \\ 0 & I_K \end{bmatrix} \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix} h_t^{\star}e_{it} \\ v_{it}e_{it} - E(v_{it}e_{it}) \end{bmatrix} + o_p(T^{-1/2}) \qquad \text{(B.28)}$$

where $e_{it} = \epsilon_{it} + \beta_i'v_{it}$. Equations (B.27) and (B.28) implies that $\hat{\Delta}_i = \Delta_i + O_p(T^{-1/2})$ and $\hat{\delta}_i = \delta_i + O_p(T^{-1/2})$. Given these results, together with Lemma B.6, we have

$$\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i - \Delta_i'W_i^{-1}\Delta_i = o_p(1), \qquad \hat{\Delta}_i'\hat{W}_i^{-1} - \Delta_i'W_i^{-1} = o_p(1),$$

$$(\hat{\Delta}_i - \Delta_i)\hat{W}_i^{-1}(\hat{\delta}_i - \delta_i) = O_p(T^{-1}), \qquad (\hat{\Delta}_i - \Delta_i)\hat{W}_i^{-1}(\hat{\Delta}_i - \Delta_i) = O_p(T^{-1})$$

Then we can simplify the expression of $\hat{\beta}_i^{LV} - \beta_i$ as

$$\hat{\beta}_i^{LV} - \beta_i = (\Delta_i' W_i^{-1} \Delta_i)^{-1} \Delta_i' W_i^{-1} \begin{bmatrix} R_{22\cdot 1}' & 0 \\ 0 & I_K \end{bmatrix} \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} h_t^{\star} \\ v_{it} \end{bmatrix} \epsilon_{it} + o_p(T^{-1/2}) \tag{B.29}$$

By definition of W_i, together with

$$\Delta_i = \begin{bmatrix} \Lambda_{i,22}^* - V\Lambda_{i,12}^* \\ \Sigma_{i,22} \end{bmatrix} = \begin{bmatrix} R_{22\cdot 1}'\Lambda_{i,22} \\ \Sigma_{i,22} \end{bmatrix} = \begin{bmatrix} R_{22\cdot 1}'\gamma_i^h \\ \Sigma_{i,22} \end{bmatrix}$$

we have

$$\begin{aligned} \sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) &= \left(\gamma_i^{h\prime}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})\gamma_i^h + \Omega_i\right)^{-1} \\ &\quad \times \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\gamma_i^{h\prime}(h_t - M_{hg}M_{gg}^{-1}g_t) + v_{it}\right]\epsilon_{it} + o_p(1) \end{aligned}$$

This completes the proof of Theorem 5.1. □