Theory and methods of panel data models with interactive effects

Bai, Jushan and Li, Kunpeng

Columbia University, Tsinghua University

December 2010

THEORY AND METHODS OF PANEL DATA MODELS WITH INTERACTIVE EFFECTS

By Jushan Bai, and Kunpeng Li

Columbia University and Tsinghua University

First version: December, 2010
This version: December, 2012

This paper considers the maximum likelihood estimation of panel data models with interactive effects. Motivated by applications in economics and other social sciences, a notable feature of the model is that the explanatory variables are correlated with the unobserved effects. The usual within-group estimator is inconsistent. Existing methods for consistent estimation are either designed for panel data with short time periods or are less efficient. The maximum likelihood estimator has desirable properties and is easy to implement, as illustrated by the Monte Carlo simulations. This paper develops the inferential theory for the maximum likelihood estimator, including consistency, rate of convergence and the limiting distributions. We further extend the model to include time-invariant regressors and common regressors (cross-section invariant). The regression coefficients for the time-invariant regressors are time-varying, and the coefficients for the common regressors are cross-sectionally varying.

1. Introduction. This paper studies the following panel data models with unobservable interactive effects:

$$y_{it} = \alpha_i + x_{it1}\beta_1 + \cdots + x_{itK}\beta_K + \lambda_i' f_t + e_{it}$$

$$i = 1, ..., N; t = 1, 2, ..., T$$

where y_{it} is the dependent variable; $x_{it} = (x_{it1}, ..., x_{itK})$ is a row vector of explanatory variables; α_i is an intercept; the term $\lambda_i' f_t + e_{it}$ is unobservable and has a factor structure, λ_i is an $r \times 1$ vector of factor loadings, f_t is a vector of factors, and e_{it} is the idiosyncratic error. The interactive effects $(\lambda_i' f_t)$ generalize the usual additive individual and time effects, for example, if $\lambda_i \equiv 1$, then $\alpha_i + \lambda_i' f_t = \alpha_i + f_t$.

AMS 2000 subject classifications: Primary 60F12, 60F30; secondary 60H12

Keywords and phrases: factor error structure, factors, factor loadings, maximum likelihood, principal components, within-group estimator, simultaneous equations

A key feature of the model is that the regressors x_{it} are allowed to be correlated with $(\alpha_i, \lambda_i, f_t)$. This situation is commonly encountered in economics and other social sciences, in which some of the regressors x_{it} are decision variables that are influenced by the unobserved individual heterogeneities. The practical relevance of the model will be further discussed below. The objective of this paper is to obtain consistent and efficient estimation of β in the presence of correlations between the regressors and the factor loadings and factors.

The usual pooled least squares estimator or even the within-group estimator is inconsistent for β. One method to obtain a consistent estimator is to treat $(\alpha_i, \lambda_i, f_t)$ as parameters and estimate them jointly with β. The idea is "controlling through estimating" (controlling the effects by estimating them). This is the approach used by [8], [23] and [31]. While there are some advantages, an undesirable consequence of this approach is the incidental parameters problem. There are too many parameters being estimated, and the incidental parameters bias arises (Neyman and Scott, 1948). [1], [2] and [17] consider the generalized method of moments (GMM) method. The GMM method is based on a nonlinear transformation known as quasi-differencing that eliminates the factor errors. Quasi-differencing increases the nonlinearity of the model especially with more than one factor. The GMM method works well with a small T. When T is large, the number of moment equations will be large and the so called many-moment bias arises. [27] considers an alternative method by augmenting the model with additional regressors $\bar{y}_t$ and $\bar{x}_t$, which are the cross-sectional averages of y_{it} and x_{it}. These averages provide an estimate for f_t. A further approach to controlling the correlation between the regressors and factor errors is to use the Mundlak-Chamberlain projection ([24] and [14]). The latter method projects α_i and λ_i onto the regressors such that $\lambda_i = c_0 + c_1 x_{i1} + \cdots + c_T x_{iT} + \eta_i$, where c_s $(s = 0, 1, ..., T)$ are parameters to be estimated and η_i is the projection residual (a similar projection is done for α_i). The projection residuals are uncorrelated with the regressors so that a variety of approaches can be used to estimate the model. This framework is designed for small T, and is studied by [9].

In this paper we consider the pseudo-Gaussian maximum likelihood method under large N and large T. The theory does not depend on normality. In view of the importance of the MLE in the statistical literature, it is of both practical and theoretical interest to examine the MLE in this context. We develop a rigorous theory for the MLE. We show that there is no incidental parameters bias despite large N and large T.

We allow time-invariant regressors such as education, race and gender in the model. The corresponding regression coefficients are time-dependent.

Similarly, we allow common regressors, which do not vary across individuals, such as prices and policy variables. The corresponding regression coefficients are individual-dependent so that individuals respond differently to policy or price changes. In our view, this is a sensible way to incorporate time-invariant and common regressors. For example, wages associated with education and with gender are more likely to change over time rather than remain constant. In our analysis, time invariant regressors are treated as the components of λ_i that are observable, and common regressors as the components of f_t that are observable. This view fits naturally into the factor framework in which part of the factor loadings and factors are observable, and the maximum likelihood method imposes the corresponding loadings and factors at their observed values.

While the theoretical analysis of MLE is demanding, the limiting distributions of the MLE are simple and have intuitive interpretations. The computation is also easy and can be implemented by adapting the ECM (expectation and constrained maximization) of [22]. In addition, the maximum likelihood method allows restrictions to be imposed on λ_i or on f_t to achieve more efficient estimation. These restrictions can take the form of known values, being either zeros, or other fixed values. Part of the rigorous analysis includes setting up the constrained maximization as a Lagrange multiplier problem. This approach provides insight on which kinds of restrictions are binding and which are not, shedding light on efficiency gain resulting from the restrictions.

Panel data models with interactive effects have wide applicability in economics. In macroeconomics, for example, y_{it} can be the output growth rate for country i in year t; x_{it} represents production inputs, and f_t is a vector of common shocks (technological progress, financial crises); the common shocks have heterogenous impacts across countries through the different factor loadings λ_i; e_{it} represents the country-specific unmeasured growth rates. In microeconomics, and especially in earnings studies, y_{it} is the wage rate for individual i for period t (or for cohort t), x_{it} is a vector of observable characteristics such as marital status and experience; λ_i is a vector of unobservable individual traits such as ability, perseverance, motivation and dedication; the payoff to these individual traits is not constant over time, but time varying through f_t; and e_{it} is idiosyncratic variations in the wage rates. In finance, y_{it} is stock i's return in period t, x_{it} is a vector of observable factors, f_t is a vector of unobservable common factors (systematic risks) and λ_i is the exposure to the risks; e_{it} is the idiosyncratic returns. Factor error structures are also used as a flexible trend modeling as in [20]. Most of panel data analysis assumes cross-sectional independence, e.g., [6], [12], and [18]. The factor structure is also capable of capturing the cross-sectional dependence arising from the common shocks f_t.

Throughout the paper, the norm of a vector or matrix is that of Frobenius, i.e., $\|A\| = [\text{tr}(A'A)]^{1/2}$ for matrix A; diag(A) is a column vector consisting of the diagonal elements of A when A is matrix, but diag(A) represents a diagonal matrix when A is a vector. In addition, we use $\dot{v}_t$ to denote $v_t - \frac{1}{T}\sum_{t=1}^T v_t$ 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, v_t.

2. A common shock model. In the common-shock model, we assume that both y_{it} and x_{it} are impacted by the common shocks f_t so the model takes the form

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \lambda_i' f_t + e_{it} \\
x_{itk} &= \mu_{ik} + \gamma_{ik}' f_t + v_{itk}
\end{aligned} \tag{2.1}
$$

for $k = 1, 2, ..., K$. In across-country output studies, for example, output y_{it} and inputs x_{it} (labor and capital) are both affected by the common shocks.

The parameter of interest is $\beta = (\beta_1, ..., \beta_K)'$. We also estimate $\alpha_i, \lambda_i, \mu_{ik}$ and γ_{ik} $(k = 1, 2...., K)$. By treating the latter as parameters, we also allow arbitrary correlations between (α_i, λ_i) and (μ_{ik}, γ_{ik}). Although we also treat f_t as fixed parameters, there is no need to estimate the individual f_t, but only the sample covariance of f_t. This is an advantage of the maximum likelihood method, which eliminates the incidental parameters problem in the time dimension. This kind of the maximum likelihood method was used for pure factor models in [3], [4], and [10]. By symmetry, we could also estimate individuals f_t, but then we only estimate the sample covariance of the factor loadings. The idea is that we do not simultaneously estimate the factor loadings and the factors f_t (which would be the case for the principal components method). This reduces the number of parameters considerably. If N is much smaller than T $(N \ll T)$, treating factor loadings as parameters is preferable since there are fewer number of parameters.

Because of the correlation between the regressors and regression errors in the y equation, the y and x equations form a simultaneous equation system; the MLE jointly estimates the parameters in both equations. The joint estimation avoids the Mundlak-Chamberlain projection and thus is applicable for large N and large T.

Throughout the paper, we assume the number of factors r is fixed and known. If not, the information criterions developed by [11] can be used to determine it. So λ_i and f_t are $r \times 1$ vectors. Let $x_{it} = (x_{it1}, x_{it2}, \cdots, x_{itK})$, $\gamma_{ix} = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{iK})$, $v_{itx} = (v_{it1}, v_{it2}, \cdots, v_{itK})'$ and $\mu_{ix} = (\mu_{i1}, \mu_{i2}, \cdots, \mu_{iK})'$.

The second equation of (2.1) can be written in matrix form as

$$x'_{it} = \mu_{ix} + \gamma'_{ix} f_t + v_{itx}$$

Further let $\Gamma_i = (\lambda_i, \gamma_{ix})$, $z_{it} = (y_{it}, x_{it})'$, $\varepsilon_{it} = (e_{it}, v'_{itx})'$, $\mu_i = (\alpha_i, \mu'_{ix})'$. Then model (2.1) can be written as

$$\begin{bmatrix} 1 & -\beta' \\ 0 & I_K \end{bmatrix} z_{it} = \mu_i + \Gamma'_i f_t + \varepsilon_{it}$$

Let B denote the coefficient matrix of z_{it} in the preceding equation. Let $z_t = (z'_{1t}, z'_{2t}, \cdots, z'_{Nt})'$, $\Gamma = (\Gamma_1, \Gamma_2, \cdots, \Gamma_N)'$, $\varepsilon_t = (\varepsilon'_{1t}, \varepsilon'_{2t}, \cdots, \varepsilon'_{Nt})'$ and $\mu = (\mu'_1, \mu'_2, \cdots, \mu'_N)'$. Stacking the equations over i, we have

$$(I_N \otimes B) z_t = \mu + \Gamma f_t + \varepsilon_t \tag{2.2}$$

To analyze this model, we impose the following assumptions.

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

Remark 2.1. The non-randomness assumption for f_t is not crucial. In fact, f_t can be a sequence of random variables such that $E(\|f_t\|^4) \le C < \infty$ uniformly in t and f_t is independent of ε_s for all s. The fixed f_t assumption conforms with the usual fixed effects assumption in panel data literature and, in certain sense, is more general than random f_t.

Assumption B: The idiosyncratic error terms $\varepsilon_{it} = (e_{it}, v'_{itx})'$ are assumed such that

- B.1 The e_{it} is independent and identically distributed over t and uncorrelated over i with $E(e_{it}) = 0$ and $E(e_{it}^4) \le \infty$ for all $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Let Σ_{iie} denote the variance of e_{it}.
- B.2 v_{itx} is also independent and identically distributed over t and uncorrelated over i with $E(v_{itx}) = 0$ and $E(\|v_{itx}\|^4) \le \infty$ for all $i = 1, \cdots, N$ and $t = 1, \cdots, T$. We use Σ_{iix} to denote the variance matrix of v_{itx}.
- B.3 e_{it} is independent of v_{jsx} for all (i, j, t, s). Let Σ_{ii} denote the variance matrix ε_{it}. So we have $\Sigma_{ii} = \text{diag}(\Sigma_{iie}, \Sigma_{iix})$, a block-diagonal matrix.

Remark 2.2. Let $\Sigma_{\varepsilon\varepsilon}$ denote the variance of $\varepsilon_t = (\varepsilon'_{1t}, \cdots, \varepsilon'_{Nt})'$. Due to the uncorrelatedness of ε_{it} over i, we have $\Sigma_{\varepsilon\varepsilon} = \text{diag}(\Sigma_{11}, \Sigma_{22}, \cdots, \Sigma_{NN})$, a block-diagonal matrix. Assumption B is more general than the usual assumption in the factor analysis. In a traditional factor model, the variance of the idiosyncratic error terms are assumed to be a diagonal matrix. In the present setting, the variance of ε_t is a block-diagonal matrix . Even without explanatory variables, this generalization is of interest. The factor analysis literature has a long history to explore the block-diagonal idiosyncratic variance, known as multiple battery factor analysis, see [32]. The maximum likelihood estimation theory for high dimensional factor models with block diagonal covariance matrix has not been previously studied. The asymptotic theory developed in this paper not only provides a way of analyzing the coefficient β, but also a way of analyzing the factors and loadings in the multiple battery factor models. This framework is of independent interest.

Assumption B allows cross-sectional heteroskedasticity. The maximum likelihood method will simultaneously estimate the heteroskedastic variances and other parameters. This assumption assumes the independence and homoskedasticity of the error terms over time and uncorrelatedneess over the cross section. Extension to more general heteroscedasticity and correlation patterns can be considered by our method. The model with more general error covariance structure, known as approximate factor models in the sense of [15], has been extensively investigated by the recent literature, such as [11], [7], [30] among others. This literature largely focuses on the principal components method and for pure factor models without explanatory variables. The analysis of the maximum likelihood method for our model is already challenging, the extension to approximate factor models is not considered in this paper.

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

C.1 $\|\Gamma_j\| \leq C$ for all $j = 1, \cdots, N$.
C.2 $C^{-1} \leq \tau_{min}(\Sigma_{jj}) \leq \tau_{max}(\Sigma_{jj}) \leq C$ for all $j = 1, \cdots, N$, where $\tau_{min}(\Sigma_{jj})$ and $\tau_{max}(\Sigma_{jj})$ denote the smallest and largest eigenvalues of the matrix Σ_{jj}, respectively.
C.3 there exists an $r \times r$ positive matrix Q such that $Q = \lim_{N\to\infty} N^{-1}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma$, where Γ is defined earlier.

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

Remark 2.3. Assumption D requires that part of the estimators be estimated in a compact set. This assumption is usually made for theoretical analysis, especially when dealing with nonlinear objective functions, e.g., [19], [25], and [33]. The objective function considered in this paper exhibits high nonlinearity.

2.2. *Identification restrictions.* It is a well-known result in factor analysis that the factors and loadings can only be identified up to a rotation. The models considered in this paper can be viewed as extensions of the factor models. As such they inherit the same identification problem. We show that identification conditions can be imposed on the factors and loadings without loss of generality. To see this, model (2.2) can be rewritten as

$$\begin{aligned}(I_N \otimes B)z_t &= \mu + \Gamma f_t + \varepsilon_t \\ &= (\mu + \Gamma \bar{f}) + \Gamma(f_t - \bar{f}) + \varepsilon_t \\ &= (\mu + \Gamma \bar{f}) + \Big(\Gamma M_{ff}^{1/2} R\Big)\Big(R' M_{ff}^{-1/2}(f_t - \bar{f})\Big) + \varepsilon_t,\end{aligned} \tag{2.3}$$

where R is an orthogonal matrix, which we choose to be the matrix consisting of the eigenvectors of $M_{ff}^{1/2}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma M_{ff}^{1/2}$ associated with the eigenvalues arranged in descending order. Treating $\mu + \Gamma\bar{f}$ as the new $\mu^\star$, $\Gamma M_{ff}^{1/2}R$ as the new $\Gamma^\star$ and $R' M_{ff}^{-1/2}(f_t - \bar{f})$ as the new $f_t^\star$, we have

$$(I_N \otimes B)z_t = \mu^\star + \Gamma^\star f_t^\star + \varepsilon_t$$

with $\frac{1}{T}\sum_{t=1}^T f_t^\star = 0$, $\frac{1}{T}\sum_{t=1}^T f_t^\star f_t^{\star\prime} = I_r$ and $\frac{1}{N}\Gamma^{\star\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^\star$ being a diagonal matrix. Given the above analysis, we can impose in (2.2) the following restrictions, which we refer to as IB (*Identification* restrictions for *Basic* models).

IB1. $M_{ff} = I_r$

IB2. $\frac{1}{N}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma = D$, where D is a diagonal matrix with its diagonal elements distinct and arranged in descending order.

IB3. $\bar{f} = \frac{1}{T}\sum_{t=1}^T f_t = 0$.

REMARK 2.4. The requirement that the diagonal elements of D are distinct in IB2 is not needed for the ML estimation of β, but it is needed for the identification of factors and factor loadings. Under this requirement, the orthogonal matrix R in (2.3) can be uniquely determined up to a column sign change. This assumption does simplify the analysis for the MLE of β.

2.3. *Estimation.* The objective function considered in this section is

$$\ln L = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}\Big[(I_N \otimes B)M_{zz}(I_N \otimes B')\Sigma_{zz}^{-1}\Big], \tag{2.4}$$

where $\Sigma_{zz} = \Gamma M_{ff}\Gamma' + \Sigma_{\varepsilon\varepsilon}$ and $M_{zz} = \frac{1}{T}\sum_{t=1}^{T} \dot{z}_t\dot{z}_t'$. Here Σ_{zz} is the matrix consisting of the parameters other than β, the latter is contained in B; M_{zz} is the data matrix. The objective function (2.4) can be regarded as the likelihood function (omitting a constant). Note that the determinant of $I_N \otimes B$ is 1, so the Jacobian term does not depend on B. If ε_t and f_t are independent and normally distributed, the likelihood function for the observed data has the form of (2.4). Here recall that f_t are fixed constants and ε_t are not necessarily normal, (2.4) is a pseudo-likelihood function.

For further analysis, we partition the matrix Σ_{zz} and M_{zz} as

$$\Sigma_{zz} = \begin{pmatrix} \Sigma_{zz}^{11} & \Sigma_{zz}^{12} & \cdots & \Sigma_{zz}^{1N} \\ \Sigma_{zz}^{21} & \Sigma_{zz}^{22} & \cdots & \Sigma_{zz}^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{zz}^{N1} & \Sigma_{zz}^{N2} & \cdots & \Sigma_{zz}^{NN} \end{pmatrix} \quad M_{zz} = \begin{pmatrix} M_{zz}^{11} & M_{zz}^{12} & \cdots & M_{zz}^{1N} \\ M_{zz}^{21} & M_{zz}^{22} & \cdots & M_{zz}^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ M_{zz}^{N1} & M_{zz}^{N2} & \cdots & M_{zz}^{NN} \end{pmatrix}$$

where for any (i, j), Σ_{zz}^{ij} and M_{zz}^{ij} are both $(K+1)\times(K+1)$ matrices.

Let $\hat{\beta}, \hat{\Gamma}$ and $\hat{\Sigma}_{\varepsilon\varepsilon}$ denote the MLE. The first order condition for β satisfies

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\Big\{(\dot{y}_{it} - \dot{x}_{it}\hat{\beta}) - \hat{\lambda}_i'\hat{G}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\begin{bmatrix} \dot{y}_{jt} - \dot{x}_{jt}\hat{\beta} \\ \dot{x}_{jt}' \end{bmatrix}\Big\}\dot{x}_{it} = 0 \tag{2.5}$$

where $\hat{G} = (\hat{M}_{ff}^{-1} + \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}$. The first order condition for Γ_j satisfies

$$\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Big(\hat{B}M_{zz}^{ij}\hat{B}' - \hat{\Sigma}_{zz}^{ij}\Big) = 0. \tag{2.6}$$

Post-multiplying $\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'$ on both sides of (2.6) and then taking summation over j, we have

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Big(\hat{B}M_{zz}^{ij}\hat{B}' - \hat{\Sigma}_{zz}^{ij}\Big)\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j' = 0. \tag{2.7}$$

The first order condition for Σ_{ii} satisfies

$$\hat{B}M_{zz}^{ii}\hat{B}' - \hat{\Sigma}_{zz}^{ii} = \mathbb{W}, \tag{2.8}$$

where $\mathbb{W}$ is a $(K+1)\times(K+1)$ matrix such that its upper-left 1×1 and lower-right $K\times K$ submatrices are both zero, but the remaining elements are undetermined. The undetermined elements correspond to the zero elements of Σ_{ii}. These first order conditions are needed for the asymptotic representation of the MLE.

2.4. *Asymptotic properties of the MLE.* As N tends to infinity, the number of parameters goes to infinity, which makes consistency proof more difficult. Following [10], we establish the following average consistency results which serve as the basis for subsequent analysis.

PROPOSITION 2.1 (Consistency). *Let $\hat{\theta} = (\hat{\beta}, \hat{\Gamma}, \hat{\Sigma}_{\varepsilon\varepsilon})$ be the solution by maximizing (2.4). Under Assumptions A-D and the identification conditions IB, when $N, T \to \infty$, we have*

$$\hat{\beta} - \beta \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Gamma}_i - \Gamma_i)\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i)' \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \xrightarrow{p} 0$$

The derivation of Proposition 2.1 requires considerable work. The results of $\hat{\beta} - \beta \xrightarrow{p} 0$ and $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \xrightarrow{p} 0$ can be directly derived by working with the objective function because they are free of rotational problems. To prove $\frac{1}{N}\sum_{i=1}^{N}(\hat{\Gamma}_i - \Gamma_i)\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i)' \xrightarrow{p} 0$, we have to invoke the identification conditions. In addition, the identification condition used in this section has so-called sign problem. So the estimator $\hat{\Gamma}$ having the same signs as those of Γ is assumed.

In order to derive the inferential theory, we need to strengthen Proposition 2.1. This result is stated in the following theorem.

THEOREM 2.1 (Convergence rate). *Under the assumptions of Proposition 2.1, we have*

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

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Gamma}_i - \Gamma_i\|^2 = O_p(T^{-1})$$

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

[8] considers an iterated principal components estimator for model (2.1). His derivation shows that, in the presence of heteroscedasticities over the cross section, the PC estimator for β has a bias of order $O_p(N^{-1})$. As a

comparison, Theorem 2.1 shows that the MLE is robust to the heteroscedasticities over the cross section. So if N is fixed, the estimator in [8] is inconsistent unless there is no heteroskedasticity, but the estimator here is still consistent.

Although Γ and $\Sigma_{\varepsilon\varepsilon}$ are not the parameters of interest and their asymptotic properties are not presented in this paper, Theorem 2.1 has implications for the limiting distributions of these parameters. Given that $\hat{\beta} - \beta$ has a faster convergence rate, the limiting distributions of $\text{vech}(\hat{\Gamma}_i - \Gamma_i)$ and $\text{vech}(\hat{\Sigma}_{ii} - \Sigma_{ii})$ are not affected by the estimation of β, and are the same as the case of without regressors. If we use $\hat{f}_t = (\sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \hat{\Gamma}_i')^{-1}(\sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \hat{B} z_{it})$ to estimate f_t, then the limiting distribution of $\hat{f}_t - f_t$ is also the same as in pure factor models. The asymptotic representations on these estimators are implicitly contained in the appendix.

Now we present the most important result in this section. Throughout let $\mathcal{M}(\mathbb{X})$ denote the project matrix onto the space orthogonal to $\mathbb{X}$, i.e. $\mathcal{M}(\mathbb{X}) = I - \mathbb{X}(\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$.

THEOREM 2.2 (Asymptotic representation). *Under the assumptions of Proposition 2.1, we have*

$$\hat{\beta} - \beta = \Omega^{-1} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \Sigma_{iie}^{-1} e_{it} v_{itx} + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})$$

where Ω is a $K \times K$ matrix, whose (p,q) element $\Omega_{pq} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}$ with $\Sigma_{iix}^{(p,q)}$ being the (p,q) element of matrix Σ_{iix}.

REMARK 2.5. In appendix A.3, we show that the asymptotic expression of $\hat{\beta} - \beta$ can be alternatively expressed as

$$\hat{\beta} - \beta = \begin{pmatrix} \text{tr}[\ddot{M} X_1 \mathcal{M}(\overline{\mathbb{F}}) X_1'] & \cdots & \text{tr}[\ddot{M} X_1 \mathcal{M}(\overline{\mathbb{F}}) X_K'] \\ \vdots & \vdots & \vdots \\ \text{tr}[\ddot{M} X_K \mathcal{M}(\overline{\mathbb{F}}) X_1'] & \cdots & \text{tr}[\ddot{M} X_K \mathcal{M}(\overline{\mathbb{F}}) X_K'] \end{pmatrix}^{-1} \tag{2.9}$$

$$\times \begin{pmatrix} \text{tr}[\ddot{M} X_1 \mathcal{M}(\overline{\mathbb{F}}) e'] \\ \vdots \\ \text{tr}[\ddot{M} X_K \mathcal{M}(\overline{\mathbb{F}}) e'] \end{pmatrix} + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})$$

where $X_k = (x_{itk})$ is $N \times T$ (the data matrix for the kth regressor, $k = 1, 2, \ldots, K$); $e = (e_{it})$ is $N \times T$; $\ddot{M} = \Sigma_{ee}^{-1/2} \mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda) \Sigma_{ee}^{-1/2}$ with $\Sigma_{ee} = \text{diag}\{\Sigma_{11e}, \Sigma_{22e}, \cdots, \Sigma_{NNe}\}$ and $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$; $\mathbb{F} = (f_1, f_2, \cdots, f_T)'$; $\overline{\mathbb{F}} = (1_T, \mathbb{F})$ where 1_T is a $T \times 1$ vector with all 1's.

REMARK 2.6. Theorem 2.2 shows that the asymptotic expression of $\hat{\beta}-\beta$ only involves variations in e_{it} and v_{itx}. Intuitively, this is due to the fact that the error terms of the y equation share the same factors with the explanatory variables. The variations from the common factor part of x_{itk} (i.e., $\gamma_{ik}' f_t$) do not provide information for β since this part of information is offset by the common factor part of the error terms (i.e., $\lambda_i' f_t$) in the y equation.

COROLLARY 2.1 (Limiting distribution). *Under the assumptions of Theorem 2.2, if* $\sqrt{N}/T \to 0$, *we have*

$$\sqrt{NT}(\hat{\beta}-\beta) \xrightarrow{d} N(0, \overline{\Omega}^{-1})$$

where $\overline{\Omega} = \lim_{N,T\to\infty} \Omega$, *and* $\overline{\Omega}$ *is also the limit of*

$$\overline{\Omega} = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \operatorname{tr}[\ddot{M} X_1 \mathcal{M}(\overline{\mathbb{F}}) X_1'] & \cdots & \operatorname{tr}[\ddot{M} X_1 \mathcal{M}(\overline{\mathbb{F}}) X_K'] \\ \vdots & \vdots & \vdots \\ \operatorname{tr}[\ddot{M} X_K \mathcal{M}(\overline{\mathbb{F}}) X_1'] & \cdots & \operatorname{tr}[\ddot{M} X_K \mathcal{M}(\overline{\mathbb{F}}) X_K'] \end{pmatrix}$$

REMARK 2.7. The covariance matrix $\overline{\Omega}$ can be consistently estimated by

$$\frac{1}{NT} \begin{pmatrix} \operatorname{tr}[\widehat{\ddot{M}} X_1 \mathcal{M}(\widehat{\overline{\mathbb{F}}}) X_1'] & \cdots & \operatorname{tr}[\widehat{\ddot{M}} X_1 \mathcal{M}(\widehat{\overline{\mathbb{F}}}) X_K'] \\ \vdots & \vdots & \vdots \\ \operatorname{tr}[\widehat{\ddot{M}} X_K \mathcal{M}(\widehat{\overline{\mathbb{F}}}) X_1'] & \cdots & \operatorname{tr}[\widehat{\ddot{M}} X_K \mathcal{M}(\widehat{\overline{\mathbb{F}}}) X_K'] \end{pmatrix},$$

where X_k is the $N \times T$ data matrix for the kth regressor,

$$\widehat{\ddot{M}} = \hat{\Sigma}_{ee}^{-1} - \hat{\Sigma}_{ee}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}; \tag{2.10}$$

$\widehat{\overline{\mathbb{F}}} = (1_T, \hat{\mathbb{F}})$ with $\hat{\mathbb{F}} = (\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_T)'$ and

$$\hat{f}_t = (\sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \hat{\Gamma}_i')^{-1} (\sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \hat{B} z_{it}). \tag{2.11}$$

Here $\hat{\Gamma}, \hat{\Lambda}, \hat{\Sigma}_{ii}, \hat{\Sigma}_{ee}$ and $\hat{B}$ are the maximum likelihood estimators.

REMARK 2.8. We point out that the condition $\sqrt{N}/T \to 0$ is only needed for the limiting distribution to be of this simple form. The MLE for β is still consistent under fixed N, but the limiting distribution will be different.

3. Common shock models with zero restrictions. The basic model in section 2 assumes that the explanatory variables x_{it} share the same factors with y_{it}. This section relaxes this assumption. We assume that the regressors are impacted by additional factors that do not affect the y equation. An alternative view is that some factor loadings in the y equation are restricted to be zero. Consider the following model

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \psi_i' g_t + e_{it} \\
x_{itk} &= \mu_{ik} + \gamma_{ik}^{g\prime} g_t + \gamma_{ik}^{h\prime} h_t + v_{itk}
\end{aligned}
\tag{3.1}
$$

for $k = 1, 2, \cdots, K$, where g_t is an $r_1 \times 1$ vector representing the shocks affecting both y_{it} and x_{it}, and h_t is an $r_2 \times 1$ vector representing the shocks affecting x_{it} only. Let $\lambda_i = (\psi_i', 0_{r_2\times 1}')'$, $\gamma_{ik} = (\gamma_{ik}^{g\prime}, \gamma_{ik}^{h\prime})'$ and $f_t = (g_t', h_t')'$, the above model can be written as

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \lambda_i' f_t + e_{it} \\
x_{itk} &= \mu_{ik} + \gamma_{ik}' f_t + v_{itk}
\end{aligned}
$$

which is the same as model (2.1) except that λ_i now has r_1 free parameters and the remaining ones are restricted to be zeros. For further analysis, we introduce some notations. We define

$$
\begin{aligned}
\Gamma_i^g &= (\psi_i, \gamma_{i1}^g, \ldots, \gamma_{iK}^g), & \Gamma_i^h &= (0_{r_2\times 1}, \gamma_{i1}^h, \ldots, \gamma_{iK}^h), \\
\Gamma^g &= (\Gamma_1^g, \Gamma_2^g, \ldots, \Gamma_N^g)', & \Gamma^h &= (\Gamma_1^h, \Gamma_2^h, \ldots, \Gamma_N^h)'.
\end{aligned}
$$

We also define $\mathbb{G}$ and $\mathbb{H}$ similarly as $\mathbb{F}$, i.e., $\mathbb{G} = (g_1, g_2, \ldots, g_T)'$, $\mathbb{H} = (h_1, h_2, \ldots, h_T)'$. This implies that $\mathbb{F} = (\mathbb{G}, \mathbb{H})$. The presence of zero restrictions in (3.1) requires different identification conditions from the previous model.

3.1. *Identification conditions.* Zero loading restrictions alleviate rotational indeterminacy. Instead of $r^2 = (r_1 + r_2)^2$ restrictions, we only need to impose $r_1^2 + r_1 r_2 + r_2^2$ restrictions. These restrictions are referred to as IZ restrictions (*Identification* conditions with *Zero* restrictions). They are

- IZ1 $M_{ff} = I_r$
- IZ2 $\frac{1}{N}\Gamma^{g\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^g = D_1$ and $\frac{1}{N}\Gamma^{h\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^h = D_2$, where D_1 and D_2 are both diagonal matrices with distinct diagonal elements in descending order.
- IZ3 $1_T'\mathbb{G} = 0$ and $1_T'\mathbb{H} = 0$.

In addition, we need an additional assumption for our analysis.

Assumption E: $\Psi = (\psi_1', \psi_2', \ldots, \psi_N')'$ is of full column rank.

Identification conditions IZ are less stringent than IB of the previous section. Assumption E says that the factors g_t are pervasive for the y equation. We next explain why $r_1^2 + r_1 r_2 + r_2^2$ restrictions are sufficient. Let R be an $r \times r$ invertible matrix, which we partition into

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

where R_{11} is $r_1 \times r_1$ and R_2 is $r_2 \times r_2$. The indeterminacy arises since equation (2.2) can be written as

$$(I_N \otimes B) z_t = \mu + \Gamma f_t + e_t = \mu + (\Gamma R)(R^{-1} f_t) + \varepsilon_t$$

If we treat ΓR as a new Γ and $R^{-1} f_t$ as a new f_t, we have observationally equivalent models. However, in the present context there are many zero restrictions in Γ. If ΓR is a qualified loading matrix, the same zero restrictions should be satisfied for ΓR. This leads to $\Psi R_{12} = 0$. If Ψ is of full column rank, then left-multiplying $(\Psi'\Psi)^{-1}\Psi'$ gives $R_{12} = 0$. This implies that we need $r_1^2 + r_1 r_2 + r_2^2$ restrictions for full identification since R_{11}, R_{21} and R_{22} have $r_1^2 + r_1 r_2 + r_2^2$ free parameters. As a comparison, if there are no restrictions in Γ, we need $r^2 = (r_1 + r_2)^2$ restrictions. Thus, zero loadings partially remove rotational indeterminacy. Notice IZ1 has $\frac{1}{2} r(r+1)$ restrictions and IZ2 has $\frac{1}{2} r_1(r_1 - 1) + \frac{1}{2} r_2(r_2 - 1)$ restrictions. The total number of restrictions is thus $\frac{1}{2} r(r+1) + \frac{1}{2} r_1(r_1 - 1) + \frac{1}{2} r_2(r_2 - 1) = r_1^2 + r_2^2 + r_1 r_2$, the exact number we need.

3.2. *Estimation.* The likelihood function is now maximized under three sets of restrictions, i.e. $\frac{1}{N}\Gamma^{g\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^g = D_1$, $\frac{1}{N}\Gamma^{h\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^h = D_2$ and $\Phi = 0$ where Φ denotes the zero factor loading matrix in the y equation. The likelihood function with the Lagrange multipliers is

$$\ln L = -\frac{1}{2N} \ln |\Sigma_{zz}| - \frac{1}{2N} \mathrm{tr}\Big[(I_N \otimes B) M_{zz} (I_N \otimes B') \Sigma_{zz}^{-1}\Big]$$

$$+ \mathrm{tr}\Big[\Upsilon_1 \big(\frac{1}{N}\Gamma^{g\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^g - D_1\big)\Big] + \mathrm{tr}\Big[\Upsilon_2 \big(\frac{1}{N}\Gamma^{h\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^h - D_2\big)\Big] + \mathrm{tr}[\Upsilon_3'\Phi],$$

where $\Sigma_{zz} = \Gamma\Gamma' + \Sigma_{\varepsilon\varepsilon}$; Υ_1 is $r_1 \times r_1$ and Υ_2 is $r_2 \times r_2$, both are symmetric Lagrange multipliers matrices with zero diagonal elements; Υ_3 is a Lagrange multiplier matrix of dimension $r_2 \times N$.

Let $\mathbb{U} = \hat{\Sigma}_{zz}^{-1}[(I_N \otimes \hat{B}) M_{zz} (I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1}$. Notice $\mathbb{U}$ is a symmetric matrix. The first order condition on $\hat{\Gamma}^g$ gives

$$\frac{1}{N}\hat{\Gamma}^{g\prime}\mathbb{U} + \Upsilon_1 \frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} = 0.$$

Post-multiplying $\hat{\Gamma}^g$ yields

$$\frac{1}{N}\hat{\Gamma}^{g\prime}\mathbb{U}\hat{\Gamma}^g + \Upsilon_1\frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g = 0.$$

Since $\frac{1}{N}\hat{\Gamma}^{g\prime}\mathbb{U}\hat{\Gamma}^g$ is a symmetric matrix, the above equation implies that $\Upsilon_1\frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$ is also symmetric. But $\frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$ is a diagonal matrix. So the (i,j)th element of $\Upsilon_1\frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$ is $\Upsilon_{1,ij}d_{1j}$, where $\Upsilon_{1,ij}$ is the (i,j)th element of Υ_1 and d_{1j} is the jth diagonal element of $\hat{D}_1$. Given $\Upsilon_1\frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$ is symmetric, we have $\Upsilon_{1,ij}d_{1j} = \Upsilon_{1,ji}d_{1i}$ for all $i \neq j$. However, Υ_1 is also symmetric, so $\Upsilon_{1,ij} = \Upsilon_{1,ji}$. This gives $\Upsilon_{1,ij}(d_{1j} - d_{1i}) = 0$. Since $d_{1j} \neq d_{1i}$ by $IZ2$, we have $\Upsilon_{1,ij} = 0$ for all $i \neq j$. This implies $\Upsilon_1 = 0$ since the diagonal elements of Υ_1 are all zeros.

Let $\Gamma_x^h = (\gamma_{1x}^h, \gamma_{2x}^h, \cdots, \gamma_{Nx}^h)'$ with $\gamma_{ix}^h = (\gamma_{i1}^h, \gamma_{i2}^h, \cdots, \gamma_{iK}^h)$, and $\Sigma_{xx} = \operatorname{diag}\{\Sigma_{11x}, \Sigma_{22x}, \cdots, \Sigma_{NNx}\}$, a block diagonal matrix of $NK \times NK$ dimension. We partition the matrix $\mathbb{U}$ and define the matrix $\overline{\mathbb{U}}$ as

$$\mathbb{U} = \begin{pmatrix} \mathbb{U}_{11} & \mathbb{U}_{12} & \cdots & \mathbb{U}_{1N} \\ \mathbb{U}_{21} & \mathbb{U}_{22} & \cdots & \mathbb{U}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{U}_{N1} & \mathbb{U}_{N2} & \cdots & \mathbb{U}_{NN} \end{pmatrix}, \quad \overline{\mathbb{U}} = \begin{pmatrix} \overline{\mathbb{U}}_{11} & \overline{\mathbb{U}}_{12} & \cdots & \overline{\mathbb{U}}_{1N} \\ \overline{\mathbb{U}}_{21} & \overline{\mathbb{U}}_{22} & \cdots & \overline{\mathbb{U}}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{\mathbb{U}}_{N1} & \overline{\mathbb{U}}_{N2} & \cdots & \overline{\mathbb{U}}_{NN} \end{pmatrix}$$

where $\mathbb{U}_{ij}$ is a $(K+1)\times(K+1)$ matrix and $\overline{\mathbb{U}}_{ij}$ is the lower-right $K\times K$ block of $\mathbb{U}_{ij}$. Notice $\overline{\mathbb{U}}$ is also a symmetric matrix. Then the first order condition on Γ_x^h gives

$$\frac{1}{N}\hat{\Gamma}_x^{h\prime}\overline{\mathbb{U}} + \Upsilon_2\frac{1}{N}\hat{\Gamma}_x^{h\prime}\hat{\Sigma}_{xx}^{-1} = 0.$$

Post-multiplying $\hat{\Gamma}_x^h$ yields

$$\frac{1}{N}\hat{\Gamma}_x^{h\prime}\overline{\mathbb{U}}\hat{\Gamma}_x^h + \Upsilon_2\frac{1}{N}\hat{\Gamma}_x^{h\prime}\hat{\Sigma}_{xx}^{-1}\hat{\Gamma}_x^h = 0.$$

Notice $\frac{1}{N}\hat{\Gamma}_x^{h\prime}\hat{\Sigma}_{xx}^{-1}\hat{\Gamma}_x^h = \frac{1}{N}\hat{\Gamma}^{h\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^h = \hat{D}_2$. By the similar arguments in deriving $\Upsilon_1 = 0$, we have $\Upsilon_2 = 0$. The interpretation for the zero Lagrange multipliers is that these constraints are non-binding for the likelihood. Whether or not these restrictions are imposed, the optimal value of the likelihood function is not affected, and neither is the efficiency of $\hat{\beta}$. In contrast, we cannot show Υ_3 to be zero. Thus if $\Phi = 0$ is not imposed, the optimal value of the likelihood function and the efficiency of $\hat{\beta}$ will be affected. In Section 2, we did not use the Lagrange multiplier approach to impose the identification restrictions. Had it been used, we would have obtained zero valued

Lagrange multipliers. This is another view of why these restrictions do not affect the limiting distribution of $\hat{\beta}$. See Remark 2.4.

Now the likelihood function is simplified as

$$\ln L = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}\Big[(I_N \otimes B)M_{zz}(I_N \otimes B')\Sigma_{zz}^{-1}\Big] + \mathrm{tr}[\Upsilon_3'\Phi]. \tag{3.2}$$

The first order condition on Γ is

$$\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1} = W', \tag{3.3}$$

where W is a matrix having the same dimension as Γ, whose element is zero if the counterpart of Γ is not specified to be zero, otherwise undetermined (containing the Lagrange multipliers). Post-multiplying $\hat{\Gamma}$ gives

$$\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1}\hat{\Gamma} = W'\hat{\Gamma}.$$

By the special structure of W and $\hat{\Gamma}$, it is easy to verify that $W'\hat{\Gamma}$ has the form

$$\begin{bmatrix} 0_{r_1\times r_1} & 0_{r_1\times r_2} \\ \times & 0_{r_2\times r_2} \end{bmatrix}.$$

However, the left hand side of the preceding equation is a symmetric matrix, so is the right side. It follows that the subblock "$\times$" is zero, i.e. $W'\hat{\Gamma} = 0$. Thus, $\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1}\hat{\Gamma} = 0$. (This equation would be the first order condition for M_{ff} if it were unknown.) This equality can be simplified as

$$\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = 0, \tag{3.4}$$

because $\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1} = \hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$ with $\hat{G} = (I + \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}$. Next, we partition the matrix $\hat{G} = (I + \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}$ and $\hat{H} = (\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}$ as follows

$$\hat{G} = \begin{bmatrix} \hat{G}_1 \\ \hat{G}_2 \end{bmatrix} = \begin{bmatrix} \hat{G}_{11} & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} \end{bmatrix}, \quad \hat{H} = \begin{bmatrix} \hat{H}_1 \\ \hat{H}_2 \end{bmatrix} = \begin{bmatrix} \hat{H}_{11} & \hat{H}_{12} \\ \hat{H}_{21} & \hat{H}_{22} \end{bmatrix},$$

where $\hat{G}_{11}, \hat{H}_{11}$ are $r_1 \times r_1$, while $\hat{G}_{22}, \hat{H}_{22}$ are $r_2 \times r_2$.

Notice $\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$ and $\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1} = \hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$. Substitute these results into (3.3) and use (3.4), the first order condition for ψ_i can be simplified as

$$\hat{G}_1 \sum_{i=1}^{N} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{B}M_{zz}^{ij}\hat{B}' - \hat{\Sigma}_{zz}^{ij})\hat{\Sigma}_{jj}^{-1} I_{K+1}^1 = 0, \tag{3.5}$$

where I_{K+1}^1 is the first column of the identity matrix of dimension $K+1$.

Similarly, the first order condition for $\gamma_{jx} = (\gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jK})$ is

$$\sum_{i=1}^{N} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} (\hat{B} M_{zz}^{ij} \hat{B}' - \hat{\Sigma}_{zz}^{ij}) \hat{\Sigma}_{jj}^{-1} I_{K+1}^{-} = 0, \tag{3.6}$$

where I_{K+1}^{-} is a $(K+1) \times K$ matrix, obtained by deleting the first column of the identity matrix of dimension $K+1$.

The first order condition for Σ_{jj} is

$$\begin{aligned} \hat{B} M_{zz}^{jj} \hat{B}' - \hat{\Sigma}_{zz}^{jj} - \hat{\Gamma}_j' \hat{G} \sum_{i=1}^{N} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \Big(\hat{B} M_{zz}^{ij} \hat{B}' - \hat{\Sigma}_{zz}^{ij} \Big) & \\ - \sum_{i=1}^{N} \Big(\hat{B} M_{zz}^{ji} \hat{B}' - \hat{\Sigma}_{zz}^{ji} \Big) \hat{\Sigma}_{ii}^{-1} \hat{\Gamma}_i' \hat{G} \hat{\Gamma}_j &= \mathbb{W}, \end{aligned} \tag{3.7}$$

where $\mathbb{W}$ is defined following (2.8).

The first order condition for β is

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\Sigma}_{iie}^{-1} \Big\{ (\dot{y}_{it} - \dot{x}_{it} \hat{\beta}) - \hat{\lambda}_i' \hat{G} \sum_{j=1}^{N} \hat{\Gamma}_j \hat{\Sigma}_{jj}^{-1} \begin{bmatrix} \dot{y}_{jt} - \dot{x}_{jt} \hat{\beta} \\ \dot{x}_{jt}' \end{bmatrix} \Big\} \dot{x}_{it} = 0, \tag{3.8}$$

which is the same as in Section 2.

We need an additional identity to study the properties of the MLE. Recall that, by the special structures of W and $\hat{\Gamma}$, the three submatrices of $W'\hat{\Gamma}$ can be directly derived to be zeros. The remaining submatrix is also zero, as shown earlier. However, this submatrix being zero yields the following equation (the detailed derivation is delivered in Appendix B)

$$\frac{1}{N} \hat{G}_2 \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} (\hat{B} M_{zz}^{ij} \hat{B}' - \hat{\Sigma}_{zz}^{ij}) \hat{\Sigma}_{jj}^{-1} I_{K+1}^{1} \hat{\psi}_j' = 0. \tag{3.9}$$

These identities for the MLE are used to derive the asymptotic representations.

3.3. *Asymptotic properties of the MLE.* The results on consistency and the rate of convergence are similar to those in the previous section, which are presented in Appendixes B.1 and B.2. For simplicity, we only state the asymptotic representation for the MLE here.

PROPOSITION 3.1 (Asymptotic representation). *Under Assumptions A-E and the identification restriction IZ, we have*

$$
\begin{aligned}
\mathcal{P}^0(\hat{\beta}-\beta) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ix}^{h\prime}h_t e_{it} \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i'\Pi_{\psi\psi}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\gamma_{jx}^{h\prime}\Big)h_t e_{it} \\
&\quad + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}),
\end{aligned}
$$

where $\mathcal{P}^0$ *is a* $K\times K$ *symmetric matrix with its* (p,q) *element equal to* $\mathrm{tr}(\Gamma_p^{h\prime}\ddot{M}\Gamma_q^h)+\frac{1}{N}\sum_{i=1}^N\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}$; $\Gamma_p^h=[\gamma_{1p}^h,\gamma_{2p}^h,\dots,\gamma_{Np}^h]'$; $\gamma_{jx}^h=[\gamma_{j1}^h,\dots,\gamma_{jK}^h]$; $\Pi_{\psi\psi}=\frac{1}{N}\sum_{i=1}^N\psi_i\Sigma_{iie}^{-1}\psi_i'$ *and* $\ddot{M}=\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}$.

Proposition 3.1 is derived under the identification conditions IZ. In Appendix B.3 of the supplement, we show that for any set of factors and factor loadings $(\psi_i,\gamma_{ik},g_t,h_t)$, it can always be transformed into a new set $(\psi_i^\star,\gamma_{ik}^\star,g_t^\star,h_t^\star)$, which satisfies IZ, and at the same time, leaving $\Phi=0$ intact. Given the asymptotic representation in Proposition 3.1, together with the relationship between the two sets, we have the following theorem, which doesn't depend on IZ.

THEOREM 3.1. *Under Assumptions A-E, we have*

$$
\begin{aligned}
\mathcal{P}(\hat{\beta}-\beta) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ix}^{h\prime}h_t^\star e_{it} \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i'\Pi_{\psi\psi}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\gamma_{jx}^{h\prime}\Big)h_t^\star e_{it} \\
&\quad + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})
\end{aligned}
$$

where

$$h_t^\star = \dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t;$$

$\mathcal{P}$ *is a* $K\times K$ *symmetric matrix with its* (p,q) *element equal to*

$$\frac{1}{NT}tr\Big[\ddot{M}\Gamma_q^h\mathbb{H}'\mathcal{M}(\overline{\mathbb{G}})\mathbb{H}\Gamma_p^{h\prime}\Big] + \frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}$$

where $\overline{\mathbb{G}}=(1_T,\mathbb{G})$; $\Pi_{\psi\psi}=\frac{1}{N}\sum_{i=1}^N\psi_i\Sigma_{iie}^{-1}\psi_i'$; $\ddot{M}=\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}$, $\Gamma_p^h=(\gamma_{1p}^h,\gamma_{2p}^h,\dots,\gamma_{Np}^h)'$.

Remark 3.1. In Appendix B.3, we also show that the asymptotic expression of $\hat{\beta}-\beta$ in Theorem 3.1 can be expressed alternatively as

$$\hat{\beta}-\beta=\begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_K'] \end{pmatrix}^{-1}$$

$$\times\begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})e'] \\ \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})e'] \end{pmatrix}+O_p(T^{-3/2})+O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1}),$$

where X_k and e are defined below (2.9) and $\overline{\mathbb{G}}=(1_T,\mathbb{G})$. Notice $\ddot{M}$ is defined as $\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}$, which is equal to $\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}$ since $\Lambda=(\Psi,0_{N\times r_2})$ in the present context.

Remark 3.2. The alternative expression in the preceding remark can be explained in an intuitive way. Notice that the first equation of (3.1) can be written as

$$Y=X_1\beta_1+X_2\beta_2+\cdots+X_K\beta_K+\Psi\mathbb{G}'+\alpha 1_T'+e.$$

Post-multiplying $\mathcal{M}(1_T)$, we have

$$Y\mathcal{M}(1_T)=X_1\mathcal{M}(1_T)\beta_1+\cdots+X_K\mathcal{M}(1_T)\beta_K+\Psi\mathbb{G}'\mathcal{M}(1_T)+e\mathcal{M}(1_T).$$

Pre-multiplying $\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}$ and post-multiplying $\mathcal{M}[\mathcal{M}(1_T)\mathbb{G}]$, and noticing $\mathcal{M}(1_T)\mathcal{M}[\mathcal{M}(1_T)\mathbb{G}]=\mathcal{M}(\overline{\mathbb{G}})$, we have

$$\begin{aligned}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}Y\mathcal{M}(\overline{\mathbb{G}})&=\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}X_1\mathcal{M}(\overline{\mathbb{G}})\beta_1+\cdots\\&\quad+\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}X_K\mathcal{M}(\overline{\mathbb{G}})\beta_K+\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}e\mathcal{M}(\overline{\mathbb{G}}).\end{aligned}$$

The error term of the above equation, $\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}e\mathcal{M}(\overline{\mathbb{G}})$, is asymptotically homoscedastic and uncorrelated over the cross section and over time. Applying the ordinary least squares (OLS) to the above equation, we will obtain the same limiting result. Of course, this method is infeasible because Λ,Σ_{ee} and $\mathbb{G}$ are unobservable. The maximum likelihood method amounts to making the unobservable factors and factor loadings observable.

Given Theorem 3.1 and Remark 3.1 we have the following corollary.

COROLLARY 3.1 (limiting distribution). *Under Assumptions A-E, if* $\sqrt{N}/T \to 0$, *we have*

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \overline{\mathcal{P}}^{\,-1}),$$

where $\overline{\mathcal{P}} = \lim_{N,T\to\infty} \mathcal{P}$, *and* $\overline{\mathcal{P}}$ *is also the probability limit of*

$$\overline{\mathcal{P}} = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_K'] \end{pmatrix}.$$

REMARK 3.3. Compared with the model in Section 2, $\hat{\beta}$ is more efficient under the zero loading restrictions. The reason is intuitive. In the previous model, only variations in v_{itx} provide information for β. But in the present case, variations in $\gamma_{ik}^{h\prime}h_t$ of x_{it} also provide information for β. This can also be seen by comparing the limiting variances of Corollaries 2.1 and 3.1. Notice the projection matrix now only involves $\overline{\mathbb{G}}$ instead of $\overline{\mathbb{F}}$; and $\overline{\mathbb{G}}$ is a submatrix of $\overline{\mathbb{F}}$.

REMARK 3.4. The covariance matrix $\overline{\mathcal{P}}$ can be estimated by the same method as in estimating $\overline{\Omega}$; see Remark 2.7.

REMARK 3.5. Consider additional cases of zero restrictions. If $\gamma_{ik}^{h} = 0$ for all i, then model (3.1) reduces to the common shock model of Section 2 with g_t as the common shocks only. If zero restrictions are imposed in the x equation only instead of the y equation, i.e.,

$$\begin{aligned} y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \psi_i' g_t + \phi_i' h_t + e_{it}, \\ x_{itk} &= \mu_{ik} + \gamma_{ik}^{g\prime} g_t + v_{itk}. \end{aligned}$$

it can be shown that $\hat{\beta}$ obtained by imposing $\gamma_{ik}^{h} = 0$ has the same asymptotic representation as in Theorem 2.2; the zero restriction $\gamma_{ik}^{h} = 0$ does not improve the efficiency of $\hat{\beta}$. Finally, consider the following restricted model:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \psi_i' g_t + e_{it}, \\ x_{itk} &= \mu_{ik} + \gamma_{ik}^{h\prime} h_t + v_{itk}. \end{aligned}$$

Here the y and x equations do not share the same factors, though g_t and h_t are correlated. In comparison with model (3.1), the above model imposes extra zero restrictions in the x equation. Again, it can be shown that the zero restrictions in the x equation do not improve the efficiency of $\hat{\beta}$. That is, the estimator with the additional restrictions has the same asymptotic representation as in Theorem 3.1.

4. Models with time-invariant regressors and common regressors. In this section, we extend the basic model in section 2 to include time-invariant regressors and common regressors. Examples of time-invariant regressors include gender, race and education; and examples for common regressors include price variables, unemployment rate, or macroeconomic policy variables. These types of regressors are important for empirical applications.

We first consider the model with only time-invariant regressors:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \psi_i' g_t + \phi_i' h_t + e_{it} \\ x_{itk} &= \mu_{ik} + \gamma_{ik}^{g\prime} g_t + \gamma_{ik}^{h\prime} h_t + v_{itk} \end{aligned} \tag{4.1}$$

for $k = 1, 2, \cdots, K$, where g_t is an r_1-dimensional vector and h_t is an r_2-dimensional vector. Let $f_t = (g_t', h_t')'$, an r-dimensional vector. The key point of model (4.1) is that the ϕ_i's are known (but not zeros). We treat ϕ_i as new added time-invariant regressors, whose coefficient h_t is allowed to be time-varying. The parameters of interest, besides β, include h_t for all t. The model in the previous section can be viewed as $\Phi = 0$, where $\Phi = (\phi_1, \phi_2, \cdots, \phi_N)'$. However, the earlier derivation is not applicable here because now Φ is a general matrix with full column rank, which provides more information (restrictions) on the rotation matrix. Thus the number of restrictions required to eliminate rotational indeterminacy is even fewer than in Section 3. This point can be seen in the next subsection.

We define the following notation for further analysis:

$$\begin{aligned} &\Gamma_i^g = (\psi_i, \gamma_{i1}^g, \cdots, \gamma_{iK}^g), \quad \Gamma_i^h = (\phi_i, \gamma_{i1}^h, \cdots, \gamma_{iK}^h), \quad \Gamma_i = (\Gamma_i^{g\prime}, \Gamma_i^{h\prime})', \\ &\Lambda = (\lambda_1, \lambda_2, \cdots, \lambda_N)', \quad \Psi = (\psi_1, \psi_2, \cdots, \psi_N)', \quad \lambda_i = (\psi_i', \phi_i')', \\ &\Phi = (\phi_1, \phi_2, \cdots, \phi_N)'. \end{aligned}$$

Then equation (4.1) has the same matrix expression as (2.2). Note that $\Lambda = [\Psi, \Phi]$ is the factor loading matrix for the $N \times 1$ vector $(y_{1t}, y_{2t}, ..., y_{NT})'$.

4.1. *Identification Conditions.* We make the following identification conditions, which we refer to as IO (*Identification* conditions with partial *Observable* fixed effects), to emphasize the observed fixed effects.

IO1. We partition the matrix M_{ff} as

$$M_{ff} = \begin{bmatrix} M_{gg} & M_{gh} \\ M_{hg} & M_{hh} \end{bmatrix}$$

and impose $M_{gh} = 0$ and $M_{gg} = I_{r_1}$.

IO2. $\frac{1}{N}\Gamma^{g\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^{g} = D$, where D is a diagonal matrix with its diagonal elements distinct and arranged in desecending order.

IO3. $1_T'\mathbb{G} = 0$ and $1_T'\mathbb{H} = 0$.

These restrictions can be imposed without loss of generality, as argued later and more formally in Appendix C.3. In addition, we make the following assumption.

Assumption F: The loading matrix $\Lambda = [\Psi, \Phi]$ is of full column rank.

Now we use the method in Section 3.1 to show that IO is enough to achieve full identification. Let R be the rotation matrix partitioned in the same way as in Section 3.1. If there exists some matrix $\Gamma^{\dagger}$, which shares the same structure with Γ, satisfying $\Gamma^{\dagger} = \Gamma R$, then we must have $\Psi R_{12} + \Phi R_{22} = \Phi$ because Φ is given. This is equivalent to

$$[\Psi, \Phi]\begin{bmatrix} R_{12} \\ I - R_{22} \end{bmatrix} = \Lambda \begin{bmatrix} R_{12} \\ I - R_{22} \end{bmatrix} = 0.$$

So if matrix Λ is of full column rank, then pre-multiplying $(\Lambda'\Lambda)^{-1}\Lambda'$ gives $R_{12} = 0$ and $R_{22} = I$. Thus only R_{11} and R_{12} are undetermined. This implies that we only need $r_1^2 + r_1 r_2$ restrictions. The number of restrictions implied by $IO1$ and $IO2$ is exactly $r_1^2 + r_1 r_2$. So there is further reduction in the number of restrictions to eliminate rotational indeterminacy.

4.2. *Estimation.* For clarity, in this subsection, we use Φ^* to denote the observed value for Φ. Recall that $\Sigma_{zz} = \Gamma M_{ff}\Gamma' + \Sigma_{\varepsilon\varepsilon}$, where Γ contains the factor loading coefficients (including Φ); M_{ff} contains the sub-blocks M_{gg}, M_{gh}, and M_{hh}; $\Sigma_{\varepsilon\varepsilon}$ contains the heteroskedasticity coefficients. The regression coefficient β is contained in matrix B. The maximization of the likelihood function is now subject to four sets of restrictions, $M_{gh} = 0$, $M_{gg} = I_{r_1}$, $\Phi = \Phi^*$, and $\frac{1}{N}\Gamma^{g\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^{g} = D$. The likelihood function augmented with the Lagrange multipliers is

$$\ln L = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}\Big[(I_N \otimes B)M_{zz}(I_N \otimes B')\Sigma_{zz}^{-1}\Big] + \mathrm{tr}\Big[\Upsilon_1 M_{gh}\Big] +$$

$$\mathrm{tr}\Big[\Upsilon_2(M_{gg} - I_{r_1})\Big] + \mathrm{tr}\Big[\Upsilon_3\Big(\frac{1}{N}\Gamma^{g\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^{g} - D\Big)\Big] + \mathrm{tr}\Big[\Upsilon_4(\Phi - \Phi^*)\Big],$$

where $\Upsilon_1, \Upsilon_2, \Upsilon_3$ and Υ_4 are all Lagrange multipliers matrices; Υ_1 is an $r_2 \times r_1$ matrix; Υ_2 is an $r_1 \times r_1$ symmetric matrix; Υ_3 is an $r_1 \times r_1$ symmetric matrix with all diagonal elements zeros; Υ_4 is an $r_2 \times N$ matrix; and $\Sigma_{zz} = \Gamma M_{ff}\Gamma' + \Sigma_{\varepsilon\varepsilon}$. Using the same arguments in deriving $\Upsilon_1 = 0$ in Section 3, we have $\Upsilon_3 = 0$. Then the likelihood function is simplified as

$$\ln L = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}\Big[(I_N\otimes B)M_{zz}(I_N\otimes B')\Sigma_{zz}^{-1}\Big] + \mathrm{tr}[\Upsilon_1 M_{gh}] + \mathrm{tr}[\Upsilon_2(M_{gg}-I_{r_1})] + \mathrm{tr}[\Upsilon_4(\Phi-\Phi^*)]. \tag{4.2}$$

The first order condition for Γ gives

$$\hat{M}_{ff}\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N\otimes\hat{B})M_{zz}(I_N\otimes\hat{B}')-\hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1} = W',$$

where W is defined in (3.3). Pre-multiplying $\hat{M}_{ff}^{-1}$ and post-multiplying $\hat{\Gamma}$, and by the special structures of W and $\hat{\Gamma}$, we have

$$\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N\otimes\hat{B})M_{zz}(I_N\otimes\hat{B}')-\hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1}\hat{\Gamma} = -\begin{bmatrix} 0_{r_1\times r_1} & 0_{r_1\times r_2} \\ \frac{1}{N}\hat{M}_{hh}^{-1}\Upsilon_4'\hat{\Psi} & \frac{1}{N}\hat{M}_{hh}^{-1}\Upsilon_4'\Phi \end{bmatrix}.$$

But the first order condition for M_{ff} gives

$$\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N\otimes\hat{B})M_{zz}(I_N\otimes\hat{B}')-\hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1}\hat{\Gamma} = \begin{bmatrix} \Upsilon_2 & \Upsilon_1' \\ \Upsilon_1 & 0_{r_2\times r_2} \end{bmatrix}.$$

Comparing the proceeding two results and noting that the left hand side is a symmetric matrix, we have $\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N\otimes\hat{B})M_{zz}(I_N\otimes\hat{B}')-\hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1}\hat{\Gamma} = 0$. But $\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}$ can be replaced by $\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$ (see (S.2) in Appendix). Thus

$$\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}[(I_N\otimes\hat{B})M_{zz}(I_N\otimes\hat{B}')-\hat{\Sigma}_{zz}]\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = 0. \tag{4.3}$$

The above result implies that $\Upsilon_1 = 0$, $\Upsilon_2 = 0$, $\Upsilon_4'\hat{\Psi} = 0$ and $\Upsilon_4'\Phi = 0$.

The first order condition for Σ_{ii} is the same as (3.7), i.e.

$$\hat{B}M_{zz}^{jj}\hat{B}' - \hat{\Sigma}_{zz}^{jj} - \hat{\Gamma}_j'\hat{G}\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Big(\hat{B}M_{zz}^{ij}\hat{B}' - \hat{\Sigma}_{zz}^{ij}\Big) - \sum_{i=1}^N\Big(\hat{B}M_{zz}^{ji}\hat{B}' - \hat{\Sigma}_{zz}^{ji}\Big)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{G}\hat{\Gamma}_j = \mathbb{W}, \tag{4.4}$$

where $\mathbb{W}$ is defined following (2.8).

The first order condition on β is the same as (3.8), i.e.

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\hat{\Sigma}_{iie}^{-1}\Big\{(\dot{y}_{it}-\dot{x}_{it}\hat{\beta}) - \hat{\lambda}_i'\hat{G}\sum_{j=1}^N\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\begin{bmatrix}\dot{y}_{jt}-\dot{x}_{jt}\hat{\beta} \\ \dot{x}_{jt}'\end{bmatrix}\Big\}\dot{x}_{it} = 0. \tag{4.5}$$

We need an additional identify for the theoretical analysis in the appendix. The preceding analysis shows that $\frac{1}{N}\Upsilon_4'\hat{\Psi} = 0$ and $\frac{1}{N}\Upsilon_4'\Phi = 0$. They imply

$$\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\hat{G}_2\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{B}M_{zz}^{ij}\hat{B}' - \hat{\Sigma}_{zz}^{ij})\hat{\Sigma}_{jj}^{-1}I_{K+1}^1\hat{\lambda}_j' = 0, \tag{4.6}$$

where $\hat{\lambda}_j = (\hat{\psi}_i', \phi_i')'$.

4.3. *Asymptotic properties.* The following theorem states the asymptotic representation for $\hat{\beta} - \beta$.

PROPOSITION 4.1. *Under Assumptions A-D and F, and under the identification condition IO, we have*

$$\begin{aligned}\mathcal{Q}^0(\hat{\beta} - \beta) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ix}^{h\prime}h_t e_{it}\\ &\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\lambda_i'\Pi_{\lambda\lambda}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\lambda_j'\Sigma_{jje}^{-1}\gamma_{jx}^{h\prime}\Big)h_t e_{it}\\ &\quad + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}),\end{aligned}$$

where $\mathcal{Q}^0$ *is a* $K \times K$ *symmetric matrix with its* (p,q) *element equal to* $\mathrm{tr}[M_{hh}\Gamma_p^{h\prime}\ddot{M}\Gamma_q^h] + \frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}$*;* $\ddot{M} = \Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}$*;* $\Gamma_p^h = [\gamma_{1p}^h, \gamma_{2p}^h, \ldots, \gamma_{Np}^h]'$*;* $\Pi_{\lambda\lambda} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i\Sigma_{iie}^{-1}\lambda_i'$*; and* $\gamma_{jx}^h = [\gamma_{j1}^h, \gamma_{j2}^h, \ldots, \gamma_{jK}^h]$.

Proposition 4.1 is derived under the identification conditions IO. In Appendix C.3, we show that for any set of factors and factor loadings $(\psi_i, \gamma_{ik}, g_t, h_t)$, we can always transform it to another set $(\psi_i^\star, \gamma_{ik}^\star, g_t^\star, h_t^\star)$ which satisfies IO, and at the same time, still maintains the observability of Φ (i.e., Φ is untransformed). This is in agreement with the Lagrange multiplier analysis, in which $\Upsilon_j = 0$ $(j = 1, 2, 3)$; the only binding restriction is $\Phi = \Phi^*$. Using the relationship between the two sets, we can generalize Proposition 4.1 into the following theorem, which does not depend on IO.

THEOREM 4.1. *Under Assumptions A-D and F, we have*

$$\begin{aligned}\mathcal{Q}(\hat{\beta} - \beta) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ix}^{h\prime}h_t^\star e_{it}\\ &\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\lambda_i'\Pi_{\lambda\lambda}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\lambda_j'\Sigma_{jje}^{-1}\gamma_{jx}^{h\prime}\Big)h_t^\star e_{it}\\ &\quad + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})\end{aligned}$$

where

$$h_t^\star = \dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t;$$

$\mathcal{Q}$ *is a* $K \times K$ *symmetric matrix with its* (p,q) *element equal to*

$$\mathrm{tr}[\ddot{M}\Gamma_q^h\mathbb{H}'\mathcal{M}(\overline{\mathbb{G}})\mathbb{H}\Gamma_p^{h\prime}] + \frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)};$$

and $\ddot{M}$, Γ_p^h and $\Pi_{\lambda\lambda}$ are defined in Proposition 4.1.

REMARK 4.1. In Appendix C.3 we show that the asymptotic expression of $\hat{\beta}-\beta$ in Theorem 4.1 can be expressed alternatively as

$$\hat{\beta}-\beta = \begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_K'] \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})e'] \\ \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})e'] \end{pmatrix} + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}),$$

where X_k and e are defined below (2.9) and $\overline{\mathbb{G}} = (1_T, \mathbb{G})$.

REMARK 4.2. The expression in Remark 4.1 has a similar intuitive explanation as in Section 3. The first equation of (4.1) can be written as

$$Y = X_1\beta_1 + X_2\beta_2 + \cdots + X_k\beta_K + \Psi\mathbb{G}' + \Phi\mathbb{H}' + \alpha 1_T' + e.$$

Pre-multiply $\Sigma_{ee}^{-1/2}$ to eliminate the heteroscedasticity

$$\Sigma_{ee}^{-1/2}Y = \Sigma_{ee}^{-1/2}X_1\beta_1 + \cdots + \Sigma_{ee}^{-1/2}X_k\beta_K + \Sigma_{ee}^{-1/2}\Psi\mathbb{G}' + \Sigma_{ee}^{-1/2}\Phi\mathbb{H}' + \Sigma_{ee}^{-1/2}\alpha 1_T' + \Sigma_{ee}^{-1/2}e.$$

Both Φ and 1_T are observable. Pre-multiplying $\mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)$ and post-multiplying $\mathcal{M}(1_T)$, we eliminate $\Phi\mathbb{H}'$ and $\alpha 1_T$. Thus

$$\mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}Y\mathcal{M}(1_T) = \mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}X_1\mathcal{M}(1_T)\beta_1 + \cdots + \mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)$$
$$\times\Sigma_{ee}^{-1/2}X_K\mathcal{M}(1_T)\beta_K + \mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}\Psi\mathbb{G}'\mathcal{M}(1_T) + \mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}e\mathcal{M}(1_T).$$

Notice that both $\mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}\Psi$ and $\mathbb{G}'\mathcal{M}(1_T)$ are unobservable. Pre-multiply $\mathcal{M}[\mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}\Psi]$ and post-multiply $\mathcal{M}[\mathcal{M}(1_T)\mathbb{G}]$ to eliminate the unobservable common factors. Using the result that $\mathcal{M}(1_T)\mathcal{M}[\mathcal{M}(1_T)\mathbb{G}] = \mathcal{M}(\overline{\mathbb{G}})$ and $\mathcal{M}[\mathcal{M}(\Sigma_{ee}^{-1/2}\Phi)\Sigma_{ee}^{-1/2}\Psi]\mathcal{M}(\Sigma_{ee}^{-1/2}\Phi) = \mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)$, we have

$$\mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}Y\mathcal{M}(\overline{\mathbb{G}}) = \mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}X_1\mathcal{M}(\overline{\mathbb{G}})\beta_1 + \cdots + \mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}X_K\mathcal{M}(\overline{\mathbb{G}})\beta_K + \mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}e\mathcal{M}(\overline{\mathbb{G}}).$$

The error term of the above equation, $\mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}e\mathcal{M}(\overline{\mathbb{G}})$, is asymptotically homoscedastic and uncorrelated over the cross section and over time. Applying OLS to the above equation, we have the same asymptotic expression. Again, this operation is infeasible in practice, but the MLE makes it asymptotically feasible.

From Theorem 4.1, we obtain the following corollary.

COROLLARY 4.1. *Under the conditions of Theorem 4.1, if* $\sqrt{N}/T \to 0$, *we have*

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \overline{\mathcal{Q}}^{-1}),$$

where $\overline{\mathcal{Q}} = \lim_{N,T\to\infty} \mathcal{Q}$, *which has an alternative expression*

$$\overline{\mathcal{Q}} = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT} \begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_K'] \end{pmatrix}.$$

REMARK 4.3. Compared with the model in Section 2, $\hat{\beta}$ is more efficient with observable fixed effects (time-invariant regressors). The reason is provided in Remark 3.3.

We next consider estimating h_t. It is worth emphasizing that unlike the asymptotic theory for $\hat{\beta}$, where the identification conditions in IO are inessential but facilitate the theoretical analysis, the asymptotic theory of $\hat{h}_t$ depends on the identification conditions, without which h_t is not identifiable. In what follows, we assume that the underlying parameters satisfy IO. We estimate h_t by the following formula:

$$\hat{h}_t = \Big[\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}\Phi\Big]^{-1}\Big[\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}(\dot{\mathbb{Y}}_t - \dot{\mathbb{X}}_t\hat{\beta})\Big],$$

where $\mathbb{Y}_t = (y_{1t}, y_{2t}, \cdots, y_{Nt})'$ and $\dot{\mathbb{Y}}_t = \mathbb{Y}_t - T^{-1}\sum_{t=1}^T \mathbb{Y}_t$; $\mathbb{X}_t$ is an $N \times K$ matrix with its (i,k) element x_{itk} and $\dot{\mathbb{X}}_t = \mathbb{X}_t - T^{-1}\sum_{t=1}^T \mathbb{X}_t$. Now we state the limiting result for $\hat{h}_t$.

THEOREM 4.2. *Under Assumptions A-D, F, and the identification conditions IO, if* $\sqrt{N}/T \to 0$, *then*

$$\sqrt{N}(\hat{h}_t - h_t) \xrightarrow{d} N\Big(0, \Big[\lim_{N\to\infty} \frac{1}{N}\Phi'\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\Phi\Big]^{-1}\Big).$$

Theorem 4.2 also has an intuitive explanation. Consider the first equation of (4.1), which can be written as

$$\mathbb{Y}_t = \mathbb{X}_t\beta + \Psi g_t + \Phi h_t + \alpha + e_t.$$

First remove α from the above equation, this gives (Note $1_T'\mathbb{G} = 0, 1_T'\mathbb{H} = 0$)

$$\dot{\mathbb{Y}}_t = \dot{\mathbb{X}}_t\beta + \Psi g_t + \Phi h_t + \dot{e}_t.$$

Then pre-multiplying $\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}$, we have

$$\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}(\dot{\mathbb{Y}}_t-\dot{\mathbb{X}}_t\beta) = \mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\Phi h_t+\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\dot{e}_t.$$

The error term, $\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\dot{e}_t$, is asymptotically homoscedastic and uncorrelated. Applying OLS to the above equation, we obtain the same limiting result as stated in Theorem 4.2.

Theorem 4.1 shows that $\hat{\beta}-\beta$ is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-3/2})$. But the convergence rate of $\hat{h}_t-h_t$ is only $\sqrt{N}$. So under the conditions $\sqrt{N}/T\to 0$, we can treat β as known. The limiting distribution of $\hat{h}_t-h_t$ is the same as the case of a pure multiple battery factor model (without regressors) with known loading Φ. The asymptotic representations of $\hat{\psi}_i-\psi_i$ and $\hat{g}_t-g_t$ can also be derived by the same arguments for $\hat{h}_t-h_t$. These results share the common feature that the limiting representations for one set of parameters are identical to the situation in which the remaining parameters can be treated as observable.

4.4. *Models with time-invariant regressors and common regressors.* In this subsection, we consider the joint presence of time-invariant regressors and common regressors. Consider the following model

$$\begin{aligned} y_{it} &= x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \psi_i' g_t + \phi_i' h_t + \kappa_i' d_t + e_{it} \\ x_{itk} &= \gamma_{ik}^{g\prime} g_t + \gamma_{ik}^{h\prime} h_t + \gamma_{ik}^{d\prime} d_t + v_{itk} \end{aligned} \tag{4.7}$$

for $k=1,2,\cdots,K$, where g_t, h_t and d_t are $r_1\times 1$, $r_2\times 1$ and $r_3\times 1$ vectors, respectively. A key feature of model (4.7) is that d_t and ϕ_i are observable for all i and t. We call ϕ_i the time-invariant regressors because they are invariant over time and d_t the common regressors because they are the same for all the cross-sectional units. In this model, the time-invariant regressors have time-varying coefficients, and the common regressors have heterogeneous (individual-dependent) coefficients. The parameters of interest now, besides the coefficient β, include κ_i and h_t. If $d_t\equiv 1$, κ_i plays the role of α_i in (4.1). So the model here is more general.

Similarly as the previous subsection, we make the following assumption:

Assumption G: The matrices (Ψ,Φ,K) and $(\mathbb{G},\mathbb{H},\mathbb{D})$ are both of full column rank, where $\mathrm{K}=(\kappa_1,\kappa_2,\ldots,\kappa_N)'$ and $\mathbb{D}=(d_1,d_2,\ldots,d_T)'$.

Let $\lambda_i=(\psi_i',\phi_i')'$, $\gamma_{ik}=(\gamma_{ik}^{g\prime},\gamma_{ik}^{h\prime})'$ and $\delta_i=(\kappa_i,\gamma_{ik}^{d})$. The model can be written as

$$\begin{bmatrix} 1 & -\beta' \\ 0 & I_K \end{bmatrix} z_{it} = \Gamma_i' f_t + \delta_i' d_t + \varepsilon_{it},$$

where $z_{it}, \Gamma_i, \varepsilon_{it}$ are defined in Section 2; Let $\Delta = (\delta_1, \delta_2, \cdots, \delta_N)'$, then we have

$$(I_N \otimes B)z_t - \Delta d_t = \Gamma f_t + \varepsilon_t, \tag{4.8}$$

where the symbols $\Gamma, z_t, B, \varepsilon_t$ are defiend in Section 2.

The likelihood function can be written as

$$\ln L = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2NT}\sum_{t=1}^{T}[(I_N \otimes B)z_t - \Delta d_t]'\Sigma_{zz}^{-1}[(I_N \otimes B)z_t - \Delta d_t].$$

Take Σ_{zz} and β as given, Δ maximizes the above function at

$$\hat{\Delta} = (I_N \otimes B)\Big(\sum_{s=1}^{T} z_s d_s'\Big)\Big(\sum_{s=1}^{T} d_s d_s'\Big)^{-1}.$$

Substituting $\hat{\Delta}$ into the above likelihood function, we obtain the concentrated likelihood function

$$\ln L = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2NT}\mathrm{tr}\Big[(I_N \otimes B)Z\mathcal{M}(\mathbb{D})Z'(I_N \otimes B')\Sigma_{zz}^{-1}\Big],$$

where $Z = (z_1, z_2, \cdots, z_T)$, $\mathbb{D} = (d_1, d_2, \cdots, d_T)'$, and $\mathcal{M}(\mathbb{D}) = I_T - \mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}\mathbb{D}'$, a projection matrix. Consider (4.8), which is equivalent to

$$(I_N \otimes B)Z = \Gamma\mathbb{F}' + \Delta\mathbb{D}' + \varepsilon,$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T)$. Post-multiplying $\mathcal{M}(\mathbb{D})$ on both sides, we have

$$(I_N \otimes B)Z\mathcal{M}(\mathbb{D}) = \Gamma\mathbb{F}'\mathcal{M}(\mathbb{D}) + \varepsilon\mathcal{M}(\mathbb{D}).$$

If we treat $Z\mathcal{M}(\mathbb{D})$ as the new observable data, $\mathbb{F}'\mathcal{M}(\mathbb{D})$ as the new unobservable factors, the preceding equation can be viewed as a special case of (4.1). Invoking Theorem 4.1, which does not need IO (the factors $\mathbb{F}'\mathcal{M}(\mathbb{D})$ may not satisfy IO), we have the following theorem:

THEOREM 4.3. *Under Assumptions A-D and G, the asymptotic representation of $\hat{\beta}$ in the presence of time invariant and common regressors is*

$$\begin{aligned}\mathcal{R}(\hat{\beta} - \beta) = {} & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ix}^{h\prime}h_t^{\star}e_{it} \\ & - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\lambda_i'\Pi_{\lambda\lambda}^{-1}\frac{1}{N}\sum_{j=1}^{N}\lambda_j'\Sigma_{jje}^{-1}\gamma_{jx}^{h\prime}h_t^{\star}e_{it} \\ & + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}),\end{aligned}$$

where

$$h_t^\star = h_t - \mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t - \mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t);$$

$\mathcal{R}$ *is a* $K \times K$ *symmetric matrix with its* (p,q) *element equal to*

$$\mathrm{tr}[\ddot{M}\Gamma_q^h\mathbb{H}'\mathcal{M}(\mathbb{B})\mathbb{H}\Gamma_p^{h\prime}] + \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)},$$

where $b_t = (g_t', d_t')'$ *and* $\mathbb{B} = (b_1, b_2, \cdots, b_T)' = (\mathbb{G}, \mathbb{D})$, *a matrix of* $T \times (r_1 + r_3)$ *dimension;* $\ddot{M} = \Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Lambda)\Sigma_{ee}^{-1/2}$; $\Gamma_p^h = (\gamma_{1p}^h, \gamma_{2p}^h, \ldots, \gamma_{Np}^h)'$; $\Pi_{\lambda\lambda} = \frac{1}{N}\sum_{i=1}^N \lambda_i\Sigma_{iie}^{-1}\lambda_i'$.

REMARK 4.4. The asymptotic expression of $\hat{\beta} - \beta$ can be alternatively expressed as

$$\hat{\beta} - \beta = \begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\mathbb{B})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\mathbb{B})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\mathbb{B})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\mathbb{B})X_K'] \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\mathbb{B})e'] \\ \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\mathbb{B})e'] \end{pmatrix} + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}).$$

If $\mathbb{D} = 1_T$, the above asymptotic result reduces to the one in Theorem 4.1 since $\mathbb{B} = (1_T, \mathbb{G}) = \overline{\mathbb{G}}$.

Given Theorem 4.3 and Remark 4.4, we have the following corollary.

COROLLARY 4.2. *Under Assumptions A-D and G, if* $\sqrt{N}/T \to 0$, *then*

$$\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \overline{\mathcal{R}}^{-1}),$$

where $\overline{\mathcal{R}} = \lim_{N,T\to\infty} \mathcal{R}$, *and* $\overline{\mathcal{R}}$ *can also be expressed as*

$$\overline{\mathcal{R}} = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT}\begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\mathbb{B})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\mathbb{B})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\mathbb{B})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\mathbb{B})X_K'] \end{pmatrix}.$$

We next consider the estimation of h_t and κ_i. Again, we emphasize that the estimation of h_t and κ_i is meaningful only when they are identifiable. To identify them, we impose the following IO$'$ conditions.

The identification conditions IO$'$: The identification conditions IO hold. In addition, $M_{fd} = \frac{1}{T}\sum_{t=1}^{T} f_t d_t' = 0$ with $f_t = (g_t', h_t')'$.

Under the identification conditions IO$'$, we estimate f_t by

$$\hat{f}_t = (\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}(\tilde{\mathbb{Y}}_t - \tilde{\mathbb{X}}_t\hat{\beta}),$$

where $\tilde{\mathbb{Y}}_t$ is the tth column of the matrix $Y\mathcal{M}(\mathbb{D})$, $\tilde{\mathbb{X}}_t$ is an $N \times K$ matrix with its kth column equal to the tth column of the matrix $X_k\mathcal{M}(\mathbb{D})$ and $\hat{\Lambda} = (\hat{\Psi}, \Phi)$ (note Φ is observable). The estimator $\hat{h}_t$ is included in $\hat{f}_t$. After obtaining the estimators $\hat{\beta}, \hat{\psi}_i, \hat{g}_t, \hat{h}_t$, we estimate κ_i by

$$\hat{\kappa}_i = \Big[\sum_{t=1}^{T} d_t d_t'\Big]^{-1}\Big[\sum_{t=1}^{T} d_t(y_{it} - x_{it}\hat{\beta} - \hat{\psi}_i'\hat{g}_t - \phi_i'\hat{h}_t)\Big].$$

The following Theorem states the limiting results on $\hat{h}_t$ and $\hat{\kappa}_i$.

THEOREM 4.4. *Under Assumptions A-D and G, together with the identification conditions IO$'$,*

1. *Under $\sqrt{N}/T \to 0$, we have*

$$\sqrt{N}(\hat{h}_t - h_t) \xrightarrow{d} N\Big(0, \Big[\lim_{N\to\infty}\frac{1}{N}\Phi'\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\Phi\Big]^{-1}\Big),$$

2. *Under $\sqrt{T}/N \to 0$, we have*

$$\sqrt{T}(\hat{\kappa}_i - \kappa_i) \xrightarrow{d} N\Big(0, \Sigma_{iie}\Big[\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} d_t d_t'\Big]^{-1}\Big).$$

5. Computing algorithm. To estimate the model by the maximum likelihood method, we adapt the ECM (Expectation and Conditional Maximization) procedures of [22]. The ECM procedure here can be viewed as the extension of the EM algorithm for the pure factor models considered by [29]. The E-step of the ECM algorithm is the same with the usual EM algorithm, but the M-step is broken into a sequence of maximizations instead of simultaneously maximization over the full parameter space. In the M-step, we split the parameter $\theta = (\beta, \Gamma, \Sigma_{\varepsilon\varepsilon}, M_{ff})$ into two blocks, $\theta_1 = (\Gamma, \Sigma_{\varepsilon\varepsilon}, M_{ff})$ and $\theta_2 = \beta$, and update $\theta_1^{(k)}$ to $\theta_1^{(k+1)}$ given $\theta_2^{(k)}$ and then update $\theta_2^{(k)}$ to $\theta_2^{(k+1)}$ given $\theta_1^{(k+1)}$, where $\theta^{(k)}$ is the estimated value at the kth iteration.

Because different models have different restrictions which lead to different identification conditions, we explain how to estimate each model under different identification conditions.

5.1. *Basic Model.* In this case, $M_{ff} = I_r$. So the parameters to be estimated reduce to $\theta = (\beta, \Gamma, \Sigma_{\varepsilon\varepsilon})$. Let $\theta^{(k)} = (\beta^{(k)}, \Gamma^{(k)}, \Sigma_{\varepsilon\varepsilon}^{(k)})$ be the estimated value at the kth iteration. We update $\Gamma^{(k)}$ according to

$$\Gamma^{(k+1)} = \Big[\frac{1}{T}\sum_{t=1}^{T} E(z_t f_t' | Z, \theta^{(k)})\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | Z, \theta^{(k)})\Big]^{-1}, \tag{5.1}$$

where

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | Z, \theta^{(k)}) &= I_r - \Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)} \\ &+\Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}(I_N \otimes B^{(k)})M_{zz}(I_N \otimes B^{(k)\prime})(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)},\end{aligned} \tag{5.2}$$

$$\frac{1}{T}\sum_{t=1}^{T} E(z_t f_t' | Z, \theta^{(k)}) = (I_N \otimes B^{(k)})M_{zz}(I_N \otimes B^{(k)\prime})(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}, \tag{5.3}$$

with $\Sigma_{zz}^{(k)} = \Gamma^{(k)}\Gamma^{(k)\prime} + \Sigma_{\varepsilon\varepsilon}^{(k)}$. We update $\Sigma_{\varepsilon\varepsilon}^{(k)}$ and $\beta^{(k)}$ according to

$$\begin{aligned}\Sigma_{\varepsilon\varepsilon}^{(k+1)} = \mathrm{Dg}\Big\{&\Big(I_{N(K+1)} - \Gamma^{(k+1)}\Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}\Big) \\ &\times (I_N \otimes B^{(k)})M_{zz}(I_N \otimes B^{(k)\prime})\Big\},\end{aligned} \tag{5.4}$$

$$\begin{aligned}\beta^{(k+1)} = &\Big(\sum_{i=1}^{N}\sum_{t=1}^{T} \dot{x}_{it}'(\Sigma_{iie}^{(k+1)})^{-1}\dot{x}_{it}\Big)^{-1} \\ &\times \Big(\sum_{i=1}^{N}\sum_{t=1}^{T} \dot{x}_{it}'(\Sigma_{iie}^{(k+1)})^{-1}(\dot{y}_{it} - \lambda_i^{(k+1)\prime} f_t^{(k)})\Big),\end{aligned} \tag{5.5}$$

where $f_t^{(k)}$ is the transpose of the t-th row of

$$\mathbb{F}^{(k)} = E(\mathbb{F}|Z, \theta^{(k)}) = \dot{Z}'(I_N \otimes B^{(k)\prime})(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}$$

where $\dot{Z} = (\dot{z}_1, \dot{z}_2, \cdots, \dot{z}_T)$ with $\dot{z}_t = z_t - \frac{1}{T}\sum_{s=1}^{T} z_s$; $\mathrm{Dg}(\cdot)$ is the operator that sets the entries of its argument to zeros if the counterparts of $E(\varepsilon_t\varepsilon_t')$ are zeros.

Putting together, we obtain $\theta^{(k+1)} = (\Gamma^{(k+1)}, \beta^{(k+1)}, \Sigma_{\varepsilon\varepsilon}^{(k+1)})$.

5.2. *Basic model with restrictions.* Again we need not estimate M_{ff}. We update $\Gamma^{(k)}$ by two steps. In the first step, we calculate $\Gamma^{(k+1)}$ according to (5.1). Let $T^{-1}\sum_{t=1}^{T} E(g_t g_t'|Z,\theta^{(k)})$ be the left-upper $r_1 \times r_1$ submatrix of $T^{-1}\sum_{t=1}^{T} E(f_t f_t'|Z,\theta^{(k)})$. Let $\mathcal{Z}$ be the $N \times r_1$ matrix whose i-th row is the first r_1 elements of the $\big((i-1)(K+1)+1\big)$-th row of $T^{-1}\sum_{t=1}^{T} E(z_t f_t'|Z,\theta^{(k)})$ for $i = 1, 2, \cdots, N$. Then calculate

$$\Psi^{(k+1)} = \mathcal{Z}\Big[\frac{1}{T}\sum_{t=1}^{T} E(g_t g_t'|Z,\theta^{(k)})\Big]^{-1}$$

In the second step, $\Gamma^{(k+1)}$ is obtained by replacing the $\big((i-1)(K+1)+1\big)$-th row of the first step $\Gamma^{(k+1)}$ with $(\psi_i^{(k+1)\prime}, 0_{1\times r_2})$ for $i = 1, 2, \cdots, N$, where $\psi_i^{(k+1)}$ is the transpose of the ith row of $\Psi^{(k+1)}$. We update $\Sigma_{\varepsilon\varepsilon}^{(k)}$ by (5.4) and $\beta^{(k)}$ by

$$\beta^{(k+1)} = \Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{x}_{it}'(\Sigma_{iie}^{(k+1)})^{-1}\dot{x}_{it}\Big)^{-1}\Big(\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{x}_{it}'(\Sigma_{iie}^{(k+1)})^{-1}(\dot{y}_{it} - \psi_i^{(k+1)\prime} g_t^{(k)})\Big)$$

where $\Sigma_{zz}^{(k)} = \Gamma^{(k)}\Gamma^{(k)\prime} + \Sigma_{\varepsilon\varepsilon}^{(k)}$, $g_t^{(k)}$ is the transpose of the first r_1 elements of the t-th row of $\mathbb{F}^{(k)}$ with

$$\mathbb{F}^{(k)} = E(\mathbb{F}|Z,\theta^{(k)}) = \dot{Z}'(I_N \otimes B^{(k)\prime})(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}.$$

This gives $\theta^{(k+1)} = (\Gamma^{(k+1)}, \beta^{(k+1)}, \Sigma_{\varepsilon\varepsilon}^{(k+1)})$.

5.3. *Models with the time-invariant and common regressors.* The identification restrictions are such that we only need to estimate M_{hh}. Let $\theta^{(k)} = (\beta^{(k)}, \Gamma^{(k)}, \Sigma_{\varepsilon\varepsilon}^{(k)}, M_{hh}^{(k)})$ be the estimator of the kth iteration. We update $\Gamma^{(k)}$ by two steps. First calculate $\Gamma^{(k+1)}$ as in (5.1). Notice that $f_t = (g_t', h_t')'$. Once we have obtained $T^{-1}\sum_{t=1}^{T} E(f_t f_t'|Z,\theta^{(k)})$, then

$$\frac{1}{T}\sum_{t=1}^{T} E(g_t g_t'|Z,\theta^{(k)}), \quad \frac{1}{T}\sum_{t=1}^{T} E(h_t g_t'|Z,\theta^{(k)}), \quad \frac{1}{T}\sum_{t=1}^{T} E(h_t h_t'|Z,\theta^{(k)})$$

are all known. Let $\mathcal{Z}$ be the $N \times r_1$ matrix, whose i-th row is the first r_1 elements of the $\big((i-1)(K+1)+1\big)$-th row of $\frac{1}{T}\sum_{t=1}^{T} E(z_t f_t'|Z,\theta^{(k)})$ for $i = 1, 2, \cdots, N$. Because the factor loadings Φ is known, we calculate

$$\Psi^{(k+1)} = \Big[\mathcal{Z} - \Phi\Big(\frac{1}{T}\sum_{t=1}^{T} E(h_t g_t'|Z,\theta^{(k)})\Big)\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} E(g_t g_t'|Z,\theta^{(k)})\Big]^{-1}$$

Then we obtain $\Gamma^{(k+1)}$ by replacing the $((i-1)(K+1)+1)$-th row of the first step $\Gamma^{(k+1)}$ with $\lambda_i^{(k+1)} = (\psi_i^{(k+1)\prime}, \phi_i')$ for $i = 1, 2, \cdots, N$, where $\psi_i^{(k+1)}$ is the transpose of the ith row of $\Psi^{(k+1)}$. We update $\Sigma_{\varepsilon\varepsilon}^{(k)}$ by (5.4), $\beta^{(k)}$ by (5.5) and $M_{hh}^{(k)}$ according to

$$M_{hh}^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T} E(h_t h_t' | Z, \theta^{(k)})$$

This gives $\theta^{(k+1)} = (\Gamma^{(k+1)}, \beta^{(k+1)}, \Sigma_{\varepsilon\varepsilon}^{(k+1)}, M_{hh}^{(k+1)})$.

For the model with both time-invariant and common regressors, we first post-multiply $\mathcal{M}(\mathbb{D}) = I_T - \mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}\mathbb{D}'$ on the data matrix Z and apply the preceding procedures on $Z\mathcal{M}(\mathbb{D})$. Then all the estimators are obtained.

The above iteration continues until $\|\theta^{(k+1)} - \theta^{(k)}\|$ is smaller than a preset error tolerance. For the initial values, the iterated PC estimators of [8] are used.

6. Finite sample properties. In this section, we consider the finite sample properties of the MLE. Data are generated according to

$$\begin{aligned} y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \psi_i' g_t + \phi_i' h_t + \kappa_i' d_t + e_{it}, \\ x_{itk} &= \mu_{ik} + \gamma_{ik}^{g\prime} g_t + \gamma_{ik}^{h\prime} h_t + \gamma_{ik}^{d\prime} d_t + v_{itk}, \qquad k = 1, 2. \end{aligned} \tag{6.1}$$

The dimensions of g_t, h_t, d_t are each fixed to 1. We set $\beta_1 = 1$ and $\beta_2 = 2$. We consider four types of DGP (data generating process), which correspond to the four models considered in the paper.

DGP1: $\phi_i, \kappa_i, \gamma_{ik}^h$ and γ_{ik}^d are fixed to zeros; $\alpha_i, \mu_{ik}, \psi_i$ and g_t are generated from $N(0,1)$ and $\gamma_{ik}^g = \psi_i + N(0,1)$.

DGP2: ϕ_i, κ_i and γ_{ik}^d are fixed to zeros; $\alpha_i, \mu_{ik}, \psi_i$, γ_{ik}^h, g_t and h_t are generated from $N(0,1)$; $\gamma_{ik}^g = \psi_i + N(0,1)$.

DGP3: κ_i and γ_{ik}^d are fixed to zeros; $\alpha_i, \mu_{ik}, \psi_i, \phi_i, g_t$ and h_t are generated from $N(0,1)$; $\gamma_{ik}^g = \psi_i + N(0,1)$ and $\gamma_{ik}^h = \phi_i + N(0,1)$. Here ϕ_i is observable.

DGP4: $\alpha_i, \mu_{ik}, \psi_i, \phi_i, \kappa_i, g_t$ and h_t are generated from $N(0,1)$; $d_t = 1 + N(0,1)$, $\gamma_{ik}^g = \psi_i + N(0,1)$, $\gamma_{ik}^h = \phi_i + N(0,1)$ and $\gamma_{ik}^d = \kappa_i + N(0,1)$. Here ϕ_i and d_t are observable.

Using the method of writing (2.2), we can rewrite (6.1) as

$$(I_N \otimes B) z_t = \mu + L\varsigma_t + \varepsilon_t$$

where $\varsigma_t = g_t$ for DGP1; $\varsigma_t = (g_t, h_t)'$ for DGP2 and DGP3; $\varsigma_t = (g_t, h_t, d_t)'$ for DGP4 and L is the corresponding loadings matrix. Let ι_i' be the ith row of L. We generate the cross sectional heteroscedasticity Ξ, an $N(K+1) \times 1$ vector, according to

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

where η_i is drawn from $U[u, 1-u]$ with $u = 0.1$. A similar way of generating heteroscedasticity is also used in [13] and [16]. Let $\Upsilon = \operatorname{diag}(\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_N)$ be an $N(K+1) \times N(K+1)$ block diagonal matrix, in which $\Upsilon_i = \operatorname{diag}\{1, (M_i'M_i)^{-1/2}M_i\}$ with M_i being a $K \times K$ standard normal random matrix for each i. Once Υ is generated, the error term ε_t, which is defined as $(\varepsilon_{1t}', \varepsilon_{2t}', \cdots, \varepsilon_{Nt}')'$ with $\varepsilon_{it} = (e_{it}, v_{it1}, v_{it2})'$, is calculated by $\varepsilon_t = \sqrt{\operatorname{diag}(\Xi)}\Upsilon\epsilon_t$, where ϵ_t is an $N(K+1) \times 1$ vector with all its elements being *iid* $N(0,1)$. Once ε_t is obtained, we use

$$z_t = (I_N \otimes B)^{-1}(\mu + L\varsigma_t + \varepsilon_t)$$

to yield the observable data.

Tables 1-4 report the simulation results based on 1000 repetitions. Bias and root mean square error (RMSE) are computed to measure the performance of the estimators. For the purpose of comparison, we also report the performance of the within-group (WG) estimators and Bai's iterated principal components estimators (PC).

It is seen from the simulations that the WG estimators are inconsistent. The bias of the WG estimators shows no signs of decreasing as the sample size grows. The iterated PC estimators are consistent, but biased. As the sample size becomes large, the bias decreases noticeably. However, when the sample size is moderate, the bias of the iterated PC estimators is still pronounced. In comparison, the ML estimators are consistent and unbiased. For all the sample sizes, the biases of the ML estimators are very small and negligible. In addition, the RMSEs of the ML estimators are always the smallest among the three estimators, illustrating the efficiency of the ML method. The same patten is observed for all of the four models considered.

7. Conclusion. This paper considers estimating panel data models with interactive effects, in which explanatory variables are correlated with the unobserved effects. Standard panel data methods (such as the within-group estimator) are not suitable for this type of models. We study the maximum likelihood method and provide a rigorous analysis for the asymptotic theory. While the analysis is difficult, the limiting distributions of the MLE are simple and have intuitive interpretations. The maximum likelihood method can incorporate parameters restrictions to gain efficiency, a useful feature in view of the large number of parameters under large N and large T. We analyze the restrictions via the Lagrange multiplier approach. This approach can reveal what kinds of restrictions are binding and lead to efficiency gain. We allow the model to include time invariant regressors and common regressors. The coefficients of the time invariant regressors are time dependent, and the coefficients of the common regressors are cross-section dependent. This is a sensible way for modelling the effects of such variables in panel data context and fits naturally into the framework of interactive effects. The likelihood method is easy to implement and perform very well, as demonstrated by the Monte Carlo simulations.

Table 1: The performance of WG, PC and ML estimators in the basic model

		WG				PC				MLE			
		β_1		β_2		β_1		β_2		β_1		β_2	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
50	75	0.1568	0.1616	0.1571	0.1622	0.0203	0.0445	0.0194	0.0440	0.0001	0.0024	0.0000	0.0021
100	75	0.1566	0.1598	0.1545	0.1574	0.0055	0.0196	0.0051	0.0199	-0.0001	0.0011	-0.0001	0.0011
150	75	0.1546	0.1568	0.1543	0.1666	0.0025	0.0119	0.0032	0.0121	0.0000	0.0008	0.0000	0.0007
50	125	0.1557	0.1601	0.1577	0.1622	0.0165	0.0371	0.0162	0.0378	0.0000	0.0016	0.0000	0.0017
100	125	0.1554	0.1577	0.1557	0.1584	0.0043	0.0161	0.0051	0.0173	0.0000	0.0008	0.0000	0.0009
150	125	0.1546	0.1565	0.1557	0.1575	0.0022	0.0100	0.0019	0.0105	0.0000	0.0005	0.0000	0.0006

Table 2: The performance of WG, PC and ML estimators in the model with zero restrictions.

		WG				PC				MLE			
		β_1		β_2		β_1		β_2		β_1		β_2	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
50	75	0.1084	0.1122	0.1097	0.1137	0.0090	0.0223	0.0083	0.0230	0.0000	0.0011	0.0000	0.0012
100	75	0.1076	0.1097	0.1075	0.1097	0.0029	0.0109	0.0022	0.0107	0.0000	0.0006	0.0000	0.0006
150	75	0.1082	0.1100	0.1076	0.1092	0.0008	0.0064	0.0008	0.0067	0.0000	0.0004	0.0000	0.0004
50	125	0.1092	0.1127	0.1087	0.1119	0.0088	0.0209	0.0084	0.0208	0.0000	0.0009	-0.0001	0.0009
100	125	0.1084	0.1102	0.1096	0.1114	0.0024	0.0090	0.0027	0.0090	0.0000	0.0005	0.0000	0.0004
150	125	0.1074	0.1087	0.1076	0.1090	0.0011	0.0053	0.0009	0.0053	0.0000	0.0003	0.0000	0.0003

Table 3: The performance of WG, PC and ML estimators in the model with time-invariant regressors

		WG				PC				MLE			
		β_1		β_2		β_1		β_2		β_1		β_2	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
50	75	0.1568	0.1598	0.1557	0.1588	0.0170	0.0329	0.0186	0.0351	-0.0002	0.0046	0.0001	0.0045
100	75	0.1540	0.1557	0.1542	0.1560	0.0047	0.0155	0.0048	0.0148	-0.0002	0.0028	0.0001	0.0028
150	75	0.1538	0.1551	0.1541	0.1554	0.0019	0.0094	0.0024	0.0097	-0.0002	0.0021	0.0000	0.0022
50	125	0.1568	0.1595	0.1567	0.1594	0.0157	0.0294	0.0156	0.0300	0.0001	0.0042	0.0002	0.0038
100	125	0.1554	0.1568	0.1543	0.1557	0.0045	0.0132	0.0040	0.0123	0.0001	0.0022	-0.0001	0.0021
150	125	0.1553	0.1565	0.1542	0.1553	0.0015	0.0069	0.0016	0.0072	0.0000	0.0016	-0.0001	0.0016

Table 4: The performance of WG, PC and ML estimators in the model with time-invariant regressors and common regressors

		WG				PC				MLE			
		β_1		β_2		β_1		β_2		β_1		β_2	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
50	75	0.1777	0.1802	0.1796	0.1820	0.0193	0.0310	0.0194	0.0317	0.0001	0.0064	0.0002	0.0064
100	75	0.1772	0.1785	0.1880	0.1813	0.0062	0.0158	0.0065	0.0161	-0.0002	0.0035	-0.0002	0.0035
150	75	0.1775	0.1785	0.1779	0.1789	0.0032	0.0101	0.0027	0.0104	0.0002	0.0028	-0.0001	0.0029
50	125	0.1791	0.1815	0.1811	0.1833	0.0191	0.0305	0.0199	0.0320	-0.0001	0.0043	0.0000	0.0044
100	125	0.1790	0.1802	0.1787	0.1799	0.0053	0.0138	0.0060	0.0141	-0.0001	0.0030	0.0000	0.0027
150	125	0.1788	0.1796	0.1778	0.1787	0.0019	0.0068	0.0020	0.0069	0.0000	0.0023	0.0000	0.0021

References.

[1] Ahn, S. G., Y. H. Lee and P. Schmidt (2001): GMM estimation of linear panel data models with time-varying Individual effects, *Journal of Econometrics*, **101**, 219–255.

[2] Ahn, S. G., Y. H. Lee and P. Schmidt (2006): Panel data models with multiple time-varying effects, *manuscript*, Arizona State University.

[3] Amemiya Y., W. A. Fuller and S. G. Pantula (1987) The asymptotic distributions of some estimators for a factor analysis model, *Journal of Multivariate Analysis*, **22:1**, 51–64.

[4] Anderson, T.W. and Y. Amemiya (1988): The asymptotic normal distribution of estimators in factor analysis under general conditions, *The Annals of Statistics*, **16:2**, 759–771.

[5] Anderson, T.W. and H. Rubin (1956): Statistical inference in factor analysis, In *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability: Contributions to The Theory of Statistics*, University of California Press.

[6] Arellano, M. (2003): *Panel Data Econometrics*, Oxford University Press.

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

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

[9] Bai, J. (2009b): Likelihood approach to small T dynamic panel models with interactive effects. *manuscript*, Columbia University.

[10] Bai, J. and K. Li (2012): Statistical analysis of factor models of high dimension. *Annals of Statistics*, **40(1)**, 436–465.

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

[12] Baltagi, B.H. (2005): *Econometric Analysis of Panel Data*. Chichester: Wiley.

[13] Breitung, J. and J. Tenhofen (2011): GLS estimation of dynamic factor models, *Journal of the American Statistical Association*, **106(3)**, 1150–1166.

[14] Chamberlain, G. (1984): Panel data, in Z. Griliches and M. Intriligator (eds.), *in Handbook of Econometrics*, Vol.2, 1247–1318, Amsterdam: North-Holland.

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

[16] Doz, C., D. Giannone and L. Reichlin (2012): A qausi-maximum likelihood approach for large approximate dynamic factor models, *Review of Economics and Statistics*, **94(4)**, 1014–1024.

[17] Holtz-Eakin, D., W., Newey and H.S. Rosen (1988): Estimating vector autoregressions with panel data, *Econometrica*, **56(6)**, 1371–1395.

[18] Hsiao, C. (2003): Analysis of Panel Data, *NewYork: Cambridge University Press.*

[19] Jenrich, R.I. (1969) Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics*, **40(2)**, 633–643.

[20] Kneip, A., R.C., Sickles and W. Song (2012): A new panel data treatment for heterogeneity in time trends, *Econometric theory*, **28(3)**, 590–628.

[21] Lawley D. N. and A. E. Maxwell (1971): *Factor Analysis as a Statistical Method*, New York: American Elsevier Publishing Company.

[22] Meng X.L. and D.B. Rubin (1993): Maximum likelihood estimation via the ECM algorithm: A general framework, *Biometrika*, **80(2)**, 267–278.

[23] Moon, H. and M., Weidner (2009): Likelihood expansion for panel regression models with factors, *manuscipt, University of Southern California.*

[24] Mundlak, Y. (1978): On the pooling of time series and cross section data, *Econo-*

metrica, **46(1)**, 69–85.
[25] Newey, W. and D. McFadden (1994): Large Sample Estimation and Hypothesis Testing, in Engle, R.F. and D. McFadden (eds.) *Handbook of Econometrics,* North Holland.
[26] Neyman, J. and E.L. Scott (1948): Consistent estimates based on partially consistent observations. *Econometrica*, **16(1)**, 1-32.
[27] Pesaran, M.H. (2006): Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica*, **74(4)**, 967–1012.
[28] Ross, S.A. (1976): The arbitrage theory of capital asset pricing, *Journal of Economic Theory*,**13(3)**, 341–360.
[29] Rubin, D.B. and D.T. Thayer (1982): EM algorithms for ML factor analysis, *Psychometrika*,**47(1)**, 69–76.
[30] Stock, J.H. and M.W. Watson (2002): Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association* , **97**, 1167–1179.
[31] Su, L., S., Jin and Y., Zhang (2012): Specification Test for Panel Data Models with Interactive Fixed Effects, *manuscript*, Singapore Management University.
[32] Tucker, L.R. (1958): An inter-battery method of factor analysis, *Psychometrika*, **23(2)**, 111–136.
[33] Wu, C.F.J. (1983): On the convergence properties of the EM algorithm, *The Annals of Statistics*, 95–103.

Jushan Bai
Department of Economics
Columbia University
420 West 118th Street
New York, NY 10027
E-mail:Jushan.bai@columbia.edu

Kunpeng Li
Department of Economics
School of Economics and Management
Tsinghua University
Beijing, China, 100084
E-mail:likp.07@sem.tsinghua.edu.cn

SUPPLEMENT TO "THEORY AND METHODS OF PANEL DATA MODELS WITH INTERACTIVE EFFECTS"

This supplement provides the detailed proofs for the propositions and theorems in the main text.

We first introduce the symbols to be used in this supplement.

Table 1: The symbols used in the supplement

H	$= (\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma)^{-1}$	$\hat{H}$	$= (\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}$
		$\hat{H}_N$	$= N \cdot \hat{H}$
G	$= (M_{ff}^{-1} + \Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma)^{-1}$	$\hat{G}$	$= (\hat{M}_{ff}^{-1} + \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}$
		$\hat{G}_N$	$= N \cdot \hat{G}$
A	$= (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$		
$\dot{l}_{it}$	$= (\dot{x}_{it}', 0_{K\times K})'$		
ω_{pq}	$= N^{-1}\sum_{i=1}^N \gamma_{ip}\Sigma_{iie}^{-1}\gamma_{iq}'$		
υ_p	$= N^{-1}\sum_{i=1}^N \lambda_i\Sigma_{iie}^{-1}\gamma_{ip}'$	$\hat{\upsilon}_p$	$= N^{-1}\sum_{i=1}^N \lambda_i\hat{\Sigma}_{iie}^{-1}\gamma_{ip}'$
ξ_t	$= N^{-1}\sum_{i=1}^N \Gamma_i\Sigma_{ii}^{-1}\dot{l}_{it}$	$\hat{\xi}_t$	$= N^{-1}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\dot{l}_{it}$
χ_t	$= N^{-1}\sum_{i=1}^N \Gamma_i\Sigma_{ii}^{-1}\varepsilon_{it}$	$\hat{\chi}_t$	$= N^{-1}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{it}$

From $(A+B)^{-1} = A^{-1} - A^{-1}B(A+B)^{-1}$, we have $\hat{H} = \hat{G}(I - \hat{M}_{ff}^{-1}\hat{G})^{-1}$. From $\Sigma_{zz} = \Gamma M_{ff}\Gamma' + \Sigma_{\varepsilon\varepsilon}$, we have

$$\Sigma_{zz}^{-1} = \Sigma_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1}\Gamma(M_{ff}^{-1} + \Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma)^{-1}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1} = \Sigma_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1}\Gamma G\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}, \tag{S.1}$$

and

$$\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1} = \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}(\hat{M}_{ff}^{-1} + \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} = \hat{M}_{ff}^{-1}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}. \tag{S.2}$$

APPENDIX A: TECHNICAL MATERIALS FOR SECTION 2

APPENDIX A.1: PROOF OF PROPOSITION 2.1

The following six lemmas are useful for the proof of Proposition 2.1. The proofs of these six lemmas do not involve the special structure of the factor loadings and the identification conditions, they still hold in the context of Sections 3 and 4.

LEMMA A.1. *Let Y and Z be two real symmetric matrices, which have the orthogonal reduction*

$$P'YP = D_Y \qquad Q'ZQ = D_Z$$

where D_Y (D_Z) is a diagonal matrix with its diagonal elements arranged in an increasing (descending) order. Then we have $\mathrm{tr}(YZ) \geq \mathrm{tr}(D_Y D_Z)$.

LEMMA A.2. *Let Q be an $r \times r$ matrix satisfying*

$$
\begin{aligned}
QQ' &= I_r \\
Q'VQ &= D
\end{aligned}
$$

where V is an $r \times r$ diagonal matrix with strictly positive and distinct elements, arranged in decreasing order, and D is also diagonal. Then Q must be a diagonal matrix with elements either -1 or 1 and $V = D$.

Lemma A.1 is given in Theobald (1975) and Lemma A.2 is given in [10].

To prove Proposition 2.1, we use a superscript "*" to denote the true parameters, for example $\Gamma^*, \Sigma_{\varepsilon\varepsilon}^*$, etc. The variables without the superscript "*" denote the function arguments (input variables) in the likelihood function.

LEMMA A.3. *Let $\theta = (\beta, \Gamma, \Sigma_{\varepsilon\varepsilon})$ and Θ be the parameter set such that Assumption D is satisfied and β is in a compact set. Then we have*

$$
(a) \quad \sup_{\theta\in\Theta} \frac{1}{N}\mathrm{tr}\Big[\mathcal{B}\Gamma^* \frac{1}{T}\sum_{t=1}^{T} f_t^* \varepsilon_t' \mathcal{B}' \Sigma_{zz}^{-1}\Big] = o_p(1)
$$

$$
(b) \quad \sup_{\theta\in\Theta} \frac{1}{N}\mathrm{tr}\Big[\mathcal{B}\frac{1}{T}\sum_{t=1}^{T}[\dot{\varepsilon}_t\dot{\varepsilon}_t' - \Sigma_{\varepsilon\varepsilon}^*]\mathcal{B}'\Sigma_{zz}^{-1}\Big] = o_p(1)
$$

where $\mathcal{B} = (I_N \otimes B)(I_N \otimes B^*)^{-1}$.

PROOF OF LEMMA A.3. Consider (a). Recall

$$
B = \begin{pmatrix} 1 & -\beta' \\ 0 & I_K \end{pmatrix} \quad \text{so} \quad B^{-1} = \begin{pmatrix} 1 & \beta' \\ 0 & I_K \end{pmatrix}. \quad \text{Let} \quad L = \begin{pmatrix} 0 & (\beta-\beta^*)' \\ 0 & 0 \end{pmatrix}.
$$

Then we have $\mathcal{B} = (I_N \otimes B)(I_N \otimes B^*)^{-1} = I_N \otimes (I_{K+1} - L)$. Now the left hand side of (a) is equal to

$$
\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\Gamma^* f_t^* \varepsilon_t' \Sigma_{zz}^{-1}\Big] - 2\,\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(I_N\otimes L)\Gamma^* f_t^*\varepsilon_t'\Sigma_{zz}^{-1}\Big]
$$

$$
+\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(I_N\otimes L)\Gamma^* f_t^*\varepsilon_t'(I_N\otimes L)'\Sigma_{zz}^{-1}\Big]
$$

Consider the first term of the above equation. Notice that

$$
\sum_{i=1}^{N}\Big\|H^{1/2}\Gamma_i\Sigma_{ii}^{-1/2}\Big\|^2 = \sum_{i=1}^{N}\mathrm{tr}\Big(H^{1/2}\Gamma_i\Sigma_{ii}^{-1}\Gamma_i' H^{1/2}\Big) \tag{A.1}
$$

$$= \operatorname{tr}\Big[H^{1/2}\Big(\sum_{i=1}^{N}\Gamma_i\Sigma_{ii}^{-1}\Gamma_i'\Big)H^{1/2}\Big] = \operatorname{tr}(H^{1/2}H^{-1}H^{1/2}) = r$$

The last equation is due to the definition of H. So Lemma A.2 of [10] holds in the present context. Using their arguments, we can prove $\sup_{\theta\in\Theta}\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}$ $\Gamma^* f_t^*\varepsilon_t'\Sigma_{zz}^{-1}\Big] = o_p(1)$. The second term can be proved to be $o_p(1)$ uniformly on Θ similarly as the first since we can treat $(I_N\otimes L)\Gamma^*$ as a new Γ^* because of the boundedness of β. Now consider the third term, which is equivalent to

$$\sum_{p=1}^{K}\sum_{q=1}^{K}(\beta_p-\beta_p^*)(\beta_q-\beta_q^*)\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}v_{itp}\gamma_{iq}^{*\prime}f_t^*\Big)$$

$$-\sum_{p=1}^{K}\sum_{q=1}^{K}(\beta_p-\beta_p^*)(\beta_q-\beta_q^*)\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}v_{itp}\Sigma_{iie}^{-1}\lambda_i'G\lambda_j\Sigma_{jje}^{-1}\gamma_{jq}^{*\prime}f_t^*\Big) = a_1+a_2,$$

Consider a_1. By the Cauchy-Schwarz inequality, the term $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}$ $v_{itp}\gamma_{iq}^{*\prime}f_t^*$ is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Sigma_{iie}^{-1}\|^2\cdot\|\gamma_{iq}^*\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}v_{itp}f_t^*\Big\|^2\Big)^{1/2}$$

The first factor is $O_p(1)$ due to Assumptions C and D. The second factor is $O_p(T^{-1/2})$. So the above expression is $O_p(T^{-1/2})$. Noticing K is a finite number and $|\beta_p-\beta_p^*|$ is bounded, we have $a_1 = O_p(T^{-1/2})$.

Consider a_2. We will show

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}v_{itp}\Sigma_{iie}^{-1}\lambda_i'G\lambda_j\Sigma_{jje}^{-1}\gamma_{jq}^{*\prime}f_t^* = O_p(T^{-1/2}) \tag{A.2}$$

By the Cauchy-Schwarz inequality, the left hand side above is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}v_{itp}f_t^*\Big\|^2\Big)^{1/2}\Big(\sum_{i=1}^{N}\Big\|\Sigma_{iie}^{-1}\lambda_i'H^{1/2}\Big\|^2\Big)\Big(\frac{1}{N}\sum_{j=1}^{N}\|\gamma_{jq}^*\|^2\Big)^{1/2}\|\mathcal{U}\|$$

with $\mathcal{U} = (I_r + H^{1/2}M_{ff}^{-1}H^{1/2})^{-1}$. The first factor is $O_p(T^{-1/2})$, the third is $O(1)$ and $\|\mathcal{U}\|\le 1$. Consider the term $\sum_{i=1}^{N}\|\Sigma_{iie}^{-1}\lambda_i'H^{1/2}\|^2$, which, due to the boundedness of Σ_{iie} by Assumption D, is bounded by $C\sum_{i=1}^{N}\|\Sigma_{iie}^{-1/2}\lambda_i'H^{1/2}\|^2$, which is further bounded by $C\sum_{i=1}^{N}\|\Sigma_{ii}^{-1/2}\Gamma_i'H^{1/2}\|^2 = Cr$ by (A.1). Given these results, we prove (A.2). Furthermore $|\beta_p - \beta_p^*|$ is bounded, so we have $a_2 = O_p(T^{-1/2})$. Then (a) follows.

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

$$\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon}^*)\Sigma_{zz}^{-1}\Big] - \mathrm{tr}\Big[\frac{1}{N}\bar\varepsilon\bar\varepsilon'\Sigma_{zz}^{-1}\Big] - 2\ \mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(I_N\otimes L)(\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon}^*)\Sigma_{zz}^{-1}\Big]$$

$$+2\ \mathrm{tr}\Big[\frac{1}{N}(I_N\otimes L)\bar\varepsilon\bar\varepsilon'\Sigma_{zz}^{-1}\Big] + \mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(I_N\otimes L)(\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon}^*)(I_N\otimes L')\Sigma_{zz}^{-1}\Big]$$

$$-\mathrm{tr}\Big[\frac{1}{N}(I_N\otimes L)\bar\varepsilon\bar\varepsilon'(I_N\otimes L')\Sigma_{zz}^{-1}\Big] = b_1 - b_2 - 2b_3 + 2b_4 + b_5 - b_6 \tag{A.3}$$

where L is defined in result (a). The term b_1 is equal to

$$\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon}^*)\Sigma_{\varepsilon\varepsilon}^{-1}\Big] - \mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}(\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon}^*)\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma G\Big]$$

The first term is $\mathrm{tr}[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it}\varepsilon_{it}' - \Sigma_{ii}^*)\Sigma_{ii}^{-1}]$, which is bounded in norm by (ignore the trace)

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Sigma_{ii}^{-1}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_{it}\varepsilon_{it}' - \Sigma_{ii}^*)\Big\|^2\Big)^{1/2} = O_p(T^{-1/2})$$

uniformly on Θ. The second term is equal to the trace of

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}H^{1/2}\Gamma_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{ij,t}\Sigma_{jj}^{-1}\Gamma_j'H^{1/2}\mathcal{U}$$

with $\epsilon_{ij,t} = \varepsilon_{it}\varepsilon_{jt} - E(\varepsilon_{it}\varepsilon_{jt}')$ and $\mathcal{U} = (I_r + H^{1/2}M_{ff}^{-1}H^{1/2})^{-1}$. The above expression is bounded in norm by

$$\Big(\sum_{i=1}^{N}\|\Sigma_{jj}^{-1}\Gamma_j'H^{1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{ij,t}\Big\|^2\Big)^{1/2}\|\mathcal{U}\|$$

The term $\sum_{i=1}^{N}\|\Sigma_{jj}^{-1}\Gamma_j'H^{1/2}\|^2$ is bounded by $C\sum_{i=1}^{N}\|\Sigma_{jj}^{-1/2}\Gamma_j'H^{1/2}\|^2 = Cr$ by (A.1). So the above expression is $O_p(T^{-1/2})$ uniformly on Θ. This shows that b_1 is $O_p(T^{-1/2})$ uniformly on Θ.

Consider b_2. By $\Sigma_{zz}^{-1} \le \Sigma_{\varepsilon\varepsilon}^{-1}$, b_2 is bounded by $\mathrm{tr}[\frac{1}{N}\bar\varepsilon\bar\varepsilon'\Sigma_{\varepsilon\varepsilon}^{-1}]$, which is equivalent to

$$\frac{1}{T}\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_{it}'\Big)\Sigma_{ii}^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\varepsilon_{is}\Big)$$

which is $O_p(T^{-1})$ uniformly on Θ due to the boundedness of Σ_{ii}. So b_2 is $O_p(T^{-1})$ uniformly on Θ.

Consider b_3, which is equal to

$$\sum_{p=1}^{K}(\beta_p-\beta_p^*)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}$$

$$-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[L\varepsilon_{it}\varepsilon_{jt}'-E(L\varepsilon_{it}\varepsilon_{jt}')]\Sigma_{jj}^{-1}\Gamma_j'G\Big]$$

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

$$\sum_{p=1}^{K}|\beta_p-\beta_p^*|\Big(\frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-2}\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}e_{it}v_{itp}\Big|^2\Big)^{1/2}$$

which is $O_p(T^{-1/2})$ due to the boundedness of Σ_{iie} and β. The term inside the trace operator of the second term can be written alternatively as

$$\sum_{p=1}^{K}(\beta_p-\beta_p^*)\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}H^{1/2}\Gamma_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\mathcal{L}_{ijtp}-E(\mathcal{L}_{ijtp})]\Sigma_{jj}^{-1}\Gamma_j'H^{1/2}\mathcal{U}$$

with

$$\mathcal{L}_{ijtp}=\begin{pmatrix} v_{itp}e_{jt} & v_{itp}v_{jt1} & \cdots & v_{itp}v_{jtK}\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 0\end{pmatrix}$$

which is bounded in norm by

$$\sum_{p=1}^{K}|\beta_p-\beta_p^*|\Big(\sum_{i=1}^{N}\|\Sigma_{jj}^{-1}\Gamma_j'H^{1/2}\|^2\Big)\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\mathcal{L}_{ijtp}-E(\mathcal{L}_{ijtp})]\Big\|^2\Big]^{1/2}\|\mathcal{U}\|$$

with $\mathcal{U}=(I_r+H^{1/2}M_{ff}^{-1}H^{1/2})^{-1}$. The term $\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}[\mathcal{L}_{ijtp}-E(\mathcal{L}_{ijtp})]\|^2$ is $O_p(T^{-1})$. From this, the above term is $O_p(T^{-1/2})$ uniformly on Θ. This gives $b_3=O_p(T^{-1/2})$ uniformly on Θ.

Consider b_4, which is equal to

$$\frac{1}{T}\sum_{p=1}^{K}(\beta_p-\beta_p^*)\Big[\frac{1}{N}\sum_{j=1}^{N}\Sigma_{jje}^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_{jt}\Big)\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T}v_{jtp}\Big)\Big]$$

$$-\mathrm{tr}\Big[\sum_{p=1}^{K}(\beta_p-\beta_p^*)\frac{1}{N}\sum_{i=1}^{N}\lambda_i'\Sigma_{iie}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}v_{itp}\Big)\Big(\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{jt}\Big)'\Sigma_{jj}^{-1}\Gamma_j'\Big)G\Big]$$

The first term is $O_p(T^{-1})$ uniformly on Θ. The second term inside the trace operator is bounded in norm by

$$\sum_{p=1}^{K}|\beta_p-\beta_p^*|\Big(\sum_{i=1}^{N}\|H^{1/2}\Gamma_i\Sigma_{ii}^{-1}\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}v_{itp}\Big|^2\Big)^{1/2}$$

$$\times\Big(\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{jt}\|^2\Big)^{1/2}\|\mathcal{U}\|$$

with $\mathcal{U}=(I_r+H^{1/2}M_{ff}^{-1}H^{1/2})^{-1}$, which is $O_p(T^{-1})$ uniformly on Θ. So b_4 is $O_p(T^{-1})$ uniformly on Θ.

Consider b_5, which is equal to

$$\sum_{p=1}^{K}\sum_{q=1}^{K}(\beta_p-\beta_p^*)(\beta_q-\beta_q^*)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}[v_{itp}v_{itq}-E(v_{itp}v_{itq})]$$

$$-\sum_{p=1}^{K}\sum_{q=1}^{K}(\beta_p-\beta_p^*)(\beta_q-\beta_q^*)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\Sigma_{iie}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{itp}v_{jtq}-E(v_{itp}v_{jtq})]\Sigma_{jje}^{-1}\lambda_j'G\Big]$$

The first term is bounded in norm by

$$\sum_{p=1}^{K}\sum_{q=1}^{K}\Big\{|\beta_p-\beta_p^*|\cdot|\beta_q-\beta_q^*|\Big(\frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-2}\Big)^{1/2}$$

$$\times\Big(\frac{1}{N}\sum_{i=1}^{T}\Big(\frac{1}{T}\sum_{t=1}^{T}[v_{itp}v_{itq}-E(v_{itp}v_{itq})]\Big)^2\Big)^{1/2}\Big\}$$

which is $O_p(T^{-1/2})$ by the boundedness of β. For the second term, the trace is equal to the trace of

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}H^{1/2}\lambda_i\Sigma_{iie}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{itp}v_{jtq}-E(v_{itp}v_{jtq})]\Sigma_{jje}^{-1}\lambda_j'H^{1/2}\mathcal{U}$$

which is bounded in norm by

$$\Big(\sum_{j=1}^{N}\|\Sigma_{jje}^{-1}\lambda_j'H^{1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}[v_{itp}v_{jtq}-E(v_{itp}v_{jtq})]\Big)^2\Big)^{1/2}\|\mathcal{U}\|$$

where $\mathcal{U}=(I_r+H^{1/2}M_{ff}^{-1}H^{1/2})^{-1}$. The above is $O_p(T^{-1/2})$ uniformly on Θ. So the second term is $O_p(T^{-1/2})$ because β is bounded. Thus $b_5=O_p(T^{-1/2})$ uniformly on Θ.

Consider b_6. Since $\Sigma_{zz}^{-1} \leq \Sigma_{\varepsilon\varepsilon}^{-1}$, we have

$$b_6 \leq \operatorname{tr}\Big[\frac{1}{N}(I_N \otimes L)\bar{\varepsilon}\bar{\varepsilon}'(I_N \otimes L')\Sigma_{\varepsilon\varepsilon}^{-1}\Big] = \frac{1}{N}\sum_{j=1}^{N}\Sigma_{jje}^{-1}\Big(\sum_{p=1}^{K}(\beta_p - \beta_p^*)\frac{1}{T}\sum_{t=1}^{T} v_{jtp}\Big)^2$$

which is $O_p(T^{-1})$ uniformly on Θ.

Summarizing all the results, we obtain (b). □

LEMMA A.4. *Under Assumptions A-D, if* $\|\hat{\beta} - \beta^*\| = o_p(1)$*, then*

$$\frac{1}{N}\operatorname{tr}\Big[\hat{\mathcal{B}}\Gamma^* M_{ff}^*\Gamma^{*\prime}\hat{\mathcal{B}}'\hat{\Sigma}_{zz}^{-1}\Big] = \frac{1}{N}\operatorname{tr}\Big[\Gamma^* M_{ff}^*\Gamma^{*\prime}\hat{\Sigma}_{zz}^{-1}\Big] + o_p(1)$$

where $\hat{\mathcal{B}} = (I_N \otimes \hat{B})(I_N \otimes B^*)^{-1} = I_N \otimes (\hat{B}B^{*-1})$.

PROOF OF LEMMA A.4. Using the notations in result (a) of Lemma A.3, the left hand side is equal to

$$\frac{1}{N}\operatorname{tr}\Big[\Gamma^* M_{ff}^*\Gamma^{*\prime}\hat{\Sigma}_{zz}^{-1}\Big] - 2\operatorname{tr}\Big[(I_N \otimes \hat{L})\Gamma^* M_{ff}^*\Gamma^{*\prime}\hat{\Sigma}_{zz}^{-1}\Big]$$
$$+\operatorname{tr}\Big[(I_N \otimes \hat{L})\Gamma^* M_{ff}^*\Gamma^{*\prime}(I_N \otimes \hat{L}')\hat{\Sigma}_{zz}^{-1}\Big]$$

Given $\|\hat{\beta} - \beta^*\| = o_p(1)$, it suffices to prove

$$-2\operatorname{tr}\Big[(I_N\otimes\hat{L})\Gamma^* M_{ff}^*\Gamma^{*\prime}\hat{\Sigma}_{zz}^{-1}\Big]+\operatorname{tr}\Big[(I_N\otimes\hat{L})\Gamma^* M_{ff}^*\Gamma^{*\prime}(I_N\otimes\hat{L}')\hat{\Sigma}_{zz}^{-1}\Big] = O_p(\|\hat{\beta}-\beta^*\|)$$

We use c_1 and c_2 to denote the two terms of the above expression. Consider c_1, which is equal to (omitting -2)

$$\operatorname{tr}\Big[\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)\frac{1}{N}\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}\lambda_j^*\gamma_{jp}^{*\prime}M_{ff}^*\Big]$$

$$-\operatorname{tr}\Big[\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i\gamma_{ip}^{*\prime}M_{ff}^*\sum_{j=1}^{N}\Gamma_j^*\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{G}\Big]$$

The first term of the above equation is $O_p(\|\hat{\beta} - \beta\|)$ by the boundedness of $\lambda_j^*, \gamma_{jp}^*, \hat{\Sigma}_{jje}$. The second term is equivalent to

$$\operatorname{tr}\Big[\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)\frac{1}{N}\sum_{i=1}^{N}\hat{H}^{1/2}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i\gamma_{ip}^{*\prime}M_{ff}^*\sum_{j=1}^{N}\Gamma_j^*\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}^{1/2}\mathcal{U}\Big]$$

with $\mathcal{U} = (I + \hat{H}^{1/2}\hat{M}_{ff}^{-1}\hat{H}^{1/2})^{-1}$. Ignore the trace, the term is bounded in norm by

$$\sum_{p=1}^{K}\Big\{|\hat{\beta}_p - \beta_p^*|\Big(\sum_{i=1}^{N}\Big\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Big\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\|\gamma_{ip}^*\|^2\Big)^{1/2}$$

$$\times\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Gamma_j^*\|^2\Big)^{1/2}\|\mathcal{U}\|\cdot\|M_{ff}^*\|\Big\}$$

where we use the fact that $\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\| \leq \|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\|$. The above expression is $O_p(\|\hat{\beta} - \beta^*\|)$. So we have $c_1 = O_p(\|\hat{\beta} - \beta^*\|)$.

Consider c_2. Since $\hat{\Sigma}_{zz}^{-1} \leq \hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$, we only need to explore the term $\operatorname{tr}[(I_N \otimes \hat{L})\Gamma^* M_{ff}^*\Gamma^{*\prime}(I_N \otimes \hat{L}')\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}]$, which is equivalent to

$$\operatorname{tr}\Big[\sum_{p=1}^{K}\sum_{q=1}^{K}(\hat{\beta}_p - \beta_p^*)(\hat{\beta}_q - \beta_q^*)\frac{1}{N}\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^*\gamma_{jq}^{*\prime}M_{ff}^*\Big].$$

The above is $O_p(\|\hat{\beta} - \beta^*\|^2)$ due to the boundedness of $\gamma_{jp}^*, M_{ff}^*, \Sigma_{jje}$. So we have $c_2 = O_p(\|\hat{\beta} - \beta^*\|^2)$.

Combining the results on c_1 and c_2, we have Lemma A.4. □

Lemma A.5. *Under Assumptions A-D,*

$$(a)\quad H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t^{*\prime}\Big](I - A) = \|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(T^{-1/2})$$

$$(b)\quad \hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\epsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(T^{-1/2})$$

$$(c)\quad \hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii}^*)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}$$

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

where $\epsilon_{ij,t} = \varepsilon_{it}\varepsilon_{jt}' - E(\varepsilon_{it}\varepsilon_{jt}')$. *The symbols of* $A, \hat{H}_N$ *and* $\hat{\chi}_t$ *are defined in Table 1.*

Proof of Lemma A.5. Consider (a). By the definitions of $A, \hat{H}_N$ and $\hat{\chi}_t$, the left hand side of (a) is equivalent to

$$\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{it}f_t^{*\prime}\sum_{j=1}^{N}\Gamma_j^*\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}$$

By the Cauchy-Schwarz inequality, term $\sum_{j=1}^N \Gamma_j^* \hat{\Sigma}_{jj}^{-1} \hat{\Gamma}_j' \hat{H}$ is bounded in norm by

$$\Big(\frac{1}{N}\sum_{j=1}^N \|\Gamma_j^* \hat{\Sigma}_{jj}^{-1/2}\|^2\Big)^{1/2} \Big(\sum_{j=1}^N \|\hat{\Sigma}_{jj}^{-1/2} \hat{\Gamma}_j' \hat{H}^{1/2}\|^2\Big)^{1/2} \cdot \|N^{1/2}\hat{H}^{1/2}\|$$

The first factor $\frac{1}{N}\sum_{j=1}^N \|\Gamma_j^* \hat{\Sigma}_{jj}^{-1/2}\|^2 = O_p(1)$ by the boundedness of $\hat{\Sigma}_{jj}$. The second factor is $\sqrt{r}$ by (A.1). So we have $\sum_{j=1}^N \Gamma_j^* \hat{\Sigma}_{jj}^{-1} \hat{\Gamma}_j' \hat{H} = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(1)$.

Also, the term $\hat{H}\frac{1}{T}\sum_{i=1}^N \sum_{t=1}^T \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \varepsilon_{it} f_t^{*\prime}$ is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\| \cdot \Big(\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2} \Big(\frac{1}{N}\sum_{i=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T f_t^* \varepsilon_{it}'\Big\|^2\Big)^{1/2}$$

which is $\|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1/2})$. Combining results, we have (a).

Consider (b). The left hand side of (b), by the Cauchy-Schwarz inequality, is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \Big(\sum_{j=1}^N \|\hat{H}^{1/2}\hat{\Gamma}_j \hat{\Sigma}_{jj}^{-1/2}\|^2\Big) \Big(\frac{1}{N^2}\sum_{i=1}^N \sum_{j=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T \epsilon_{ij,t}\Big\|^2\Big)^{1/2}$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$. So (b) follows.

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

$$\|\hat{H}^{1/2}\|^2 \sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|$$

$$\leq C\|\hat{H}^{1/2}\|^2 \sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|$$

By $\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\|^2 = r$, we have, for all i, $\|\hat{H}^{1/2}\hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\| \leq \sqrt{r}$. So the above equation is bounded by $C\sqrt{r}\|\hat{H}^{1/2}\|^2 \sum_{i=1}^N \|\hat{H}^{1/2}\hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\| \cdot \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|$. By the Cauchy-Schwarz inequality, This term is bounded by

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

which is $\|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(N^{-1/2}[\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2})$. So (c) follows. □

LEMMA A.6. *Let $\mathcal{E} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(\|\hat{\beta} - \beta^*\|)$. Under Assumptions A-D,*

$$(a)\quad (I - A)'\Big[\frac{1}{T}\sum_{t=1}^{T} f_t^*(\hat{\beta} - \beta^*)\hat{\xi}_t'\Big]\hat{H}_N = \mathcal{E}$$

$$(b)\quad \hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T} \hat{\chi}_t(\hat{\beta} - \beta^*)'\hat{\xi}_t'\Big]\hat{H}_N = \mathcal{E} \cdot o_p(1)$$

$$(c)\quad \hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T} \hat{\xi}_t(\hat{\beta} - \beta^*)(\hat{\beta} - \beta^*)'\hat{\xi}_t'\Big]\hat{H}_N = \mathcal{E} \cdot O_p(\|\hat{\beta} - \beta^*\|)$$

where $\hat{\chi}_t, \hat{\xi}_t, \hat{H}_N$ and A are defined in Table 1.

PROOF OF LEMMA A.6. Consider (a). By the definitions of $\hat{\xi}_t$ and A , the left hand side of (a) is equivalent to

$$\hat{H}\Big(\sum_{i=1}^{N} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\Big)\Big(\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T} f_t^*(\hat{\beta} - \beta^*)'\dot{l}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\Big)$$

The term $\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}$ has already been proved to be $\|N^{1/2}\hat{H}^{1/2}\|\cdot O_p(1)$ in Lemma A.5(a). By the definition of $\dot{l}_{jt}$ and the fact $\dot{x}_{jtp} = \gamma_{ip}^{*\prime} f_t^* + \dot{v}_{itp}$,

$$\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T} f_t^*(\hat{\beta} - \beta^*)'\dot{l}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H} = \sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)M_{ff}^*\sum_{j=1}^{N}\gamma_{jp}^*\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}$$

$$+\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T} f_t^* v_{jtp}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H} = d_1 + d_2$$

where $\dot{v}_{jtp}$ can be replaced by v_{jtp} since $\frac{1}{T}\sum_{t=1}^{T} f_t^* = 0$. Term d_1 is bounded in norm by

$$\|M_{ff}^*\| \cdot \sum_{p=1}^{K}|\hat{\beta}_p - \beta_p^*| \cdot \Big(\frac{1}{N}\sum_{j=1}^{N}\|\gamma_{jp}^*\hat{\Sigma}_{jje}^{-1/2}\|^2\Big)^{1/2}$$

$$\times\Big(\sum_{j=1}^{N}\|\hat{\Sigma}_{jje}^{-1/2}\hat{\lambda}_j'\hat{H}^{1/2}\|^2\Big)^{1/2}\|N^{1/2}\hat{H}^{1/2}\|$$

Notice

$$\sum_{j=1}^{N}\|\hat{\Sigma}_{jje}^{-1/2}\hat{\lambda}_j'\hat{H}^{1/2}\|^2 = \sum_{j=1}^{N}\mathrm{tr}[\hat{H}^{1/2}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}^{1/2}] = \mathrm{tr}\Big[\Big(\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\Big)\hat{H}\Big]$$

$$= \mathrm{tr}\Big[(\sum_{j=1}^{N} \hat{\lambda}_j \hat{\Sigma}_{jje}^{-1} \hat{\lambda}_j') (\sum_{j=1}^{N} \hat{\lambda}_j \hat{\Sigma}_{jje}^{-1} \hat{\lambda}_j' + \sum_{j=1}^{N} \hat{\gamma}_{jx} \hat{\Sigma}_{jjx}^{-1} \hat{\gamma}_{jx}')^{-1} \Big] < r. \tag{A.4}$$

But $\frac{1}{N}\sum_{j=1}^{N} \|\gamma_{jp}^* \hat{\Sigma}_{jje}^{-1/2}\|^2 = O_p(1)$ by the boundedness of $\hat{\Sigma}_{jje}$, we have $d_1 = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(\hat{\beta} - \beta)$. Consider d_2. Ignore $\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)$, the remaining expression is bounded in norm by

$$C\Big(\sum_{j=1}^{N} \|\hat{\Sigma}_{jje}^{-1/2} \hat{\lambda}_j' \hat{H}^{1/2}\|\Big)^{1/2} \Big(\frac{1}{N}\sum_{j=1}^{N} \Big\| \frac{1}{T}\sum_{t=1}^{N} f_t^* v_{jtp} \Big\|^2\Big)^{1/2} \|N^{1/2}\hat{H}^{1/2}\|$$

which is $\|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1/2})$. So d_2 is bounded in norm by

$$\sum_{p=1}^{K} |\hat{\beta}_p - \beta_p^*| \cdot \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1/2}) = \|N^{1/2}\hat{H}^{1/2}\| \cdot o_p(\|\hat{\beta} - \beta^*\|)$$

Given the results on d_1 and d_2, (a) follows.

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

$$\hat{H}\frac{1}{T}\sum_{t=1}^{T} \Big(\sum_{i=1}^{N} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \varepsilon_{it}\Big)(\hat{\beta} - \beta^*)' \Big(\sum_{j=1}^{N} \dot{l}_{jt}' \hat{\Sigma}_{jj}^{-1} \hat{\Gamma}_j'\Big)\hat{H}$$

by the same arguments in (a), the above expression is equal to

$$\sum_{p=1}^{K} (\hat{\beta}_p - \beta_p^*) \hat{H} \frac{1}{T} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \varepsilon_{it} f_t^{*\prime} \sum_{j=1}^{N} \gamma_{jp}^* \hat{\Sigma}_{jje}^{-1} \hat{\lambda}_j' \hat{H}$$

$$+ \sum_{p=1}^{K} (\hat{\beta}_p - \beta_p^*) \hat{H} \frac{1}{T} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \varepsilon_{it} v_{jtp} \hat{\Sigma}_{jje}^{-1} \hat{\lambda}_j' \hat{H} = d_3 + d_4, \quad \text{say}$$

where we neglect term

$$\sum_{p=1}^{K} (\hat{\beta}_p - \beta_p^*) \hat{H} \frac{1}{T} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \varepsilon_{it} \Big(\sum_{j=1}^{N} \bar{v}_{jp} \hat{\Sigma}_{jje}^{-1} \hat{\lambda}_j' \hat{H}\Big)$$

since it is of smaller order than d_4. Consider d_3. Ignore $\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p^*)$, the remaining expression is bounded in norm by

$$C\Big(\sum_{i=1}^{N} \|\hat{H}^{1/2} \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2} \Big(\sum_{j=1}^{N} \|\hat{\Sigma}_{jje}^{-1/2} \hat{\lambda}_j' \hat{H}\|^2\Big)^{1/2}$$

$$\times\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T f_t^*\varepsilon_{it}\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^N\|\gamma_{jp}^*\hat{\Sigma}_{jje}^{-1/2}\|^2\Big)^{1/2}\|N^{1/2}\hat{H}^{1/2}\|^2$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(1)$. Given this result, it follows $d_3=\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(\|\hat{\beta}-\beta^*\|)$. Consider d_4. Ignore $\sum_{p=1}^K(\hat{\beta}_p-\beta_p^*)$, the remaining expression is equal to

$$\hat{H}\sum_{i=1}^N\sum_{j=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}v_{jtp}-E(\varepsilon_{it}v_{itp})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}+\hat{H}\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}E(\varepsilon_{it}v_{itp})\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{H}$$

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

$$C\Big(\sum_{i=1}^N\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\sum_{j=1}^N\|\hat{\Sigma}_{jje}^{-1/2}\hat{\lambda}_j'\hat{H}\|^2\Big)^{1/2}$$

$$\times\Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}v_{jtp}-E(\varepsilon_{it}v_{itp})]\Big\|^2\Big)^{1/2}\|N^{1/2}\hat{H}^{1/2}\|^2$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(T^{-1/2})$. The second term is bounded in norm by

$$C\frac{1}{N}\Big(\sup_{i\le N}\|E(\varepsilon_{it}v_{itp})\|\Big)\Big(\sum_{i=1}^N\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}$$

$$\times\Big(\sum_{j=1}^N\|\hat{\Sigma}_{jje}^{-1/2}\hat{\lambda}_j'\hat{H}^{1/2}\|^2\Big)^{1/2}\|N^{1/2}\hat{H}^{1/2}\|^2$$

Notice $\sup_{i\le N}\|E(\varepsilon_{it}v_{itp})\|\le[\sup_{i\le N}(E\|\varepsilon_{it}\|^2)^{1/2}][\sup_{i\le N}(Ev_{jtp}^2)^{1/2}]$, which is bounded by Assumption B. So the above expression is $\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(N^{-1})$. Thus,

$$\hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{it}v_{jtp}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}=\|N^{1/2}\hat{H}^{1/2}\|^2\cdot[O_p(N^{-1})+O_p(T^{-1/2})]$$

Given this result, it follows $d_4=\|N^{1/2}\hat{H}^{1/2}\|^2\cdot o_p(\|\hat{\beta}-\beta^*\|)$. Combining the results on d_3 and d_4, we have (b).

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

$$\hat{H}\frac{1}{T}\sum_{t=1}^T\Big(\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\dot{l}_{it}\Big)(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\Big(\sum_{j=1}^N\dot{l}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\Big)\hat{H}$$

which is equal to

$$\sum_{p=1}^{K}\sum_{q=1}^{K}(\hat{\beta}_p-\beta_p^*)(\hat{\beta}_q-\beta_q^*)\Big\{\hat{H}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{*\prime}M_{ff}^*\sum_{j=1}^{N}\gamma_{jq}^*\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}$$

$$+\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}v_{itp}f_t^{*\prime}\sum_{j=1}^{N}\gamma_{jq}^*\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}$$

$$+\hat{H}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{*\prime}\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t^*v_{jtq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}$$

$$+\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{itp}v_{jtq}-E(v_{itp}v_{jtq})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}$$

$$+\hat{H}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}E(v_{itp}v_{itq})\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{H}\Big\}$$

$$=\sum_{p=1}^{K}\sum_{q=1}^{K}(\hat{\beta}_p-\beta_p^*)(\hat{\beta}_q-\beta_q^*)(d_5+d_6+\cdots+d_9),\quad \text{say}$$

where we neglect term

$$\sum_{p=1}^{K}\sum_{q=1}^{K}(\hat{\beta}_p-\beta_p^*)(\hat{\beta}_q-\beta_q^*)\Big[\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\bar{v}_{ip}\bar{v}_{jq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{H}\Big]$$

since it is of smaller order than $\sum_{p=1}^{K}\sum_{q=1}^{K}(\hat{\beta}_p-\beta_p^*)(\hat{\beta}_q-\beta_q^*)d_8$.

Consider d_5, which is bounded in norm by

$$\|M_{ff}^*\|\cdot\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\gamma_{ip}^*\hat{\Sigma}_{iie}^{-1/2}\|\Big)^{1/2}$$

$$\times\Big(\frac{1}{N}\sum_{j=1}^{N}\|\gamma_{jq}^*\hat{\Sigma}_{jje}^{-1/2}\|^2\Big)^{1/2}\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)=\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(1).$$

Consider d_6, which is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\gamma_{ip}^*\hat{\Sigma}_{iie}^{-1/2}\|\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}f_t^*v_{jtq}\|^2\Big)^{1/2}$$

$$\times\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)=\|N^{1/2}\hat{H}^{1/2}\|^2\cdot O_p(T^{-1/2}).$$

The term d_7 is also $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$, similar to d_6.

Term d_8 is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{H}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)$$

$$\times\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\big[\frac{1}{T}\sum_{t=1}^{T}v_{itp}v_{jtq}-E(v_{itp}v_{jtq})\big]^2\Big)^{1/2},$$

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

Consider d_9, which is bounded in norm by

$$C\frac{1}{N}\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)\cdot\sup_{i\le N}|E(v_{itp}v_{itq})|$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(N^{-1})$.

Summarizing all the results, we have (c). □

Proof of Proposition 2.1. We consider the objective function,

$$\ln L = -\frac{1}{N}\ln|\Sigma_{zz}| - \frac{1}{N}\mathrm{tr}\Big[(I_N\otimes B)M_{zz}(I_N\otimes B')\Sigma_{zz}^{-1}\Big] + K + \frac{1}{N}\ln|\Sigma_{zz}^*|$$

Here we add a centering constant to (2.4). Throughout the paper, we assume $\hat{\beta}$ is in a compact set. Since the likelihood function is a quadratic form of $\hat{\beta}$, it cannot achieve the maximum value at too large $\hat{\beta}$. So this assumption doesn't loss generality.

Let $\Gamma^\dagger = (I_N\otimes B)^{-1}\Gamma$ and $\Sigma_{\varepsilon\varepsilon}^\dagger = (I_N\otimes B)^{-1}\Sigma_{\varepsilon\varepsilon}(I_N\otimes B')^{-1}$, then $\Sigma_{zz}^\dagger = \Gamma^\dagger M_{ff}\Gamma^{\dagger\prime} + \Sigma_{\varepsilon\varepsilon}^\dagger = (I_N\otimes B)^{-1}\Sigma_{zz}(I_N\otimes B')^{-1}$. Notice $\ln|I_N\otimes B| = 0$ by the definition of B. So the likelihood function can be written as

$$\ln L = -\frac{1}{N}\ln|\Sigma_{zz}^\dagger| - \frac{1}{N}\mathrm{tr}\Big[M_{zz}\Sigma_{zz}^{\dagger-1}\Big] + K + \frac{1}{N}\ln|\Sigma_{zz}^*| \tag{A.5}$$

By $(I_N\otimes B^*)\dot{z}_t = \Gamma^* f_t^* + \dot{\varepsilon}_t$, we have

$$(I_N\otimes B^*)M_{zz}(I_N\otimes B^{*\prime}) = \Gamma^* M_{ff}^*\Gamma^{*\prime} + \Sigma_{\varepsilon\varepsilon}^* + \Gamma^*\frac{1}{T}\sum_{t=1}^{T}f_t^*\varepsilon_t'$$

$$+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t f_t^{*\prime}\Gamma^{*\prime} + \frac{1}{T}\sum_{t=1}^{T}[\dot{\varepsilon}_t\dot{\varepsilon}_t' - \Sigma_{\varepsilon\varepsilon}^*]$$

Let $\Gamma^{*\dagger} = (I_N \otimes B^*)^{-1}\Gamma^*$, $\Sigma_{\varepsilon\varepsilon}^{*\dagger} = (I_N \otimes B^*)^{-1}\Sigma_{\varepsilon\varepsilon}^*(I_N \otimes B^{*\prime})^{-1}$ and $\Sigma_{zz}^{*\dagger} = \Gamma^{*\dagger}M_{ff}^*\Gamma^{*\dagger\prime} + \Sigma_{\varepsilon\varepsilon}^{*\dagger}$. Then the proceeding equation can be rewritten as

$$M_{zz} = \Sigma_{zz}^{*\dagger} + \Gamma^{*\dagger}\frac{1}{T}\sum_{t=1}^T f_t^*\varepsilon_t'(I_N \otimes B^{*\prime-1}) + (I_N \otimes B^{*-1})\frac{1}{T}\sum_{t=1}^T \varepsilon_t f_t^{*\prime}\Gamma^{*\dagger\prime}$$

$$+(I_N \otimes B^{*-1})\frac{1}{T}\sum_{t=1}^T[\dot{\varepsilon}_t\dot{\varepsilon}_t' - \Sigma_{\varepsilon\varepsilon}^*](I_N \otimes B^{*\prime-1})$$

Equation (A.5) can be rewritten as

$$\ln L = \overline{L}(\theta) + R(\theta) \tag{A.6}$$

where

$$\overline{L}(\theta) = -\frac{1}{N}\ln|\Sigma_{zz}^\dagger| - \frac{1}{N}\mathrm{tr}\Big[\Sigma_{zz}^{*\dagger}\Sigma_{zz}^{\dagger-1}\Big] + K + \frac{1}{N}\ln|\Sigma_{zz}^*|$$

and

$$R(\theta) = -\frac{1}{N}\mathrm{tr}\Big[(M_{zz} - \Sigma_{zz}^{*\dagger})\Sigma_{zz}^{\dagger-1}\Big]$$

Lemma A.3 implies that $\sup_{\theta\in\Theta}|R(\theta)| = o_p(1)$. So we have $|R(\theta^*) - R(\hat{\theta})| \leq 2\sup_{\theta\in\Theta}|R(\theta)| = o_p(1)$. Since $\hat{\theta}$ maximizes $\ln L$, it follows that $\overline{L}(\hat{\theta}) + R(\hat{\theta}) \geq \overline{L}(\theta^*) + R(\theta^*)$. This yields $\overline{L}(\hat{\theta}) \geq \overline{L}(\theta^*) + R(\theta^*) - R(\hat{\theta}) \geq \overline{L}(\theta^*) - |o_p(1)| = -|o_p(1)|$, the last equation uses the fact that $\overline{L}(\theta^*) = 0$ by the centering. However, $\overline{L}(\theta)$ is maximized at $\theta = \theta^*$ and $\overline{L}(\theta^*) = 0$. This yields $\overline{L}(\hat{\theta}) \leq 0$. Combining $\overline{L}(\hat{\theta}) \geq -|o_p(1)|$ and $\overline{L}(\hat{\theta}) \leq 0$, we have $\overline{L}(\hat{\theta}) = o_p(1)$.

Consider $\overline{L}(\hat{\theta})$, which is equivalent to

$$\overline{L}(\hat{\theta}) = -\frac{1}{N}\ln\Big|\hat{\Gamma}^\dagger\hat{M}_{ff}\hat{\Gamma}^{\dagger\prime} + \hat{\Sigma}_{\varepsilon\varepsilon}^\dagger\Big| - \frac{1}{N}\mathrm{tr}\Big[\Gamma^{*\dagger}M_{ff}^*\Gamma^{*\dagger\prime}\hat{\Sigma}_{zz}^{\dagger-1}\Big]$$

$$-\frac{1}{N}\mathrm{tr}\Big[\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{zz}^{\dagger-1}\Big] + K + \frac{1}{N}\ln|\Sigma_{zz}^*|$$

Notice that $\hat{\Sigma}_{zz}^{\dagger-1} = \hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1} - \hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\hat{\Gamma}^\dagger\hat{G}^\dagger\hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}$ with $\hat{G}^\dagger = (\hat{M}_{ff}^{-1} + \hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\hat{\Gamma}^\dagger)^{-1}$. So $\frac{1}{N}\mathrm{tr}[\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{zz}^{\dagger-1}]$ is equal to $\frac{1}{N}\mathrm{tr}[\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}] - \frac{1}{N}\mathrm{tr}[\hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\hat{\Gamma}^\dagger\hat{G}^\dagger]$. The term $\frac{1}{N}\mathrm{tr}[\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}]$, by the definitions of $\Sigma_{\varepsilon\varepsilon}^{*\dagger}$ and $\hat{\Sigma}_{\varepsilon\varepsilon}^\dagger$, is equivalent to $\frac{1}{N}\sum_{i=1}^N(\hat{\beta} - \beta^*)'\Sigma_{iix}^*(\hat{\beta} - \beta^*)\hat{\Sigma}_{iie}^{-1} + \frac{1}{N}\sum_{i=1}^N\mathrm{tr}[\Sigma_{ii}^*\hat{\Sigma}_{ii}^{-1}]$. The term $\frac{1}{N}\mathrm{tr}[\hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\hat{\Gamma}^\dagger\hat{G}^\dagger]$ is equal to $\mathrm{tr}[\frac{1}{N}\sum_{i=1}^N\hat{\Gamma}_i^\dagger\hat{\Sigma}_{ii}^{\dagger-1}\Sigma_{ii}^{*\dagger}\hat{\Sigma}_{ii}^{\dagger-1}\hat{\Gamma}_i^{\dagger\prime}\hat{G}^\dagger]$, where $\hat{\Sigma}_{ii}^\dagger = \hat{B}^{-1}\hat{\Sigma}_{ii}\hat{B}^{-1\prime}$, $\Sigma_{ii}^{*\dagger} = B^{*-1}\Sigma_{ii}^*B^{*-1\prime}$ and $\hat{\Gamma}_i^\dagger = \hat{B}^{-1}\hat{\Gamma}_i$. The latter term is bounded in norm by (ignore the trace)

$$\frac{1}{N}\sup_{i\leq N}(\|\Sigma_{ii}^{*\dagger}\|\cdot\|\hat{\Sigma}_{ii}^{\dagger-1/2}\|^2)\cdot\Big(\sum_{i=1}^N\|\hat{H}^{\dagger 1/2}\hat{\Gamma}_i^\dagger\hat{\Sigma}_{ii}^{\dagger-1/2}\|^2\Big)\cdot\|I + \hat{H}^{\dagger-1/2}\hat{M}_{ff}^{-1}\hat{H}^{\dagger-1/2}\|$$

where $\hat{H}^\dagger = (\hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\hat{\Gamma}^\dagger)^{-1}$. The first factor $\sup_{i\le N}(\|\Sigma_{ii}^{*\dagger}\|\cdot\|\hat{\Sigma}_{ii}^{\dagger-1/2}\|^2)$ is bounded by Assumptions C and D. Since

$$\sum_{i=1}^N \|\hat{H}^{\dagger 1/2}\hat{\Gamma}_i^\dagger\hat{\Sigma}_{ii}^{\dagger-1/2}\|^2 = \sum_{i=1}^N \mathrm{tr}(\hat{H}^{\dagger 1/2}\hat{\Gamma}_i^\dagger\hat{\Sigma}_{ii}^{\dagger-1}\hat{\Gamma}_i^{\dagger\prime}\hat{H}^{\dagger 1/2}) = \mathrm{tr}(\hat{H}^\dagger\hat{H}^{\dagger-1}) = r,$$

the second factor is r. So we have $\frac{1}{N}\mathrm{tr}[\hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\Sigma_{\varepsilon\varepsilon}^{*\dagger}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}\hat{\Gamma}^\dagger\hat{G}^\dagger] = O_p(N^{-1})$. Using the arguments in proving their Proposition 1 of [10], we can show that $\frac{1}{N}\ln|\hat{\Gamma}^\dagger\hat{M}_{ff}\hat{\Gamma}^{\dagger\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{\dagger-1}+I_N|$ and $\frac{1}{N}\ln|\Gamma^{*\dagger}M_{ff}^*\Gamma^{*\dagger\prime}\ \Sigma_{\varepsilon\varepsilon}^{*\dagger-1}+I_N|$ are both $O_p(\frac{\ln N}{N})$. Given these results, in combination with $\ln|\hat{\Sigma}_{\varepsilon\varepsilon}^\dagger| = \ln|\hat{\Sigma}_{\varepsilon\varepsilon}|$ by $\det(I_N\otimes B)=1$, we have

$$\overline{L}(\hat{\theta}) = -\Big(\frac{1}{N}\sum_{i=1}^N\ln|\hat{\Sigma}_{ii}| + \frac{1}{N}\sum_{i=1}^N\mathrm{tr}[\Sigma_{ii}^*\hat{\Sigma}_{ii}^{-1}] - K - \frac{1}{N}\sum_{i=1}^N\ln|\Sigma_{ii}^*|\Big) + O_p(N^{-1})$$

$$-\frac{1}{N}\mathrm{tr}\Big[\Gamma^{*\dagger}M_{ff}^*\Gamma^{*\dagger\prime}\hat{\Sigma}_{zz}^{\dagger-1}\Big] - \frac{1}{N}\sum_{i=1}^N(\hat{\beta}-\beta^*)'\Sigma_{iix}^*(\hat{\beta}-\beta^*)\hat{\Sigma}_{iie}^{-1} + O_p\Big(\frac{\ln N}{N}\Big).$$

The main three terms of the above equation are all non-positive. By $\overline{L}(\hat{\theta}) = o_p(1)$, each of the three terms must be $o_p(1)$. That is,

$$\frac{1}{N}\sum_{i=1}^N(\hat{\beta}-\beta^*)'\Sigma_{iix}^*(\hat{\beta}-\beta^*)\hat{\Sigma}_{iie}^{-1} = o_p(1) \tag{A.7}$$

$$\frac{1}{N}\sum_{i=1}^N\ln|\hat{\Sigma}_{ii}| + \frac{1}{N}\sum_{i=1}^N\mathrm{tr}[\Sigma_{ii}^*\hat{\Sigma}_{ii}^{-1}] - K - \frac{1}{N}\sum_{i=1}^N\ln|\Sigma_{ii}^*| = o_p(1) \tag{A.8}$$

$$\frac{1}{N}\mathrm{tr}\Big[\Gamma^{*\dagger}M_{ff}^*\Gamma^{*\dagger\prime}\hat{\Sigma}_{zz}^{\dagger-1}\Big] = o_p(1) \tag{A.9}$$

Consider (A.7). The matrix Σ_{iix}^* is definite positive matrix for all i and $\hat{\Sigma}_{iie}$ is a scaler which is bounded by $[C^{-1}, C]$. So we have

$$\hat{\beta}-\beta^* = o_p(1) \tag{A.10}$$

Consider (A.8). Let $\hat{\omega}_{i1},\hat{\omega}_{i2},\cdots,\hat{\omega}_{iK}$ be the eigenvalues of the matrix $\hat{\Sigma}_{ii}$ which are arranged in descending order. Similarly, $\omega_{i1}^*,\omega_{i2}^*,\cdots,\omega_{iK}^*$ are the eigenvalues of the matrix Σ_{ii}^* in descending order. Consider the following function

$$f(\hat{\Sigma}_{ii}) = \ln|\hat{\Sigma}_{ii}| + \mathrm{tr}[\hat{\Sigma}_{ii}^{-1}\Sigma_{ii}^*] - K - \ln|\Sigma_{ii}^*| - b\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2$$

$$= \ln|\hat{\Sigma}_{ii}| + \mathrm{tr}[\hat{\Sigma}_{ii}^{-1}\Sigma_{ii}^*] - K - \ln|\Sigma_{ii}^*| - b\ \mathrm{tr}[\hat{\Sigma}_{ii}^2] + 2b\ \mathrm{tr}[\hat{\Sigma}_{ii}\Sigma_{ii}^*] - b\ \mathrm{tr}[\Sigma_{ii}^{*2}]$$

Since the eigenvalues of $\hat{\Sigma}_{ii}^2$ are $\hat{\omega}_{i1}^2, \hat{\omega}_{i2}^2, \cdots, \hat{\omega}_{iK}^2$, So we have $tr[\hat{\Sigma}_{ii}^2] = \sum_{p=1}^K \hat{\omega}_{ip}^2$. The same argument applying to Σ_{ii}^{*2} leads to $tr[\Sigma_{ii}^{*2}] = \sum_{p=1}^K \omega_{ip}^{*2}$. So the above equation is equal to

$$f(\hat{\Sigma}_{ii}) = \sum_{p=1}^K [(\ln\hat{\omega}_{ip} - 1 - \ln\omega_{ip}^*) - b(\hat{\omega}_{ip}^2 + \omega_{ip}^{*2})] + tr[(\hat{\Sigma}_{ii}^{-1} + 2b\hat{\Sigma}_{ii})\Sigma_{ii}^*]$$

The matrix $\hat{\Sigma}_{ii}^{-1} + 2b\hat{\Sigma}_{ii}$ has the eigenvalues $\hat{\omega}_{ip}^{-1} + 2b\hat{\omega}_{ip}$ for $p = 1, 2, \cdots, K$. We can choose b small enough to guarantee the order of $\hat{\omega}_{i1}^{-1} + 2b\hat{\omega}_{i1}, \hat{\omega}_{i2}^{-1} + 2b\hat{\omega}_{i2}, \cdots, \hat{\omega}_{iK}^{-1} + 2b\hat{\omega}_{iK}$ is the same as the order of $\hat{\omega}_{i1}^{-1}, \hat{\omega}_{i2}^{-1}, \cdots, \hat{\omega}_{iK}^{-1}$ because all $\hat{\omega}_{ip}$ are bounded by Assumption D. So we have

$$f(\hat{\Sigma}_{ii}) \geq \sum_{p=1}^K \Big\{(\ln\hat{\omega}_{ip} - 1 - \ln\omega_{ip}^*) - b(\hat{\omega}_{ip}^2 + \omega_{ip}^{*2}) + \hat{\omega}_{ip}^{-1}\omega_{ip}^* + 2b\hat{\omega}_{ip}\omega_{ip}^*\Big\}$$

by Lemma A.1. Using the arguments of [10], there exists a constant b small enough such that $f(\hat{\Sigma}_{ii}) \geq 0$. So we have

$$o_p(1) = \frac{1}{N}\sum_{i=1}^N (\ln|\hat{\Sigma}_{ii}| + \mathrm{tr}[\hat{\Sigma}_{ii}^{-1}\Sigma_{ii}^*] - K - \ln|\Sigma_{ii}^*|) \geq b\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2$$

This implies that

$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2 \xrightarrow{p} 0 \tag{A.11}$$

Consider (A.9), which is equal to $\frac{1}{N}\mathrm{tr}\Big([I_N \otimes (\hat{B}B^{*-1})]\Gamma^* M_{ff}^* \Gamma^{*\prime}[I_N \otimes (\hat{B}B^{*-1})^\prime]\hat{\Sigma}_{zz}^{-1}\Big) = o_p(1)$. By (A.10) and Lemma A.4 , we have

$$\frac{1}{N}\mathrm{tr}\Big[\Gamma^* M_{ff}^* \Gamma^{*\prime}\hat{\Sigma}_{zz}^{-1}\Big] = o_p(1)$$

Since $\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{G}\hat{\Gamma}^\prime\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$ and $\hat{G} = \hat{H} - \hat{H}\hat{M}_{ff}^{-1}\hat{G}$, the left hand side of the above equation can be written as

$$\mathrm{tr}\Big[\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* M_{ff}^* - \frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}(\hat{\Gamma}^\prime\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}^\prime\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* M_{ff}^*\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}(\hat{\Gamma}^\prime\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}\hat{M}_{ff}^{-1}\hat{G}\hat{\Gamma}^\prime\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* M_{ff}^*\Big]$$

The matrix inside each of the trace operator is semi-definitely positive. Given $\frac{1}{N}\mathrm{tr}\Big[\Gamma^* M_{ff}^* \Gamma^{*\prime} \hat{\Sigma}_{zz}^{-1}\Big] = o_p(1)$ together with the fact that $M = o_p(1)$ if $\mathrm{tr}(M) = o_p(1)$ for a semi-positive definite matrix M, we have

$$\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* - \frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}(\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* = o_p(1) \tag{A.12}$$

and

$$\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}(\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}\hat{M}_{ff}^{-1}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* = o_p(1) \tag{A.13}$$

Notice that $\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* = \frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^* + \frac{1}{N}\Gamma^{*\prime}(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{*-1})\Gamma^*$. Since the term $\frac{1}{N}\Gamma^{*\prime}(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{*-1})\Gamma^*$ is equal to $\frac{1}{N}\sum_{i=1}^N \Gamma_i^*\hat{\Sigma}_{ii}^{-1}(\Sigma_{ii}^* - \hat{\Sigma}_{ii})\Sigma_{ii}^{*-1}\Gamma_i^{*\prime}$, which is bounded in norm by $\frac{1}{N}\sum_{i=1}^N \|\Gamma_i^*\hat{\Sigma}_{ii}^{-1}\| \cdot \|\Sigma_{ii}^* - \hat{\Sigma}_{ii}\| \cdot \|\Sigma_{ii}^{*-1}\Gamma_i^{*\prime}\|$ and is further bounded by $C\frac{1}{N}\sum_{i=1}^N \|\Sigma_{ii}^* - \hat{\Sigma}_{ii}\|$ for a large C by the boundedness of $\|\Gamma_i^*\|, \|\hat{\Sigma}_{ii}\|, \|\Sigma_{ii}^*\|$ for all i in view of Assumptions C and D. By the Cauchy-Shwarz inequality, $C\frac{1}{N}\sum_{i=1}^N \|\Sigma_{ii}^* - \hat{\Sigma}_{ii}\|$ is bounded by $C\big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2\big)^{1/2}$. So we have

$$\begin{aligned} \|\frac{1}{N}\Gamma^{*\prime}(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{*-1})\Gamma^*\| &= \|\frac{1}{N}\sum_{i=1}^N \Gamma_i^*\hat{\Sigma}_{ii}^{-1}(\Sigma_{ii}^* - \hat{\Sigma}_{ii})\Sigma_{ii}^{*-1}\Gamma_i^{*\prime}\| \\ &\le C\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2\Big)^{1/2} \end{aligned} \tag{A.14}$$

Then we have $\frac{1}{N}\Gamma^{*\prime}(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{*-1})\Gamma^* = o_p(1)$ by (A.11). So $\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* = \frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^* + o_p(1)$. Given this result, together with (A.12), it follows

$$\frac{1}{N}\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}(\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* \xrightarrow{p} C^*$$

where C^* is the limit of $\frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^*$. Comparing the above result with (A.13), we have $\hat{M}_{ff}^{-1}\hat{G} = o_p(1)$. So $\hat{G} = o_p(1)$ due to the boundedness of $\hat{M}_{ff}$ by Assumption D. Since $\hat{G} = \hat{H}(I_r - \hat{M}_{ff}^{-1}\hat{G})$, we also have $\hat{H} = o_p(1)$. We summarize these results as

$$\hat{G} = o_p(1); \qquad \hat{H} = o_p(1) \tag{A.15}$$

Let $A = (\hat{\Gamma} - \Gamma^*)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$, then $\Gamma^{*\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H} = I_r - A$. By (A.14) and (A.11), equation (A.12) can be written, in terms of A, as

$$\frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^* - (I_r - A)\Big(\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\Big)(I_r - A)' = o_p(1) \tag{A.16}$$

Equation (A.16) can be written alternatively as

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Gamma}_i-\Gamma_i^*)\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i^*)'-A\Big(\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\Big)A'=o_p(1) \tag{A.17}$$

Now we turn to the first order condition for the proof of consistency. Notice that

$$\begin{aligned}\begin{bmatrix}\dot{y}_{jt}-\dot{x}_{jt}\hat{\beta}\\ \dot{x}_{jt}'\end{bmatrix}&=\begin{bmatrix}\dot{y}_{jt}-\dot{x}_{jt}\beta^*\\ \dot{x}_{jt}'\end{bmatrix}-\begin{bmatrix}\dot{x}_{jt}\\ 0_{K\times K}\end{bmatrix}(\hat{\beta}-\beta^*)\\ &=\Gamma_j^{*\prime}f_t^*+\dot{\varepsilon}_{jt}-\dot{l}_{jt}(\hat{\beta}-\beta^*)\end{aligned} \tag{A.18}$$

where $\dot{l}_{jt}=(\dot{x}_{jt}',0_{K\times K})'$ is a $(K+1)\times K$ matrix. Also notice that

$$\hat{B}M_{zz}^{ij}\hat{B}'=\frac{1}{T}\sum_{t=1}^{T}\hat{B}\dot{z}_{it}\dot{z}_{jt}'\hat{B}'=\frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\dot{y}_{it}-\dot{x}_{it}\hat{\beta}\\ \dot{x}_{it}'\end{bmatrix}\Big[\dot{y}_{jt}-\dot{x}_{jt}\hat{\beta},\ \dot{x}_{jt}\Big] \tag{A.19}$$

So we have

(A.20)

$$\begin{aligned}\hat{B}M_{zz}^{ij}\hat{B}'-\hat{\Sigma}_{zz}^{ij}&=\Gamma_i^{*\prime}M_{ff}^*\Gamma_j^*+\Gamma_i^{*\prime}\frac{1}{T}\sum_{t=1}^{T}f_t^*\dot{\varepsilon}_{jt}'+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}f_t^{*\prime}\Gamma_j^*+\frac{1}{T}\sum_{t=1}^{T}\epsilon_{ij,t}\\ &-\Gamma_i^{*\prime}\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta}-\beta^*)'\dot{l}_{jt}'-\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}(\hat{\beta}-\beta^*)'\dot{l}_{jt}'-\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta^*)f_t^{*\prime}\Gamma_j^*\\ &-\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta^*)\dot{\varepsilon}_{jt}'+\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\dot{l}_{jt}'-\hat{\Gamma}_i'\hat{M}_{ff}\hat{\Gamma}_j-1(i=j)(\hat{\Sigma}_{jj}-\Sigma_{jj}^*)\end{aligned}$$

where $1(i=j)=1$ if $i=j$, 0 otherwise, and $\epsilon_{ij,t}=\varepsilon_{it}\dot{\varepsilon}_{jt}'-E(\varepsilon_{it}\dot{\varepsilon}_{jt}')$. For simplicity, we neglect the smaller order term $\bar{\varepsilon}_i\bar{\varepsilon}_j=T^{-2}(\sum_{t=1}^{T}\varepsilon_{it})(\sum_{j=1}^{T}\varepsilon_{jt})$. Using (A.20), equation (2.7) is equal to

$$\begin{aligned}\hat{M}_{ff}-M_{ff}^*&=-A'M_{ff}^*-M_{ff}^*A+A'M_{ff}^*A+(I-A)'\Big[\frac{1}{T}\sum_{t=1}^{T}f_t^*\hat{\chi}_t'\Big]\hat{H}_N\\ &+\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_tf_t^{*\prime}\Big](I-A)+\hat{H}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\epsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\Big]\hat{H}\\ &-(I-A)'\Big[\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta}-\beta^*)'\hat{\xi}_t'\Big]\hat{H}_N-\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta^*)'\hat{\xi}_t'\Big]\hat{H}_N\end{aligned} \tag{A.21}$$

$$-\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)f_t^{*\prime}\Big](I-A)-\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)\hat{\chi}_t'\Big]\hat{H}_N$$

$$+\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\hat{\xi}_t'\Big]\hat{H}_N-\hat{H}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Sigma}_{\varepsilon\varepsilon}-\Sigma_{\varepsilon\varepsilon}^*)\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$$

where $\hat{\chi}_t, \hat{\xi}_t, \hat{H}_N, \hat{H}$ and A are defined in Table 1 and $\epsilon_{ij,t} = \varepsilon_{it}\varepsilon_{jt}' - E(\varepsilon_{it}\varepsilon_{jt}')$.

Consider (A.21). The 4th-6th and the last terms of the right hand side of (A.21) are summarized in Lemma A.5. The 7th-11th terms are summarized in Lemma A.6. Since we have already proved that $\hat{\beta}-\beta^* = o_p(1)$ and $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}^*\|^2 = o_p(1)$, (A.21) can be written as

$$\hat{M}_{ff} = (I_r-A)'M_{ff}^*(I_r-A)+\|N^{1/2}\hat{H}^{1/2}\|^2\cdot o_p(1) \tag{A.22}$$

However, (A.16) indicates that $N\hat{H} = (I_r-A)'(\frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^*)^{-1}(I_r-A) + o_p(\|I_r-A\|^2)$. So $\|N^{1/2}\hat{H}^{1/2}\|^2 = \mathrm{tr}(N\hat{H}) = \mathrm{tr}[(I_r-A)'(\frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^*)^{-1}(I_r-A)]+o_p(\|I-A\|^2)$. Given this result, we can show $A = O_p(1)$. Otherwise, if A is stochastically unbounded, the first term of the right hand side of (A.22), which dominates the second term, will diverge to infinity. But the left hand side of (A.22) is stochastically bounded by Assumption D (it is I_r under our identification condition). A contradiction is obtained. Given $A = O_p(1)$, we immediately obtain

$$N\hat{H} = O_p(1) \tag{A.23}$$

Then (A.22) can be simplified as $\hat{M}_{ff} = (I_r-A)'M_{ff}^*(I_r-A)+o_p(1)$. Notice that the identification condition requires $M_{ff} = \hat{M}_{ff} = I_r$. This yields

$$(I_r-A)'(I_r-A) = I_r + o_p(1) \tag{A.24}$$

By (A.16) and (A.24), applying Lemma A.2 with $Q = (I_r-A)'$, $V = \frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}$ and $D = \frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^*$, it follows that $I_r - A$ converges in probability to a diagonal matrix with diagonal elements either 1 or -1. By assuming $\hat{\Gamma}$ and Γ^* have the same column signs, we rule out -1 as the diagonal element. So $A = o_p(1)$. Then (A.17) implies the second result of Proposition 2.1.

This completes the proof of Proposition 2.1. □

COROLLARY A.1. *Under the assumptions of Proposition 2.1, we have*

(*a*) $\displaystyle \frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^* - \frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^* = o_p(1)$

(*b*) $\displaystyle \frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} - \frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^* = o_p(1)$

(*c*) $A = (\hat{\Gamma}-\Gamma^*)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H} = o_p(1)$

APPENDIX A2: PROOFS OF THEOREMS 2.1 AND 2.2

Given consistency, we now drop the superscript "*" from the true parameters for notational simplicity. The following lemmas are useful to prove Theorem 2.1.

LEMMA A.7. *Under Assumptions A-D,*

$$(a)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|(I-A)'\frac{1}{T}\sum_{t=1}^{T}f_t\varepsilon_{jt}'\Big\|^2 = O_p(T^{-1})$$

$$(b)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)\Gamma_j\Big\|^2 = O_p(T^{-1})$$

$$(c)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{ij,t}\Big\|^2 = O_p(T^{-1})$$

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

where $\epsilon_{ij,t} = \varepsilon_{it}\varepsilon_{jt}' - E(\varepsilon_{it}\varepsilon_{jt}')$.

PROOF OF LEMMA A.7. Consider (a). The left hand side of (a) is bounded by

$$\|I-A\|^2\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}f_t\varepsilon_{jt}'\|^2$$

which is $O_p(T^{-1})$ because $\frac{1}{T}\sum_{t=1}^{T}f_t\varepsilon_{jt}' = O_p(T^{-1/2})$ and $A = o_p(1)$ by Corollary A.1(c).

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

$$\frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}f_t'\Gamma_j\Big\|^2$$

which is bounded by

$$C\cdot\|N\hat{H}\|^2\cdot\Big(\frac{1}{N}\sum_{j=1}^{N}\|\Gamma_j\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}f_t\varepsilon_{it}'\|^2\Big)$$

which is $O_p(T^{-1})$ by Corollary A.1(b) and $N\hat{H} = O_p(1)$.

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

$$C\|N\hat{H}\|^2\cdot\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}\epsilon_{ij,t}\|^2\Big)$$

which is $O_p(T^{-1})$ by the same reason as (b) and Assumption B.

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

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

which is $O_p(N^{-2}\sum_{j=1}^N\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2)$ by the same reason as (b). □

LEMMA A.8. *Under Assumptions A-D,*

$$(a)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|(I-A)'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big\|^2 = O_p(\|\hat{\beta}-\beta\|^2)$$

$$(b)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big)\Big\|^2 = o_p(\|\hat{\beta}-\beta\|^2)$$

$$(c)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)\Gamma_j\Big\|^2 = O_p(\|\hat{\beta}-\beta\|^2)$$

$$(d)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\varepsilon_{jt}'\Big)\Big\|^2 = o_p(\|\hat{\beta}-\beta\|^2)$$

$$(e)\quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big)\Big\|^2 = O_p(\|\hat{\beta}-\beta\|^4)$$

PROOF OF LEMMA A.8. Consider (a). The left hand side of (a) is bounded in norm by

$$\|I-A\|^2 \cdot \Big(\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\|^2\Big)$$

The first factor is $O_p(1)$ by Corollary A.1(c). Consider the second factor, which is equal to

$$\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)M_{ff}\gamma_{jp}+\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\frac{1}{T}\sum_{t=1}^{T}f_t v_{jtp}\Big\|^2$$

The above expression is bounded by

$$2\frac{1}{N}\sum_{j=1}^{N}\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)M_{ff}\gamma_{jp}\|^2+2\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\frac{1}{T}\sum_{t=1}^{T}f_t v_{jtp}\Big\|^2$$

Notice

$$|\hat{\beta}_p - \beta_p| \le \|\hat{\beta} - \beta\| \qquad \text{for } \ \forall p \tag{A.25}$$

The first term is $O_p(\|\hat{\beta} - \beta\|^2)$, the second term is $O_p(T^{-1}\|\hat{\beta} - \beta\|^2)$. Thus (a) follows.

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

$$\frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big\|^2$$

which is equal to

$$\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\dot{x}_{jtp}\Big\|^2$$

By $\dot{x}_{jtp} = \gamma_{jp}' f_t + \dot{v}_{jtp}$, the above expression is bounded by

$$4\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{it}f_t'\gamma_{jp}\Big\|^2$$

$$+4\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}v_{jtp}-E(\varepsilon_{it}v_{jtp})]\Big\|^2$$

$$+4\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}E(\varepsilon_{it}v_{jtp})\Big\|^2$$

$$+4\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{it}\bar{v}_{jp}\Big\|^2 = e_1+e_2+e_3+e_4, \quad \text{say}$$

Consider e_1. Ignore the factor 4, it is bounded, due to (A.25), by

$$CK\|\hat{\beta}-\beta\|^2\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)$$

$$\times\Big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}f_t'\|^2\Big)\sum_{p=1}^{K}\frac{1}{N}\sum_{j=1}^{N}\|\gamma_{jp}\|^2$$

So $e_1 = O_p(T^{-1}\|\hat{\beta}-\beta\|^2)$ by $N\hat{H} = O_p(1)$ and $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$.

Consider e_2. Ignore the factor 4, it is bounded by

$$CK\|\hat{\beta}-\beta\|^2\cdot\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)$$

$$\times\Big(\sum_{p=1}^{K}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}v_{jtp}-E(\varepsilon_{it}v_{jtp})]\Big\|^2\Big)$$

So $e_2 = O_p(T^{-1}\|\hat{\beta}-\beta\|^2)$ by the similar arguments as e_1.

Consider e_3, which is bounded in norm by

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

$$\times\sum_{p=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\|E(\varepsilon_{it}v_{jtp})\|^2$$

Notice $E(\varepsilon_{it}v_{jtp})=0$ if $i\neq j$. So $e_3 = O_p(N^{-1}\|\hat{\beta}-\beta\|^2)$. e_4 is of smaller order than e_2. Given these results, (b) follows.

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

$$\frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta)f_t'\Gamma_j\Big\|^2$$

which is bounded in norm by

$$\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta)f_t'\Big\|^2\Big(\frac{1}{N}\sum_{j=1}^{N}\|\Gamma_j\|^2\Big)$$

Since $\frac{1}{N}\sum_{j=1}^{N}\|\Gamma_j\|^2=O(1)$, we only need to consider $\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta)f_t'$, which is equal to

$$\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\gamma_{ip}'M_{ff}$$

$$+\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}v_{itp}f_t' = e_5+e_6 \quad \text{say}$$

The term $\hat{H}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\gamma_{ip}'M_{ff}$ is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|\Big(\sum_{i=1}^{N}\|\hat{H}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iie}^{-1/2}\gamma_{ip}\|^2\Big)^{1/2}\|M_{ff}\|$$

which is $O_p(1)$ by (A.4). So $e_5 = O_p(\|\hat{\beta}-\beta\|)$. Consider e_6. The term $\hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\lambda}_i\,\hat{\Sigma}_{iie}^{-1}v_{itp}f_t'$ is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)^{\frac{1}{2}}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}f_tv_{itp}\|^2\Big)^{\frac{1}{2}}$$

which is $O_p(T^{-1/2})$. So $e_6 = o_p(\|\hat\beta - \beta\|)$. Given these results, (c) follows.

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

$$\frac{1}{N}\sum_{j=1}^N \Big\| \hat H \sum_{i=1}^N \hat\Gamma_i \hat\Sigma_{ii}^{-1} \frac{1}{T}\sum_{t=1}^T \dot l_{it}(\hat\beta - \beta)\varepsilon_{jt}' \Big\|^2$$

which is equal to

$$\frac{1}{N}\sum_{j=1}^N \Big\| \sum_{p=1}^K (\hat\beta_p - \beta_p) \hat H \frac{1}{T}\sum_{i=1}^N \sum_{t=1}^T \hat\lambda_i \hat\Sigma_{iie}^{-1} \dot x_{itp} \varepsilon_{jt}' \Big\|^2$$

By $\dot x_{itp} = \gamma_{ip}' f_t + \dot v_{itp}$, the above expression is bounded by

$$4\frac{1}{N}\sum_{j=1}^N \Big\| \sum_{p=1}^K (\hat\beta_p - \beta_p) \hat H \sum_{i=1}^N \hat\lambda_i \hat\Sigma_{iie}^{-1} \gamma_{ip}' \frac{1}{T}\sum_{t=1}^T f_t \varepsilon_{jt}' \Big\|^2$$
$$+4\frac{1}{N}\sum_{j=1}^N \Big\| \sum_{p=1}^K (\hat\beta_p - \beta_p) \hat H \frac{1}{T}\sum_{i=1}^N \sum_{t=1}^T \hat\lambda_i \hat\Sigma_{iie}^{-1} [v_{itp}\varepsilon_{jt}' - E(v_{itp}\varepsilon_{jt}')] \Big\|^2$$
$$+4\frac{1}{N}\sum_{j=1}^N \Big\| \sum_{p=1}^K (\hat\beta_p - \beta_p) \hat H \sum_{i=1}^N \hat\lambda_i \hat\Sigma_{iie}^{-1} E(v_{itp}\varepsilon_{jt}') \Big\|^2$$
$$+4\frac{1}{N}\sum_{j=1}^N \Big\| \sum_{p=1}^K (\hat\beta_p - \beta_p) \hat H \frac{1}{T}\sum_{i=1}^N \sum_{t=1}^T \hat\lambda_i \hat\Sigma_{iie}^{-1} \varepsilon_{jt}' \bar v_{jp} \Big\|^2 = e_7 + e_8 + e_9 + e_{10}, \quad \text{say}$$

Using (A.25), term e_7 is bounded by (ignoring the factor 4)

$$K\|\hat\beta - \beta\|^2 \cdot \|N^{1/2}\hat H^{1/2}\|^2 \Big(\sum_{i=1}^N \|\hat H^{1/2}\hat\lambda_i \hat\Sigma_{iie}^{-1/2}\|^2 \Big)$$
$$\times \Big(\sum_{p=1}^K \frac{1}{N}\sum_{i=1}^N \|\hat\Sigma_{iie}^{-1/2}\gamma_{ip}\|^2 \Big) \frac{1}{N}\sum_{j=1}^N \|\frac{1}{T}\sum_{t=1}^T f_t \varepsilon_{jt}'\|^2$$

which is $O_p(T^{-1}\|\hat\beta - \beta\|^2)$ by (A.4).

Consider e_8, which is bounded by (ignoring the factor 4)

$$CK\|\hat\beta - \beta\|^2 \cdot \|N^{1/2}\hat H^{1/2}\|^2 \Big(\sum_{i=1}^N \|\hat H^{1/2}\hat\lambda_i \hat\Sigma_{iie}^{-1/2}\|^2 \Big)$$
$$\times \sum_{p=1}^K \frac{1}{N^2}\sum_{i=1}^N \sum_{j=1}^N \Big\| \frac{1}{T}\sum_{t=1}^T [v_{itp}\varepsilon_{lt}' - E(v_{itp}\varepsilon_{lt}')] \Big\|^2$$

which is $O_p(T^{-1}\|\hat{\beta}-\beta\|^2)$ by the similar arguments as e_7.

Consider e_9. which is bounded by (ignoring the factor 3)

$$CK\frac{1}{N}\|\hat{\beta}-\beta\|^2\cdot\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^N\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)\sum_{p=1}^K\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\|E(v_{itp}\varepsilon_{jt})\|^2$$

which is $O_p(N^{-1}\|\hat{\beta}-\beta\|^2)$ by $E(v_{itp}\varepsilon_{jt})=0$ for $i\neq j$. Term e_{10} is of smaller order than e_8. Given these results, we have (d).

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

$$\frac{1}{N}\sum_{j=1}^N\Big\|\hat{H}\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T\dot{l}_{it}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big\|^2$$

which is equal to

$$\frac{1}{N}\sum_{j=1}^N\Big\|\sum_{p=1}^K\sum_{q=1}^K(\hat{\beta}_p-\beta_p)(\hat{\beta}_q-\beta_q)\hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\dot{x}_{itp}\dot{x}_{jtq}\Big\|^2$$

By (A.25), the above expression is bounded by

$$K^2\|\hat{\beta}-\beta\|^4\sum_{p=1}^K\sum_{q=1}^K\frac{1}{N}\sum_{j=1}^N\|\hat{H}\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\dot{x}_{itp}\dot{x}_{jtq}\Big\|^2$$

which is further bounded by

$$CK^2\|\hat{\beta}-\beta\|^4\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^N\|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)$$
$$\times\sum_{p=1}^K\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\Big(\frac{1}{T}\sum_{t=1}^T\dot{x}_{itp}\dot{x}_{jtp}\Big)^2$$

Notice, for $\forall i,t,p,\ E(\dot{x}_{itp}^4)\leq C$ for some sufficiently large constant C by Assumptions B and C. So the above expression is $O_p(\|\hat{\beta}-\beta\|^4)$, obtaining (e). $\square$

PROPOSITION A.1. *Under Assumptions A-D and the identification conditions IB, we have*

$$A\equiv\sum_{i=1}^N(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}=O_p(T^{-1/2})+O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big]^{1/2}\Big)$$
$$+O_p\Big(\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2\Big)+O_p(\|\hat{\beta}-\beta\|)$$

Proof of Proposition A.1. Notice the equation

$$\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} - \frac{1}{N}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma = \frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma)$$

$$+\frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} - \frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma) + \frac{1}{N}\Gamma'(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1})\Gamma$$

Since $\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}$ and $\frac{1}{N}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma$ are both diagonal matrices, we have

$$\begin{aligned}&\mathrm{ndiag}\Big\{\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma) + \frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\Big\}\\ &\quad= \mathrm{ndiag}\Big\{\frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma) - \frac{1}{N}\Gamma'(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1})\Gamma\Big\}\end{aligned} \tag{A.26}$$

where ndiag denotes the non-diagonal elements. However, by Lemmas A.5 and A.6, and noting $\hat{M}_{ff} = M_{ff} = I_r$, equation (A.21) can be simplified as

$$\hat{H}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma) + (\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H} = O_p(T^{-1/2}) + O_p(\|\hat{\beta}-\beta\|)$$

which can be written as

$$\hat{H}_N\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma) + \frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}_N = O_p(T^{-1/2}) + O_p(\|\hat{\beta}-\beta\|) \tag{A.27}$$

where $\hat{H}_N = N\hat{H}$ is a diagonal matrix. We use $\hat{q}_1, \hat{q}_2, \cdots, \hat{q}_r$ to denote its diagonal elements.

Equation (A.26) puts $\frac{1}{2}r(r-1)$ restrictions and equation (A.27) puts $\frac{1}{2}r(r+1)$ restrictions on the matrix $\frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}$. So it can be uniquely determined by the equation system (A.26) and (A.27). Solving this equation system, we have

$$\Big[\frac{1}{N}(\hat{\Gamma}-\Gamma)\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\Big]_{ii} = \frac{1}{2\hat{q}_i}\Big[O_p(T^{-1/2}) + O_p(\|\hat{\beta}-\beta\|)\Big]$$

for $i = 1, 2, \cdots, r$, and

$$\Big[\frac{1}{N}(\hat{\Gamma}-\Gamma)\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\Big]_{ij} = \frac{1}{\hat{q}_j - \hat{q}_i}\Big\{\Big[O_p(T^{-1/2}) + O_p(\|\hat{\beta}-\beta\|)\Big]$$

$$-\hat{q}_i\Big[\frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma) - \frac{1}{N}\Gamma'(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1})\Gamma\Big]\Big\}$$

for $i \neq j$. However, Corollary A.1(b) shows $\hat{q}_j - q_j \xrightarrow{p} 0$, where q_j is the jth diagonal element of Q defined in Assumption C.3. So we have

$$(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H} = O_p(T^{-1/2}) + O_p(\|\hat{\beta}-\beta\|)$$

$$+O_p\Big(\frac{1}{N}(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma)\Big)+O_p\Big(\frac{1}{N}\Gamma'(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}-\Sigma_{\varepsilon\varepsilon}^{-1})\Gamma\Big)$$

which implies Proposition A.1 because $O_p[\frac{1}{N}\Gamma'(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}-\Sigma_{\varepsilon\varepsilon}^{-1})\Gamma]=O_p([\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2]^{1/2})$ by (A.14).

This completes the proof of Proposition A.1 □

Proposition A.2. *Under the assumptions of Proposition 2.1, we have*

$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2=O_p(T^{-1})+O_p(\|\hat{\beta}-\beta\|^2)$$

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

Proof of Proposition A.2. Consider the first order condition (2.6). Some algebra computation shows that (2.6) is equal to

$$\hat{\Gamma}_j-\Gamma_j=-\hat{M}_{ff}^{-1}(\hat{M}_{ff}-M_{ff})\Gamma_j-\hat{M}_{ff}^{-1}A'M_{ff}\Gamma_i \tag{A.28}$$

$$+\hat{M}_{ff}^{-1}(I-A)'\frac{1}{T}\sum_{t=1}^T f_t\varepsilon_{jt}'+\hat{M}_{ff}^{-1}\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\chi}_t f_t'\Big]\Gamma_j+\hat{M}_{ff}^{-1}\hat{H}\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T\epsilon_{ij,t}$$

$$-\hat{M}_{ff}^{-1}(I-A)'\Big[\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big]-\hat{M}_{ff}^{-1}\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\chi}_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big]$$

$$-\hat{M}_{ff}^{-1}\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big]\Gamma_j-\hat{M}_{ff}^{-1}\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)\varepsilon_{jt}'\Big]$$

$$+\hat{M}_{ff}^{-1}\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big]-\hat{M}_{ff}^{-1}\hat{H}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})$$

where $\epsilon_{ij,t}=\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')$. Notice that the identification condition requires $\hat{M}_{ff}=M_{ff}=I_r$. So the above result can be simplified as

$$\hat{\Gamma}_j-\Gamma_j=-A'\Gamma_j+(I-A)'\frac{1}{T}\sum_{t=1}^T f_t\varepsilon_{jt}'+\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\chi}_t f_t'\Big]\Gamma_j$$

$$+\hat{H}\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T\epsilon_{ij,t}-(I-A)'\Big[\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big]-\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T\hat{\chi}_t(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big]$$

$$
\text{(A.29)} \qquad -\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big]\Gamma_j-\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\varepsilon_{jt}'\Big]
$$

$$
+\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{l}_{jt}'\Big]-\hat{H}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})
$$

The right hand side of (A.29) has 10 terms. We use $a_{i1}, a_{i2}, \cdots, a_{i10}$ to denote them. So we have $\|\hat{\Gamma}_i-\Gamma_i\|\le\|a_{i1}\|+\|a_{i2}\|+\cdots+\|a_{i10}\|$. Using the fact $(\sum_{p=1}^{10}\|a_p\|)^2\le 10\sum_{p=1}^{10}\|a_p\|^2$, we have

$$
\text{(A.30)} \qquad \begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2 &\le C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Gamma}_i-\Gamma_i\|^2\\ &\le 10C\frac{1}{N}\sum_{i=1}^{N}(\|a_{i1}\|^2+\|a_{i2}\|^2+\cdots+\|a_{i10}\|^2)\end{aligned}
$$

The first inequality uses the fact that $\|\hat{\Sigma}_{ii}^{-1}\|$ is bounded by some C. Consider the first term, which is equal to $\frac{1}{N}\sum_{j=1}^{N}\|A'\Gamma_j\|^2$. This term is bounded by $\|A\|^2\cdot\frac{1}{N}\sum_{j=1}^{N}\|\Gamma_j\|^2$. Since $A=O_p(T^{-1/2})+O_p([\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2]^{1/2})+O_p(\|\hat{\beta}-\beta\|)$ by Propositions A.1, together with $\frac{1}{N}\sum_{j=1}^{N}\|\Gamma_j\|^2=O(1)$, we have $\frac{1}{N}\sum_{i=1}^{N}\|a_{i1}\|^2=O_p(T^{-1})+O_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2)+O_p(\|\hat{\beta}-\beta\|^2)$. The 2nd-4th and 10th terms are summarized in Lemma A.7. The 5th-9th terms are summarized in Lemma A.8. From these results, we have

$$
\text{(A.31)} \qquad \begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2 &= O_p(T^{-1})+O_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2)\\ &\quad +O_p(\|\hat{\beta}-\beta\|^2)\end{aligned}
$$

Now we turn to the first order condition (2.8). By (A.18) and (A.19), equation (2.8) is equivalent to

$$
\hat{\Sigma}_{jje}-\Sigma_{jje}=-(\hat{\lambda}_j-\lambda_j)'\hat{M}_{ff}(\hat{\lambda}_j-\lambda_j)-2(\hat{\lambda}_j-\lambda_j)'\hat{M}_{ff}\lambda_j-\lambda_j'(\hat{M}_{ff}-M_{ff})\lambda_j
$$

$$
\text{(A.32)} \quad +2\frac{1}{N}\sum_{t=1}^{T}\lambda_j'f_te_{jt}+\frac{1}{T}\sum_{t=1}^{T}(e_{jt}^2-\Sigma_{jje})-2\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\frac{1}{T}\sum_{t=1}^{T}\lambda_j'f_t\dot{x}_{jtp}
$$

$$
-2\sum_{p=1}^{K}(\hat{\beta}_p-\beta_p)\Big[\frac{1}{T}\sum_{t=1}^{T}e_{jt}\dot{x}_{jtp}\Big]+\sum_{p=1}^{K}\sum_{q=1}^{K}\dot{x}_{jtp}(\hat{\beta}_p-\beta_p)(\hat{\beta}_q-\beta_q)\dot{x}_{jtq}
$$

and

$$\hat{\Sigma}_{jjx} - \Sigma_{jjx} = \frac{1}{T}\sum_{t=1}^{T}[v_{jtx}v_{jtx}' - \Sigma_{jjx}] - (\hat{\gamma}_{jx} - \gamma_{jx})'\hat{M}_{ff}(\hat{\gamma}_{jx} - \gamma_{jx})$$

$$-(\hat{\gamma}_{jx} - \gamma_{jx})'\hat{M}_{ff}\gamma_{jx} - \gamma_{jx}'\hat{M}_{ff}(\hat{\gamma}_{jx} - \gamma_{jx}) - \gamma_{jx}'(\hat{M}_{ff} - M_{ff})\gamma_{jx} \tag{A.33}$$

$$+\frac{1}{T}\sum_{t=1}^{T}\gamma_{jx}' f_t v_{jtx}' + \frac{1}{T}\sum_{t=1}^{T} v_{jtx} f_t \gamma_{jx}$$

Consider (A.32). The second term of right hand side of (A.32) involves $\hat{\lambda}_j - \lambda_j$. The third term involves $\hat{M}_{ff} - M_{ff}$. But the expression of $\hat{\lambda}_j - \lambda_j$ is given in (A.29) (the first column) and $\hat{M}_{ff} - M_{ff}$ is given in (A.21)[1]. Using (A.29) and (A.21) to replace $\hat{\lambda}_j - \lambda_j$ and $\hat{M}_{ff} - M_{ff}$ from (A.32), we have

$$\hat{\Sigma}_{jje} - \Sigma_{jje} = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \Sigma_{jje}) - (\hat{\lambda}_j - \lambda_j)'(\hat{\lambda}_j - \lambda_j) + \lambda_j' A' A \lambda_j \tag{A.34}$$

$$+2\lambda_j' A' \frac{1}{T}\sum_{t=1}^{T} f_t e_{jt} - 2\lambda_j' A' \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t \varepsilon_{it}' \hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}\lambda_j$$

$$+\lambda_j'\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\lambda_j - \lambda_j'\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\Sigma_{ii}^{-1}\hat{\Gamma}'\hat{H}\lambda_j$$

$$-2\lambda_j'\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})] + 2\lambda_j'\hat{H}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje} - \Sigma_{jje}) + \mathcal{O}_{j1}$$

where $A = (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$, $\epsilon_{ij,t} = \varepsilon_{it}\varepsilon_{jt} - E(\varepsilon_{it}\varepsilon_{jt})$ and $\mathcal{O}_{j1}$ is defined as

$$\mathcal{O}_{j1} = -2\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p)\lambda_j' A'\Big[\frac{1}{T}\sum_{t=1}^{T} f_t \dot{x}_{jtp}\Big] - 2\lambda_j'\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta} - \beta)'\hat{\xi}_t'\Big]\hat{H}_N\lambda_j$$

$$-2\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p)\Big[\frac{1}{T}\sum_{t=1}^{T} e_{jt}\dot{x}_{jtp}\Big] + \sum_{p=1}^{K}\sum_{q=1}^{K}\Big[\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{itp}(\hat{\beta}_p - \beta_p)(\hat{\beta}_q - \beta_q)\dot{x}_{jtq}\Big]$$

$$+2\lambda_j' A'\Big[\frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\beta} - \beta)'\hat{\xi}_t'\Big]\hat{H}_N\lambda_j + 2\sum_{p=1}^{K}(\hat{\beta}_p - \beta_p)\lambda_j'\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t\dot{x}_{jtp}\Big]$$

[1]Under the identification condition $\hat{M}_{ff} = M_{ff} = I_r$, we immediately obtain that $\hat{M}_{ff} - M_{ff} = 0$. However, this result, while simple, is not as useful as it looks. If we use this result, we would face the term $\hat{H}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma} - \Gamma) + (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$ in the subsequent expression, which still requires invoke (A.21).

$$-2\sum_{p=1}^{K}(\hat\beta_p-\beta_p)\lambda_j'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)\dot x_{jtp}\Big]+2\lambda_j'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)e_{jt}\Big]$$

$$+\lambda_j'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)(\hat\beta-\beta)'\hat\xi_t'\Big]\hat H_N\lambda_j \tag{A.35}$$

Applying similar arguments to (A.33), we have

$$\hat\Sigma_{jjx}-\Sigma_{jjx}=\frac{1}{T}\sum_{t=1}^{T}(v_{jtx}v_{jtx}'-\Sigma_{jjx})-\gamma_{jx}'\hat H\sum_{i=1}^{N}\hat\Gamma_i\hat\Sigma_{ii}^{-1}(\hat\Sigma_{ii}-\Sigma_{ii})\Sigma_{ii}^{-1}\hat\Gamma_i'\hat H\gamma_{jx}$$

$$-\gamma_{jx}'A'\Big[\frac{1}{T}\sum_{t=1}^{T}f_t\hat\chi_t'\Big]\hat H_N\gamma_{jx}-\gamma_{jx}'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\chi_t f_t'\Big]A\gamma_{jx}+\gamma_{jx}'A'\frac{1}{T}\sum_{t=1}^{T}f_t v_{jtx}'$$

$$+\frac{1}{T}\sum_{t=1}^{T}v_{jtx}f_t'A\gamma_{jx}-\gamma_{jx}'\hat H\sum_{i=1}^{N}\hat\Gamma_i\hat\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_{ij,t}-\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\eta_{ij,t}'\hat\Sigma_{ii}^{-1}\hat\Gamma_i'\hat H\gamma_{jx}$$

$$+\gamma_{jx}'\hat H\hat\gamma_{jx}\hat\Sigma_{jjx}^{-1}(\hat\Sigma_{jjx}-\Sigma_{jjx})+(\hat\Sigma_{jjx}-\Sigma_{jjx})\hat\Sigma_{jjx}^{-1}\hat\gamma_{jx}'\hat H\gamma_{jx} \tag{A.36}$$

$$+\gamma_{jx}'\hat H\frac{1}{T}\sum_{t=1}^{T}\hat\Gamma'\hat\Sigma_{\varepsilon\varepsilon}^{-1}(\varepsilon_t\varepsilon_t'-\Sigma_{\varepsilon\varepsilon})\hat\Sigma_{\varepsilon\varepsilon}^{-1}\hat\Gamma\hat H\gamma_{jx}+\gamma_{jx}'A'A\gamma_{jx}-(\hat\gamma_{jx}-\gamma_{jx})'(\hat\gamma_{jx}-\gamma_{jx})+\mathcal{O}_{j2}$$

where $\eta_{ij,t}=\varepsilon_{it}v_{jtx}'-E(\varepsilon_{it}v_{jtx}')$ and $\mathcal{O}_{j2}$ is defined as

$$\mathcal{O}_{j2}=\gamma_{jx}'A'\Big[\frac{1}{T}\sum_{t=1}^{T}f_t(\hat\beta-\beta)'\hat\xi_t'\Big]\hat H_N\gamma_{jx}+\gamma_{jx}'\hat H_N\Big[\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)f_t'\Big]A\gamma_{jx}$$

$$-\gamma_{jx}'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\chi_t(\hat\beta-\beta)'\hat\xi_t'\Big]\hat H_N\gamma_{jx}-\gamma_{jx}'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)\hat\chi_t'\Big]\hat H_N\gamma_{jx}$$

$$+\gamma_{jx}'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)v_{jtx}'\Big]+\Big[\frac{1}{T}\sum_{t=1}^{T}v_{jtx}(\hat\beta-\beta)'\hat\xi_t'\Big]\hat H_N\gamma_{jx}$$

$$+\gamma_{jx}'\hat H_N\Big[\frac{1}{T}\sum_{t=1}^{T}\hat\xi_t(\hat\beta-\beta)(\hat\beta-\beta)'\hat\xi_t'\Big]\hat H_N\gamma_{jx} \tag{A.37}$$

Terms $\mathcal{O}_{j1}$ and $\mathcal{O}_{j2}$ depend on $\hat\beta-\beta$. By the special structure of Σ_{ii}, we have

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat\Sigma_{jj}-\Sigma_{jj}\|^2=\frac{1}{N}\sum_{j=1}^{N}(\hat\Sigma_{jje}-\Sigma_{jje})^2+\frac{1}{N}\sum_{j=1}^{N}\|\hat\Sigma_{jjx}-\Sigma_{jjx}\|^2$$

Consider $\frac{1}{N}\sum_{j=1}^{N}(\hat{\Sigma}_{jje}-\Sigma_{jje})^2$. The right hand side of (A.34) has 10 terms and we use $b_{j1}, b_{j2}, \cdots, b_{j9}$ to denote the first 9 terms. By the Cauchy-Schwarz inequality, we have

$$\frac{1}{N}\sum_{j=1}^{N}(\hat{\Sigma}_{jje}-\Sigma_{jje})^2 \leq 10\frac{1}{N}\sum_{j=1}^{N}(\|b_{j1}\|^2+\cdots+\|b_{j9}\|^2+\|\mathcal{O}_{j1}\|^2)$$

The first term is $O_p(T^{-1})$ can be easily verified. For the second term, we need to use (A.29) (the first column) to substitute $\hat{\lambda}_j-\lambda_j$. By a little tedious computation, we have $\frac{1}{N}\sum_{j=1}^{N}\|b_{j2}\|^2 = o_p(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2) + O_p(T^{-2})+O_p(\|\hat{\beta}-\beta\|^4)$. By Lemmas A.5, A.7 and Proposition A.1, we have

$$\frac{1}{N}\sum_{j=1}^{N}(\|b_{j3}\|^2+\cdots+\|b_{j9}\|^2) = O_p(T^{-1})+o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2\Big)$$

The above result, together with Lemma A.9 below, implies

$$\frac{1}{N}\sum_{j=1}^{N}(\hat{\Sigma}_{jje}-\Sigma_{jje})^2 = O_p(T^{-1})+o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2\Big)+O_p(\|\hat{\beta}-\beta\|^2)$$

Similar arguments are applicable for $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jjx}-\Sigma_{jjx}\|^2$. So we have

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2 &= O_p(T^{-1})+o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2\Big) \\ &\quad + O_p(\|\hat{\beta}-\beta\|^2)\end{aligned} \tag{A.38}$$

Substituting (A.31) into (A.38), we have $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2 = O_p(T^{-1}) + O_p(\|\hat{\beta}-\beta\|^2)$. Substituting this result into (A.31), we get the remaining result of Proposition A.2. This completes the proof of Proposition A.2. □

COROLLARY A.2. *Under Assumptions A-D and the identification conditions IB, we have*

$$\sum_{i=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H} = O_p(T^{-1/2})+O_p(\|\hat{\beta}-\beta\|)$$

LEMMA A.9. *Under Assumptions A-D,*

(*a*) $\frac{1}{N}\sum_{j=1}^{N}\|\mathcal{O}_{j1}\|^2 = o_p(\|\hat{\beta}-\beta\|^2)$

(*b*) $\frac{1}{N}\sum_{j=1}^{N}\|\mathcal{O}_{j2}\|^2 = o_p(\|\hat{\beta}-\beta\|^2)$

where $\mathcal{O}_{j1}$ is defined in (A.35) and $\mathcal{O}_{j2}$ defined in (A.37) above.

Proof of Lemma A.9. Consider (a). There are 9 terms on the left hand side of (A.34). We use $c_{j1}, c_{j2}, \cdots c_{j9}$ to denote them. By the Cauchy-Schwarz inequality, we have

$$\frac{1}{N}\sum_{j=1}^{N}\|O_{j1}\|^2 \leq 9\frac{1}{N}\sum_{j=1}^{N}(\|c_{j1}\|^2 + \|c_{j2}\|^2 + \cdots + \|c_{j9}\|^2)$$

Now we check the terms one by one. Consider $\frac{1}{N}\sum_{j=1}^{N}\|c_{j1}\|^2$. Term $\lambda_j' A' \frac{1}{T}\sum_{t=1}^{T} f_t \dot{x}_{jtp}$ is equal to $\operatorname{tr}[A'\frac{1}{T}\sum_{t=1}^{T} f_t \dot{x}_{jtp}\lambda_j']$. The term inside of the trace operator is bounded in norm by $\|A\| \cdot \|\frac{1}{T}\sum_{t=1}^{T} f_t \dot{x}_{jtp}\lambda_j'\|$. So we have

$$\frac{1}{N}\sum_{j=1}^{N}\|c_{j1}\|^2 \leq 4K\|\hat{\beta}-\beta\|^2 \cdot \|A\|^2 \frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T} f_t \dot{x}_{jtp}\lambda_j'\|^2$$

But $\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T} f_t \dot{x}_{jtp}\lambda_j'\|^2$ is $O_p(1)$ and $\|A\| = o_p(1)$. Thus $\frac{1}{N}\sum_{j=1}^{N}\|c_{j1}\|^2 = o_p(\|\hat{\beta}-\beta\|^2)$.

The second term $\frac{1}{N}\sum_{j=1}^{N}\|c_{j2}\|^2$ and the third term $\frac{1}{N}\sum_{j=1}^{N}\|c_{j3}\|^2$ are both $o_p(\|\hat{\beta}-\beta\|^2)$ given that $\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jtp}\dot{x}_{jtq}\|^2 = O_p(1)$ and $\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T} e_{jt}\dot{x}_{jtp}\|^2 = o_p(1)$.

The 4th, 5th and 7th terms are all $o_p(\|\hat{\beta}-\beta\|^2)$ which are implicitly implied by the results in Lemma A.6 and Proposition 2.1.

Consider c_{j6}. Term $\lambda_j'\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\dot{x}_{jtp}$ is equal to $\operatorname{tr}[\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\dot{x}_{jtp}\lambda_j']$. So we have

$$\frac{1}{N}\sum_{j=1}^{N}\|c_{j6}\|^2 \leq 4K\sum_{p=1}^{K}\|\hat{\beta}-\beta\|^2\frac{1}{N}\sum_{j=1}^{N}\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\dot{x}_{jtp}\lambda_j'\|^2$$

By $\dot{x}_{jtp} = \gamma_{jp}' f_t + \dot{v}_{jtp}$, the above term is further bounded by

$$\begin{aligned}
&16K\sum_{p=1}^{K}\|\hat{\beta}-\beta\|^2 \cdot \Big\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}f_t'\Big\|^2(\frac{1}{N}\sum_{j=1}^{N}\|\gamma_{jp}\|^2 \cdot \|\lambda_j\|^2) \\
&+16K\sum_{p=1}^{K}\|\hat{\beta}-\beta\|^2\frac{1}{N}\sum_{j=1}^{N}\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}v_{jtp} - E(\varepsilon_{it}v_{jtp})]\lambda_j'\|^2 \\
&+16K\sum_{p=1}^{K}\|\hat{\beta}-\beta\|^2\frac{1}{N}\sum_{j=1}^{N}\|\hat{H}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}E(\varepsilon_{jt}v_{jtp})\lambda_j'\|^2
\end{aligned}$$

$$+16K\sum_{p=1}^{K}\|\hat{\beta}-\beta\|^2\frac{1}{N}\sum_{j=1}^{N}\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\bar{v}_{jp}\lambda_j'\|$$

The first term of the above is $o_p(\|\hat{\beta}-\beta\|^2)$ by $\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}f_t' = O_p(T^{-1/2})$ which is implied by Lemma A.6(a). The second term is also $o_p(\|\hat{\beta}-\beta\|^2)$ because term $\frac{1}{N}\sum_{j=1}^{N}\|\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}v_{jtp}-E(\varepsilon_{it}v_{jtp})]\lambda_j'\|^2$ is bounded by

$$C\|\hat{H}_N\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{j=1}^{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}v_{jtp}-E(\varepsilon_{it}v_{jtp})]\lambda_j'\|^2\Big)$$

which is $O_p(T^{-1})$. The third term is also $o_p(\|\hat{\beta}-\beta\|^2)$ since $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$. The last term is of smaller order than the second one. Given these results, we have $\frac{1}{N}\sum_{j=1}^{N}\|c_{j6}\|^2 = o_p(\|\hat{\beta}-\beta\|^2)$.

The 8th term is $o_p(\|\hat{\beta}-\beta\|^2)$ which can be proved similarly as the 6th term. The last term is $o_p(\|\hat{\beta}-\beta\|^2)$, which can be proved similarly as the 7th term. We thus obtain (a).

Result (b) can be proved similarly as (a). □

Proof of Theorem 2.1. Consider the first order condition (2.5). Its pth element $(p=1,2,\cdots,K)$ is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\Big[(\dot{y}_{it}-\dot{x}_{it}\hat{\beta})-\hat{\lambda}_i'\hat{G}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\begin{bmatrix}\dot{y}_{jt}-\dot{x}_{jt}\hat{\beta}\\ \dot{x}_{jt}'\end{bmatrix}\Big]\dot{x}_{itp}=0$$

Using $\dot{y}_{it}-\dot{x}_{it}\hat{\beta}=\lambda_i'f_t+\dot{e}_{it}-\dot{x}_{it}(\hat{\beta}-\beta)$ and (A.18), the above equation is

$$\text{(A.39)}\quad \vartheta_{p1}(\hat{\beta}_1-\beta_1)+\vartheta_{p2}(\hat{\beta}_2-\beta_2)+\cdots+\vartheta_{pK}(\hat{\beta}_K-\beta_K)$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\lambda_i'f_t\dot{x}_{itp}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}e_{it}\dot{x}_{itp}$$
$$-\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\dot{x}_{itp}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\Gamma_j'f_t-\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\dot{x}_{itp}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\varepsilon_{jt}$$

where

(A.40)

$$\vartheta_{mn}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\dot{x}_{itm}\dot{x}_{itn}-\frac{1}{NT}\sum_{t=1}^{T}\Big(\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\dot{x}_{itm}\hat{\lambda}_i'\Big)\hat{G}\Big(\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j\dot{x}_{jtn}\Big)$$

for all $m, n = 1, 2, \cdots, K$. By $\dot{x}_{jtp} = \gamma_{jp}' f_t + \dot{v}_{jtp}$, equation (A.39) is equivalent to

$$
\begin{aligned}
&\vartheta_{p1}(\hat{\beta}_1 - \beta_1) + \vartheta_{p2}(\hat{\beta}_2 - \beta_2) + \cdots + \vartheta_{pK}(\hat{\beta}_K - \beta_K) \\
&= -\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip}(\hat{\lambda}_i - \lambda_i)' M_{ff}\Big] + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1}\lambda_i' f_t v_{itp} \\
&+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1}\gamma_{ip}' f_t e_{it} - \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip}\hat{\lambda}_i'\hat{G}\sum_{j=1}^N\sum_{t=1}^T \hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\varepsilon_{jt} f_t'\Big] \\
&+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip}\hat{\lambda}_i'\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma} - \Gamma)M_{ff}\Big] - \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1} f_t\hat{\lambda}_i' v_{itp}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\Big] \\
&+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1} e_{it} v_{itp} + \mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii}) I_{K+1}^{p+1}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big] \\
&-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}(\hat{\gamma}_{ip} - \gamma_{ip})\hat{\lambda}_i'\hat{G}\Big] - \mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip}\hat{\lambda}_i'\hat{G}\hat{M}_{ff}^{-1}(\hat{M}_{ff} - M_{ff})\Big] \\
&-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt} v_{itp} - E(\varepsilon_{jt} v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]
\end{aligned} \tag{A.41}
$$

where I_{K+1}^{p+1} is the $(p+1)$th column of the identity matrix I_{K+1}. We neglect term $\frac{1}{NT}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\bar{e}_i\bar{v}_{ip}$ and $\mathrm{tr}[\frac{1}{N}T\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\varepsilon_{jt}\bar{v}_{ip}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}]$ since the former is of smaller order than the 7th term and the latter is of smaller order than the last term. The first term on the right involves $\hat{\lambda}_j - \lambda_j$. However, the expression of $\hat{\lambda}_j - \lambda_j$ has been given in (A.29) (the first column). Using (A.29) to replace $\hat{\lambda}_j - \lambda_j$ from (A.41), we have

$$
\begin{aligned}
&\vartheta_{p1}(\hat{\beta}_1 - \beta_1) + \vartheta_{p2}(\hat{\beta}_2 - \beta_2) + \cdots + \vartheta_{pK}(\hat{\beta}_K - \beta_K) \\
&= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1} e_{it} v_{itp} - \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt} v_{itp} - E(\varepsilon_{jt} v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big] \\
&+\mathrm{tr}\Big[\hat{H}\frac{1}{T}\sum_{t=1}^T \hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon})\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}\hat{v}_p\Big] - \mathrm{tr}\Big[(I - A)'\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta} - \beta)'\hat{\xi}_t'\hat{H}_N\hat{v}_p\Big] \\
&-\mathrm{tr}\Big[\hat{H}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T \hat{\Sigma}_{jje}^{-1}[\varepsilon_{it} e_{jt} - E(\varepsilon_{it} e_{jt})]\gamma_{jp}'\Big] \\
&+\mathrm{tr}\Big[(I - A)'\sum_{q=1}^K(\hat{\beta}_q - \beta_q)\Big(\frac{1}{T}\sum_{t=1}^T f_t\frac{1}{N}\sum_{j=1}^N \dot{x}_{jtq}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big)\Big]
\end{aligned} \tag{A.42}
$$

$$+\mathcal{J}_{p1}+\mathcal{J}_{p2}$$

where $\hat{\upsilon}_p$ is defined in Table 1 and $\mathcal{J}_{p1}$ and $\mathcal{J}_{p2}$ are defined by

(A.43)

$$\mathcal{J}_{p1} = \operatorname{tr}\Big[\hat{G}^2(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\frac{1}{N}\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j\gamma_{jp}'\Big] + \operatorname{tr}\Big[A\frac{1}{N}\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}(\hat{\lambda}_j-\lambda_j)\gamma_{jp}'\Big]$$
$$-\operatorname{tr}\Big[\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)A\hat{\upsilon}_p\Big] + \operatorname{tr}\Big[\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}f_t\hat{\lambda}_j'v_{jtp}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma)\Big]$$
$$+\operatorname{tr}\Big[\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}f_t\hat{\lambda}_j'v_{jtp}\hat{G}\Big] - \operatorname{tr}\Big[\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}f_t(\hat{\lambda}_j-\lambda_j)'v_{jtp}\Big]$$
$$+\operatorname{tr}\Big[A'\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}f_t\gamma_{jp}'e_{jt}\Big] - \operatorname{tr}\Big[\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)\gamma_{ip}'\hat{G}\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\Big]$$
$$+\operatorname{tr}\Big[\hat{\upsilon}_p\hat{G}\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\Big] + \operatorname{tr}\Big[A'A\hat{\upsilon}_pM_{ff}\Big] - \operatorname{tr}\Big[A'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\hat{H}_N\hat{\upsilon}_p\Big]$$
$$-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\gamma}_{ip}-\gamma_{ip})\hat{\lambda}_i'\hat{G}\Big] + \operatorname{tr}\Big[\frac{1}{N}\hat{H}\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje}-\Sigma_{jje})\hat{\Sigma}_{jje}^{-1}\gamma_{jp}\Big]$$
$$+\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})I_{K+1}^{p+1}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big] - \operatorname{tr}\Big[\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}\hat{\upsilon}_p\Big]$$

and

$$\mathcal{J}_{p2} = -\operatorname{tr}\Big[\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\upsilon}_p\Big] - \operatorname{tr}\Big[\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\hat{\chi}_t'\Big)\hat{H}_N\hat{\upsilon}_p\Big]$$
$$+\operatorname{tr}\Big[\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\upsilon}_p\Big] + \operatorname{tr}\Big[\hat{H}_N\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'A\hat{\upsilon}_t\Big]$$
$$+\operatorname{tr}\Big[\hat{H}_N\sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t\frac{1}{N}\sum_{j=1}^{N}\dot{x}_{jtq}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big)\Big] \tag{A.44}$$
$$+\operatorname{tr}\Big[\hat{H}_N\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\Big(\frac{1}{N}\sum_{j=1}^{N}e_{jt}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big)\Big]$$
$$-\operatorname{tr}\Big[\hat{H}_N\sum_{q=1}^{K}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}_q-\beta_q)\Big(\frac{1}{N}\sum_{j=1}^{N}\dot{x}_{jtq}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big)\Big]$$

Equation (A.42) is dealt with in detail by the three lemmas below. The terms involving $\hat{\beta}-\beta$ are analyzed in Lemma A.10. The terms without $\hat{\beta}-\beta$ are summarized in Lemma A.12. The residual terms $\mathcal{J}_{p1}$ and $\mathcal{J}_{p2}$ are treated in Lemma A.11.

Let $\Omega_{pq} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}$, by Lemmas A.10, A.11 and A.12, we have

$$\begin{aligned}&\Omega_{p1}(\hat\beta_1-\beta_1)+\Omega_{p2}(\hat\beta_2-\beta_2)+\cdots+\Omega_{pK}(\hat\beta_K-\beta_K)\\&=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})+O_p(N^{-1}T^{-1/2})\end{aligned}$$

The above equation hold for all $p=1,2,\cdots,K$. So we have

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

Given the above result, in combination with Proposition A.2, we obtain the remaining three results of the theorem. This completes the proof of Theorem 2.1 □

LEMMA A.10. *Under Assumptions A-D, we have*

(a) $\vartheta_{pq} = \mathrm{tr}(\omega_{qp}) - \mathrm{tr}(\upsilon_p' Q^{-1}\upsilon_q) + \Omega_{pq} + o_p(1)$

$$(b)\quad \mathrm{tr}\Big[(I-A)'\Big(\frac{1}{T}\sum_{t=1}^T f_t(\hat\beta-\beta)'\hat\xi_t'\Big)\hat H_N\hat\upsilon_p\Big] = \sum_{q=1}^K(\hat\beta_q-\beta_q)\mathrm{tr}(\upsilon_q'Q^{-1}\upsilon_p)+o_p(\|\hat\beta-\beta\|)$$

$$(c)\quad \mathrm{tr}\Big[(I-A)'\frac{1}{T}\sum_{t=1}^T f_t(\hat\beta-\beta)'\Big(\frac{1}{N}\sum_{j=1}^N \dot x_{jt}'\hat\Sigma_{jje}^{-1}\gamma_{jp}'\Big)\Big] = \sum_{q=1}^K(\hat\beta_q-\beta_q)\mathrm{tr}(\omega_{qp})+o_p(\|\hat\beta-\beta\|)$$

where ϑ_{pq} is defined in (A.40), $\omega_{pq} = \frac{1}{N}\sum_{i=1}^N \gamma_{ip}\Sigma_{iie}^{-1}\gamma_{iq}'$, $\upsilon_p = \frac{1}{N}\sum_{i=1}^N \lambda_i\Sigma_{iie}^{-1}\gamma_{ip}'$ and $\Omega_{pq} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}$ with $\Sigma_{iix}^{(p,q)}$ being the (p,q) entry of the matrix Σ_{iix}.

PROOF OF LEMMA A.10. Consider (a). By the definition of ϑ_{pq},

$$\vartheta_{pq} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat\Sigma_{iie}^{-1}\dot x_{itp}\dot x_{itq} - \frac{1}{NT}\sum_{t=1}^T\Big(\sum_{i=1}^N\hat\Sigma_{iie}^{-1}\dot x_{itp}\hat\lambda_i'\Big)\hat G\Big(\sum_{j=1}^N\hat\Sigma_{jje}^{-1}\hat\lambda_j\dot x_{jtq}\Big)$$

Consider the first term of the above expression. It can be split into two terms (note that both $\hat\Sigma_{iie}$ and Σ_{iie} are scalars)

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\dot x_{itp}\dot x_{itq} - \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{\hat\Sigma_{iie}-\Sigma_{iie}}{\hat\Sigma_{iie}\Sigma_{iie}}\dot x_{itp}\dot x_{itq} = a_1 - a_2, \quad \text{say}$$

The term a_1 is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\gamma_{ip}'\gamma_{iq}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}'f_t v_{itq}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{iq}'f_t v_{itq}$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}(v_{itp}v_{itq}-\Sigma_{iix}^{(p,q)})+\frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}-\frac{1}{N}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\bar{v}_{ip}\bar{v}_{iq}$$

The first term of the above is $\mathrm{tr}(\omega_{qp})$ by the definition of ω_{qp}. The 2th-4th terms are $O_p(N^{-1/2}T^{-1/2})$. The 5th term is Ω_{pq}. The last term is $O_p(T^{-1})$. Thus, $a_1=\mathrm{tr}(\omega_{qp})+\Omega_{pq}+o_p(1)$.

The term a_2 is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}\gamma_{ip}'\gamma_{iq}+\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}\gamma_{ip}'\frac{1}{T}\sum_{t=1}^{T}f_t v_{itq}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}\Sigma_{iix}^{(p,q)}+\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}\frac{1}{T}\sum_{t=1}^{T}(v_{itp}v_{itq}-\Sigma_{iix}^{(p,q)})$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}\gamma_{iq}'\frac{1}{T}\sum_{t=1}^{T}f_t v_{itp}-\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}\bar{v}_{ip}\bar{v}_{iq}$$

By the boundedness of $\hat{\Sigma}_{iie},\Sigma_{iie},\Sigma_{iix}^{(p,q)},\gamma_{ip}$ and M_{ff}, the first and 3rd terms of the above expression are bounded in norm by

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

which is $o_p(1)$ by Proposition 2.1. The second term, by similar arguments, is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t v_{itq}\Big\|^2\Big)^{1/2}$$

which is $o_p(T^{-1/2})$. The 4th term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}v_{itp}v_{itq}-\Sigma_{iix}^{(p,q)}\Big|^2\Big)^{1/2}$$

which is $O_p(T^{-1})$. The 5th term can be proved to be $o_p(1)$ similarly as the second one. The last term is of smaller order than the 4th term. Given these results, we have $a_2=o_p(1)$.

Consider the term $\frac{1}{NT}\sum_{t=1}^{T}(\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\dot{x}_{itp}\hat{\lambda}_i')\hat{G}(\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j\dot{x}_{jtq})$, which is equal to

$$\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}\gamma_{jq}'\Big]+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{itp}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}\gamma_{jq}'\Big]$$

$$+\mathrm{tr}\Big[\sum_{i=1}^{N}\gamma_{jp}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}v_{jtq}f_t'\Big]+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\Sigma_{iix}^{(p,q)}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{itp}v_{jtq}-E(v_{itp}v_{jtq})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{G}\Big]$$

$$-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\bar{v}_{jp}\bar{v}_{jq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{G}\Big]=a_3+\cdots+a_8,\quad\text{say}$$

Consider a_3. The term $\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i$ is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\Sigma_{iie}^{-1}\lambda_i'-\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}(\Sigma_{iie}^{-1}-\hat{\Sigma}_{iie}^{-1})\lambda_i'+\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'$$

The first term is υ_p'. The second term can be proved to be bounded in norm by $C[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2]^{1/2}$ similarly as (A.14). The third term is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\gamma_{ip}\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\|\hat{\lambda}_i-\lambda_i\|^2\Big)^{1/2}$$

$$\leq\Big(\frac{1}{N}\sum_{i=1}^{N}\|\gamma_{ip}\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Sigma_{ii}^{-1}\|\|\hat{\Gamma}_i-\Gamma_i\|^2\Big)^{1/2}$$

which is $O_p(T^{-1/2})+O_p(\|\hat{\beta}-\beta\|)$ by Proposition A.2. Thus,

$$\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i=\upsilon_p+o_p(1),\qquad\qquad\hat{\upsilon}_p=\upsilon_p+o_p(1)\tag{A.45}$$

Note $\hat{G}_N-Q^{-1}=o_p(1)$, so we have $a_3=\mathrm{tr}[\upsilon_p'Q^{-1}\upsilon_q]+o_p(1)$.

Consider a_4. The term $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t v_{itp}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'$ is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{N}\sum_{t=1}^{N}f_t v_{itp}\Big\|^2\Big)^{1/2}$$

which is $O_p(T^{-1/2})$. So a_4 is $O_p(T^{-1/2})$ by (A.45) and $\hat{G}_N - Q^{-1} = o_p(1)$. The term a_5 can be proved similarly as a_4.

Consider a_6, which is bounded in norm by

$$C \sup_{i\le N} |\Sigma_{iix}^{(p,q)}| \frac{1}{N} \mathrm{tr}\Big[\Big(\sum_{i=1}^N \|\hat{H}^{1/2}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big) \cdot \|(\hat{H}^{1/2}\hat{M}_{ff}^{-1}\hat{H}^{1/2}+I)^{-1}\|\Big]$$

which is $O_p(N^{-1})$.

Consider a_7, the term inside the trace operator is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \big(\frac{1}{T}\sum_{t=1}^T [v_{itp}v_{jtq} - E(v_{itp}v_{jtq})]\big)^2\Big)^{1/2} \|\hat{G}_N\|$$

The above term is $O_p(T^{-1/2})$ by $\hat{G}_N - Q^{-1} = o_p(1)$. So we have $a_7 = O_p(T^{-1/2})$. Term a_8 is of smaller order than a_7.

Summarizing all the results, we obtain (a).

Consider (b). The term $\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\xi}_t'$ is equal to

$$\sum_{q=1}^K (\hat{\beta}_q - \beta_q)\Big(\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t\dot{x}_{jtq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\Big)$$

which can be split into

$$\sum_{q=1}^K (\hat{\beta}_q - \beta_q)\Big(\frac{1}{N}\sum_{j=1}^N \gamma_{jq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\Big) + \sum_{q=1}^K (\hat{\beta}_q - \beta_q)\Big(\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t v_{jtq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\Big)$$

The term $\frac{1}{N}\sum_{j=1}^N \gamma_{jq}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'$ is $\upsilon_q' + o_p(1)$ by (A.45). The second term is bounded in norm by

$$C\sum_{q=1}^K \|\hat{\beta}_q - \beta_q\| \Big(\frac{1}{N}\sum_{j=1}^N \|\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T f_t v_{jtq}\Big\|^2\Big)^{1/2}$$

which is $o_p(\|\hat{\beta}-\beta\|)$. So we have

$$\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\xi}_t' = \sum_{q=1}^K (\hat{\beta}_q-\beta_q)\upsilon_q' + o_p(\|\hat{\beta}-\beta\|)$$

Notice $(I-A)' \xrightarrow{p} I_r$ by Corollary A.1(a). Given this result, together with $\hat{H}_N \xrightarrow{p} Q^{-1}$ by Corollary A.1(b), result (b) follows.

Consider (c). The term $\sum_{p=1}^K(\hat\beta_p-\beta_p)\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t\dot x_{jtp}\hat\Sigma_{jje}^{-1}\gamma_{jp}'$ can be split into

$$\sum_{q=1}^K(\hat\beta_q-\beta_q)\frac{1}{N}\sum_{j=1}^N\gamma_{jq}\hat\Sigma_{jje}^{-1}\gamma_{jp}'+\sum_{q=1}^K(\hat\beta_q-\beta_q)\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_tv_{jtq}\hat\Sigma_{jje}^{-1}\gamma_{jp}'$$

The term $\frac{1}{N}\sum_{j=1}^N\gamma_{jq}\hat\Sigma_{jje}^{-1}\gamma_{jp}'$ has already been proved to be $\frac{1}{N}\sum_{j=1}^N\gamma_{jq}\Sigma_{jje}^{-1}\gamma_{jp}'$ $+o_p(1)=\omega_{qp}+o_p(1)$ in result (a). So the first term is $\sum_{q=1}^K(\hat\beta_q-\beta_q)\mathrm{tr}(\omega_{qp})+$ $o_p(1)$. The second term is bounded in norm by

$$C\sum_{q=1}^K\|\hat\beta_q-\beta_q\|\Big(\frac{1}{N}\sum_{j=1}^N\|\gamma_{jp}\hat\Sigma_{jje}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T f_tv_{jtq}\Big\|^2\Big)^{1/2}$$

which is $o_p(\|\hat\beta-\beta\|)$. Given these results, we have

$$\sum_{p=1}^K(\hat\beta_p-\beta_p)\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t\dot x_{jtp}\hat\Sigma_{jje}^{-1}\gamma_{jp}'=\sum_{q=1}^K(\hat\beta_q-\beta_q)\mathrm{tr}(\omega_{qp})+o_p(\|\hat\beta-\beta\|)$$

Notice $(I-A)'\xrightarrow{p}I_r$ by Corollary A.1(a), then (c) follows. □

LEMMA A.11. *Under Assumptions A-D, we have*

(*a*) $\mathcal{J}_{p1}=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat\beta-\beta\|)$

(*b*) $\mathcal{J}_{p2}=o_p(\|\hat\beta-\beta\|)$

for all p, where $\mathcal{J}_{p1}$ and $\mathcal{J}_{p2}$ are defined in (A.43) and (A.44), respectively.

PROOF OF LEMMA A.11. Consider (a). By definition, $\mathcal{J}_{p1}$ is composed of 15 terms which we denote by $b_1,b_2,\cdots,b_{15}$. We put aside the 2nd, 3rd and 10th terms temporarily.

Consider b_1. Notice that $\frac{1}{N}\sum_{i=1}^N\hat\Sigma_{iie}^{-1}\hat\lambda_i\gamma_{jp}'=O_p(1)$ by (A.45). So b_1 is $O_p(N^{-1}T^{-1/2})+o_p(\|\hat\beta-\beta\|)$ by Corollary A.2 and $\hat G=O_p(N^{-1})$. Consider b_4. The term $\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T\hat\Sigma_{jje}^{-1}\hat\lambda_j'f_tv_{jtp}$ is equal to

$$\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T\hat\Sigma_{jje}^{-1}(\hat\lambda_j-\lambda_j)'f_tv_{jtp}-\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T\frac{\hat\Sigma_{jje}-\Sigma_{jje}}{\hat\Sigma_{jje}\Sigma_{jje}}\lambda_j'f_tv_{jtp}$$

$$+\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T\Sigma_{jje}^{-1}\lambda_j'f_tv_{jtp}$$

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

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{jje}^{-1}\|\hat{\lambda}_j-\lambda_j\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_tv_{jtp}\Big\|^2\Big)^{1/2}$$

Note that $\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{jje}^{-1}\|\hat{\lambda}_j-\lambda_j\|^2$ is bounded by $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2$. So the first term is $O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$ by Proposition A.2. The second term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{j=1}^{N}(\hat{\Sigma}_{jje}-\Sigma_{jje})^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_tv_{jtp}\Big\|^2\Big)^{1/2}$$

Note that $\frac{1}{N}\sum_{j=1}^{N}(\hat{\Sigma}_{jje}-\Sigma_{jje})^2$ is bounded by $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2$. So the second term is also $O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$ by Proposition A.2. The third term is $O_p(N^{-1/2}T^{-1/2})$. Given these results, we have $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'f_tv_{jtp}$ $=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$. So b_4 is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$ by Corollary A.2.

Using the result in b_4, we have $b_5=O_p(N^{-3/2}T^{-1/2})+O_p(N^{-1}T^{-1})+o_p(\|\hat{\beta}-\beta\|)$. Consider b_6. Term $\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_tf_t'$, which is equal to $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{it}f_t'$, can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$ similarly as $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'f_tv_{jtp}$. So b_6 is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$ by Corollary A.2.

Since

$$\Big\|\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)\gamma_{ip}'\hat{G}\Big\|\le\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\|\hat{\lambda}_i-\lambda_i\|^2\Big)^{\frac{1}{2}}\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\|\gamma_{ip}\|^2\Big)^{\frac{1}{2}}\|\hat{G}_N\|,$$

we have $\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)\gamma_{ip}'\hat{G}=O_p(T^{-1/2})+O_p(\|\hat{\beta}-\beta\|)$. Given this result, together with $\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\lambda_i\gamma_{ip}'=O_p(1)$ by the boundedness of $\hat{\Sigma}_{iie},\lambda_i$ and γ_{ip}, by the similar arguments, we can prove that $b_8=O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$, $b_9=O_p(N^{-3/2}T^{-1/2})+O_p(N^{-1}T^{-1})+o_p(\|\hat{\beta}-\beta\|)$ and $b_{11}=O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$,

Consider the term $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}f_t\gamma_{jp}'e_{jt}$, which can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$ similarly as the term $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'f_tv_{jtp}$. Given this result, we have $b_7=O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$ by Corollary A.2.

Notice that $\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\gamma}_{ip}-\gamma_{ip})\hat{\lambda}_i'$ is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\gamma}_{ip}-\gamma_{ip}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\|\hat{\lambda}_i\|^2\Big)^{1/2}$$

which is $O_p(T^{-1/2}) + O_p(\|\hat{\beta} - \beta\|)$. Given this result, it follows that b_{12} is $O_p(N^{-1}T^{-1/2}) + o_p(\|\hat{\beta} - \beta\|)$.

The term $\frac{1}{N}\sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'$ is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Gamma}_i\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}$$

Notice that $\frac{1}{N}\sum_{i=1}^N \|\hat{\Gamma}_i\|^4 \leq 2^4(\frac{1}{N}\sum_{i=1}^N \|\Gamma_i\|^4 + \frac{1}{N}\sum_{i=1}^N \|\hat{\Gamma}_i - \Gamma_i\|^4)$. Using (A.29), it is easy to check that $\frac{1}{N}\sum_{i=1}^N \|\hat{\Gamma}_i - \Gamma_i\|^4 = o_p(1)$. So we have $\frac{1}{N}\sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i' = O_p(T^{-1/2}) + o_p(\|\hat{\beta} - \beta\|)$. From this, we have $b_{15} = O_p(N^{-1}T^{-1/2}) + o_p(\|\hat{\beta} - \beta\|)$. The terms b_{13} and b_{14} can be proved to be $O_p(N^{-1}T^{-1/2}) + o_p(\|\hat{\beta} - \beta\|)$ similarly as b_{15}.

We now consider the 2nd, 3rd and 10th terms. Using the first column of (A.29), the 3rd term can be verified to be $O_p(T^{-3/2}) + O_p(N^{-1/2}T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$. For the 2nd and 10th terms, substituting the first column of (A.29) into the 2nd term, we obtain an expression which is composed of 10 terms. The first of which is canceled out with the 10th term. The remaining expression can be proved to be $O_p(T^{-3/2}) + O_p(N^{-1/2}T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$. So we have that the 2nd, 3rd and 10th terms are $O_p(T^{-3/2}) + O_p(N^{-1/2}T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$. This completes the proof of (a).

The proof of (b) is quite similar to the results in Lemmas A.6, A.8 and A.10 and hence omitted. $\square$

LEMMA A.12. *Under Assumptions A-D, we have*

$$(a)\quad \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1} e_{it} v_{itp} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$$

$$(b)\quad \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{\Gamma}_j \hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt}v_{itp} - E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$$

$$(c)\quad \mathrm{tr}\Big[\hat{H}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_t\varepsilon_t' - \Sigma_{\varepsilon\varepsilon}]\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}\hat{v}_p\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$$

$$(d)\quad \mathrm{tr}\Big[\hat{H}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T \hat{\Sigma}_{jje}^{-1}[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\gamma_{jp}'\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$$

Proof of Lemma A.12. Consider (a). The left hand side of (a) is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{\Sigma}_{iie}-\Sigma_{iie}}{\hat{\Sigma}_{iie}\Sigma_{iie}}e_{it}v_{itp} \tag{A.46}$$

The first term is $O_p(N^{-1/2}T^{-1/2})$ and the second term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\Sigma}_{iie}-\Sigma_{iie})^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}e_{it}v_{itp}\Big|^2\Big)^{1/2}$$

which is $O_p(T^{-1})+O_p(T^{-1/2}\|\hat{\beta}-\beta\|)$ by Proposition A.2. Given these results, (a) follows.

The remaining three results are can be proved similarly as Lemma C.1(d) in [10] and hence omitted. □

Lemma A.12 is used to derive Theorem 2.1. Given Theorem 2.1, we can strengthen the results in Lemma A.12, which are stated in the following lemma. These results are helpful to derive Theorem 2.2.

Lemma A.13. *Under Assumptions A-D, we have*

(a)
$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}e_{it}v_{itp}$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

(b)
$$tr\Big[\hat{H}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_t\varepsilon_t'-E(\varepsilon_t\varepsilon_t')]\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}\hat{v}_j\Big]$$
$$=O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

(c)
$$tr\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$
$$=O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

(d)
$$tr\Big[\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big]$$
$$=O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

PROOF. Consider (a). Using (A.46), We show

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat\Sigma_{iie}-\Sigma_{iie}}{\hat\Sigma_{iie}\Sigma_{iie}}e_{it}v_{itp}=O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat\beta-\beta\|)$$

The left hand side is equal to

$$-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\hat\Sigma_{iie}\Sigma_{iie}}(e_{is}^2-\Sigma_{iie})e_{it}v_{itp}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\mathrm{Res}\{\hat\Sigma_{iie}-\Sigma_{iie}\}}{\hat\Sigma_{iie}\Sigma_{iie}}e_{it}v_{itp}$$

where $\mathrm{Res}\{\hat\Sigma_{iie}-\Sigma_{iie}\}$ is the right hand side expression of (A.34) excluding the first term. We separate the first term from the remaining expressions because the first term is $O_p(T^{-1/2})$ and the remainings are all $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. Consider the 2nd term of the above expression. By the boundedness of $\hat\Sigma_{ii}$ and Σ_{ii}, the 2nd term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{j=1}^{N}\|\mathrm{Res}\{\hat\Sigma_{iie}-\Sigma_{iie}\}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}e_{it}v_{itp}\Big|^2\Big)^{1/2}$$

Term $\frac{1}{N}\sum_{j=1}^{N}\|\mathrm{Res}\{\hat\Sigma_{iie}-\Sigma_{iie}\}\|^2$ is $O_p(T^{-2})+O_p(N^{-1}T^{-1})$ which can be verified term by term. Since this process is quite easy, we omit it. Thus the 2nd term is $O_p(T^{-3/2})+O_p(N^{-1/2}T^{-1})$ in view of $\frac{1}{T}\sum_{t=1}^{T}e_{it}v_{itp}=O_p(T^{-1/2})$.

The 1st term can be rewritten as

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\Sigma_{iie}^2}(e_{is}^2-\Sigma_{iie})e_{it}v_{itp}-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\hat\Sigma_{iie}-\Sigma_{iie}}{\hat\Sigma_{iie}\Sigma_{iie}^2}(e_{is}^2-\Sigma_{iie})e_{it}v_{itp}$$

The first expression is $O_p(N^{-1/2}T^{-1})$ and the second is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat\Sigma_{iie}-\Sigma_{iie})^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}(e_{is}^2-\Sigma_{iie})e_{it}v_{itp}\Big)^2\Big)^{1/2}$$

The above expression is $O_p(T^{-3/2})$. Thus the second expression is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$. So (a) follows.

Consider (b). Notice $\hat v_p=v_p+o_p(1)$ by (A.45) and $\hat H_N=Q^{-1}+o_p(1)$, then it suffices to consider the term

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat\Gamma_i\hat\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat\Sigma_{jj}^{-1}\hat\Gamma_j'$$

which can be rewritten as

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}(\hat{\Gamma}_j-\Gamma_j)'$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i(\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1})\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\Gamma_j'$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i\Sigma_{ii}^{-1}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')](\hat{\Sigma}_{jj}^{-1}-\Sigma_{jj}^{-1})\Gamma_j'$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i\Sigma_{ii}^{-1}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Sigma_{jj}^{-1}\Gamma_j' \tag{A.47}$$

The last term of (A.47) is $O_p(N^{-1}T^{-1/2})$. Consider the fourth term of (A.47), which can be written as

$$\frac{1}{N}\sum_{j=1}^{N}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Gamma_i\Sigma_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}-E(\varepsilon_{it}\varepsilon_{jt}')]\Big)(\hat{\Sigma}_{jj}^{-1}-\Sigma_{jj}^{-1})\Gamma_j'$$

By the boundedness of $\hat{\Sigma}_{jj},\Sigma_{jj},\Gamma_j$, the above is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Gamma_i\Sigma_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\Big)^{1/2}$$

which is $O_p(N^{-1/2}T^{-1})$ since $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2=O_p(T^{-1})$ by Theorem 2.1.

Consider the third term of (A.47), which is equal to

$$\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i(\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Sigma_{jj}^{-1}\Gamma_j'$$

$$+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i(\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')](\hat{\Sigma}_{jj}^{-1}-\Sigma_{jj})\Gamma_j'$$

The first expression is the same as the 4th term and hence $O_p(N^{-1/2}T^{-1})$. The second expression is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Gamma_i\|^2\cdot\|\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Big\|^2\Big)^{1/2}$$

Term $\frac{1}{N}\sum_{i=1}^{N}\|\Gamma_i\|^2\cdot\|\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1}\|^2$ is bounded in norm by $C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2=O_p(T^{-1})$ by the boundedness of $\Sigma_{ii},\hat{\Sigma}_{ii}$ and Γ_i. So the second expression is $O_p(T^{-3/2})$. Thus the third term of (A.47) is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$.

Consider the second term of (A.47), which is equal to

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i\Sigma_{ii}^{-1}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}(\hat{\Gamma}_j-\Gamma_j)'$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\Gamma_i(\hat{\Sigma}_{ii}^{-1}-\Sigma_{ii}^{-1})\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}(\hat{\Gamma}_j-\Gamma_j)'$$

By the boundedness of $\hat{\Sigma}_{jj},\Gamma_i$, the second expression is bounded in norm by

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

$$\times\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Big\|^2\Big)^{1/2}$$

Terms $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2$ and $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}^{-1}\|\cdot\|\hat{\Gamma}_j-\Gamma_j\|^2$ are both $O_p(T^{-1})$ by Theorem 2.1. So the second expression is $O_p(T^{-3/2})$. The first expression can be written as

$$\frac{1}{N}\sum_{j=1}^{N}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Gamma_i\Sigma_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Big)\hat{\Sigma}_{jj}^{-1}(\hat{\Gamma}_j-\Gamma_j)'$$

which is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Gamma_i\Sigma_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}^{-1}\|\cdot\|\hat{\Gamma}_j-\Gamma_j\|^2\Big)^{1/2}$$

which is $O_p(N^{-1/2}T^{-1})$ since $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}^{-1}\|\cdot\|\hat{\Gamma}_j-\Gamma_j\|^2=O_p(T^{-1})$ by Theorem 2.1. Given these two results, we have that the second term of (A.47) is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$.

Now consider the 1st term of (A.47), which is equal to

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}(\hat{\Gamma}_j-\Gamma_j)'$$

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\Gamma_j'$$

The second expression is the same as the 2nd term of (A.47) and hence $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$. The first expression is bounded in norm by

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

which is $O_p(T^{-3/2})$. Thus the 1st term of (A.47) is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$.

Summarizing all the results, we obtain (b).

Notice $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iie}-\Sigma_{iie}\|^2 \le \frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 = O_p(T^{-1})$ and $\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\|\hat{\lambda}_i-\lambda_i\|^2 \le \frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2 = O_p(T^{-1})$. So results (c) and (d) can be proved in the same way as result (b). This completes the proof of Lemma A.13.

□

Proofs of Theorem 2.2. Consider (A.42). Terms on the left hand side are summarized in Lemma A.10. The first three and the fifth terms on the right hand side are shown in Lemma A.13. The fourth and sixth are also covered by Lemma A.10. Terms $\mathcal{J}_{p1},\mathcal{J}_{p2}$ are given in Lemma A.11.Given these results, we have

(A.48)

$$\Omega(\hat{\beta}-\beta)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx}+O_p(T^{-3/2})+O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})$$

This proves Theorem 2.2. □

APPENDIX A.3: PROOF OF THE ALTERNATIVE EXPRESSION OF THEOREM 2.2

In this section, we show that the asymptotic expression of (2.9) is equal to the one in Theorem 2.2. We first introduce some notations. Let

$$\overline{\Gamma}_p=\begin{pmatrix}\mu_{1p} & \gamma_{1p}'\\ \mu_{2p} & \gamma_{2p}'\\ \vdots & \vdots\\ \mu_{Np} & \gamma_{Np}'\end{pmatrix}\quad \mathbb{F}=\begin{pmatrix}1 & f_1'\\ 1 & f_2'\\ \vdots & \vdots\\ 1 & f_T'\end{pmatrix}\quad \mathbb{V}_p=\begin{pmatrix}v_{11p} & v_{12p} & \cdots & v_{1Tp}\\ v_{21p} & v_{22p} & \cdots & v_{2Tp}\\ \vdots & \vdots & \ddots & \vdots\\ v_{N1p} & v_{N2p} & \cdots & v_{NTp}\end{pmatrix}$$

Then, by the second equation of (2.1), the p-th regressor can be written as:

$$X_p = \overline{\Gamma}_p \overline{\mathbb{F}}' + \mathbb{V}_p, \quad (p = 1, 2, \ldots, K) \tag{A.49}$$

Now consider term $\frac{1}{NT}\mathrm{tr}[\ddot{M} X_p \mathcal{M}(\overline{\mathbb{F}}) X_q']$, where $p, q = 1, 2, \cdots, K$. By (A.49), this term is equal to

$$\frac{1}{NT}\mathrm{tr}[\ddot{M} X_p \mathcal{M}(\overline{\mathbb{F}}) X_q'] = \frac{1}{NT}\mathrm{tr}[\ddot{M} \mathbb{V}_p \mathcal{M}(\overline{\mathbb{F}}) \mathbb{V}_q'] = \frac{1}{NT}\mathrm{tr}[\mathbb{V}_q' \Sigma_{ee}^{-1} \mathbb{V}_p]$$

$$-\frac{1}{NT}\mathrm{tr}[\Lambda' \Sigma_{ee}^{-1} \mathbb{V}_p \mathbb{V}_q' \Sigma_{ee}^{-1} \Lambda (\Lambda' \Sigma_{ee}^{-1} \Lambda)^{-1}] - \frac{1}{NT}\mathrm{tr}[\overline{\mathbb{F}}' \mathbb{V}_q' \Sigma_{ee}^{-1} \mathbb{V}_p \overline{\mathbb{F}} (\overline{\mathbb{F}}' \overline{\mathbb{F}})^{-1}]$$

$$+\frac{1}{NT}\mathrm{tr}[\overline{\mathbb{F}}' \mathbb{V}_q' \Sigma_{ee}^{-1} \Lambda (\Lambda' \Sigma_{ee}^{-1} \Lambda)^{-1} \Lambda' \Sigma_{ee}^{-1} \mathbb{V}_p \overline{\mathbb{F}} (\overline{\mathbb{F}}' \overline{\mathbb{F}})^{-1}] \tag{A.50}$$

The first term on the right hand side of (A.50) is equal to $\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \Sigma_{iie}^{-1} v_{itp} v_{itq}$, which is equal to

$$\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \Sigma_{iie}^{-1}[v_{itp} v_{itq} - \Sigma_{iix}^{(p,q)}] + \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1} \Sigma_{iix}^{(p,q)} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iie} \Sigma_{iix}^{(p,q)} + o_p(1)$$

because $\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \Sigma_{iie}^{-1}[v_{itp} v_{itq} - \Sigma_{iix}^{(p,q)}] = O_p(N^{-1/2}T^{-1/2})$.

Consider the 2nd term of (A.50), which is equal to $\mathrm{tr}[\frac{1}{NT}\Lambda' \Sigma_{ee}^{-1} \mathbb{V}_p \mathbb{V}_q' \Sigma_{ee}^{-1} \Lambda (\Lambda' \Sigma_{ee}^{-1} \Lambda)^{-1}]$. The term inside the trace operator is equal to

$$\Big(\frac{1}{N^2 T}\sum_{i=1}^N \sum_{j=1}^N \frac{1}{\Sigma_{iie}\Sigma_{jje}} \lambda_i \lambda_j' \sum_{t=1}^T v_{itp} v_{jtq}\Big)\Big(\frac{1}{N}\Lambda' \Sigma_{ee}^{-1} \Lambda\Big)^{-1}$$

which can be rewritten as

$$\Big(\frac{1}{N^2 T}\sum_{i=1}^N \sum_{j=1}^N \frac{1}{\Sigma_{iie}\Sigma_{jje}} \lambda_i \lambda_j' \sum_{t=1}^T [v_{itp} v_{jtq} - E(v_{itp} v_{jtq})]\Big)\Big(\frac{1}{N}\Lambda' \Sigma_{ee}^{-1} \Lambda\Big)^{-1}$$

$$+\Big(\frac{1}{N^2}\sum_{i=1}^N \frac{1}{\Sigma_{iie}^2} \lambda_i \lambda_i' \Sigma_{iix}^{(p,q)}\Big)\Big(\frac{1}{N}\Lambda' \Sigma_{ee}^{-1} \Lambda\Big)^{-1}$$

The first expression is $O_p(N^{-1}T^{-1/2})$ and the second is $O_p(N^{-1})$. So the 2nd term of (A.50) is $O_p(N^{-1}T^{-1/2}) + O_p(N^{-1})$.

Consider the 3rd term of (A.50), which is $\mathrm{tr}[\frac{1}{NT}\overline{\mathbb{F}}' \mathbb{V}_q' \Sigma_{ee}^{-1} \mathbb{V}_p \overline{\mathbb{F}} (\overline{\mathbb{F}}' \overline{\mathbb{F}})^{-1}]$. The term inside the trace operator is equal to

$$\Big(\frac{1}{NT^2}\sum_{i=1}^N \sum_{t=1}^T \sum_{s=1}^T \overline{f}_s \overline{f}_t' v_{itp} v_{isq}\Big)\Big(\frac{1}{T}\overline{\mathbb{F}}' \overline{\mathbb{F}}\Big)^{-1}$$

where $\overline{f}_t = (1, f_t')'$. The above expression can be rewritten as

$$\Big(\frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T \overline{f}_s\overline{f}_t'[v_{itp}v_{isq} - E(v_{itp}v_{isq})]\Big)\Big(\frac{1}{T}\overline{\mathbb{F}}'\overline{\mathbb{F}}\Big)^{-1} + \frac{1}{NT}\sum_{i=1}^N \Sigma_{iix}^{(p,q)}$$

The first expression is $O_p(N^{-1/2}T^{-1})$ and the second is $O_p(T^{-1})$. So the third term of (A.50) is $O_p(N^{-1/2}T^{-1}) + O_p(T^{-1})$.

Consider the last term of (A.50). Notice $\frac{1}{NT}\Lambda'\Sigma_{ee}^{-1}\mathbb{V}_p\overline{\mathbb{F}}$ is equal to

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\Sigma_{iie}}\lambda_i\overline{f}_t'v_{itp}$$

which is $O_p(N^{-1/2}T^{-1/2})$. Thus the last term of (A.50) is $O_p(N^{-1}T^{-1})$.

Given these results, we have

$$\frac{1}{NT}\mathrm{tr}[\ddot{M}X_p\mathcal{M}(\overline{\mathbb{F}})X_q'] = \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)} + o_p(1)$$

Then it follows

$$\frac{1}{NT}\begin{pmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{F}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{F}})X_K'] \\ \vdots & \vdots & \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{F}})X_1'] & \cdots & \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{F}})X_K'] \end{pmatrix} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iie}^{-1}\Sigma_{iix} + o_p(1) \tag{A.51}$$

Now we turn attention to $\frac{1}{NT}\mathrm{tr}[\ddot{M}X_p\mathcal{M}(\overline{\mathbb{F}})e']$, where $p = 1, 2, \cdots K$. By (A.49), we have

$$\frac{1}{NT}\mathrm{tr}[\ddot{M}X_p\mathcal{M}(\overline{\mathbb{F}})e'] = \frac{1}{NT}\mathrm{tr}[\ddot{M}\mathbb{V}_p\mathcal{M}(\overline{\mathbb{F}})e'] = \frac{1}{NT}\mathrm{tr}[e'\Sigma_{ee}^{-1}\mathbb{V}_p]$$
$$-\frac{1}{NT}\mathrm{tr}[\Lambda'\Sigma_{ee}^{-1}\mathbb{V}_pe'\Sigma_{ee}^{-1}\Lambda(\Lambda'\Sigma_{ee}^{-1}\Lambda)^{-1}] - \frac{1}{NT}\mathrm{tr}[\overline{\mathbb{F}}'e'\Sigma_{ee}^{-1}\mathbb{V}_p\overline{\mathbb{F}}(\overline{\mathbb{F}}'\overline{\mathbb{F}})^{-1}]$$
$$+\frac{1}{NT}\mathrm{tr}[\overline{\mathbb{F}}'e'\Sigma_{ee}^{-1}\Lambda(\Lambda'\Sigma_{ee}^{-1}\Lambda)^{-1}\Lambda'\Sigma_{ee}^{-1}\mathbb{V}_p\overline{\mathbb{F}}(\overline{\mathbb{F}}'\overline{\mathbb{F}})^{-1}]$$

The first term on the right hand side is equal to $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\Sigma_{iie}}v_{itp}e_{it}$, which is $O_p(N^{-1/2}T^{-1/2})$. The second term can be proved to be $O_p(N^{-1}T^{-1/2})$, similarly as the 2nd term of (A.50). The third term can be proved to be

$O_p(N^{-1/2}T^{-1})$ similarly as the 3rd term of (A.50). The last term can be proved to be $O_p(N^{-1}T^{-1})$ similarly as the 4th term of (A.50). So we have

$$\frac{1}{NT}\mathrm{tr}[\ddot{M}X_p\mathcal{M}(\overline{\mathbb{F}})e'] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\Sigma_{iie}}v_{itp}e_{it} + O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2})$$

Then it follows

$$\begin{bmatrix} \mathrm{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{F}})e'] \\ \vdots \\ \mathrm{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{F}})e'] \end{bmatrix} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\Sigma_{iie}}v_{itx}e_{it} + O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) \tag{A.52}$$

Combining (A.51) and (A.52), we obtain the same asymptotic expression as in Theorem 2.2. Taking limit on both sides of (A.51), we obtain Corollary 2.1.

It is interesting to note that the limiting distribution in Corollary 2.1 can be obtained by a generalized principal components method. Rewrite the model as $\mathbb{Y}_t = \mathbb{X}_t\beta + \Lambda f_t + e_t$, where $\mathbb{Y}_t = (y_{1t}, y_{2t}, \cdots, y_{Nt})'$, $\mathbb{X}_t$ is an $N \times K$ matrix with its (i,k) element x_{itk}. Recall that $\Sigma_{ee} = E(e_te_t')$. Suppose that Σ_{ee} is known. Consider the following objective function

$$SSR(\beta, \Lambda, F) = \sum_{t=1}^{T}(\mathbb{Y}_t - \mathbb{X}_t\beta - \Lambda f_t)'\Sigma_{ee}^{-1}(\mathbb{Y}_t - \mathbb{X}_t\beta - \Lambda f_t)$$

$$= \mathrm{tr}[(Y - X\beta - \Lambda F')'\Sigma_{ee}^{-1}(Y - X\beta - \Lambda F)] = \mathrm{tr}[(Y^\dagger - X^\dagger\beta - \Lambda^\dagger F')'(Y^\dagger - X^\dagger\beta - \Lambda^\dagger F)]$$

where X is a three dimensional data matrix such that $X\beta = X_1\beta_1 + \cdots + X_K\beta_K$ (with X_k being $N \times T$ for $k \leq K$); $Y^\dagger = \Sigma_{ee}^{-1/2}Y$, and $X^\dagger\beta = X_1^\dagger\beta_1 + \cdots + X_K^\dagger\beta_K$ with $X_k^\dagger = \Sigma_{ee}^{-1/2}X_k$; $\Lambda^\dagger = \Sigma_{ee}^{-1/2}\Lambda$. If α is also present, we use $\dot{\mathbb{Y}}_t$ and $\dot{\mathbb{X}}_t$, etc. This objective function is similar to that of [8], which uses $\Sigma_{ee} = I_N$. If Σ_{ee} is known, we can treat $Y^\dagger$ and $X^\dagger$ as the data and use the same estimation method as in [8]. Using the asymptotic representation in [8], it is not difficult to verify that the estimator obtained by minimizing the above objective function has the same asymptotic representation as in Corollary 2.1. Because Σ_{ee} is unknown, this leads naturally to an iterated two-step procedure. The first step estimates (β, Λ, F) the same way as in [8]. The second step constructs an estimate of Σ_{ee} based on the residuals and then reestimate (β, Λ, F); these two steps are iterated for a number of times. Despite the iteration, this is a two-step procedure (or a sequential procedure). In contrast, the maximum likelihood procedure is a joint procedure (simultaneous maximization). Also, the maximum likelihood procedure has much better finite sample properties.

APPENDIX B: TECHNICAL MATERIALS FOR SECTION 3

In addition to Table 1, we define the following notations, which will be used in the subsequent proof.

The symbols used in the Appendix B

υ_p^g	$= N^{-1}\sum_{i=1}^N \lambda_i \Sigma_{iie}^{-1}\gamma_{ip}^{g\prime}$	$\hat{\upsilon}_p^g$	$= N^{-1}\sum_{i=1}^N \lambda_i \hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{g\prime}$
ϱ_p	$= N^{-1}\sum_{i=1}^N \psi_i \Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}$	$\hat{\varrho}_p$	$= N^{-1}\sum_{i=1}^N \psi_i \hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}$
$\Pi_{\lambda\lambda}$	$= N^{-1}\sum_{i=1}^N \lambda_i \Sigma_{iie}^{-1}\lambda_i'$	$\hat{\Pi}_{\lambda\lambda}$	$= N^{-1}\sum_{i=1}^N \lambda_i \hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'$
$\Pi_{\lambda\psi}$	$= N^{-1}\sum_{i=1}^N \lambda_i \Sigma_{iie}^{-1}\psi_i'$	$\hat{\Pi}_{\lambda\psi}$	$= N^{-1}\sum_{i=1}^N \lambda_i \hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'$
$\Pi_{\psi\psi}$	$= N^{-1}\sum_{i=1}^N \psi_i \Sigma_{iie}^{-1}\psi_i'$	$\hat{\Pi}_{\psi\psi}$	$= N^{-1}\sum_{i=1}^N \psi_i \hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'$
$\mathcal{F}_t$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \Sigma_{iie}^{-1}\lambda_i'$	$\hat{\mathcal{F}}_t$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'$
$\mathcal{H}_t$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \Sigma_{iie}^{-1}\psi_i'$	$\hat{\mathcal{H}}_t$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'$
$\mathcal{G}_{tp}$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \Sigma_{iie}^{-1}\gamma_{ip}'$	$\hat{\mathcal{G}}_{tp}$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \hat{\Sigma}_{iie}^{-1}\gamma_{ip}'$
$\mathcal{G}_{tp}^g$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \Sigma_{iie}^{-1}\gamma_{ip}^{g\prime}$	$\hat{\mathcal{G}}_{tp}^g$	$= N^{-1}\sum_{i=1}^N \dot{x}_{it}' \hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{g\prime}$
ω_{pq}^{gg}	$= N^{-1}\sum_{i=1}^N \gamma_{ip}^g \Sigma_{iie}^{-1}\gamma_{iq}^{g\prime}$	$\hat{G}_{1N}$	$= N \cdot G_1$
ω_{pq}^{hh}	$= N^{-1}\sum_{i=1}^N \gamma_{ip}^h \Sigma_{iie}^{-1}\gamma_{iq}^{h\prime}$	$\hat{G}_{2N}$	$= N \cdot G_2$

where G_1 and G_2 are defined in the main context. Note that not all variables in the right column have "hat".

We then derive (3.9) in the main text:

$$\frac{1}{N}\hat{G}_2 \sum_{i=1}^N \sum_{j=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{B} M_{zz}^{ij} \hat{B}' - \hat{\Sigma}_{zz}^{ij})\hat{\Sigma}_{jj}^{-1} I_{K+1}^1 \hat{\psi}_j' = 0$$

The first order condition on Γ gives

$$\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{zz}^{-1} = W'$$

Notice $\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1} = \hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$ and $\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}$, we have

$$\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}][\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}] = W'$$

However, $\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = 0$ by (3.4). So the above equation can be simplified as

$$\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}[(I_N \otimes \hat{B})M_{zz}(I_N \otimes \hat{B}') - \hat{\Sigma}_{zz}]\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} = W'$$

By $\frac{1}{N}W'\hat{\Gamma} = 0$, we have

$$\frac{1}{N}\hat{G}_2 \sum_{i=1}^N \sum_{j=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{B} M_{zz}^{ij} \hat{B}' - \hat{\Sigma}_{zz}^{ij})\hat{\Sigma}_{jj}^{-1} I_{K+1}^1 \hat{\psi}_j' = 0$$

APPENDIX B1: PROOF OF CONSISTENCY

Again, for consistency, we use the superscript "*" to denote the true parameters.

PROPOSITION B.1. *Let $\hat{\theta} = (\hat{\beta}, \hat{\Gamma}, \hat{\Sigma}_{\varepsilon\varepsilon})$ be the solution by maximizing (3.2). Under Assumptions A-E, together with the identification conditions IZ, when $N, T \to \infty$, we have*

$$\hat{\beta} - \beta \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Gamma}_i - \Gamma_i)\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i)' \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \xrightarrow{p} 0$$

To prove Proposition B.1, we need the following lemmas.

LEMMA B.1. *Under Assumptions A-E,*

$$(a)\quad \hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t^* e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' = O_p(T^{-1/2})$$

$$(b)\quad \hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t^{*\prime}\Big)\hat{\Pi}_{\lambda\psi} = O_p(T^{-1/2})$$

$$(c)\quad \hat{G}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' = O_p(T^{-1/2})$$

$$(d)\quad \hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\Big(\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta} - \beta^*)'\hat{\mathcal{H}}_t\Big) = O_p(\|\hat{\beta} - \beta^*\|)$$

$$(e)\quad \hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta} - \beta^*)'\hat{\mathcal{H}}_t\Big) = o_p(\|\hat{\beta} - \beta^*\|)$$

$$(f)\quad \hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta} - \beta^*)f_t^{*\prime}\Big)\hat{\Pi}_{\lambda\psi} = O_p(\|\hat{\beta} - \beta^*\|)$$

$$(g)\quad \hat{G}_N\Big(\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta} - \beta^*)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big) = o_p(\|\hat{\beta} - \beta^*\|)$$

$$(h)\quad \hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta} - \beta^*)(\hat{\beta} - \beta^*)'\hat{\mathcal{H}}_t\Big) = O_p(\|\hat{\beta} - \beta^*\|^2)$$

$$(i)\quad \frac{1}{N}\hat{G}\sum_{i=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje}-\Sigma_{jje}^{*})\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' = o_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}^{*}\|^2\Big)$$

Proof of Lemma B.1. Consider (a). The left hand side of (a) is bounded, by the Cauchy-Schwarz inequality, in norm by

$$C\|\hat{H}^{1/2}\|\cdot\|(\hat{H}+I)^{-1}\|\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Gamma_i^{*\prime}\|^2\Big)^{1/2}$$

$$\cdot\Big(\sum_{j=1}^{N}\|\hat{\Sigma}_{jje}^{-1/2}\hat{\psi}_j\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t^{*}e_{jt}\Big\|^2\Big)^{1/2}$$

Note that $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$ by (A.1), $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\Gamma_i^{*\prime}\|^2 = O_p(1)$ by the boundedness of $\hat{\Sigma}_{ii}$, $\frac{1}{N}\sum_{j=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T}f_t^{*}e_{jt}\|^2 = O_p(T^{-1})$ and $\|(\hat{H}+I)^{-1}\|\le\sqrt{r}$. Further notice that $\|\hat{H}^{1/2}\|(\sum_{j=1}^{N}\|\hat{\Sigma}_{jje}^{-1/2}\hat{\psi}_j\|^2)^{1/2}$ is equal to

$$\Big[\mathrm{tr}(\hat{H})\mathrm{tr}\Big(\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{jje}^{-1}\hat{\psi}_i'\Big)\Big]^{1/2}\le 1$$

by the definition of $\hat{H}$. Given these results, we have (a).

Using the arguments in proving Lemmas A.5 and A.6, we can prove the remaining results similarly as (a). To save space, the detailed proofs are omitted. □

We point out that Lemmas A.3-A.6 continue to hold in the present context because the proofs don't involve the special structure of the factor loadings and the identification conditions.

Proof for Proposition B.1. The proof of Proposition 2.1, from the beginning to equation (A.24) except equations (A.28), continues to hold in the present context because this part of proof doesn't involve the priori restrictions as well as the identification conditions. In addition, (2.7) holds for the same reason. The following proof are based on these results.

By (A.24), together with $\hat{M}_{ff} = M_{ff} = I_r$, we have

$$I = (I-A)'(I-A) + o_p(1) \tag{B.1}$$

The above result, in combination with (A.16) indicates that

$$\|N^{1/2}\hat{H}^{1/2}\| = O_p(1) \tag{B.2}$$

Noticing $\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i\hat{\lambda}_i' + \frac{1}{N}\sum_{i=1}^N \hat{\gamma}_{ix}\hat{\Sigma}_{iix}^{-1}\hat{\gamma}_{ix}'$, we have $\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i\hat{\lambda}_i' = O_p(1)$, which implies

$$\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\hat{\psi}_i\hat{\psi}_i' = O_p(1) \tag{B.3}$$

Now consider the first order condition (3.5). Noticing $\hat{H} = \hat{G} + \hat{H}\hat{G}$, by (A.20), equation (3.5) is equivalent to

$$\begin{aligned}
\hat{\psi}_j - \psi_j^* = &-\hat{H}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i^*)'\lambda_j^* + \hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{T}\sum_{t=1}^T f_t^* e_{jt} \\
&+\hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t f_t^{*\prime}\Big)\lambda_j^* + \hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})] \\
&-\hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{T}\sum_{t=1}^T f_t^*\dot{x}_{jt}(\hat{\beta}-\beta) - \hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t\dot{x}_{jt}\Big)(\hat{\beta}-\beta) \\
&-\hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta^*)f_t^{*\prime}\Big)\lambda_j^* - \hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T \dot{l}_{it}(\hat{\beta}-\beta^*)e_{jt} \\
&+\hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\dot{x}_{jt}'\Big) - \hat{G}_1\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje} - \Sigma_{jje}^*) \\
&+\hat{H}_1\hat{G}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i^*)'\lambda_j^* + \hat{H}_1\hat{G}\hat{H}^{-1}(\hat{\lambda}_j - \lambda_j^*)
\end{aligned} \tag{B.4}$$

where we neglect the smaller order term $\hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\bar{\varepsilon}_i\bar{e}_j$. Furthermore, equation (3.9) can be written as

$$\begin{aligned}
\hat{H}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i^*)'\hat{\Pi}_{\lambda\psi} = &\,\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t f_t^{*\prime}\Big)\hat{\Pi}_{\lambda\psi} \\
&+\frac{1}{NT}\hat{G}_2\sum_{i=1}^N\sum_{j=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' \\
&+\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t^* e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' - \hat{G}_{2N}\Big(\sum_{t=1}^T \hat{\chi}_t(\hat{\beta}-\beta^*)'\hat{\mathcal{H}}_t\Big) \\
&-\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta^*)f_t^{*\prime}\Big)\hat{\Pi}_{\lambda\psi} + \hat{H}_2\hat{G}\hat{H}^{-1}\frac{1}{N}\sum_{j=1}^N(\hat{\lambda}_j - \lambda_j^*)\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'
\end{aligned} \tag{B.5}$$

$$-\hat{G}_{2N}\Big(\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big)+\hat{H}_2\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i^*)'\hat{\Pi}_{\lambda\psi}$$

$$+\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\hat{\mathcal{H}}_t\Big)-\frac{1}{N}\hat{G}_2\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje}-\Sigma_{jje}^*)\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$$

$$-\hat{G}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta}-\beta^*)'\hat{\mathcal{H}}_t$$

where $\hat{\Pi}_{\lambda\psi}=\frac{1}{N}\sum_{j=1}^{N}\lambda_j^*\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$ and $\hat{\mathcal{H}}_t=\frac{1}{N}\sum_{j=1}^{N}\dot{x}_{jt}'\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$.

We continue to use notation $A=(\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$. Except the 6th and 8th terms on the right hand side of (B.5), all the remaining are given in Lemma B.1. However, the 6th term is of smaller order than the left hand side and hence negligible. For the 8th term, it is bounded in norm by

$$CN^{-1/2}\|\hat{H}^{1/2}\|\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\lambda}_i\Sigma_{iie}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\lambda_i^*\|^2\Big)^{1/2}\|\hat{G}\hat{H}^{-1}\|$$

which is $o_p(1)$ by $\hat{H}=o_p(1)$. So we can write equation (B.5) alternatively, in terms of A, as

$$A_{12}'\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'=o_p(1) \tag{B.6}$$

where $A_{12}'=\hat{H}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}^g-\Gamma^{g*})'$ is the $(1,2)$th submatrix of A.

Consider (B.4). First post-multiplying $\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}$, then taking summation over index j from 1 to N, then dividing both sides by N, and by the similar arguments in proving Lemma B.1, we have $\frac{1}{N}\sum_{j=1}^{N}(\hat{\psi}_j-\psi_j^*)\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}=-A_{11}\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}+o_p(1)$, which can be written alternatively as

$$\frac{1}{N}\sum_{j=1}^{N}\hat{\psi}_j\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}=(I-A_{11})\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}+o_p(1)$$

Similarly, post-multiplying (B.4) by $\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$ instead of $\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}$, we have

$$\frac{1}{N}\sum_{j=1}^{N}\hat{\psi}_j\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'=(I-A_{11})\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'+o_p(1)$$

Combining the preceding two equations, we have

$$\frac{1}{N}\sum_{j=1}^{N}\hat{\psi}_j\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'=(I-A_{11})\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}(I-A_{11})'+o_p(1) \tag{B.7}$$

Notice both $\frac{1}{N}\sum_{j=1}^{N}\hat{\psi}_j\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$ and $\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}$ are of full rank, then $I-A_{11}$ is of full rank. Given this result, in combination with $\frac{1}{N}\sum_{j=1}^{N}\hat{\psi}_j\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime} = (I-A_{11})\frac{1}{N}\sum_{j=1}^{N}\psi_j^*\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime} + o_p(1)$, we have $\frac{1}{N}\sum_{j=1}^{N}\hat{\psi}_j\hat{\Sigma}_{jje}^{-1}\psi_j^{*\prime}$ is of full rank. Given this result, in combination with (B.6), we have $A_{12} = o_p(1)$. By $A_{12} = o_p(1)$, together with (B.1), we have $A_{21} = o_p(1)$ and $(I-A_{11})(I-A_{11})' = I + o_p(1)$ and $(I-A_{22})(I-A_{22})' = I + o_p(1)$. Given these results, in conjunction with the identification conditions $IZ2$, assuming the column signs are known, by Lemma A.2, we have $A_{11} = o_p(1)$ and $A_{22} = o_p(1)$. Then $A = o_p(1)$. The remaining proof is same as Proposition 2.1 and hence omitted. This completes the proof for consistencies of the estimates. □

Corollary B.1. *Under Assumptions A-E, together with the identification conditions IZ, we have*

$$(a)\quad \frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} - \frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^* = o_p(1); \qquad \frac{1}{N}(\hat{\Gamma}-\Gamma^*)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = o_p(1)$$

$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i' - \frac{1}{N}\sum_{i=1}^{N}\psi_i^*\Sigma_{iie}^{-1}\psi_i^{*\prime} = o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\psi}_i-\psi_i^*)\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i' = o_p(1)$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{ix}\hat{\Sigma}_{iix}^{-1}\hat{\gamma}_{ix}' - \frac{1}{N}\sum_{i=1}^{N}\gamma_{ix}^*\Sigma_{iix}^{*-1}\gamma_{ix}^{*\prime} = o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\gamma}_{ix}-\gamma_{ix}^*)\hat{\Sigma}_{iix}^{-1}\hat{\gamma}_{ix}' = o_p(1)$$

Proof of Corollary B.2. Results (a) and (b) have already been proved in the proof of Proposition B.1. Notice that

$$\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_j' + \frac{1}{N}\sum_{i=1}^{N}\hat{\gamma}_{ix}\hat{\Sigma}_{iix}^{-1}\hat{\gamma}_{ix}'$$

a similar expression also holds for $\frac{1}{N}\Gamma^{*\prime}\Sigma_{\varepsilon\varepsilon}^{*-1}\Gamma^*$. Using results (a) and (b), the first part of (c) follows immediately. The second par of (c) can be proved in the same way. So Corollary B.2 follows. □

APPENDIX B2: PROOF OF THE CONVERGENCE RATES

From now on, we drop the superscript "*" from the true parameters. Symbols with a hat represents the MLE estimators. Those without a hat denote the true parameters. To derive the convergence rates, we need the following lemma.

LEMMA B.2. *Let* $\mathcal{E} = O_p(T^{-1/2}) + O_p\big[\big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big)^{1/2}\big] + O_p\big(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Gamma}_i - \Gamma_i\|^2\big) + O_p(\|\hat{\beta} - \beta\|)$. *Under Assumptions A-E, we have*

$$A \equiv \sum_{i=1}^N (\hat{\Gamma}_i - \Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H} = \mathcal{E}$$

PROOF OF LEMMA B.2. By equation (3.4), (A.21) continues to hold in the present context. As remarked at the beginning of Appendix A, Lemmas A.5 and A.6 still hold. By Lemmas A.5 and A.6, in combination with Corollary B.2(a), we have

$$A + A' - AA' = \mathcal{E} \tag{B.8}$$

where $\mathcal{E}$ is defined earlier, which emphasize the order of magnitude. We partition the matrix A as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Then equation (B.8) is equivalent to

$$A_{11} + A_{11}' - A_{11}A_{11}' - A_{12}A_{12}' = \mathcal{E} \tag{B.9}$$
$$A_{12} + A_{21}' - A_{11}A_{21}' - A_{12}A_{22}' = \mathcal{E} \tag{B.10}$$
$$A_{21} + A_{12}' - A_{21}A_{11}' - A_{22}A_{12}' = \mathcal{E} \tag{B.11}$$
$$A_{22} + A_{22}' - A_{22}A_{22}' - A_{21}A_{21}' = \mathcal{E} \tag{B.12}$$

Again $\mathcal{E}$ signifies the order of magnitude. Equation (B.11) is equal to

$$A_{21}(I - A_{11}') + (I - A_{22})A_{12}' = \mathcal{E} \tag{B.13}$$

Lemma B.1 shows that the right hand side of (B.5) is $\mathcal{E}$. The left hand side of (B.5) is

$$A_{12}'\frac{1}{N}\sum_{j=1}^N \psi_j\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' = A_{12}'\Big[\frac{1}{N}\sum_{j=1}^N \psi_j\Sigma_{jje}^{-1}\psi_j' + o_p(1)\Big]$$

by Corollary B.2(b). Because $\frac{1}{N}\sum_{j=1}^N \psi_j\Sigma_{jje}^{-1}\psi_j' > 0$, we have $A_{12} = \mathcal{E}$. Notice $I - A_{11} \xrightarrow{p} I$ and $I - A_{22} \xrightarrow{p} I$ by $A \xrightarrow{p} 0$. Equation (B.13) implies $A_{21} = \mathcal{E}$.

Given $A_{12} = \mathcal{E}$ and $A_{21} = \mathcal{E}$, and $A_{11} \xrightarrow{p} 0, A_{22} \xrightarrow{p} 0$, equations (B.9) and (B.12) imply

$$A_{11} + A_{11}' = \mathcal{E}$$
$$A_{22} + A_{22}' = \mathcal{E}$$

Let $\hat{Q} = \frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}$ (recall in Assumption C, Q is defined as $\lim_{N\to\infty} \frac{1}{N}\Gamma'\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma$), $\hat{Q}_{11} = \frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$, $\hat{Q}_{12} = \frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^h$, $\hat{Q}_{21} = \frac{1}{N}\hat{\Gamma}^{h\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$, and $\hat{Q}_{22} = \frac{1}{N}\hat{\Gamma}^{h\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^h$. By the definition of $\hat{Q}$ and $\hat{H}_N$ ($H_N = N\hat{H}$), we have

$$\hat{H}_N = \begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix}^{-1}$$

This gives

$$\hat{H}_N = \begin{bmatrix} \hat{Q}_{11}^{-1} + \hat{Q}_{11}^{-1}\hat{Q}_{12}\hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{21}\hat{Q}_{11}^{-1} & -\hat{Q}_{11}^{-1}\hat{Q}_{12}\hat{Q}_{22\cdot 1}^{-1} \\ -\hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{21}\hat{Q}_{11}^{-1} & \hat{Q}_{22\cdot 1}^{-1} \end{bmatrix}$$

where $\hat{Q}_{22\cdot 1} = \hat{Q}_{22} - \hat{Q}_{21}\hat{Q}_{11}^{-1}\hat{Q}_{12}$. Furthermore, let $\mathcal{A} = \frac{1}{N}(\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}$ and $\mathcal{A}_{11} = \frac{1}{N}(\hat{\Gamma}^g - \Gamma^g)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$, $\mathcal{A}_{12} = \frac{1}{N}(\hat{\Gamma}^g - \Gamma^g)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^h$, $\mathcal{A}_{21} = \frac{1}{N}(\hat{\Gamma}^h - \Gamma^h)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g$, $\mathcal{A}_{22} = \frac{1}{N}(\hat{\Gamma}^h - \Gamma^h)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^h$. From $A = \mathcal{A}\hat{H}_N$, we have

$$A_{12} = -\mathcal{A}_{11}\hat{Q}_{11}^{-1}\hat{Q}_{12}\hat{Q}_{22\cdot 1}^{-1} + \mathcal{A}_{12}\hat{Q}_{22\cdot 1}^{-1}$$

$$A_{11} = \mathcal{A}_{11}(\hat{Q}_{11}^{-1} + \hat{Q}_{11}^{-1}\hat{Q}_{12}\hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{21}\hat{Q}_{11}^{-1}) + \mathcal{A}_{12}\hat{Q}_{22\cdot 1}^{-1}\hat{Q}_{21}\hat{Q}_{11}^{-1}$$

So we have $A_{11} = \mathcal{A}_{11}\hat{Q}_{11}^{-1} - A_{12}\hat{Q}_{21}\hat{Q}_{11}^{-1}$. By Corollary B.2(a), we have $\hat{Q}_{21} \xrightarrow{p} Q_{21}, \hat{Q}_{11} \xrightarrow{p} Q_{11}$. Given $A_{12} = \mathcal{E}$, we have

$$A_{11} = \mathcal{A}_{11}\hat{Q}_{11}^{-1} + \mathcal{E} \tag{B.14}$$

Substituting (B.14) into $A_{11} + A_{11}' = \mathcal{E}$, we have

$$\mathcal{A}_{11}\hat{Q}_{11}^{-1} + \hat{Q}_{11}^{-1\prime}\mathcal{A}_{11}' = \mathcal{E} \tag{B.15}$$

By similar aruguments, we also have

$$\mathcal{A}_{22}\hat{Q}_{22}^{-1} + \hat{Q}_{22}^{-1\prime}\mathcal{A}_{22}' = \mathcal{E}$$

The identification condition $IZ2$ is equal to

$$\begin{aligned} &\operatorname{ndiag}\Big\{\frac{1}{N}(\hat{\Gamma}^g - \Gamma^g)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}^g + \frac{1}{N}\hat{\Gamma}^{g\prime}\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}^g - \Gamma^g)\Big\} \\ &\quad = \operatorname{ndiag}\Big\{\frac{1}{N}(\hat{\Gamma}^g - \Gamma^g)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}^g - \Gamma^g) - \frac{1}{N}\Gamma^{g\prime}(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1})\Gamma^g\Big\} \end{aligned} \tag{B.16}$$

Notice $\frac{1}{N}(\hat{\Gamma}^g - \Gamma^g)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}^g - \Gamma^g)$ is $O_p(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Gamma}_i - \Gamma_i\|^2)$, and $\frac{1}{N}\Gamma^{g\prime}(\hat{\Sigma}_{\varepsilon\varepsilon}^{-1} - \Sigma_{\varepsilon\varepsilon}^{-1})\Gamma^g$ is $O_p([\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2})$. So (B.16) can be written, in terms of $\mathcal{A}$, as

$$\text{ndiag}(\mathcal{A}_{11} + \mathcal{A}_{11}') = \mathcal{E}$$

This equation together with (B.15) implies $\mathcal{A}_{11} = \mathcal{E}$ by using the arguments of Proposition A.1. Matrix A_{22} can be proved to be $\mathcal{E}$ similarly as A_{11}. Given the results of $A_{11}, A_{22}, A_{12}, A_{21}$, Lemma B.2 follows. □

Proposition B.2. *Under Assumptions A-E, together with the identification conditions IZ, we have*

$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Gamma}_i - \Gamma_i\|^2 = O_p(T^{-1}) + O_p(\|\hat{\beta} - \beta\|^2)$$

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

Proof of Proposition B.2. The first order condition (3.6) is equal to

$$\hat{\gamma}_{jx} - \gamma_{jx} = -A'\gamma_{jx} + \hat{H}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i'\frac{1}{T}\sum_{t=1}^T f_t v_{jtx}' - \hat{H}\hat{\gamma}_{jx}\hat{\Sigma}_{jjx}^{-1}(\hat{\Sigma}_{jjx} - \Sigma_{jjx})$$

$$+\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t f_t'\Big)\gamma_{jx} + \hat{H}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}v_{jtx}' - E(\varepsilon_{it}v_{jtx}')] \tag{B.17}$$

$$-\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta} - \beta)f_t'\Big)\gamma_{jx} - \hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta} - \beta)v_{jtx}'\Big)$$

where $A = (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$. We use $c_{j1}, c_{j2}, \cdots, c_{j7}$ to denote the 7 terms on the right hand side. Given $\hat{\Sigma}_{iix}^{-1}$ is bounded in a compact set by Assumption D, we have

$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{iix}^{-1}\| \cdot \|\hat{\gamma}_{ix} - \gamma_{ix}\|^2 \le C\frac{1}{N}\sum_{i=1}^N \|\hat{\gamma}_{ix} - \gamma_{ix}\|^2$$

$$\le 7C\frac{1}{N}\sum_{i=1}^N (\|c_{j1}\|^2 + \|c_{j2}\|^2 + \cdots + \|c_{j7}\|^2)$$

We can derive the bound for each term of the above equation. This process is similar to the derivation of (A.31) in Appendix A. Thus we state the result without writing out the details.

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iix}^{-1}\|\cdot\|\hat{\gamma}_{ix}-\gamma_{ix}\|^2 = O_p(T^{-1})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)+O_p(\|\hat{\beta}-\beta\|^2)$$

Also, By (B.4) we can derive a similar result for $\hat{\psi}_j - \psi_j$,

$$\frac{1}{N}\sum_{i=1}^{N}|\hat{\Sigma}_{iie}^{-1}|\cdot\|\hat{\psi}_i-\psi_i\|^2 = O_p(T^{-1})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)+O_p(\|\hat{\beta}-\beta\|^2)$$

Combining the above results and noticing that

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2 \le \frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iix}^{-1}\|\cdot\|\hat{\gamma}_{ix}-\gamma_{ix}\|^2+\frac{1}{N}\sum_{i=1}^{N}|\hat{\Sigma}_{iie}^{-1}|\cdot\|\hat{\psi}_i-\psi_i\|^2$$

we have

$$\text{(B.18)}\qquad \frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2 = O_p(T^{-1})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big) + O_p(\|\hat{\beta}-\beta\|^2)$$

Next, we study the first order condition (3.7). Using the fact that

$$\hat{B}M_{zz}^{ij}\hat{B}'-\hat{\Sigma}_{zz}^{ij} = \Gamma_i'\Gamma_j+\Gamma_i'\frac{1}{T}\sum_{t=1}^{T}f_t\varepsilon_{jt}'+\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}f_t'\Gamma_j+\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]$$

$$-\Gamma_i'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\dot{l}_{jt}'-\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}(\hat{\beta}-\beta)'\dot{l}_{jt}'-\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta)f_t'\Gamma_j$$

$$-\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta)\varepsilon_{jt}'+\frac{1}{T}\sum_{t=1}^{T}\dot{l}_{it}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{l}_{jt}'-\hat{\Gamma}_i'\hat{\Gamma}_j-1\{i=j\}(\hat{\Sigma}_{jj}-\Sigma_{jj})$$

we have

$$\hat{\Sigma}_{jje}-\Sigma_{jje} = \frac{1}{T}\sum_{t=1}^{T}(e_{jt}^2-\Sigma_{jje})+(\hat{\lambda}_j-\lambda_j)'(\hat{\lambda}_j-\lambda_j)-2\hat{\lambda}_j'\hat{G}(\hat{\lambda}_j-\lambda_j)$$

$$-2(\hat{\lambda}_j-\lambda_j)'\frac{1}{T}\sum_{t=1}^{T}f_te_{jt}+2\hat{\lambda}_j'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\frac{1}{T}\sum_{t=1}^{T}f_te_{jt}+2\hat{\lambda}_j'\hat{G}\frac{1}{T}\sum_{t=1}^{T}f_te_{jt}$$

$$-2\hat{\lambda}_j'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]+2\hat{\lambda}_j'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\lambda_j$$
$$-2\hat{\lambda}_j'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)\lambda_j+2\hat{\lambda}_j'\hat{G}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje}-\Sigma_{jje})$$
$$+\mathcal{V}_{j1}$$

where

$$\mathcal{V}_{j1}=-2\lambda_j'\frac{1}{T}\sum_{t=1}^{T}f_t\dot{x}_{jt}(\hat{\beta}-\beta)+\frac{1}{T}\sum_{t=1}^{T}\dot{x}_{jt}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{x}_{jt}'$$
$$+2\hat{\lambda}_j'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i'\frac{1}{T}\sum_{t=1}^{T}f_t\dot{x}_{jt}(\hat{\beta}-\beta)+2\hat{\lambda}_j'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t\dot{x}_{jt}\Big)(\hat{\beta}-\beta)$$
$$+2\hat{\lambda}_j'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)\lambda_j+2\hat{\lambda}_j'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)e_{jt}\Big)$$
$$-2\frac{1}{T}\sum_{t=1}^{T}e_{jt}\dot{x}_{jt}(\hat{\beta}-\beta)-2\hat{\lambda}_j'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\mathcal{F}}_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\dot{x}_{jt}'\Big)$$

We also have

$$\hat{\Sigma}_{jjx}-\Sigma_{jjx}=\frac{1}{T}\sum_{t=1}^{T}(v_{jtx}v_{jtx}'-\Sigma_{jjx})+(\hat{\gamma}_{jx}-\gamma_{jx})'(\hat{\gamma}_{jx}-\gamma_{jx})$$
$$+\gamma_{jx}'\sum_{i=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{G}\hat{\gamma}_{jx}+\hat{\gamma}_{jx}'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\gamma_{jx}-(\hat{\gamma}_{jx}-\gamma_{jx})'\hat{G}\hat{\gamma}_{jx}$$
$$-\hat{\gamma}_{jx}'\hat{G}(\hat{\gamma}_{jx}-\gamma_{jx})-(\hat{\gamma}_{jx}-\gamma_{jx})'\frac{1}{T}\sum_{t=1}^{T}f_tv_{jtx}'-\frac{1}{T}\sum_{t=1}^{T}v_{jtx}f_t'(\hat{\gamma}_{jx}-\gamma_{jx})$$
$$+\hat{\gamma}_{jx}'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\frac{1}{T}\sum_{t=1}^{T}f_tv_{jtx}'+\frac{1}{T}\sum_{t=1}^{T}v_{jtx}f_t'\sum_{i=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{G}\hat{\gamma}_{jx}$$
$$+\hat{\gamma}_{jx}'\hat{G}\sum_{t=1}^{T}f_tv_{jtx}'+\frac{1}{T}\sum_{t=1}^{T}v_{jtx}f_t'\hat{G}\hat{\gamma}_{jx}-\hat{\gamma}_{jx}'\hat{G}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}v_{jtx}'-E(\varepsilon_{it}v_{jtx}')]$$
$$-\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}[v_{jtx}\varepsilon_{it}'-v_{jtx}\varepsilon_{it}']\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{G}\hat{\gamma}_{jx}+(\hat{\Sigma}_{jjx}-\Sigma_{jjx})\hat{\Sigma}_{jjx}^{-1}\hat{\gamma}_{jx}'\hat{G}\hat{\gamma}_{jx}$$

$$+\hat{\gamma}_{jx}'\hat{G}\hat{\gamma}_{jx}\hat{\Sigma}_{jjx}^{-1}(\hat{\Sigma}_{jjx}-\Sigma_{jjx})-\hat{\gamma}_{jx}'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^T\hat{\chi}_tf_t'\Big)\gamma_{jx}-\gamma_{jx}'\Big(\frac{1}{T}\sum_{t=1}^Tf_t\hat{\chi}_t'\Big)\hat{G}_N\hat{\gamma}_{jx}$$
$$+\mathcal{V}_{j2}$$

where

$$\mathcal{V}_{j2}=\hat{\gamma}_{jx}'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)\gamma_{jx}+\gamma_{jx}'\Big(\frac{1}{T}\sum_{t=1}^Tf_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{G}_N\hat{\gamma}_{jx}$$

$$+\hat{\gamma}_{jx}'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)v_{jtx}'\Big)+\Big(\frac{1}{T}\sum_{t=1}^Tv_{jtx}(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{G}_N\hat{\gamma}_{jx}$$

Similarly as in Appendix A, we can prove $\frac{1}{N}\sum_{i=1}^N(\hat{\Sigma}_{iie}-\Sigma_{iie})^2=O_p(T^{-1})+o_p(\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2)+O_p(\|\hat{\beta}-\beta\|^2)$ and $\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{iix}-\Sigma_{iix}\|^2=O_p(T^{-1})+o_p(\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2)+O_p(\|\hat{\beta}-\beta\|^2)$. Moreover,

$$\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2=\frac{1}{N}\sum_{i=1}^N(\hat{\Sigma}_{iie}-\Sigma_{iie})^2+\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{iix}-\Sigma_{iix}\|^2.$$

We have

$$\frac{1}{N}\sum_{j=1}^N\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2=O_p(T^{-1})+o_p\Big(\frac{1}{N}\sum_{j=1}^N\|\hat{\Sigma}_{jj}^{-1}\|\cdot\|\hat{\Gamma}_j-\Gamma_j\|^2\Big)+O_p(\|\hat{\beta}-\beta\|^2) \tag{B.19}$$

Substituting (B.19) into (B.18), we get

$$\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{jj}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2=O_p(T^{-1})+O_p(\|\hat{\beta}-\beta\|^2)$$

Substituting the above result into (B.19), we have

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

This completes the proof of Proposition B.2. □

PROPOSITION B.3. *Under Assumptions A-E, we have*

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

To prove Proposition B.3, we need the following lemmas.

LEMMA B.3. *Under Assumptions A-E, together with the identification conditions IZ, we have*

(a) $\vartheta_{pq} = \mathrm{tr}(\omega_{qp}) - \mathrm{tr}(\upsilon_p' Q^{-1}\upsilon_q) + \Omega_{pq} + o_p(1)$

(b) $\mathrm{tr}\Big[\hat{G}_N \frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\hat{\upsilon}_p\Big] = \sum_{q=1}^{K}(\hat{\beta}_q - \beta_q)tr[\upsilon_p Q^{-1}\upsilon_q] + o_p(\|\hat{\beta}-\beta\|)$

(c) $\mathrm{tr}\Big[\hat{G}_1\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^{g}\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)tr(\omega_{pq}^{gg}) + o_p(\|\hat{\beta}-\beta\|)$

(d) $\mathrm{tr}\Big[(I-A)'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\hat{\upsilon}_p'\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)tr[\upsilon_p Q^{-1}\upsilon_q] + o_p(\|\hat{\beta}-\beta\|)$

(e) $\mathrm{tr}\Big[\hat{H}_N\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'(I-A)\hat{\upsilon}_p'\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)tr[\upsilon_p' Q^{-1}\upsilon_q] + o_p(\|\hat{\beta}-\beta\|)$

(f) $\mathrm{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]$

$$= \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)tr\Big[\varrho_q'\Pi_{\psi\psi}^{-1}\varrho_p\Big] + o_p(\|\hat{\beta}-\beta\|)$$

where the symbols $\omega_{pq}, \Pi_{\psi\psi}, \hat{\Pi}_{\psi\psi}, \upsilon_p, \hat{\upsilon}_p, \hat{\mathcal{H}}_t, \hat{\mathcal{G}}_{tp}^{g}$ *and* $\hat{\xi}_t$ *are all defined in Table 1 and 2, and*

$$\vartheta_{pq} = \frac{1}{NT}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\dot{x}_{itp}\dot{x}_{itq} - \frac{1}{NT}\sum_{t=1}^{T}\Big(\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\dot{x}_{itp}\hat{\lambda}_i'\Big)\hat{G}\Big(\sum_{j=1}^{N}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j\dot{x}_{itq}\Big).$$

PROOF OF LEMMA B.3. The proof of this lemma is quite similar to that of Lemma A.10. To save space, we only prove (b) and (f).

Consider (b). First we show $\hat{\upsilon}_p = \upsilon_p + o_p(1)$. Notice

$$\frac{1}{N}\sum_{j=1}^{N}\lambda_j\hat{\Sigma}_{jje}^{-1}\gamma_{jp}' = \frac{1}{N}\sum_{j=1}^{N}\lambda_j\Sigma_{jje}^{-1}\gamma_{jp}' + \frac{1}{N}\sum_{j=1}^{N}\lambda_j(\hat{\Sigma}_{jje}^{-1} - \Sigma_{jje}^{-1})\gamma_{jp}'$$

The second term is bounded in norm by $C[\frac{1}{N}\sum_{i=1}^{N}(\hat{\Sigma}_{iie}-\Sigma_{iie})^2]^{1/2}$, which is $o_p(1)$ by Proposition B.1. By the definition of υ_p we have $\hat{\upsilon}_p = \upsilon_p + o_p(1)$.

However, $\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'$ is equal to $\sum_{q=1}^{K}\upsilon_q(\hat{\beta}_q-\beta_q) + o_p(\|\hat{\beta}-\beta\|)$, which is given in result (b) in Lemma A.10. By $\hat{G}_N - Q^{-1} \xrightarrow{p} 0$, we have (b).

Consider (f). First, we can show $\hat{\varrho}_p = \varrho_p + o_p(1)$ similarly as $\hat{\upsilon}_p = \upsilon_p + o_p(1)$. Then for the term $\hat{\Pi}_{\psi\psi}$, by its definition, it's equal to

$$\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}\psi_i' + \frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}(\hat{\psi}_i - \psi_i)' + \frac{1}{N}\sum_{i=1}^{N}\psi_i(\hat{\Sigma}_{iie}^{-1} - \Sigma_{iie}^{-1})\hat{\psi}_i'$$

The second expression is bounded by $C[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i - \Gamma_i\|^2]^{1/2}$, and the third is bounded by $C[\frac{1}{N}\sum_{i=1}^{N}(\hat{\Sigma}_{iie} - \Sigma_{iie})^2]^{1/2}$. By Proposition B.1, these two terms are both $o_p(1)$. It follows that $\hat{\Pi}_{\psi\psi} = \Pi_{\psi\psi} + o_p(1)$.

Now consider the term $\frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\beta} - \beta)'\hat{\mathcal{H}}_t$. By the definition of $\hat{\mathcal{H}}_t$, it can be written as

$$\sum_{q=1}^{K}(\hat{\beta}_q - \beta_q)\Big(\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} f_t\dot{x}_{jtq}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big)$$

By $\dot{x}_{jtq} = \gamma_{jq}' f_t + \dot{\upsilon}_{jtq}$, the above expression is equal to

$$\sum_{q=1}^{K}(\hat{\beta}_q - \beta_q)\Big(\frac{1}{N}\sum_{j=1}^{N}\gamma_{jq}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big) + \sum_{q=1}^{K}(\hat{\beta}_q - \beta_q)\Big(\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} f_t\upsilon_{jtq}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big)$$

where $\dot{\upsilon}_{jtp}$ can be replaced with υ_{jtp} since $\frac{1}{T}\sum_{t=1}^{T} f_t = 0$. Notice $\frac{1}{N}\sum_{j=1}^{N}\gamma_{jq}$ $\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' = \frac{1}{N}\sum_{j=1}^{N}\gamma_{jq}\Sigma_{jje}^{-1}\psi_j' + o_p(1)$, which can be proved similarly as $\hat{\Pi}_{\psi\psi} = \Pi_{\psi\psi} + o_p(1)$, and $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} f_t\upsilon_{jtq}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j' = o_p(1)$. Given these results, we have

$$\frac{1}{T}\sum_{t=1}^{T} f_t(\hat{\beta} - \beta)'\hat{\mathcal{H}}_t = \sum_{q=1}^{K}(\hat{\beta}_q - \beta_q)\Big(\frac{1}{N}\sum_{j=1}^{N}\gamma_{jq}\Sigma_{jje}^{-1}\psi_j'\Big) + o_p(\|\hat{\beta} - \beta\|)$$

The above result, together with $\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma \xrightarrow{p} I_2 + o_p(1)$, $\hat{\varrho}_p = \varrho_p + o_p(1)$ and $\hat{\Pi}_{\psi\psi} = \Pi_{\psi\psi} + o_p(1)$, we have (f). $\square$

LEMMA B.4. *Under Assumptions A-E, together with the identification conditions IZ, we have*

$$(a)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it} + O_p(T^{-1}) + o_p(\|\hat{\beta} - \beta\|)$$

$$(b)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}e_{it}v_{itp}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$$

$$(c)\quad \hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_te_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p$$
$$=\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}h_te_{jt}\Sigma_{jje}^{-1}\psi_j'\Pi_{\psi\psi}^{-1}\varrho_p+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$$

$$(e)\quad \hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$$

$$(f)\quad \hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$$

$$(g)\quad \frac{1}{NT}\hat{G}_2\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p$$
$$=O_p(N^{-1}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$$

Proof of Lemma B.4. The proofs of the results in this lemma are quite similar to those of Lemma A.12. To save the space, we prove (c) as an illustration.

Notice $\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma=I_2+o_p(1)$ by Corollary B.2(a) and $\frac{1}{N}\sum_{j=1}^{N}\psi_j\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$ $=\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\psi_j'+o_p(1)$ by Lemma B.3(f). Term $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_te_{jt}$ $\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$ is $\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t\, e_{jt}\Sigma_{jje}^{-1}\psi_j'+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$. Term $\frac{1}{N}\sum_{j=1}^{N}$ $\psi_j\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{h\prime}$ is $\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\gamma_{jp}^{h\prime}+o_p(1)$. Given this result, (c) follows. □

Proof of Proposition B.3. By (3.8), equation (A.41) still holds in the present context. Consider the first term on the right hand side of (A.41). Notice $M_{ff}=I_r$ and $\hat{\lambda}_j=(\hat{\psi}_j',0_{r_2\times1}')'$. The first term is equal to $-\mathrm{tr}[\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}^{g}$ $\hat{\Sigma}_{iie}^{-1}(\hat{\psi}_i-\psi_i)']$, using (B.4) to replace $\hat{\psi}_i-\psi_i$, we can rewrite (A.41) as

$$\sum_{q=1}^{K}\vartheta_{pq}(\hat{\beta}_q-\beta_q)=-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\gamma_{ip}'f_t\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}_N\hat{\chi}_t \tag{B.20}$$

$$-\mathrm{tr}\Big[\hat G_1\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}\Gamma_i'\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_te_{jt}\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]-\mathrm{tr}\Big[\hat G_{1N}\Big(\frac{1}{T}\sum_{t=1}^T\hat\chi_tf_t'\Big)\hat\upsilon_p^g\Big]$$

$$+\mathrm{tr}\Big[\hat G_1\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}(\hat\Gamma_i-\Gamma_i)'\hat\upsilon_p^g\Big]+\mathrm{tr}\Big[\hat G_{1N}\frac{1}{T}\sum_{t=1}^T\hat\xi_t(\hat\beta-\beta)f_t'\hat\upsilon_p^g\Big]$$

$$+\mathrm{tr}\Big[\hat G_1\hat\Gamma'\hat\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma\Big(\frac{1}{T}\sum_{t=1}^Tf_t(\hat\beta-\beta)'\hat{\mathcal G}_{tp}^g\Big)\Big]+\mathrm{tr}\Big[\hat G_{1N}\Big(\frac{1}{T}\sum_{t=1}^T\hat\chi_t(\hat\beta-\beta)'\hat{\mathcal G}_{tp}^g\Big)\Big]$$

$$+\mathrm{tr}\Big[\hat G_{1N}\Big(\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T\hat\xi_t(\hat\beta-\beta)e_{jt}\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big)\Big]+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\gamma_{ip}'f_t\hat\Sigma_{iie}^{-1}e_{it}$$

$$-\mathrm{tr}\Big[\hat G_{1N}\frac{1}{T}\sum_{t=1}^T\hat\xi_t(\hat\beta-\beta)(\hat\beta-\beta)'\hat{\mathcal G}_{tp}^g\Big]+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^Tv_{itp}\hat\Sigma_{iie}^{-1}e_{it}$$

$$+\mathrm{tr}\Big[\hat G_1\frac{1}{N}\sum_{j=1}^N\hat\lambda_j\hat\Sigma_{jje}^{-1}(\hat\Sigma_{jje}-\Sigma_{jje})\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]-\mathrm{tr}\Big[\hat H_1\hat G\hat H^{-1}\frac{1}{N}\sum_{j=1}^N(\hat\lambda_j-\lambda_j)\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N\gamma_{ip}\hat\Sigma_{iie}^{-1}\hat\lambda_i'\mathbb A'\Big]-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^Tf_tv_{itp}\hat\Sigma_{iie}^{-1}(\hat\lambda_i-\lambda_i)'\Big]$$

$$-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N(\hat\gamma_{ip}-\gamma_{ip})\hat\Sigma_{iie}^{-1}\hat\lambda_i'\hat G\Big]+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^Tf_tv_{itp}\hat\Sigma_{iie}^{-1}\hat\lambda_i'\mathbb A'\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^Tf_tv_{itp}\hat\Sigma_{iie}^{-1}\hat\lambda_i'\hat G\Big]+\mathrm{tr}\Big[\frac{1}{N}\hat G\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}(\hat\Sigma_{ii}-\Sigma_{ii})I_{K+1}^{p+1}\hat\Sigma_{iie}^{-1}\hat\lambda_i'\Big]$$

$$-\mathrm{tr}\Big[\hat G_1\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\hat\Gamma_j\hat\Sigma_{jj}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat\Sigma_{iie}^{-1}\hat\lambda_i'\hat G\Big]$$

where $\hat\upsilon_p^g,\hat{\mathcal G}_{tp}^g,\hat\chi_t$ and $\hat\xi_t$ are defined in Tables 1 and 2; $\mathbb A=(\hat\Gamma-\Gamma)'\hat\Sigma_{\varepsilon\varepsilon}^{-1}\hat\Gamma\hat G$. All terms except the 4th and 14th term can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})+O_p(\|\hat\beta-\beta\|)$. The 4th and 14th terms are each $O_p(T^{-1/2})$ by Lemma B.2. But they share common components that are offset each other. To see this, by $\hat G=\hat H-\hat H\hat G$, the sum of the 4th and 14th terms is

$$\mathrm{tr}\Big[\hat G_1\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}(\hat\Gamma_i-\Gamma_i)'\hat\upsilon_p^g\Big]+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N\gamma_{ip}\hat\Sigma_{iie}^{-1}\hat\lambda_i'\mathbb A'\Big]\tag{B.21}$$

$$= \mathrm{tr}\Big[\mathbb{A}'\hat{\upsilon}_p\Big] - \mathrm{tr}\Big[\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\frac{1}{N}\sum_{j=1}^N \lambda_j\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{h\prime}\Big]$$

$$+ \mathrm{tr}[\hat{\upsilon}_p'\mathbb{A}'] + \mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \gamma_{ip}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'\mathbb{A}'\Big]$$

$$= \mathrm{tr}[(A+A')\hat{\upsilon}_p] - \mathrm{tr}[\hat{H}\mathbb{A}\hat{\upsilon}_p] + \mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \gamma_{ip}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'\mathbb{A}'\Big]$$

$$- \mathrm{tr}[A\hat{G}\hat{\upsilon}_p] - \mathrm{tr}\Big[\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\frac{1}{N}\sum_{j=1}^N \lambda_j\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{h\prime}\Big]$$

where $A = (\hat{\Gamma}-\Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$. The first term involves $A+A'$, However, by (A.21) and noting $\hat{M}_{ff} = M_{ff} = I_r$, we have

$$A + A' = A'A + (I-A)'\Big[\frac{1}{T}\sum_{t=1}^T f_t\hat{\chi}_t'\Big]\hat{H}_N \tag{B.22}$$

$$+\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t f_t'\Big](I-A) + \hat{H}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\epsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\Big]\hat{H}$$

$$-(I-A)'\Big[\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big]\hat{H}_N - \hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big]\hat{H}_N$$

$$-\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big](I-A) - \hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta)\hat{\chi}_t'\Big]\hat{H}_N$$

$$+\hat{H}_N\Big[\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\xi}_t'\Big]\hat{H}_N - \hat{H}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Sigma}_{\varepsilon\varepsilon}-\Sigma_{\varepsilon\varepsilon})\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}.$$

The last term of (B.21) is equal to $\mathrm{tr}[\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i^g-\Gamma_i^g)'\hat{\varrho}_p]$, which involves $\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i^g-\Gamma_i^g)'$. This term is implicitly given in (B.5). To see this, notice that the 8th term of (B.5) from the right hand side to the left, then the left hand side now is

$$\hat{H}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\hat{\Pi}_{\lambda\psi} - \hat{H}_2\hat{G}\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\hat{\Pi}_{\lambda\psi}$$

Notice $\hat{G}_2 = \hat{H}_2 - \hat{H}_2\hat{G}$ by $\hat{G} = \hat{H} - \hat{H}\hat{G}$. The above expression is

$$\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i)'\hat{\Pi}_{\lambda\psi} = \hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i^g-\Gamma_i^g)'\frac{1}{N}\sum_{j=1}^N \psi_j\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$$

which is $\hat{G}_2 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i^g - \Gamma_i^g)'\hat{\Pi}_{\psi\psi}$. Given this result, by (B.5), we have

$$
\begin{aligned}
\text{(B.23)} \quad \hat{G}_2 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i^g - \Gamma_i^g)' &= \hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\Big)\hat{\Pi}_{\psi\psi}^{-1} \\
&+\frac{1}{NT}\hat{G}_2 \sum_{i=1}^N\sum_{j=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1} \\
&+\hat{G}_2\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i'\Big(\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big)\hat{\Pi}_{\psi\psi}^{-1} - \hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)g_t'\Big) \\
&-\hat{G}_{2N}\Big(\sum_{t=1}^T \hat{\chi}_t(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\Big)\hat{\Pi}_{\psi\psi}^{-1} + \hat{H}_2\hat{G}\hat{H}^{-1}\frac{1}{N}\sum_{j=1}^N(\hat{\lambda}_j - \lambda_j)\hat{\Sigma}_{jje}^{-1}\hat{\phi}_j'\hat{\Pi}_{\psi\psi}^{-1} \\
&-\hat{G}_{2N}\Big(\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big)\hat{\Pi}_{\psi\psi}^{-1} \\
&+\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^T\hat{\chi}_t g_t'\Big) - \frac{1}{N}\hat{G}_2\sum_{j=1}^N\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje} - \Sigma_{jje})\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1} \\
&-\hat{G}_2\sum_{i=1}^N\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i'\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\hat{\Pi}_{\psi\psi}^{-1}
\end{aligned}
$$

Substituting (B.22) and (B.23) into (B.21), we obtain an alternative expression of $\mathrm{tr}[\hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i)'\hat{\upsilon}_p^g] + \mathrm{tr}[\frac{1}{N}\sum_{i=1}^N \gamma_{ip}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\mathbb{A}']$. Then equation (B.20) can be rewritten as

$$
\begin{aligned}
\vartheta_{p1}(\hat{\beta}_1 - \beta_1) + \vartheta_{p2}(\hat{\beta}_2-\beta_2) + \cdots + \vartheta_{pK}(\hat{\beta}_K - \beta_K) &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it} \\
&+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\hat{\Sigma}_{iie}^{-1}e_{it}v_{itp} - \mathrm{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\Big(\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_t e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\Big)\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big] \\
&+\mathrm{tr}\Big[\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)\hat{\upsilon}_p\Big] + \mathrm{tr}\Big[\hat{G}_1\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\Big(\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^g\Big)\Big] \\
&-\mathrm{tr}\Big[(I-A)'\Big(\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\upsilon}_p'\Big] \\
&-\mathrm{tr}\Big[\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^T\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)(I-A)\hat{\upsilon}_p'\Big] \\
\text{(B.24)} \quad &+\mathrm{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\Big(\frac{1}{T}\sum_{t=1}^T f_t(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\Big)\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]
\end{aligned}
$$

$$+\mathcal{S}_{p1}+\mathcal{S}_{p2}+\mathcal{S}_{p3}$$

where the symbols $\hat{\upsilon}_p, \hat{\upsilon}_p^g, \hat{\varrho}_p, \hat{\xi}_t, \hat{\Pi}_{\psi\psi}, \hat{\mathcal{H}}_t$ and $\mathcal{G}_{tp}^g$ are defined in Tables 1 and 2 and $\mathcal{S}_{p1}, \mathcal{S}_{p2}, \mathcal{S}_{p3}$ are given by

$$\mathcal{S}_{p1}=\operatorname{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}f_t\hat{\lambda}_i'v_{itp}\hat{G}\Big]+\operatorname{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}f_t\hat{\lambda}_i'v_{itp}\mathbb{A}'\Big]$$

$$-\operatorname{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'f_tv_{itp}\Big]+\operatorname{tr}\Big[\hat{H}_1\mathbb{A}'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}f_te_{it}\gamma_{ip}^{g\prime}\Big]$$

$$+\operatorname{tr}\Big[\hat{G}_1\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}f_te_{it}\gamma_{ip}^{g\prime}\Big]+\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})I_{K+1}^{p+1}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$

$$-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\gamma}_{ip}-\gamma_{ip})\hat{\lambda}_i'\hat{G}\Big]+\operatorname{tr}\Big[\hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}(\hat{\Sigma}_{iie}-\Sigma_{iie})\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{g\prime}\Big]$$

$$-\operatorname{tr}\Big[\hat{H}_1\hat{G}\hat{H}^{-1}\frac{1}{N}\sum_{i=1}^{N}(\hat{\lambda}_i-\lambda_i)\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{g\prime}\Big]-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}(\hat{\lambda}_i-\lambda_i)'\mathbb{A}'\Big]$$

$$-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}(\hat{\lambda}_i-\lambda_i)'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_tf_t'\Big)\Big]-\operatorname{tr}\Big[\hat{\upsilon}_p\hat{G}A'\Big]$$

$$+\operatorname{tr}\Big[\hat{\upsilon}_p\hat{H}\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_tf_t'\Big)\Big]-\operatorname{tr}\Big[\hat{\upsilon}_p\mathbb{A}\hat{H}\Big]+\operatorname{tr}\Big[\hat{\upsilon}_pA'A\Big] \tag{B.25}$$

$$+\operatorname{tr}\Big[\hat{\upsilon}_p\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\hat{H}_N\hat{G}\Big]-\operatorname{tr}\Big[\hat{\upsilon}_p\hat{H}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}\Big]$$

$$-\operatorname{tr}\Big[\hat{\upsilon}_pA'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\hat{H}_N\Big]-\operatorname{tr}\Big[\hat{\upsilon}_p\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_tf_t'\Big)A\Big]$$

$$+\operatorname{tr}\Big[I_1A'\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_te_{jt}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]-\operatorname{tr}\Big[\hat{H}_2\hat{G}\hat{H}^{-1}\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_j-\lambda_j)\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]$$

$$+\operatorname{tr}\Big[\frac{1}{N}\hat{G}_2\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje}-\Sigma_{jje})\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]$$

and

$$\mathcal{S}_{p2} = \operatorname{tr}\Big[\hat{G}_{1N}\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^{g}\Big] + \operatorname{tr}\Big[\hat{G}_1\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\sum_{j=1}^{N}e_{jt}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$-\operatorname{tr}\Big[\hat{G}_{1N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^{g}\Big] - \operatorname{tr}\Big[\hat{v}_p\hat{H}_N\frac{1}{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\Big]$$

$$-\operatorname{tr}\Big[\hat{v}_p\hat{H}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\hat{\chi}_t'\Big)\hat{H}_N\Big] + \operatorname{tr}\Big[\hat{v}_p\hat{H}_N\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\Big]$$

$$+\operatorname{tr}\Big[\hat{G}_2\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]$$

$$-\operatorname{tr}\Big[\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\Big)\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big] \tag{B.26}$$

$$+\operatorname{tr}\Big[\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\mathcal{H}}_t\Big)\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]$$

and

$$\mathcal{S}_{p3} = -\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{jt}v_{itp} - E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$

$$-\operatorname{tr}\Big[\hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big] \tag{B.27}$$

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

$$-\operatorname{tr}\Big[\frac{1}{NT}\hat{G}_2\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p\Big]$$

The expressions $\mathcal{S}_{p1}$ and $\mathcal{S}_{p2}$ are dealt with in Lemma B.5 below. The last four terms on the right hand side of (B.24) and Π_{pq} are summarized in Lemma B.3. The first three terms on the right hand side of (B.24) and $\mathcal{S}_{p3}$ are dealt with in Lemma B.4. Given these results, we have

$$\mathcal{P}_{p1}(\hat{\beta}_1-\beta_1) + \mathcal{P}_{p2}(\hat{\beta}_2-\beta_2) + \cdots + \mathcal{P}_{pK}(\hat{\beta}_K-\beta_K)$$

$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp} + \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it}$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i'\Pi_{\psi\psi}^{-1}\varrho_p h_t e_{it} + O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$$

for any $p = 1, 2, \cdots, K$. The above result is equivalent to

$$\mathcal{P}(\hat{\beta}-\beta) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ix}^{h\prime}h_t e_{it} \tag{B.28}$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i'\Pi_{\psi\psi}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\gamma_{jx}^{h\prime}\Big)h_t e_{it} + O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$$

which implies that $\hat{\beta}-\beta = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. This completes the proof of Proposition B.3. $\square$

Corollary B.2. *Under Assumptions A-E, together with the identification conditions IZ, we have*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2 = O_p(T^{-1})$$

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

This corollary is a direct result of Propositions B.2 and B.3.

Lemma B.5. *Under Assumptions A-E, together with the identification conditions IZ, we have*

(a) $\mathcal{S}_{p1} = O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$

(b) $\mathcal{S}_{p2} = o_p(\|\hat{\beta}-\beta\|)$

where $\mathcal{S}_{p1}$ and $\mathcal{S}_{p2}$ are defined in (B.25) and (B.26), respectively.

Proof of Lemma B.5. Using Proposition B.2, we can prove, just like Lemma A.11, all the terms in S_{p1} are $O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$. It is worth pointing out that, to prove $\operatorname{tr}[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}(\hat{\lambda}_i-\lambda_i)'\mathbb{A}']$ and $\operatorname{tr}[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}\lambda_i'A'A]$ to be $O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$, we need to strengthen Proposition B.2 to $A \equiv \sum_{i=1}^{N}(\hat{\Gamma}_i-\Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H} =$

$O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})+o_p(\|\hat{\beta}-\beta\|)$. Since our identification conditions are similar as IC3 in [10], this result can be proved by the same way as in their paper. We omit the details.

Result (b) is easier to prove. The details are omitted. □

APPENDIX B3: PROOF OF THEOREM 3.1 AND ITS ALTERNATIVE EXPRESSION

Note that (B.28) is close to Proposition 3.1 except that the remainder term $O_p(T^{-1}) + O_p(N^{-1}T^{-1/2})$ needs to be strengthened to $O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2})$. The strengthened results are stated in the following lemma.

LEMMA B.6. *Under Assumptions A-E, together with the identification conditions IZ, we have*

$$(a)\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it}$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it}+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

$$(b)\ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}e_{it}v_{itp}$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

$$(c)\ \hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p=\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}h_t e_{jt}\Sigma_{jje}^{-1}\psi_j'\Pi_{\psi\psi}^{-1}\varrho_p$$
$$+O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

$$(d)\ \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

$$(e)\ \hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

$$(f)\ \hat{H}\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

$$(g)\quad \frac{1}{NT}\hat{G}_2\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'\hat{\Pi}_{\psi\psi}^{-1}\hat{\varrho}_p$$
$$=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\beta}-\beta\|)$$

Proof of Lemma B.6. The whole proof of Lemma B.6 is very similar to that of Lemma A.13, we omit the proof to avoid repetition. □

Proof of Proposition 3.1. Given (B.24) and Lemmas B.3, B.5 and B.6, the proof of Proposition 3.1 is almost the same as that of Theorem 2.2. □

Proof of Theorem 3.1. To prove Theorem 3.1, we first show that for any Γ and f_t we can always transform them into new $\Gamma^\star$ and $f_t^\star$, which satisfy the identification conditions IZ. Consider Model (3.1), which we write out below for ease of reading:

$$y_{it}=\alpha_i+x_{it1}\beta_1+x_{it2}\beta_2+\cdots+x_{itK}\beta_K+\psi_i'g_t+e_{it}$$
$$x_{itk}=\mu_{ik}+\gamma_{ik}^{g\prime}g_t+\gamma_{ik}^{h\prime}h_t+v_{itk}$$

Now we rewrite the second equation of the above model as the following way

$$\begin{aligned}x_{itk}&=(\mu_{ik}+\gamma_{ik}^{g\prime}\bar{g}+\gamma_{ik}^{h\prime}\bar{h})+\gamma_{ik}^{g\prime}\dot{g}_t+\gamma_{ik}^{h\prime}\dot{h}_t+v_{itk}\\
&=(\mu_{ik}+\gamma_{ik}^{g\prime}\bar{g}+\gamma_{ik}^{h\prime}\bar{h})+\Big(\gamma_{ik}^{g\prime}+\gamma_{ik}^{h\prime}(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\Big)\dot{g}_t\\
&\quad+\gamma_{ik}^{h\prime}\Big(\dot{h}_t-(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t\Big)\\
&=(\mu_{ik}+\gamma_{ik}^{g\prime}\bar{g}+\gamma_{ik}^{h\prime}\bar{h})+\Big\{\Big(\gamma_{ik}^{g\prime}+\gamma_{ik}^{h\prime}(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\Big)(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q_1\Big\}\\
&\quad\times\Big\{Q_1'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1/2}\dot{g}_t\Big\}+\Big\{\gamma_{ik}^{h\prime}\Big[\dot{\mathbb{H}}'\dot{\mathbb{H}}-(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}(\dot{\mathbb{G}}'\dot{\mathbb{H}})\Big]^{1/2}Q_2\Big\}\\
&\quad\times\Big\{Q_2'\Big[\dot{\mathbb{H}}'\dot{\mathbb{H}}-(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}(\dot{\mathbb{G}}'\dot{\mathbb{H}})\Big]^{-1/2}\Big(\dot{h}_t-(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t\Big)\Big\}\end{aligned}$$

where Q_1 is an $r_1\times r_1$ and Q_2 is an $r_2\times r_2$ orthogonal matrix defined below. Similarly we can rewrite the first equation as

$$\begin{aligned}y_{it}&=\alpha_i+\sum_{k=1}^{K}x_{itk}\beta_k+\psi_i'g_t+e_{it}\\
&=(\alpha_i+\psi_i'\bar{g})+\sum_{k=1}^{K}x_{itk}\beta_k+\psi_i'\dot{g}_t+e_{it}\\
&=(\alpha_i+\psi_i'\bar{g})+\sum_{k=1}^{K}x_{itk}\beta_k+\Big\{\psi_i'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q_1\Big\}\Big\{Q_1'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1/2}\dot{g}_t\Big\}\end{aligned}$$

Let

(B.29) $\quad g_t^\star = Q_1'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1/2}\dot{g}_t$

(B.30) $\quad \psi_i^{\star\prime} = \psi_i'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q_1$

(B.31) $\quad \gamma_{ik}^{g\star\prime} = \Big(\gamma_{ik}^{g\prime} + \gamma_{ik}^{h\prime}(\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\Big)(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q_1$

(B.32) $\quad \gamma_{ik}^{h\star\prime} = \gamma_{ik}^{h\prime}\Big[\dot{\mathbb{H}}'\dot{\mathbb{H}} - (\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}(\dot{\mathbb{G}}'\dot{\mathbb{H}})\Big]^{1/2}Q_2$

(B.33) $\quad h_t^\star = Q_2'\Big[\dot{\mathbb{H}}'\dot{\mathbb{H}} - (\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}(\dot{\mathbb{G}}'\dot{\mathbb{H}})\Big]^{-1/2}\Big(\dot{h}_t - (\dot{\mathbb{H}}'\dot{\mathbb{G}})(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t\Big)$

Let $\Gamma^{g\star}, \Gamma^{h\star}, \Gamma^\star$ be defined similarly as $\Gamma^g, \Gamma^h, \Gamma$ in the main text. If matrix Q_1 is chosen to be the eigenvector matrix of $\Gamma^{g\star\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^{g\star}$ with the associated eigenvalues in descending order and Q_2 to be the eigenvector matrix of $\Gamma^{h\star\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^{h\star}$, we can easily verity $\Gamma^\star$ and $f_t^\star = (g_t^{\star\prime}, h_t^{\star\prime})'$ satisfy IZ.

Since $(\Gamma^\star, f_t^\star, \beta)$ satisfy IZ, by Proposition 3.1, we have

$$\begin{aligned}\mathcal{P}^\star(\hat{\beta}-\beta) &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Sigma_{iie}^{-1}e_{it}v_{itx} + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Sigma_{iie}^{-1}\gamma_{ix}^{h\star\prime}h_t^\star e_{it} \\ &- \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Sigma_{iie}^{-1}\psi_i^{\star\prime}\Pi_{\psi\psi}^{\star-1}\Big(\frac{1}{N}\sum_{j=1}^N \psi_j^\star\Sigma_{jje}^{-1}\gamma_{jx}^{h\star\prime}\Big)h_t^\star e_{it} \\ &+ O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})\end{aligned}$$

Substituting (B.29)–(B.33) into the above expression, we obtain the result as Theorem 3.1 state. □

Deriving the alternative expression of $\hat{\beta}-\beta$. We first introduce the following notations for ease of exposition. Let

$$\overline{\Gamma}_p^g = \begin{pmatrix} \mu_{1p} & \gamma_{1p}^{g\prime} \\ \mu_{2p} & \gamma_{2p}^{g\prime} \\ \vdots & \vdots \\ \mu_{Np} & \gamma_{Np}^{g\prime} \end{pmatrix} \quad \Gamma_p^h = \begin{pmatrix} \gamma_{1p}^{h\prime} \\ \gamma_{2p}^{h\prime} \\ \vdots \\ \gamma_{Np}^{h\prime} \end{pmatrix} \quad \mathbb{H} = \begin{pmatrix} h_1' \\ h_2' \\ \vdots \\ h_T' \end{pmatrix}$$

Then the second equation of (3.1) can be written as

(B.34) $\qquad X_p = \overline{\Gamma}_p^g\overline{\mathbb{G}}\,' + \Gamma_p^h\mathbb{H}\,' + \mathbb{V}_p, \quad p = 1, 2, \ldots, K$

Now consider term $\frac{1}{NT}\mathrm{tr}[\ddot{M}X_q\mathcal{M}(\overline{\mathbb{G}})X_p']$, where $p, q = 1, 2, \cdots, K$. By (B.34), this term is equal to

$$\frac{1}{NT}\mathrm{tr}\Big[\ddot{M}\Gamma_q^h\mathbb{H}\,'\mathcal{M}(\overline{\mathbb{G}})\mathbb{H}\Gamma_p^{h\prime}\Big] + \frac{1}{NT}\mathrm{tr}\Big[\ddot{M}\mathbb{V}_q\mathcal{M}(\overline{\mathbb{G}})\mathbb{H}\Gamma_p^{h\prime}\Big]$$

$$+\frac{1}{NT}\text{tr}\Big[\ddot{M}\Gamma_q^h\mathbb{H}'\mathcal{M}(\overline{\mathbb{G}})\mathbb{V}_p\Big]+\frac{1}{NT}\text{tr}\Big[\ddot{M}\mathbb{V}_q\mathcal{M}(\overline{\mathbb{G}})\mathbb{V}_p'\Big] \tag{B.35}$$

By the similar way to deal with $\text{tr}[\frac{1}{NT}\ddot{M}\mathbb{V}_p\mathcal{M}(\overline{\mathbb{F}})\mathbb{V}_q']$ in Theorem 2.2, we can prove the last term of (B.35) is equal to $\frac{1}{N}\sum_{i=1}^N\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}+o_p(1)$. Consider the second term of (B.35), which is equal to

$$\text{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\gamma_{ip}^h h_t' v_{itq}\Big]-\text{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\gamma_{ip}^h\overline{g}_t' v_{itq}\Big)(\overline{\mathbb{G}}'\overline{\mathbb{G}})^{-1}\overline{\mathbb{G}}'\mathbb{H}\Big]$$

$$-\text{tr}\Big[(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\psi_i h_t' v_{itq}\Big)\Big]$$

$$+\text{tr}\Big[(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\psi_i\overline{g}_t' v_{itq}\Big)(\overline{\mathbb{G}}'\overline{\mathbb{G}})^{-1}\overline{\mathbb{G}}'\mathbb{H}\Big]$$

The four terms of the above expression are all $O_p(N^{-1/2}T^{-1/2})$ by $(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}=O(1)$ and $(\overline{\mathbb{G}}'\overline{\mathbb{G}})^{-1}\overline{\mathbb{G}}'\mathbb{H}=O(1)$. So the second term of (B.35) is $O_p(N^{-1/2}T^{-1/2})$. The third term can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the second term. Given these results, we have

$$\frac{1}{NT}\text{tr}[\ddot{M}X_q\mathcal{M}(\overline{\mathbb{G}})X_p']=\frac{1}{NT}\text{tr}\Big[\ddot{M}\Gamma_q^h\mathbb{H}'\mathcal{M}(\overline{\mathbb{G}})\mathbb{H}\Gamma_p^{h\prime}\Big]+\frac{1}{N}\sum_{i=1}^N\Sigma_{iie}^{-1}\Sigma_{iix}^{(p,q)}+o_p(1)$$

Then it follows

$$\frac{1}{NT}\begin{pmatrix}\text{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \text{tr}[\ddot{M}X_1\mathcal{M}(\overline{\mathbb{G}})X_K'] \\ \vdots & \vdots & \vdots \\ \text{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_1'] & \cdots & \text{tr}[\ddot{M}X_K\mathcal{M}(\overline{\mathbb{G}})X_K']\end{pmatrix}=\mathcal{P}+o_p(1) \tag{B.36}$$

Now we turn attention to $\text{tr}[\frac{1}{NT}\ddot{M}X_p\mathcal{M}(\overline{\mathbb{G}})e']$, which is equal to

$$\frac{1}{NT}\text{tr}\Big[\ddot{M}\Gamma_p^h\mathbb{H}'\mathcal{M}(\overline{\mathbb{G}})e'\Big]+\frac{1}{NT}\text{tr}\Big[\ddot{M}\mathbb{V}_p\mathcal{M}(\overline{\mathbb{G}})e'\Big]. \tag{B.37}$$

The first term of the above expression is equal to

$$\text{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\gamma_{ip}^h h_t' e_{it}\Big]-\text{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\Sigma_{iie}^{-1}\gamma_{ip}^h\overline{g}_t' e_{it}\Big)(\overline{\mathbb{G}}'\overline{\mathbb{G}})^{-1}\overline{\mathbb{G}}'\mathbb{H}\Big]$$

$$-\text{tr}\Big[(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}\Big(\frac{1}{NT}\sum_{i=1}^N\Sigma_{iie}^{-1}\psi_i h_t' e_{it}\Big)\Big]$$

$$+\mathrm{tr}\Big[(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i\overline{g}_t'e_{it}\Big)(\overline{\mathbb{G}}'\overline{\mathbb{G}})^{-1}\overline{\mathbb{G}}'\mathbb{H}\Big]$$

Using $h_t' - \overline{g}_t'(\overline{\mathbb{G}}'\overline{\mathbb{G}})^{-1}\overline{\mathbb{G}}'\mathbb{H} = \dot{h}_t' - \dot{g}_t'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{\mathbb{G}}'\dot{\mathbb{H}}$, we can rewrite the above expression as

$$\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h}\dot{h}_t'e_{it}\Big] - \mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h}\dot{g}_t'e_{it}\Big)(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{\mathbb{G}}'\dot{\mathbb{H}}\Big]$$

$$-\mathrm{tr}\Big[(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\Sigma_{iie}^{-1}\psi_i\dot{h}_t'e_{it}\Big)\Big]$$

$$+\mathrm{tr}\Big[(\Gamma_p^{h\prime}\Sigma_{ee}^{-1}\Psi)(\Psi'\Sigma_{ee}^{-1}\Psi)^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i\dot{g}_t'e_{it}\Big)(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{\mathbb{G}}'\dot{\mathbb{H}}\Big]$$

which is equivalent to

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}(\dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t)e_{it}$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i'\Pi_{\psi\psi}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\gamma_{jp}^{h\prime}\Big)(\dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t)e_{it}$$

From this result and

$$\frac{1}{NT}\mathrm{tr}\Big[\ddot{M}\mathbb{V}_p\mathcal{M}(\overline{\mathbb{G}})e'\Big] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}+O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1}),$$

which has been shown in Appendix A.3, we have

$$\mathrm{tr}[\frac{1}{NT}\ddot{M}X_p\mathcal{M}(\overline{\mathbb{G}})e'] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp} \tag{B.38}$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}(\dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t)e_{it}$$

$$-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\psi_i'\Pi_{\psi\psi}^{-1}\Big(\frac{1}{N}\sum_{j=1}^{N}\psi_j\Sigma_{jje}^{-1}\gamma_{jp}^{h\prime}\Big)(\dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t)e_{it}$$

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

Combining (B.36) and (B.38), we can see that the alternative asymptotic expression of $\hat{\beta} - \beta$ is equivalent to the one in Theorem 3.1. In addition, Corollary 3.1 is an immediate result of (B.36). □

APPENDIX C: PROOFS OF RESULTS FOR MODELS WITH TIME-INVARIANT AND COMMON REGRESSORS

APPENDIX C1: PROOF OF CONSISTENCY

Again, for consistency, we use the superscript "*" to denote the true parameters.

Proposition C.1. *Let $\hat{\theta} = (\hat{\beta}, \hat{\Gamma}, \hat{\Sigma}_{\varepsilon\varepsilon}, \hat{M}_{ff})$ be the solution by maximizing (4.2). Under Assumptions A-D plus F, together with the identification conditions IO, when $N, T \to \infty$, we have*

$$\hat{\beta} - \beta^* \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Gamma}_i - \Gamma_i^*)\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i^*)' \xrightarrow{p} 0$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2 \xrightarrow{p} 0$$

$$\|\hat{M}_{ff} - M_{ff}^*\| \xrightarrow{p} 0$$

Proof of Proposition C.1. First note that $\hat{\beta} - \beta^* \xrightarrow{p} 0$ and $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2 \xrightarrow{p} 0$ continue to hold as they do not rely on priori restrictions. To prove the remaining results, it is sufficient to prove $A = (\hat{\Gamma} - \Gamma^*)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H} \xrightarrow{p} 0$. Equation (4.6) can be written as

$$\hat{G}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t^* e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j' + \hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t^{*\prime}\hat{\Pi}_{\lambda\lambda} \tag{C.1}$$

$$+\hat{G}_2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt} - E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j' - \hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)f_t^{*\prime}\hat{\Pi}_{\lambda\lambda}$$

$$-\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta^*)'\mathcal{F}_t\Big) - \hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^*\Big(\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta}-\beta^*)'\hat{\mathcal{F}}_t\Big)$$

$$-\hat{G}_2\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j' + \hat{G}_{2N}\Big(\frac{1}{T}\sum_{j=1}^{N}\hat{\xi}_t(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\hat{\mathcal{F}}_t\Big)$$

$$-\frac{1}{N}\hat{G}_2\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}(\hat{\Sigma}_{iie} - \Sigma_{iie}^*)\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i' + \hat{G}_2\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_j - \lambda_j^*)\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'$$

$$-\hat{G}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i-\Gamma_i^*)'M_{ff}^*\hat{\Pi}_{\lambda\lambda}-\hat{G}_2\hat{H}^{-1}(\hat{M}_{ff}-M_{ff}^*)\hat{\Pi}_{\lambda\lambda}=0$$

where $\hat{\Pi}_{\lambda\lambda}, \hat{\xi}_t$ and $\hat{\mathcal{F}}_t$ are defined in Tables 1 and 2. The last term involves $\hat{M}_{ff}-M_{ff}^*$. Notice (2.7) continues to hold, so (A.21) is applicable. By (A.21), we can rewrite (C.1) as

$$\hat{G}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i^{*\prime}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t^*e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'+\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_tf_t^{*\prime}\Big)\hat{\Pi}_{\lambda\lambda}$$

$$+\hat{G}_2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'-I_2\mathbb{A}'M_{ff}^*A\hat{\Pi}_{\lambda\lambda}$$

$$-\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta^*)'\hat{\mathcal{F}}_t\Big)-\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^*\Big(\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta}-\beta^*)'\hat{\mathcal{F}}_t\Big)$$

$$-\hat{G}_2\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'+\hat{G}_{2N}\frac{1}{T}\sum_{j=1}^{N}\hat{\xi}_t(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\hat{\mathcal{F}}_t$$

$$-\frac{1}{N}\hat{G}_2\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}(\hat{\Sigma}_{iie}-\Sigma_{iie}^*)\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'+\hat{G}_2\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_j-\lambda_j^*)\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'$$

$$-\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)f_t^{*\prime}\Big)\hat{\Pi}_{\lambda\lambda}-\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^*\Big(\frac{1}{T}\sum_{t=1}^{T}f_t^*\hat{\chi}_t'\Big)\hat{H}_N\hat{\Pi}_{\lambda\lambda} \tag{C.2}$$

$$-\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_tf_t^{*\prime}\Big)(I_r-A)\hat{\Pi}_{\lambda\lambda}-\hat{G}_2\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\epsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{\Pi}_{\lambda\lambda}$$

$$+\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma^*\Big(\frac{1}{T}\sum_{t=1}^{T}f_t^*(\hat{\beta}-\beta^*)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\Pi}_{\lambda\lambda}+\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta^*)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\Pi}_{\lambda\lambda}$$

$$+\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)f_t^{*\prime}\Big)(I_r-A)\hat{\Pi}_{\lambda\lambda}+\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)\hat{\chi}_t'\Big)\hat{H}_N\hat{\Pi}_{\lambda\lambda}$$

$$-\hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta^*)(\hat{\beta}-\beta^*)'\hat{\xi}_t'\hat{H}_N\hat{\Pi}_{\lambda\lambda}-\hat{G}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii}^*)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}\hat{\Pi}_{\lambda\lambda}$$

$$-\hat{H}_2\hat{M}_{ff}^{-1}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}\hat{\Gamma}M_{ff}^*A\hat{\Pi}_{\lambda\lambda}+I_2M_{ff}^*A\hat{\Pi}_{\lambda\lambda}=0$$

where $\mathbb{A}=(\hat{\Gamma}-\Gamma^*)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{G}$ and $A=(\hat{\Gamma}-\Gamma^*)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}$. There are 22 terms on the left hand side of (C.2). We can prove that all the terms except the 11th and 22th terms are $o_p(1)$, given the fact that $\hat{\beta}-\beta^* \xrightarrow{p} 0$ and $\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}^*\|^2 \xrightarrow{p} 0$, whose proofs, as pointed out in Appendix B, involve no priori restrictions and still hold in present context. By the definitions of $\mathbb{A}$ and A, the 11th term is equal to

$$\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma^*)M_{ff}^*A\hat{\Pi}_{\lambda\lambda} = \hat{H}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma^*)M_{ff}^*A\hat{\Pi}_{\lambda\lambda}$$
$$-\hat{H}_2\hat{M}_{ff}^{-1}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}(\hat{\Gamma}-\Gamma^*)M_{ff}^*A\hat{\Pi}_{\lambda\lambda}$$

where we have used $\hat{G} = \hat{H} - \hat{H}\hat{M}_{ff}^{-1}\hat{G}$. Since (A.21) continues to hold in Section 4, we have $A = O_p(1)$. By (A.16), we have $\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = O_p(1)$. By $\frac{1}{N}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma} = O_p(1)$, it is easy to verify that $\hat{\Pi}_{\lambda\lambda} = \frac{1}{N}\sum_{j=1}^N \lambda_j^*\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j' = O_p(1)$. Given $\hat{\Pi}_{\lambda\lambda} = O_p(1)$ and $A = O_p(1)$, together with $\hat{G}\hat{H}^{-1} = O_p(1), \hat{M}_{ff}^{-1} = O_p(1), \hat{H} = o_p(1)$, we have

$$I_2(I_r - A)'M_{ff}^*A\hat{\Pi}_{\lambda\lambda} \xrightarrow{p} 0$$

However, term $\hat{\Pi}_{\lambda\lambda} = \frac{1}{N}\sum_{j=1}^N \lambda_j^*\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'$ can be proved to be invertible similarly as $\frac{1}{N}\sum_{j=1}^N \psi_j^*\hat{\Sigma}_{jje}^{-1}\hat{\psi}_j'$ in Appendix B. So we have

$$I_2(I_r - A)'M_{ff}^*A \xrightarrow{p} 0$$

From this result, in combination with $\hat{M}_{ff} = (I_r - A)'M_{ff}^*(I_r - A) + o_p(1)$, we obtain $A_{12} = o_p(1)$ and $A_{22} = o_p(1)$. These results, in combination with $\hat{M}_{ff} = (I_r - A)'M_{ff}^*(I_r - A) + o_p(1)$ again and $IO2$, we have $A_{21} = o_p(1)$ and $A_{11} = o_p(1)$. The remaining proof is the same as that of Proposition 2.1. This completes the proof of Proposition C.1. □

APPENDIX C2: PROOFS OF THE CONVERGENCE RATES AND PROPOSITION 4.1

Now we drop "*" from the true value of the parameters for notational simplicity. The following lemma is useful for deriving the rates of convergence.

LEMMA C.1. *Under Assumptions A-D plus G, in combination with the identification conditions IO, we have*

$$\sum_{i=1}^N(\hat{\Gamma}_i - \Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H} = O_p(T^{-1/2}) + O_p([\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2})$$
$$+O_p(\frac{1}{N}\sum_{i=1}^N\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i - \Gamma_i\|^2) + O_p(\|\hat{\beta}-\beta\|)$$

PROOF OF LEMMA C.1. The proof of Lemma C.1 is similar to that of Lemma B.2. We partition matrix A into $A_{11}, A_{12}, A_{21}, A_{22}$ and prove each submatrix is $O_p(T^{-1/2}) + O_p([\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2}) + O_p(\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Gamma}_i - \Gamma_i\|^2) + O_p(\|\hat{\beta} - \beta\|)$. The details are omitted. □

PROPOSITION C.2. *Under Assumptions A-D plus G, together with the identification conditions IO, we have*

$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Gamma}_i - \Gamma_i\|^2 = O_p(T^{-1}) + O_p(\|\hat{\beta} - \beta\|^2)$$

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

$$\|\hat{M}_{ff} - M_{ff}\|^2 = O_p(T^{-1}) + O_p(\|\hat{\beta} - \beta\|^2)$$

PROOF OF PROPOSITION C.2. The first order condition with respect to ψ_j is identical to the one in the last section. By (B.4), we have

$$\begin{aligned}
\hat{\psi}_j - \psi_j = & -\hat{G}_1 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} (\hat{\Gamma}_i - \Gamma_i)' M_{ff} \lambda_j + \hat{G}_1 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \Gamma_i' \frac{1}{T}\sum_{t=1}^T f_t e_{jt} \\
& + \hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t f_t'\Big)\lambda_j + \hat{G}_1 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \frac{1}{T}\sum_{t=1}^T [\varepsilon_{it} e_{jt} - E(\varepsilon_{it} e_{jt})] \\
& - \hat{G}_1 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1} \Gamma_i' \frac{1}{T}\sum_{t=1}^T f_t (\hat{\beta} - \beta)' \dot{x}_{jt}' - \hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\chi}_t (\hat{\beta} - \beta)' \dot{x}_{jt}'\Big) \\
& - \hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t (\hat{\beta} - \beta) f_t'\Big)\lambda_j - \hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t (\hat{\beta} - \beta) e_{jt}\Big) \\
& + \hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^T \hat{\xi}_t (\hat{\beta} - \beta)(\hat{\beta} - \beta)' \dot{x}_{jt}'\Big) - \hat{G}_1 \hat{\lambda}_j \hat{\Sigma}_{jje}^{-1} (\hat{\Sigma}_{jje} - \Sigma_{jje}) \\
& - \hat{H}_1 \hat{M}_{ff}^{-1} \hat{G} \hat{H}^{-1} (\hat{M}_{ff} - M_{ff}) \hat{\lambda}_j - \hat{H}_1 \hat{M}_{ff}^{-1} \hat{G} \hat{H}^{-1} M_{ff} (\hat{\lambda}_j - \lambda_j)
\end{aligned} \tag{C.3}$$

The remaining proof is the same as that of Proposition B.2 and the details are hence omitted. □

PROPOSITION C.3. *Under Assumption A-D plus G, we have*

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

To prove Proposition C.3, we need the following lemmas.

LEMMA C.2. *Under Assumptions A-D plus G, together with the identification conditions IO, we have*

$$(a)\quad \operatorname{tr}\Big[\hat{G}_1\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^{g}\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)\operatorname{tr}[I_1M_{ff}\omega_{qp}I_1'] + o_p(\|\hat{\beta}-\beta\|)$$

$$(b)\quad \operatorname{tr}\Big[I_1(I-A)'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\hat{\upsilon}_p^{g}\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)\operatorname{tr}[I_1M_{ff}\upsilon_q'Q^{-1}\upsilon_pI_1'] + o_p(\|\hat{\beta}-\beta\|)$$

$$(c)\quad \operatorname{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\hat{\upsilon}_p^{h}\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)\operatorname{tr}[I_2M_{ff}\upsilon_q'Q^{-1}\upsilon_pI_2'] + o_p(\|\hat{\beta}-\beta\|)$$

$$(d)\quad \operatorname{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\mathcal{F}}_t\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^{h}\Big] = \sum_{q=1}^{K}(\hat{\beta}_q-\beta_q)\operatorname{tr}\Big[I_2M_{ff}\upsilon_q'\Pi_{\lambda\lambda}^{-1}\upsilon_pI_2'\Big] + o_p(\|\hat{\beta}-\beta\|)$$

where $\hat{\Pi}_{\lambda\lambda}, \Pi_{\lambda\lambda}, \hat{\upsilon}_p^g, \hat{\upsilon}_p^h, \hat{\upsilon}_p, \upsilon_p, \omega_{pq}, \mathcal{G}_{tp}^g, \hat{\xi}_t$ *are defined in Tables 1, 2 and 3.* I_1 *denotes the first* r_1 *rows and* I_2 *denotes the remaining* r_2 *rows of the identity matrix* I_r.

LEMMA C.3. *Under Assumptions A-D plus G, together with the identification conditions IO, we have*

$(a)\quad \mathbb{S}_{p1} = o_p(N^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$

$(b)\quad \mathbb{S}_{p2} = o_p(\|\hat{\beta}-\beta\|)$

where $\mathbb{S}_{p1}$ *and* $\mathbb{S}_{p2}$ *are defined in (C.6) below.*

LEMMA C.4. *Under Assumptions A-D plus G, together with the identification conditions IO, we have*

$$(a)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(b)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}e_{it}v_{itp} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(c)\quad \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}h_t e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{v}_p^h$$
$$= O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(d)\quad \operatorname{tr}\Big[\hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$
$$= O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(e)\quad \operatorname{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$
$$= O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(f)\quad \operatorname{tr}\Big[\hat{H}_1\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{v}_p^g\Big]$$
$$= O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(g)\quad \operatorname{tr}\Big[\hat{G}_2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{v}_p^h\Big]$$
$$= O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(h)\quad \operatorname{tr}\Big[\hat{G}_2\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{v}_p^h\Big]$$
$$= O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat{\beta}-\beta\|)$$

The proofs of Lemmas C.2, C.3 and C.4 are quite similar as the counterparts in Appendix B.

PROOF OF PROPOSITION C.3. As argued in Section 4, equation (2.5) still holds. So equation (A.41) continues to hold. Consider the first term on the right hand side of (A.41). Notice that $-\operatorname{tr}[\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'M_{ff}] = -\operatorname{tr}[\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}^{g}\hat{\Sigma}_{iie}^{-1}(\hat{\psi}_i-\psi_i)']$. Using (C.3) to replace $\hat{\psi}_i-\psi_i$ from the expression, we have

$$\vartheta_{p1}(\hat\beta_1-\beta_1)+\vartheta_{p2}(\hat\beta_2-\beta_2)+\cdots+\vartheta_{pK}(\hat\beta_K-\beta_K)$$

$$=\operatorname{tr}\Big[\hat G_1\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}(\hat\Gamma_i-\Gamma_i)'M_{ff}\hat\upsilon_p^g\Big]-\operatorname{tr}\Big[\hat G_1\hat\Gamma'\hat\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma\frac{1}{NT}\sum_{j=1}^N\sum_{t=1}^T f_te_{jt}\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$-\operatorname{tr}\Big[\hat G_{1N}\Big(\frac{1}{T}\sum_{t=1}^T\hat\chi_tf_t'\Big)\hat\upsilon_p^g\Big]+\operatorname{tr}\Big[\hat G_1\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}\Gamma_i'\frac{1}{T}\sum_{t=1}^T f_t(\hat\beta-\beta)'\hat{\mathcal G}_{tp}^g\Big]$$

$$+\operatorname{tr}\Big[\hat G_{1N}\Big(\frac{1}{T}\sum_{t=1}^T\hat\chi_t(\hat\beta-\beta)'\hat{\mathcal G}_{tp}^g\Big)\Big]+\operatorname{tr}\Big[\hat G_{1N}\Big(\frac{1}{T}\sum_{t=1}^T\hat\xi_t(\hat\beta-\beta)f_t'\Big)\hat\upsilon_p^g\Big]$$

$$+\operatorname{tr}\Big[\hat G_1\frac{1}{T}\sum_{j=1}^N\sum_{t=1}^T\hat\xi_t(\hat\beta-\beta)e_{jt}\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]-\operatorname{tr}\Big[\hat G_{1N}\Big(\frac{1}{T}\sum_{t=1}^T\hat\xi_t(\hat\beta-\beta)(\hat\beta-\beta)'\hat{\mathcal G}_{tp}^{g\prime}\Big)\Big]$$

$$+\operatorname{tr}\Big[\frac{1}{N}\hat G_1\sum_{j=1}^N\hat\lambda_j\hat\Sigma_{jje}^{-1}(\hat\Sigma_{jje}-\Sigma_{jje})\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\hat\Sigma_{iie}^{-1}\lambda_i'f_tv_{itp}$$

$$+\operatorname{tr}\Big[\hat H_1\hat M_{ff}^{-1}\hat G\hat H^{-1}(\hat M_{ff}-M_{ff})\frac{1}{N}\sum_{j=1}^N\hat\lambda_j\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\hat\Sigma_{iie}^{-1}\gamma_{ip}'f_te_{it}$$

$$+\operatorname{tr}\Big[\hat H_1\hat M_{ff}^{-1}\hat G\hat H^{-1}M_{ff}\frac{1}{N}\sum_{j=1}^N(\hat\lambda_j-\lambda_j)\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]+\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\hat\Sigma_{iie}^{-1}e_{it}v_{itp}$$

$$+\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^N\hat\Sigma_{iie}^{-1}\gamma_{ip}\hat\lambda_i'\mathbb{A}'M_{ff}\Big]-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^N\hat\Sigma_{iie}^{-1}\gamma_{ip}\hat\lambda_i'\hat G_N\Big(\frac{1}{T}\sum_{t=1}^T\hat\chi_tf_t'\Big)\Big]$$

$$-\operatorname{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\hat\Sigma_{iie}^{-1}f_t\hat\lambda_i'v_{itp}\hat G\hat\Gamma'\hat\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma\Big]-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^N\hat\Sigma_{iie}^{-1}\gamma_{ip}\hat\lambda_i'\hat G\hat M_{ff}^{-1}(\hat M_{ff}-M_{ff})\Big]$$

$$-\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^N\hat\Sigma_{iie}^{-1}(\hat\gamma_{ip}-\gamma_{ip})\hat\lambda_i'\hat G\Big]+\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}(\hat\Sigma_{ii}-\Sigma_{ii})I_{K+1}^{p+1}\hat\Sigma_{iie}^{-1}\hat\lambda_i'\hat G\Big]$$

$$-\operatorname{tr}\Big[\hat G_1\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N\hat\Gamma_i\hat\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat\Sigma_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$-\operatorname{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T\hat\Gamma_j\hat\Sigma_{jj}^{-1}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat\Sigma_{iie}^{-1}\hat\lambda_i'\hat G\Big] \tag{C.4}$$

The right hand side of the above expression has 22 terms. We pick out the first term $\mathrm{tr}[\hat{G}_1 \sum_{i=1}^N \hat{\Gamma}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i)' M_{ff} \hat{\upsilon}_p^g]$ and the 15th term $\mathrm{tr}[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip} \hat{\lambda}_i' \mathbb{A}' M_{ff}]$ for consideration since these two terms are each $O_p(T^{-1/2})$ due to Proposition C.1, which violates the claim of this proposition. But we show that the sum of the 1st and 15th terms satisfies the proposition. By $\hat{G} = \hat{H} - \hat{H}\hat{M}_{ff}^{-1}\hat{G}$, term $\mathbb{A} = (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{G}$ is equal to $(\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H} - (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{\Gamma}\hat{H}\hat{M}_{ff}^{-1}\hat{G} = A - A\hat{M}_{ff}^{-1}\hat{G}$. So the 15th term $\mathrm{tr}[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip} \hat{\lambda}_i' \mathbb{A}' M_{ff}]$ is equal to

$$\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip}\hat{\lambda}_i'\mathbb{A}'M_{ff}\Big] = \mathrm{tr}\Big[M_{ff}\mathbb{A}\frac{1}{N}\sum_{i=1}^N \hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}\gamma_{ip}'\Big]$$

$$= \mathrm{tr}\Big[M_{ff}\mathbb{A}\frac{1}{N}\sum_{i=1}^N(\hat{\lambda}_i - \lambda_i)\hat{\Sigma}_{iie}^{-1}\gamma_{ip}'\Big] + \mathrm{tr}\Big[M_{ff}\mathbb{A}\hat{\upsilon}_p\Big]$$

$$= \mathrm{tr}\Big[M_{ff}\mathbb{A}\frac{1}{N}\sum_{i=1}^N(\hat{\lambda}_i - \lambda_i)\hat{\Sigma}_{iie}^{-1}\gamma_{ip}'\Big] + \mathrm{tr}\Big[M_{ff}A\hat{\upsilon}_p\Big] - \mathrm{tr}\Big[M_{ff}A\hat{M}_{ff}^{-1}\hat{G}\hat{\upsilon}_p\Big]$$

So the sum of the 1st and 15th terms can be written as

$$\text{(C.5)} \qquad \mathrm{tr}\Big[\hat{G}_1\sum_{i=1}^N \hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Gamma}_i - \Gamma_i)'M_{ff}\hat{\upsilon}_p^g\Big] + \mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^N \hat{\Sigma}_{iie}^{-1}\gamma_{ip}\hat{\lambda}_i'\mathbb{A}'M_{ff}\Big]$$

$$= \mathrm{tr}\Big[I_1A'M_{ff}\hat{\upsilon}_p^g\Big] + \mathrm{tr}\Big[I_1M_{ff}A\hat{\upsilon}_p^g\Big] + \mathrm{tr}\Big[I_2M_{ff}A\hat{\upsilon}_p^h\Big]$$

$$-\mathrm{tr}\Big[\hat{H}_1\hat{M}_{ff}^{-1}\mathbb{A}'M_{ff}\hat{\upsilon}_p^g\Big] - \mathrm{tr}\Big[M_{ff}A\hat{M}_{ff}^{-1}\hat{G}\hat{\upsilon}_p\Big] + \mathrm{tr}\Big[M_{ff}\mathbb{A}\frac{1}{N}\sum_{j=1}^N(\hat{\lambda}_j - \lambda_j)\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big]$$

The first two terms can be written as $\mathrm{tr}[I_1(A'M_{ff} + M_{ff}A)\hat{\upsilon}_p^g]$. Under the identification condition IZ, the first r_1 rows of $\hat{M}_{ff} - M_{ff}$ is zero. Thus the expression of $I_1(A'M_{ff} + M_{ff}A)$ is implicitly given in (A.21). The third term involves $I_2M_{ff}A$. Notice the last term of the left hand side of (C.2) is $I_2M_{ff}A\hat{\Pi}_{\lambda\lambda}$. Shifting $I_2M_{ff}A\hat{\Pi}_{\lambda\lambda}$ from the left to the right, then post-multiplying $-\hat{\Pi}_{\lambda\lambda}^{-1}$, we obtain the expression of $I_2M_{ff}A$. Substituting these two expressions into (C.5), we can rewrite the first three terms of (C.5). This allows us to rewrite (C.4) as

$$\text{(C.6)} \qquad \vartheta_{p1}(\hat{\beta}_1 - \beta_1) + \vartheta_{p2}(\hat{\beta}_2 - \beta_2) + \cdots + \vartheta_{pK}(\hat{\beta}_K - \beta_K)$$
$$= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1}e_{it}v_{itp} + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}h_te_{it}$$

$$
\begin{aligned}
&-\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}h_t e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h \\
&+\mathrm{tr}\Big[\hat{G}_1\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\Gamma_i'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^{g}\Big)\Big] \\
&-\mathrm{tr}\Big[I_1(I-A)'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^g\Big] \\
&+\mathrm{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\Big(\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\mathcal{F}}_t\Big)\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h\Big] \\
&-\mathrm{tr}\Big[\hat{G}_2\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Gamma\Big(\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^h\Big] \\
&+\mathbb{S}_{p1}+\mathbb{S}_{p2}+\mathbb{S}_{p3}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbb{S}_{p1}=\;&\mathrm{tr}\Big[\hat{G}_1\hat{M}_{ff}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t e_{jt}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]+\mathrm{tr}\Big[I_1\mathbb{A}'\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t e_{jt}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big] \\
&-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'f_t v_{itp}+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}f_t\hat{\lambda}_i'v_{itp}\mathbb{A}'\Big] \\
&+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}f_t\hat{\lambda}_i'v_{itp}\hat{G}\hat{M}_{ff}^{-1}\Big]-\mathrm{tr}\Big[I_1A'\frac{1}{T}\Big(\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^g\Big] \\
&-\mathrm{tr}\Big[\hat{H}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)A\hat{\upsilon}_p^g\Big]+\mathrm{tr}\Big[\hat{H}_{1N}\hat{M}_{ff}^{-1}\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)\hat{\upsilon}_p^g\Big] \\
&-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\gamma_{ip}\hat{\Sigma}_{iie}^{-1}(\hat{\lambda}_i-\lambda_i)'\hat{G}_N\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)\Big]-\mathrm{tr}\Big[\hat{H}_1\hat{M}_{ff}^{-1}\mathbb{A}'M_{ff}\hat{\upsilon}_p^g\Big] \\
&+\mathrm{tr}\Big[\hat{H}\hat{M}_{ff}^{-1}\hat{G}_N\frac{1}{T}\Big(\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)\hat{\upsilon}_p'\Big]-\mathrm{tr}\Big[M_{ff}A\hat{M}_{ff}^{-1}\hat{G}\hat{\upsilon}_p\Big] \\
&-\mathrm{tr}\Big[I_2\mathbb{A}'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^h\Big]+\mathrm{tr}\Big[M_{ff}\mathbb{A}\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_j-\lambda_j)\hat{\Sigma}_{jje}^{-1}\gamma_{jp}'\Big] \\
&-\mathrm{tr}\Big[\hat{G}_2\hat{M}_{ff}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\hat{\chi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^h\Big]+\mathrm{tr}\Big[I_1A'M_{ff}A\hat{\upsilon}_p^g\Big]
\end{aligned}
$$

$$-\mathrm{tr}\Big[\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t f_t'\Big)A\hat{\upsilon}_p^h\Big]+\mathrm{tr}\Big[\frac{1}{N}\hat{G}_1\sum_{j=1}^{N}\hat{\lambda}_j\hat{\Sigma}_{jje}^{-1}(\hat{\Sigma}_{jje}-\Sigma_{jje})\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$+\mathrm{tr}\Big[\hat{H}_1\hat{M}_{ff}^{-1}\hat{G}\hat{H}^{-1}M_{ff}\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_j-\lambda_j)\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}(\hat{\gamma}_{ip}-\gamma_{ip})\hat{\lambda}_i'\hat{G}\Big]$$

$$+\mathrm{tr}\Big[\hat{H}_1\hat{M}_{ff}^{-1}\hat{G}\hat{H}^{-1}(\hat{M}_{ff}-M_{ff})\hat{\upsilon}_p^g\Big]-\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}\hat{\lambda}_i'\hat{G}\hat{M}_{ff}^{-1}(\hat{M}_{ff}-M_{ff})\Big]$$

$$+\mathrm{tr}\Big[I_2\mathbb{A}'M_{ff}A\hat{\upsilon}_p^h\Big]+\mathrm{tr}\Big[\hat{G}_2\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}\hat{\upsilon}_p^h\Big]$$

$$+\mathrm{tr}\Big[\hat{H}_2\hat{M}_{ff}^{-1}\hat{G}\hat{\Gamma}'\hat{\Sigma}_{\varepsilon\varepsilon}\hat{\Gamma}M_{ff}A\hat{\upsilon}_p^h\Big]+\mathrm{tr}\Big[I_2\mathbb{A}'\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_te_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})I_{K+1}^{p+1}\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]-\mathrm{tr}\Big[\hat{G}_2\frac{1}{N}\sum_{j=1}^{N}(\hat{\lambda}_j-\lambda_j)\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h\Big]$$

$$-\mathrm{tr}\Big[\hat{H}_1\sum_{i=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H}\hat{\upsilon}_p^g\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{N}\hat{G}_2\sum_{i=1}^{N}\hat{\lambda}_i\hat{\Sigma}_{iie}^{-1}(\hat{\Sigma}_{iie}-\Sigma_{iie})\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h\Big]$$

$$+\mathrm{tr}\Big[\hat{G}_2\hat{M}_{ff}^{-1}\frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}f_te_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h\Big]$$

and

$$\mathbb{S}_{p2}=\mathrm{tr}\Big[\hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)\hat{\upsilon}_p^g\Big]-\mathrm{tr}\Big[\hat{H}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)(I-A)\hat{\upsilon}_p^g\Big]$$

$$+\mathrm{tr}\Big[\hat{G}_{2N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)f_t'\Big)A\hat{\upsilon}_p^h\Big]+\mathrm{tr}\Big[\hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^g\Big)\Big]$$

$$+\mathrm{tr}\Big[\hat{G}_1\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)e_{jt}\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]-\mathrm{tr}\Big[\hat{G}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\mathcal{G}}_{tp}^g\Big)\Big]$$

$$-\mathrm{tr}\Big[\hat{H}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\xi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^g\Big]-\mathrm{tr}\Big[\hat{H}_{1N}\Big(\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\hat{\chi}_t'\Big)\hat{H}_N\hat{\upsilon}_p^g\Big]$$

$$+\mathrm{tr}\Big[\hat{H}_{1N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\hat{v}_p^g\Big]+\mathrm{tr}\Big[\hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\mathcal{F}}_t\hat{\Pi}_{\lambda\lambda}^{-1}\hat{v}_p^h\Big]$$

$$+\mathrm{tr}\Big[\hat{G}_2\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{v}_p^h\Big]-\mathrm{tr}\Big[\hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\chi}_t(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\hat{v}_p^h\Big]$$

$$-\mathrm{tr}\Big[\hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\mathcal{F}}_t\hat{\Pi}_{\lambda\lambda}^{-1}\hat{v}_p^h\Big]-\mathrm{tr}\Big[\hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)\hat{\chi}_t'\hat{H}_N\hat{v}_p^h\Big]$$

$$+\mathrm{tr}\Big[\hat{G}_{2N}\frac{1}{T}\sum_{t=1}^{T}\hat{\xi}_t(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\hat{\xi}_t'\hat{H}_N\hat{v}_p^h\Big]$$

and

$$\mathbb{S}_{p3}=-\mathrm{tr}\Big[\hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$

$$-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$

$$+\mathrm{tr}\Big[\hat{H}_1\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{v}_p^g\Big]$$

$$-\mathrm{tr}\Big[\hat{G}_2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{v}_p^h\Big]$$

$$+\mathrm{tr}\Big[\hat{G}_2\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{v}_p^h\Big]$$

Consider (C.6). Terms $\mathbb{S}_{p1}, \mathbb{S}_{p2}$ are dealt with in Lemma C.3. The terms on the left hand side and the 4th-7th terms on the right hand side are summarized in Lemma C.2. The first three terms on the right hand side and $\mathbb{S}_{p3}$ are given in Lemma C.4. Using the results in Lemma C.2, C.3 and C.4, we have

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

This leads to Proposition C.3. □

Corollary C.1. *Under Assumptions A-D plus G, together with the identification conditions IO, we have*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Gamma}_i-\Gamma_i\|^2=O_p(T^{-1})$$

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

This corollary is an immediate result of Proposition C.2 and C.3.

To prove Proposition 4.1, we need the following lemma.

LEMMA C.5. *Under the assumptions of Theorem 4.1, we have*

$$(a)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}\gamma_{ip}^{h\prime}h_t e_{it}$$
$$+ O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(b)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Sigma}_{iie}^{-1}e_{it}v_{itp} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Sigma_{iie}^{-1}e_{it}v_{itp}$$
$$+ O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(c)\quad \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}h_t e_{jt}\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h = \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T}h_t e_{jt}\Sigma_{jje}^{-1}\lambda_j'\Pi_{\lambda\lambda}^{-1}\upsilon_p^h$$
$$+ O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(d)\quad \mathrm{tr}\Big[\hat{G}_1\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\gamma_{jp}^{g\prime}\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(e)\quad \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_j\hat{\Sigma}_{jj}^{-1}[\varepsilon_{jt}v_{itp}-E(\varepsilon_{jt}v_{itp})]\hat{\Sigma}_{iie}^{-1}\hat{\lambda}_i'\hat{G}\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(f)\quad \mathrm{tr}\Big[\hat{H}_1\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{\upsilon}_p^g\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(g)\quad \mathrm{tr}\Big[\hat{G}_2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\varepsilon_{it}e_{jt}-E(\varepsilon_{it}e_{jt})]\hat{\Sigma}_{jje}^{-1}\hat{\lambda}_j'\hat{\Pi}_{\lambda\lambda}^{-1}\hat{\upsilon}_p^h\Big]$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta}-\beta\|)$$

$$(h)\quad \mathrm{tr}\Big[\hat{G}_2\frac{1}{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Gamma}_i\hat{\Sigma}_{ii}^{-1}[\varepsilon_{it}\varepsilon_{jt}'-E(\varepsilon_{it}\varepsilon_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Gamma}_j'\hat{H}\hat{\upsilon}_p^h\Big]$$

$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\beta} - \beta\|)$$

PROOF OF LEMMA C.5. Proof of Lemma C.5 is quite similar to that of Lemma A.13 and hence is omitted. □

PROOFS OF PROPOSITION 4.1. Given (C.6) and Lemmas C.2, C.3 and C.5, the proof of Proposition 4.1 is identical to that of Proposition 3.1. The details are omitted. □

APPENDIX C3: PROOFS OF THEOREMS 4.1 AND 4.3

PROOF OF THEOREM 4.1. To prove Theorem 4.1, we first transform the parameters set (Γ, f_t, β) into $(\Gamma^\star, f_t^\star, \beta)$ which satisfies the identification condition IO. For ease of reading, we rewrite model (4.1) below

$$\begin{aligned} y_{it} &= \alpha_i + x_{it1}\beta_1 + x_{it2}\beta_2 + \cdots + x_{itK}\beta_K + \psi_i' g_t + \phi_i' h_t + e_{it} \\ x_{itk} &= \mu_{ik} + \gamma_{ik}^{g\prime} g_t + \gamma_{ik}^{h\prime} h_t + v_{itk} \end{aligned}$$

The first equation of the above can be rewritten as

$$\begin{aligned} y_{it} = (\alpha_i + \psi_i'\bar{g} + \phi_i'\bar{h}) + \sum_{k=1}^K x_{itk}\beta_k + \Big\{[\psi_i' + \phi_i'\dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}](\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q\Big\} \\ \times \Big\{Q'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1/2}\dot{g}_t\Big\} + \phi_i'\Big\{\dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t\Big\} + e_{it} \end{aligned}$$

and the second equation can be rewritten as

$$\begin{aligned} x_{itk} = (\mu_{ik} + \gamma_{ik}^{g\prime}\bar{g} + \gamma_{ik}^{h\prime}\bar{h}) + \Big\{[\gamma_{ik}^{g\prime} + \gamma_{ik}^{h\prime}\dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}](\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q\Big\} \\ \times \Big\{Q'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1/2}\dot{g}_t\Big\} + \gamma_{ik}^{h\prime}\Big\{\dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t\Big\} + v_{itk} \end{aligned}$$

where Q is an $r_1 \times r_1$ orthogonal matrix which is defined below. Let

$$\psi_i^{\star\prime} = [\psi_i' + \phi_i'\dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}](\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q \tag{C.7}$$

$$\gamma_{ik}^{g\star\prime} = [\gamma_{ik}^{g\prime} + \gamma_{ik}^{h\prime}\dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}](\dot{\mathbb{G}}'\dot{\mathbb{G}})^{1/2}Q \tag{C.8}$$

$$\gamma_{ik}^{h\star\prime} = \gamma_{ik}^{h\prime} \tag{C.9}$$

$$g_t^\star = Q'(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1/2}\dot{g}_t \tag{C.10}$$

$$h_t^\star = \dot{h}_t - \dot{\mathbb{H}}'\dot{\mathbb{G}}(\dot{\mathbb{G}}'\dot{\mathbb{G}})^{-1}\dot{g}_t \tag{C.11}$$

Let $\Gamma^{g\star}, \Gamma^{h\star}, \Gamma^\star$ be defined similarly as $\Gamma^g, \Gamma^h, \Gamma$. If we choose Q to be the eigenvector matrix of $\Gamma^{g\star\prime}\Sigma_{\varepsilon\varepsilon}^{-1}\Gamma^{g\star}$ with the associated eigenvalues in descending order, we can easily verify that the parameters set $(\Gamma^\star, f_t^\star, \beta)$ satisfy the identification conditions IO. Then by Proposition 4.1, we have

$$\begin{aligned}\mathcal{Q}^\star(\hat\beta-\beta) &= \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Sigma_{iie}^{-1} e_{it} v_{itx} + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Sigma_{iie}^{-1}\gamma_{ix}^{h\star\prime} h_t^\star e_{it} \\ &\quad -\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \Sigma_{iie}^{-1}\lambda_i^{\star\prime}\Pi_{\lambda\lambda}^{\star-1}\Big(\frac{1}{N}\sum_{j=1}^N \lambda_j^{\star\prime}\Sigma_{jje}^{-1}\gamma_{jx}^{h\star\prime}\Big)h_t^\star e_{it} \\ &\quad + O_p(T^{-3/2}) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})\end{aligned}$$

Substituting (C.7)$-$(C.11) into the above expression, we have the same asymptotic expression as stated in Theorem 4.1.

The proof of the alternative expression in Theorem 4.1 is the same with that of Theorem 3.1 and hence omitted.

Corollary 4.2 is a consequence of the alternative expression of $\hat\beta-\beta$. $\square$

Proof of Theorem 4.3. To prove Theorem 4.3, notice the first equation of (4.7) can always be written as

$$\begin{aligned} y_{it} &= \sum_{k=1}^K x_{itk}\beta_k + \Big\{\Big(\psi_i' + \phi_i'\mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1}\Big)[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{1/2}Q\Big\} \\ &\qquad \times\Big\{Q'[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1/2}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t)\Big\} \\ &\quad + \phi_i'\Big\{h_t - \mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t - \mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t)\Big\} \\ &\quad + \Big\{\kappa_i' + \psi_i'\mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1} + \phi_i'\mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}\Big\}d_t + e_{it}\end{aligned}$$

The second equation of (4.7) can always be written as

$$\begin{aligned} x_{itk} &= \Big\{\Big(\gamma_{ik}^{g\prime} + \gamma_{ik}^{h\prime}\mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1}\Big)[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{1/2}Q\Big\} \\ &\qquad \times\Big\{Q'[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1/2}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t)\Big\} \\ &\quad + \gamma_{ik}^{h\prime}\Big\{h_t - \mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t - \mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t)\Big\} \\ &\quad + \Big\{\gamma_{ik}^{d\prime} + \gamma_{ik}^{g\prime}\mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1} + \gamma_{ik}^{h\prime}\mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}\Big\}d_t + v_{itk}\end{aligned}$$

Let

$$\begin{aligned}\psi_i^{\star\prime} &= (\psi_i' + \phi_i'\mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1})[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{1/2}Q \\ g_t^\star &= Q'[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1/2}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t) \\ \gamma_{ik}^{g\star\prime} &= (\gamma_{ik}^{g\prime} + \gamma_{ik}^{h\prime}\mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1})[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{1/2}R\end{aligned}$$

$$h_t^\star = h_t - \mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t - \mathbb{H}'\mathcal{M}(\mathbb{D})\mathbb{G}[\mathbb{G}'\mathcal{M}(\mathbb{D})\mathbb{G}]^{-1}(g_t - \mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}d_t)$$
$$\kappa_i^{\star\prime} = \gamma_{ik}^{d\prime} + \gamma_{ik}^{g\prime}\mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1} + \gamma_{ik}^{h\prime}\mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}$$
$$\gamma_{ik}^{d\star\prime} = \gamma_{ik}^{d\prime} + \gamma_{ik}^{g\prime}\mathbb{G}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1} + \gamma_{ik}^{h\prime}\mathbb{H}'\mathbb{D}(\mathbb{D}'\mathbb{D})^{-1}$$

After appropriately choosing the orthogonal matrix Q, we can make the parameters $(\Gamma^\star, f_t^\star, \beta)$ satisfy the identification condition IO$'$. Using the same method in deriving Theorem 4.1, we can prove Theorem 4.3. The details are omitted. □

APPENDIX C4: PROOFS OF THEOREM 4.2 AND 4.4

The following proposition is useful to derive Theorem 4.2 and 4.4.

PROPOSITION C.4. *Under Assumptions A-D, in combination with the identification conditions IO and IO$'$, we have*

$$A \equiv \sum_{i=1}^{N}(\hat{\Gamma}_i - \Gamma_i)\hat{\Sigma}_{ii}^{-1}\hat{\Gamma}_i'\hat{H} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

PROOF OF PROPOSITION C.4. Using Corollary C.1 and (C.2), we can prove that A_{12} and A_{22} are both $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. However, the identification condition IO1 implies that $\hat{M}_{gh} = M_{gh} = 0$. Given this result, by (A.21), we have $A_{21} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. Now we only need to prove $A_{11} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. This result can be proved by the same way as proving $(\hat{\Lambda} - \Lambda)'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ under IC3 in [10]. So Proposition C.4 follows. □

Theorem 4.2 can be viewed as a special case of Theorem 4.4. We only focus on proof of Theorem 4.4.

PROOF OF THEOREM 4.4. The formula to estimate f_t is

$$\hat{f}_t = (\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda})^{-1}\hat{\Lambda}'\hat{\Sigma}_{ee}^{-1}(\tilde{\mathbb{Y}}_t - \tilde{\mathbb{X}}_t\hat{\beta})$$

where $\hat{f}_t = (\hat{g}_t', \hat{h}_t')'$, $\hat{\Lambda} = (\hat{\Psi}, \Phi)$ and $\tilde{\mathbb{Y}}_t$ is the tth column of the matrix $Y\mathcal{M}(\mathbb{D})$, $\tilde{\mathbb{X}}_t$ is an $N \times K$ matrix with its kth column equal to the tth column of the matrix $X_k\mathcal{M}(\mathbb{D})$. Given the above equation, we have

(C.12)
$$\hat{h}_t = \Big[\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}\Phi\Big]^{-1}\Big[\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}(\tilde{\mathbb{Y}}_t - \tilde{\mathbb{X}}_t\hat{\beta})\Big]$$

The first equation of (4.7) can be written as

$$Y = X_1\beta_1 + \cdots + X_K\beta_K + \Psi\mathbb{G}' + \Phi\mathbb{H}' + \mathrm{K}\mathbb{D}' + e$$

Post-multiplying $\mathcal{M}(\mathbb{D})$ on both sides, by $\mathbb{G}'\mathbb{D} = 0, \mathbb{H}'\mathbb{D} = 0$, we have

$$Y\mathcal{M}(\mathbb{D}) = X_1\mathcal{M}(\mathbb{D})\beta_1 + \cdots + X_K\mathcal{M}(\mathbb{D})\beta_K + \Psi\mathbb{G}' + \Phi\mathbb{H}' + e\mathcal{M}(\mathbb{D})$$

So we have

$$\tilde{\mathbb{Y}}_t = \tilde{\mathbb{X}}_t\beta + \Psi g_t + \Phi h_t + e_t - \tilde{e}_t$$

where $\tilde{e}_{it} = (\sum_{s=1}^T e_{is}d_s')(\sum_{s=1}^T d_s d_s')^{-1}d_t$ and $\tilde{e}_t = (\tilde{e}_{1t}, \tilde{e}_{2t}, \cdots, \tilde{e}_{Nt})'$. Substituting the above equation into (C.12), we have

$$\sqrt{N}(\hat{h}_t - h_t) = -\hat{\mathcal{X}}^{-1}\Big[\frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}\tilde{\mathbb{X}}_t\sqrt{N}(\hat{\beta} - \beta)\Big]$$

$$-\hat{\mathcal{X}}^{-1}\Big[\sqrt{N}\frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}(\hat{\Psi} - \Psi)g_t\Big] \tag{C.13}$$

$$+\hat{\mathcal{X}}^{-1}\Big[\frac{1}{\sqrt{N}}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}(e_t - \tilde{e}_t)\Big]$$

where $\hat{\mathcal{X}} = \frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}\Phi$. Using the consistency result we have proved, term $\frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}\Phi$ can be showed to be

$$\frac{1}{N}\sum_{i=1}^N \phi_i\Sigma_{iie}^{-1}\phi_i' - (\frac{1}{N}\sum_{i=1}^N \phi_i\Sigma_{iie}^{-1}\psi_i')(\frac{1}{N}\sum_{i=1}^N \psi_i\Sigma_{iie}^{-1}\psi_i')^{-1}(\frac{1}{N}\sum_{i=1}^N \psi_i\Sigma_{iie}^{-1}\phi_i') + o_p(1)$$

Consider the first term on the right hand side of (C.13). Term $\frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}$ $\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}\tilde{\mathbb{X}}_t$ is equal to

$$\frac{1}{N}\sum_{i=1}^N \phi_i\hat{\Sigma}_{iie}^{-1}\tilde{x}_{it} - (\frac{1}{N}\sum_{i=1}^N \phi_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i')(\frac{1}{N}\sum_{i=1}^N \hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i')^{-1}(\frac{1}{N}\sum_{i=1}^N \hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\tilde{x}_{it})$$

where $\tilde{x}_{it}$ is the ith row of matrix $\tilde{\mathbb{X}}_t$. By the consistency result we have proved, the above expression is equal to

$$\frac{1}{N}\sum_{i=1}^N \phi_i\Sigma_{iie}^{-1}\tilde{x}_{it} - (\frac{1}{N}\sum_{i=1}^N \phi_i\Sigma_{iie}^{-1}\psi_i')(\frac{1}{N}\sum_{i=1}^N \psi_i\Sigma_{iie}^{-1}\psi_i')^{-1}(\frac{1}{N}\sum_{i=1}^N \psi_i\Sigma_{iie}^{-1}\tilde{x}_{it}) + o_p(1)$$

which is $O_p(1)$. From this, in combination with $\hat{\beta} - \beta = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-3/2})$, the first term of (C.13) is equal to $O_p(T^{-1/2}) + O_p(N^{1/2}T^{-3/2})$.

Now consider the term $\frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}(\hat{\Psi} - \Psi)g_t$, which is equal to

$$\frac{1}{N}\sum_{i=1}^N \phi_i\hat{\Sigma}_{iie}^{-1}(\hat{\psi}_i - \psi_i)'$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'\Big)^{-1}\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}(\hat{\psi}_i-\psi_i)'\Big)g_t$$

Using (C.3) to replace $\hat{\psi}_i-\psi_i$ from the above expression and using the result in Proposition C.4, we can show that

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}(\hat{\psi}_i-\psi_i)'=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

$$\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}(\hat{\psi}_i-\psi_i)'=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$$

From these, together with $\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'=\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}\psi_i'+o_p(1)$ and $\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i'=\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}\psi_i'+o_p(1)$, we have that the second term of the right hand side of (C.13) is $O_p(T^{-1/2})+O_p(N^{1/2}T^{-1})$.

Now consider the third term. We first consider $\frac{1}{N}\Phi'\hat{\Sigma}_{ee}^{-1/2}\mathcal{M}(\hat{\Sigma}_{ee}^{-1/2}\hat{\Psi})\hat{\Sigma}_{ee}^{-1/2}e_t$, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}e_{it}-(\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i')(\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i')^{-1}(\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}e_{it})$$

Using the expression for $\hat{\Sigma}_{jje}-\Sigma_{jje}$, we can show that

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}e_{it}=\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}e_{it}+O_p(T^{-1})$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\psi_i\hat{\Sigma}_{iie}^{-1}e_{it}=\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}e_{it}+O_p(T^{-1})$$

Furthermore, Using the expression for $\hat{\psi}_i-\psi_i$, we can show

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\psi}_i-\psi_i)\hat{\Sigma}_{iie}^{-1}e_{it}=O_p(N^{-1})+O_p(T^{-1})$$

So we have

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}e_{it}-(\frac{1}{N}\sum_{i=1}^{N}\phi_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i')(\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}\hat{\psi}_i')^{-1}(\frac{1}{N}\sum_{i=1}^{N}\hat{\psi}_i\hat{\Sigma}_{iie}^{-1}e_{it})$$

$$=\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}e_{it}-(\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}\psi_i')(\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}\psi_i')^{-1}(\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}e_{it})$$

$$+O_p(N^{-1}) + O_p(T^{-1})$$

Notice

$$\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}\phi_i' - (\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}\psi_i')(\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}\psi_i')^{-1}(\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}\phi_i')$$

is equal to $\frac{1}{N}\Phi'\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\Phi$. Then we have

$$\sqrt{N}(\hat{h}_t - h_t) = \Big(\frac{1}{N}\Phi'\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\Phi\Big)^{-1}$$

$$\times\Big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}e_{it} - (\frac{1}{N}\sum_{i=1}^{N}\phi_i\Sigma_{iie}^{-1}\psi_i')(\frac{1}{N}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}\psi_i')^{-1}(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi_i\Sigma_{iie}^{-1}e_{it})\Big)$$

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

If $N \to \infty, T \to \infty$, and $\sqrt{N}/T \to 0$, we have

$$\sqrt{N}(\hat{h}_t - h_t) \xrightarrow{d} N\Big(0, [\lim_{N\to\infty}\frac{1}{N}\Phi'\Sigma_{ee}^{-1/2}\mathcal{M}(\Sigma_{ee}^{-1/2}\Psi)\Sigma_{ee}^{-1/2}\Phi]^{-1}\Big),$$

This proves the first part of Theorem 4.4.

By $y_{it} = \sum_{p=1}^{K} x_{itp}\beta_p + \psi_i' g_t + \phi_i' h_t + \kappa_i' d_t + e_{it}$, we have

$$\sqrt{T}(\hat{\kappa}_i - \kappa_i) = -\sum_{p=1}^{K}\Big(\frac{1}{T}\sum_{t=1}^{T}d_t d_t'\Big)^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}d_t x_{itp}\Big)\sqrt{T}(\hat{\beta}_p - \beta_p)$$

$$-\Big(\frac{1}{T}\sum_{t=1}^{T}d_t d_t'\Big)^{-1}\Big(\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}d_t\lambda_i'(\hat{f}_t - f_t)\Big)$$

$$-\Big(\frac{1}{T}\sum_{t=1}^{T}d_t d_t'\Big)^{-1}\Big(\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}d_t(\hat{\psi}_i - \psi_i)'\hat{g}_t\Big) + \Big(\frac{1}{T}\sum_{t=1}^{T}d_t d_t'\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_t e_{it}\Big)$$

The first term is $O_p(N^{-1/2}) + O_p(T^{-1})$. The second term and the third term can be proved to be $O_p(T^{1/2}N^{-1}) + O_p(T^{-1/2})$. So we have

$$\sqrt{T}(\hat{\kappa}_i - \kappa_i) = \Big(\frac{1}{T}\sum_{t=1}^{T}d_t d_t'\Big)^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}d_t e_{it}\Big) + O_p(T^{1/2}N^{-1}) + O_p(T^{-1/2})$$

If $N \to \infty, T \to \infty$, and $\sqrt{T}/N \to 0$, we have

$$\sqrt{T}(\hat{\kappa}_i - \kappa_i) \xrightarrow{d} N\Big(0, \Sigma_{iie}(\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}d_t d_t')^{-1}\Big).$$

This completes the proof of the second part of Theorem 4.4. □

Jushan Bai
Department of Economics
Columbia University
420, West 118th Street
New York, NY 10027
E-mail: jushan.bai@columbia.edu

Kunpeng Li
Department of Economics
School of Economics and Management
Tsinghua University
Beijing, China, 100084
E-mail: likp.07@sem.tsinghua.edu.cn