Spatial panel data models with common shocks

Bai, Jushan and Li, Kunpeng

December 2013

An earlier version of this paper is available at SSRN with link:
http://dx.doi.org/10.2139/ssrn.2373628

Spatial panel data models with common shocks

Jushan Bai* and Kunpeng Li†

March 9, 2014

Abstract

Spatial effects and common-shocks effects are of increasing empirical importance. Each type of effect has been analyzed separately in a growing literature. This paper considers a joint modeling of both types. Joint modeling allows one to determine whether one or both of these effects are present. A large number of incidental parameters exist under the joint modeling. The quasi maximum likelihood method (MLE) is proposed to estimate the model. Heteroskedasticity is explicitly estimated. This paper demonstrates that the quasi-MLE is effective in dealing with the incidental parameters problem. An inferential theory including consistency, rate of convergence and limiting distributions is developed. The quasi-MLE can be easily implemented via the EM algorithm, as confirmed by the Monte Carlo simulations. The simulations further reveal the excellent finite sample properties of the quasi-MLE. Some extensions are discussed.

Key Words: Panel data models, spatial interactions, common shocks, cross-sectional dependence, incidental parameters, maximum likelihood estimation

*Department of Economics, Columbia University, New York, NY, 10027. Email: jb3064@columbia.edu
†School of International Economics and Management, Capital University of Economics and Business, Beijing, China. email: likp.07@sem.tsinghua.edu.cn.

1 Introduction

There is a large and yet still rapidly growing literature on spatial interactions and common shocks, both of which lead to cross-sectional dependence. In spatial models, the cross sectional dependence is captured by spatial weights matrices based either on physical distance, and relative position in a social network or on other types of economic distance. The cross-sectional dependence in a common-shocks model arises from the response of individuals to the shocks. The common shocks model is characterized by a common factor structure. All these models are motivated by empirical considerations.[1] The existing literature largely analyzes the two types of models separately. This paper integrates spatial interactions and common shocks. We show that the maximum likelihood method is an effective way of estimating the resulting model.

Early development of spatial models has been summarized by a number of books, including Cliff and Ord (1973), Anselin (1988), and Cressie (1993). GMM estimation of spatial models are studied by Kelijian and Prucha (1998, 1999, 2010), and Kapoor et al. (2007), among others. The maximum likelihood method is considered by Ord (1975), Anselin (1988), Lee (2004a), Yu et al. (2008) and Lee and Yu (2010), and so on. For panel data models with multiple common shocks, Ahn et al. (2013) consider the fixed-T GMM estimation. Pesaran (2006) proposes the correlated random effects method by including additional regressors, which are the cross-sectional averages of dependent and the explanatory variables. The principal components method is studied by Bai (2009) and Moon and Weidner (2009). Bai and Li (2014) consider the maximum likelihood method.

The joint presence of spatial interactions and common shocks calls for a different estimation procedure, as the existing method is not directly applicable. Under joint modeling, there exist a large number of incidental parameters. This paper also allows cross sectional heteroskedasticity, giving rise to further incidental parameters. We show that the maximum likelihood method can effectively deal with the incidental parameters.

This paper considers the following spatial panel data model with common shocks, in which both the dependent variables y_{it} and the explanatory variables x_{it} are impacted by the common shocks f_t:

$$\begin{aligned} y_{it} &= \alpha_i + \rho\sum_{j=1}^{N} w_{ij,N}y_{jt} + \sum_{p=1}^{k} x_{itp}\beta_p + \lambda_i' f_t + e_{it}, \\ x_{itp} &= \nu_{ip} + \gamma_{ip}' f_t + v_{itp}, \qquad p = 1, 2, \ldots, k; \end{aligned} \tag{1.1}$$

where y_{it} is the dependent variable; $x_{it} = (x_{it1}, x_{it2}, \ldots, x_{itk})'$ is a k-dimensional vector of explanatory variables; f_t is an r-dimensional vector of unobservable common shocks; λ_i is the corresponding heterogenous response to the common shocks; $W_N = (w_{ij,N})_{N\times N}$ is a specified spatial weights matrix whose diagonal elements $w_{ii,N}$ are 0; and e_{it} and v_{it} are the

[1]For spatial interaction and economic distance, see, e.g., Case (1991), Case et al. (1993), Conley (1999), Conley Dupor (2003), and Topa (2001); for common factors, see Ross (1976), Chamberlain and Rothschild (1983), Stock and Watson (1998), to name a few.

idiosyncratic errors. In model (1.1), the term $\lambda_i' f_t$ captures the common shock effects, and $\rho \sum_{j=1}^N w_{ij,N} y_{jt}$ captures the spatial effects. The joint modeling allows one to test which type of effects is responsible for the cross sectional dependence. We may test $\rho = 0$ while allowing common shocks. Similarly, we may determine if the number of factors is zero in a model with spatial effects. It may be possible that both effects are present.

An additional feature of the model is the allowance of cross sectional heteroskedasticity. The importance of permitting heteroskedasticity is noted by Kelejian and Prucha (2010) and Lin and Lee (2010). The heteroskedastic variances can be empirically important, e.g., Glaeser et al. (1996) and Anselin (1988). In addition, if heteroskedasticity exists but homoskedasticity is imposed, then MLE can be inconsistent. Under large N, the consistency analysis for MLE under heteroskedasticity is challenging even for spatial panel models without common shocks, owing to the simultaneous estimation of a large number of variance parameters along with (ρ, β). The existing quasi maximum likelihood studies, such as Yu et al. (2008) and Lee and Yu (2010), typically assume homoskedasticity. These authors show that the limiting variance of MLE has a sandwich formula unless normality is assumed. Interestingly, we show that the limiting variance of the MLE is not of a sandwich form if heteroskedasticity is allowed.

Spatial correlation and common shocks are also considered by Pesaran and Tosetti (2011), who specify the spatial autocorrelation on the unobservable errors e_{it} while we specify the spatial autocorrelation on the observable dependent variable y_{it}. Both specifications are of practical relevance. Spatial specification on observable data makes explicit the empirical implication of the coefficient ρ. From a theoretical perspective, the spatial interaction on the dependent variable gives rise to the endogeneity problem, while the spatial interaction on the errors, in general, does not. As a result, under the Pesaran and Tosetti setup, existing estimation methods on the common shocks models such as Pesaran (2006) and Bai (2009) can be applied to estimate the model. As a comparison, these methods cannot be directly applied to model (1.1) due to the endogeneity from the spatial interactions.

In this study, we consider the pseudo-Gaussian maximum likelihood method (MLE), which simultaneously estimates all parameters of the model, including heteroskedasticity. We give a rigorous analysis of the MLE including the consistency, the rate of convergence and limiting distributions. The asymptotic theory does not rely on normality.

In subsequent exposition, the matrix norms are defined in the following way. For any $m \times n$ matrix A, $\|A\|$ denotes the Frobenius norm of A, i.e., $\|A\| = [\text{tr}(A'A)]^{1/2}$, $\|A\|_2$ the spectral norm, i.e., $\|A\|_2 = [\lambda_{\max}(A'A)]^{1/2}$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue. In addition, $\|A\|_\infty$ is defined as $\|A\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^n |a_{ij}|$ and $\|A\|_1$ is defined as $\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^m |a_{ij}|$, where a_{ij} is the (i,j)th element of A. We use $\dot{a}_t$ to denote $\dot{a}_t = a_t - \frac{1}{T}\sum_{t=1}^T a_t$ for any column vector a_t and M_{ab} to denote $\frac{1}{T}\sum_{t=1}^T \dot{a}_t \dot{b}_t'$ for any vectors a_t and b_t.

The rest of the paper is organized as follows. Section 2 gives the matrix form of model

(1.1) and the assumptions needed for the subsequent analysis. Section 3 presents the objective function and the associated first order conditions. The asymptotic properties including the consistency, the convergence rates and the limiting distributions are derived in Section 4. Computing algorithm is discussed in Section 5. Section 6 reports simulation results. Section 7 discusses extensions of the model. The last section concludes. Technical proofs are given in a supplementary document.

2 Model description and assumptions

Let $v_{it} = (v_{it1}, v_{it2}, \ldots, v_{itk})'$ and $\nu_i = (\nu_{i1}, \nu_{i2}, \ldots, \nu_{ik})'$, and let $\gamma_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{ik})$. The x equation in model (1.1) is equivalent to

$$x_{it} = \nu_i + \gamma_i' f_t + v_{it}.$$

Now model (1.1) can be rewritten as

$$\begin{bmatrix} y_{it} - \rho \sum_{j=1}^N w_{ij,N} y_{jt} - x_{it}'\beta \\ x_{it} \end{bmatrix} = \mu_i + \Phi_i' f_t + \epsilon_{it}$$

with $\Phi_i = [\lambda_i, \gamma_i]$, $\mu_i = [\alpha_i, \nu_i']'$ and $\epsilon_{it} = [e_{it}, v_{it}']'$. Let $D(\rho, \beta)$ be an $N(k+1) \times N(k+1)$ matrix, whose (i, j) subblock, denoted by $D_{ij}(\rho, \beta)$, is equal to

$$D_{ij}(\rho, \beta) = \begin{cases} \begin{bmatrix} 1 & -\beta' \\ 0 & I_k \end{bmatrix} & \text{if } \; i = j \\ \begin{bmatrix} -\rho w_{ij,N} & 0 \\ 0 & 0 \end{bmatrix} & \text{if } \; i \neq j \end{cases} \tag{2.1}$$

Now model (1.1) can be further written as

$$D(\rho, \beta) z_t = \mu + \Phi f_t + \epsilon_t \tag{2.2}$$

where $z_t = (z_{1t}, z_{2t}, \ldots, z_{Nt})'$ with $z_{it} = (y_{it}, x_{it}')'$, $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_N)'$, $\mu = (\mu_1', \mu_2', \ldots, \mu_N')'$ and $\epsilon_t = (\epsilon_{1t}', \epsilon_{2t}', \ldots, \epsilon_{Nt}')'$. Hereafter, we suppress ρ, β from $D(\rho, \beta)$ for notational simplicity. Throughout the paper, we assume that the number of factors r is fixed and known. Determining the number of factors is discussed in Section 6, where a modified information criterion of Bai and Li (2014) is proposed. Our simulation results are based on the estimated number of factors.

To analyze model (1.1), we make the following assumptions. Comments on these assumptions are given in a number of remarks below.

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.

Assumption B: The idiosyncratic errors $\epsilon_{it} = (e_{it}, v_{it}')'$ are such that

B.1 The e_{it} is independent and identically distributed over t and uncorrelated over i with $E(e_{it}) = 0$ and $E(e_{it}^4) \leq \infty$ for all $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Let σ_i^2 denote the variance of e_{it}.

B.2 v_{it} is also independent and identically distributed over t and uncorrelated over i with $E(v_{it}) = 0$ and $E(\|v_{it}\|^4) \leq \infty$ for all $i = 1, \cdots, N$ and $t = 1, \cdots, T$. We use Σ_{iiv} to denote the variance matrix of v_{it}.

B.3 e_{it} is independent of v_{js} for all (i, j, t, s). Let Σ_{ii} denote the variance matrix ϵ_{it}. So we have $\Sigma_{ii} = \text{diag}(\sigma_i^2, \Sigma_{iiv})$, a block-diagonal matrix.

Assumption C: There exists a $C > 0$ sufficiently large such that

C.1 $\|\omega\| \leq C$, where $\omega = (\rho, \beta')'$;

C.2 $\|\Phi_j\| \leq C$ for all $j = 1, \cdots, N$;

C.3 $C^{-1} \leq \tau_{min}(\Sigma_{jj}) \leq \tau_{max}(\Sigma_{jj}) \leq C$ for all $j = 1, \cdots, N$, where $\tau_{min}(\Sigma_{jj})$ and $\tau_{max}(\Sigma_{jj})$ denote the smallest and largest eigenvalues of Σ_{jj};

C.4 there exists an $r \times r$ positive matrix Q such that $Q = \lim_{N\to\infty} N^{-1}\Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi$, where Φ is defined earlier, and $\Sigma_{\epsilon\epsilon} = \text{diag}(\Sigma_{11}, \Sigma_{22}, ..., \Sigma_{NN})$, a block diagonal matrix.

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. In addition, ρ and β are estimated in a compact set $\mathcal{A} \times \mathcal{B} \subset \mathbb{R} \times \mathbb{R}^k$.

Remark 2.1 Assumptions A-D are made in the context of factor analysis, and are used in Bai and Li (2012, 2014). Assumption A assumes that the sequence $\{f_t\}$ is fixed. If random factors are assumed instead, the analysis remains valid if we assume that f_t is independent of the errors ϵ_{is} for t and s, and f_t has finite 4th moment. The fixed f_t is consistent with the fixed effects assumption. It also allows arbitrary dynamics in f_t, either a linear or broken trend or stochastic processes. By estimating M_{ff} instead of individual f_t, we avoid incidental parameters in the time dimension. If T is much smaller than N, we could estimate individual f_t and the sample variance of λ_i (not individual λ_i) by switching the role of N and T. Assumption B assumes that the variance of idiosyncratic errors is a block-diagonal matrix. This assumption allows the k regressors to be correlated. This assumption also extends the traditional factor analysis in which the variance of errors is assumed to be diagonal. Assumption C assumes that the underlying values of parameters are in a compact set. This assumption is standard in the econometric literature. Assumption D requires that parameters be optimized in a compact set. This assumption is often made when dealing with highly nonlinear objective functions, e.g. Jennrich (1969), and Newey and McFadden (1994). Our objective function is nonlinear.

Assumption E: The weights matrix W_N satisfies that $I_N - \rho W_N$ is invertible and

$$\limsup_{N\to\infty} \|W_N\|_\infty < \infty; \qquad \limsup_{N\to\infty} \|W_N\|_1 < \infty; \tag{2.3}$$

$$\limsup_{N\to\infty} \|(I - \rho W_N)^{-1}\|_\infty < \infty; \qquad \limsup_{N\to\infty} \|(I - \rho W_N)^{-1}\|_1 < \infty. \tag{2.4}$$

In addition, all the diagonal elements of W_N are zeros.

Remark 2.2 Assumption E is standard in spatial econometrics, see Kelejian and Prucha (1998), Lee (2004a), Yu et al. (2008), Lee and Yu (2010), to name a few. Under this assumption, some key matrices, which play important roles in asymptotic analysis such as S_N in Assumption F below, can be handled in a tractable way. A set of sufficient conditions for (2.4) are $\limsup_{N\to\infty} \|W_N\|_\infty \leq 1$, $\limsup_{N\to\infty} \|W_N\|_1 \leq 1$ and $|\rho| < 1$ because

$$\limsup_{N\to\infty} \|(I - \rho W_N)^{-1}\|_\infty \leq \limsup_{N\to\infty} \sum_{j=0}^{\infty} (\|\rho W_N\|_\infty)^j \leq \frac{1}{1-\rho} < \infty,$$

$$\limsup_{N\to\infty} \|(I - \rho W_N)^{-1}\|_1 \leq \limsup_{N\to\infty} \sum_{j=0}^{\infty} (\|\rho W_N\|_1)^j \leq \frac{1}{1-\rho} < \infty.$$

Assumption F: One of the following conditions holds:

(i) $\beta \neq 0$ and, for all N,

$$\frac{1}{N}\Big(\mathrm{tr}[\Sigma_{ee}^{-1} S_N \Sigma_{ee} S_N'] + \mathrm{tr}[S_N^2] - 2\sum_{i=1}^{N} S_{ii,N}^2\Big) > \delta, \tag{2.5}$$

(ii) For all $\rho^\dagger \in \mathcal{A}$ with $\rho^\dagger \neq \rho$,

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \Big(S_{ij,N}\sigma_j^2 + S_{ji,N}\sigma_i^2 - (\rho^\dagger - \rho)\sum_{p=1}^{N} S_{ip,N} S_{jp,N} \sigma_p^2\Big)^2 \neq 0 \tag{2.6}$$

where $S_N = W_N(I_N - \rho W_N)^{-1}$ and $S_{ij,N}$ be the (i,j)th element; ρ denotes the true spatial coefficient, and $\Sigma_{ee} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$.

Remark 2.3 Assumption F makes further restrictions on the spatial weights matrix W_N to guarantee the identification of ρ. Part (i) is a local identification condition since it depends on $\beta \neq 0$. Part (ii) does not depend on β, and it can be regarded as a global identification condition for ρ. In this viewpoint, the condition (2.6) should be stronger than (2.5). It is indeed the case. To see this, note that the expression in (2.5) can be regarded as the variance of $N^{-1/2}\boldsymbol{v}'\Sigma_{ee}^{1/2} S_N^{\circ\prime} \Sigma_{ee}^{-1/2}\boldsymbol{v}$, when $\boldsymbol{v}$ is taken as a standard normal $N(0, I_N)$, where $S_N^\circ = S_N - S_N^d$ with $S_N^d = \mathrm{diag}(S_{11,N}, S_{22,N}, \ldots, S_{NN,N})$ and (see Remark 4.3 below for more related details). For $\rho^\dagger = \rho$, the expression in (2.6) is twice the variance of $N^{-1/2}\boldsymbol{v}'\Sigma_{ee} S_N^\circ \boldsymbol{v}$. So condition (2.5) can be viewed as a variant of (2.6) when $\rho^\dagger$ is restricted to ρ. Thus condition (2.6) is stronger than (2.5). The weaker condition (2.5) implies that the inclusion of explanatory variables x_{it} (with $\beta \neq 0$) helps identification.

Remark 2.4 The two sets of identification conditions in Assumption F are stated in different ways in the existing literature. Part (i) corresponds to Assumption 8 in Yu et al. (2008) but it is different from theirs because we allow heteroskedasticity. Part (ii) is related to Assumption 9 in Lee (2004a) and the condition in Theorem 2 of Yu et al. (2008). To see this, we show in Appendix A that condition (2.6) is related to the unique solution of $\mathcal{T}_{1N}(\rho^\dagger, \sigma_1^{\dagger 2}, \ldots, \sigma_N^{\dagger 2}) = 0$ with

$$\mathcal{T}_{1N}(\rho^\dagger, \sigma_1^{\dagger 2}, \ldots, \sigma_N^{\dagger 2}) = -\frac{1}{2N}\mathrm{tr}[R^\dagger \Sigma_{ee} R^{\dagger\prime} \Sigma_{ee}^{\dagger -1}] + \frac{1}{2N}\ln|R^\dagger \Sigma_{ee} R^{\dagger\prime} \Sigma_{ee}^{\dagger -1}| + \frac{1}{2},$$

where $R^\dagger = (I_N - \rho^\dagger W_N)(I_N - \rho W_N)^{-1}$, $\Sigma_{ee}^\dagger = \mathrm{diag}(\sigma_1^{\dagger 2}, \sigma_2^{\dagger 2}, \ldots, \sigma_N^{\dagger 2})$, and where ρ and $\Sigma_{ee} = \mathrm{diag}(\sigma_1^1, \sigma_2^2, ..., \sigma_N^2)$ denote the true parameters. When homoskedasticity is assumed, $\mathcal{T}_{1N}$ reduces to $T_{1,n}$ in Yu et al. (2008). After concentrating out the common variance $\sigma^{\dagger 2}$, $T_{1,n}$ leads to Assumption 9 in Lee (2004a) and the assumption of Theorem 2 in Yu et al. (2008). Because of heteroskedasticity our identification condition takes a different form.

Remark 2.5 Condition (2.5) implies that

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ij,N}^2 \geq \delta' \tag{2.7}$$

for some positive δ'. To see this, notice that the left hand side of (2.5) is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\sigma_j^2}{\sigma_i^2}S_{ij,N}^2 + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ij,N}S_{ji,N}.$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ij,N}S_{ji,N} &\leq \Big[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ij,N}^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ji,N}^2\Big]^{1/2}\\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ij,N}^2.\end{aligned}$$

Let $\Delta = \max_i \sigma_i^2 / \min_i \sigma_i^2$. By Assumption C.3, we have $\Delta < \infty$. Then

$$\delta \leq \frac{1}{N}\Big(\mathrm{tr}[\Sigma_{ee}^{-1} S_N \Sigma_{ee} S_N'] + \mathrm{tr}[S_N^2] - 2\sum_{i=1}^{N} S_{ii,N}^2\Big) \leq (\Delta+1)\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} S_{ij,N}^2\Big).$$

Letting $\delta' = \frac{1}{\Delta+1}\delta$, the condition (2.7) follows.

Also, a sufficient condition for (2.6) is, for all $\rho^\dagger \in \mathcal{A}$ and all N,

$$\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\Big(S_{ij,N}\sigma_j^2 + S_{ji,N}\sigma_i^2 - (\rho^\dagger - \rho)\sum_{p=1}^{N} S_{ip,N}S_{jp,N}\sigma_p^2\Big)^2 > \delta,$$

for some positive δ.

Identification conditions (IC hereafter). It is known in the factor literature that the loadings Φ can only be identified up to a rotation. To remove the rotational indeterminacy, we impose the following normalization restrictions: (a) $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t = 0$; (b) $M_{ff} =$

$\frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})' = I_r$; (c) $\frac{1}{N}\Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi$ is diagonal with the diagonal elements being distinct and arranged in descending order.

The above normalization is used by the maximum likelihood method in classical factor analysis, e.g., Anderson (2003, Chapter 14); also see Bai and Li (2012). Under this normalization, there is no need to estimate the sample variance matrix M_{ff}, and the analysis is also simpler. If $M_{ff} \neq I_r$, we redefine the factor loading Φ as $\Phi^\dagger = \Phi M_{ff}^{1/2}$, then the corresponding $M_{ff}^\dagger$ will be I_r. There exist other normalization restrictions to fix the rotational indeterminacy. Different normalizations will give different estimates of Φ and M_{ff}, but the estimation of key parameters $\omega = (\rho, \beta)$ and $\Sigma_{\epsilon\epsilon}$ is invariant to the different normalization restrictions.

3 Objective function and First order conditions

Let $\theta = (\omega, \Phi, \Sigma_{\epsilon\epsilon})$ be the parameters to be estimated. The objective function considered in this paper is

$$\mathcal{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\mathrm{tr}[DM_{zz}D'\Sigma_{zz}^{-1}] \tag{3.1}$$

where $\Sigma_{zz} = \Phi\Phi' + \Sigma_{\epsilon\epsilon}$; $D = D(\rho, \beta)$ is given in equation (2.1); and $M_{zz} = \frac{1}{T}\sum_{t=1}^{T}\dot{z}_t\dot{z}_t'$ is the data matrix. The above objective function is the likelihood function when f_t and ϵ_t are assumed to be iid normal and are independent. Such assumptions are not necessary; in fact, f_t does not have to be random, and ϵ_t does not have to be normal, as is demonstrated in our theoretical analysis, as well as in the simulation analysis. The maximizer $\hat{\theta}$, defined by

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

is referred to as the quasi maximum likelihood estimator or MLE, where Θ is the parameters space specified by Assumption D. By the definition of $D = D(\rho, \beta)$, the determinant of D is equal to the determinant of $I_N - \rho W_N$, so $\ln|D| = \ln|I_N - \rho W_N|$. Thus the Jacobian term is relatively easy to handle. The more difficult part is that D also appears in the second term of the likelihood, where it also depends on both ρ and β. We can rewrite the objective function as

$$\mathcal{L}(\omega, \Phi, \Sigma_{\epsilon\epsilon}) = -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|I_N - \rho W_N| - \frac{1}{2N}\mathrm{tr}[DM_{zz}D'\Sigma_{zz}^{-1}].$$

The first order condition for Φ is

$$\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{D}M_{zz}\hat{D}' - \hat{\Sigma}_{zz}) = 0 \tag{3.2}$$

where $\hat{D} = D(\hat{\rho}, \hat{\beta})$. The first order condition for $\Sigma_{\epsilon\epsilon}$ is

$$\hat{D}M_{zz}\hat{D}' - \hat{\Sigma}_{zz} = \mathbb{W} \tag{3.3}$$

where $\mathbb{W}$ is an $N(k+1) \times N(k+1)$ matrix whose ith $(k+1) \times (k+1)$ diagonal subblock is such that the upper-left 1×1 and lower-right $k \times k$ submatrices are both zeros and the rest elements of $\mathbb{W}$ are unspecified. The unspecified elements of $\mathbb{W}$ correspond to the zero elements of $\Sigma_{\epsilon\epsilon}$. The first order condition for ρ is

$$\begin{aligned} -\frac{1}{N}\mathrm{tr}\big[(I_N - \hat{\rho} W_N)^{-1} W_N\big] + \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2}(\dot{y}_{it} - \hat{\rho}\ddot{y}_{it} - x_{it}'\hat{\beta})\ddot{y}_{it} \\ - \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i'\hat{G}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t\ddot{y}_{it} = 0 \end{aligned} \tag{3.4}$$

where $\ddot{y}_{it} = \sum_{j=1}^N w_{ij,N}\dot{y}_{jt}$ and $\hat{G} = (I_r + \hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi})^{-1}$. The first order condition for β is

$$\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2}\dot{x}_{it}(\dot{y}_{it} - \hat{\rho}\ddot{y}_{it} - \dot{x}_{it}'\hat{\beta}) - \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2}\dot{x}_{it}\hat{\lambda}_i'\hat{G}\hat{\Phi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\hat{D}\dot{z}_t = 0 \tag{3.5}$$

These first order conditions are useful when deriving the rate of convergence and the limiting distributions. They are not used for the consistency proof, nor are they used in computing the MLE. That is, MLE does not need to solve for the first order conditions. The MLE obtained by the EM algorithm automatically satisfies the first order conditions. This is proved in Appendix E, and is also confirmed by numerical simulations.

4 Asymptotic properties of the MLE

In this section, we first show that the MLE is consistent, we then derive the convergence rates, the asymptotic representation and the limiting distributions.

Proposition 4.1 *Under Assumptions A-F, when $N, T \to \infty$, we have, for $\omega = (\rho, \beta')'$,*

$$\hat{\omega} - \omega = o_p(1);$$
$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = o_p(1).$$

In addition, if IC holds, we also have

$$\frac{1}{N}\sum_{i=1}^N \|\hat{\Sigma}_{ii}^{-1}\| \cdot \|\hat{\Phi}_i - \Phi_i\|^2 = o_p(1).$$

Remark 4.1 In the analysis of panel data models with common shocks but without spatial effects, a difficult problem is to establish consistency. The parameters of interest β is simultaneously estimated with high dimensional nuisance parameters Φ and $\Sigma_{\epsilon\epsilon}$. The analysis has to deal with these nuisance parameters. The presence of spatial effects further compounds the difficult, partly due to the transformation matrix D and spatial endogeneity. As shown in appendix A, we need D^{-1} for further theoretical analysis. The expression of D^{-1} is given in Lemma A.1 of the supplement.

The consistency result allows us to further derive the rates of convergence.

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

$$\hat{\omega} - \omega = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-3/2});$$

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

In addition, if IC holds, we also have

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

Remark 4.2 An implication of Theorem 4.1 is that the MLE of ω is $\sqrt{T}$-consistent when N is finite. This means that the ML method considered in this paper is still applicable in the "finite N" setting. But the asymptotic expression and limiting distribution will be different.

Theorem 4.1 also has implications for asymptotic properties of $\hat{\Phi}_i$ and $\hat{\Sigma}_{ii}$. Given that $\hat{\omega} - \omega$ has a faster convergence rate, the limiting distributions of $\operatorname{vech}(\hat{\Phi}_i - \Phi_i)$ and $\operatorname{vech}(\hat{\Sigma}_{ii} - \Sigma_{ii})$ are not affected by the estimation of ω, and are the same as the case of without regressors.

To provide the asymptotic representation of $\hat{\omega} = (\hat{\rho}, \hat{\beta}')'$, we define

$$\Omega = \frac{1}{N}\begin{bmatrix} \operatorname{tr}[S_N^2] + \operatorname{tr}[\Sigma_{ee}^{-1}S_N\Psi S_N'] - 2\sum_{i=1}^N S_{ii,N}^2 & \beta'(\sum_{i=1}^N S_{ii,N}\Sigma_{iiv}/\sigma_i^2) \\ (\sum_{i=1}^N S_{ii,N}\Sigma_{iiv}/\sigma_i^2)\beta & \sum_{i=1}^N \Sigma_{iiv}/\sigma_i^2 \end{bmatrix}$$

where Ψ is a diagonal matrix with the ith diagonal element being $\sigma_i^2 + \beta'\Sigma_{iiv}\beta$, that is, $\Psi = \operatorname{diag}(\sigma_1^2 + \beta'\Sigma_{11v}\beta, \ldots, \sigma_N^2 + \beta'\Sigma_{NNv}\beta)$. Then we have the following theorem.

Theorem 4.2 *Under Assumptions A-F, when $N, T \to \infty$ and $\sqrt{N}/T \to 0$, we have*

$$\sqrt{NT}(\hat{\omega} - \omega) = \Omega^{-1}\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}\begin{bmatrix} \sum_{j=1}^N e_{it}\eta_{ij,t}S_{ij,N} \\ e_{it}v_{it} \end{bmatrix} + o_p(1).$$

where $\eta_{ij,t} = v_{jt}'\beta + 1(i \neq j)e_{jt}$.

Remark 4.3 Ignoring Ω^{-1}, we can write the asymptotic expression in Theorem 4.2 alternatively as

$$\frac{1}{\sqrt{NT}}\begin{bmatrix} \sum_{t=1}^T e_t'S_N^{\circ\prime}\Sigma_{ee}^{-1}e_t + \sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}e_{it}(\sum_{j=1}^N S_{ij,N}v_{jt}')\beta \\ \sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}e_{it}v_{it} \end{bmatrix} \tag{4.1}$$

where $S_N^\circ = S_N - S_N^d$ with $S_N^d = \operatorname{diag}(S_{11,N}, S_{22,N}, \ldots, S_{NN,N})$. To obtain the variance of (4.1), let

$$\boldsymbol{\varepsilon}_a = \frac{1}{\sqrt{NT}}\sum_{t=1}^T e_t'S_N^{\circ\prime}\Sigma_{ee}^{-1}e_t, \qquad \boldsymbol{\varepsilon}_b = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}e_{it}(\sum_{j=1}^N S_{ij,N}v_{jt}')\beta,$$

$$\boldsymbol{\varepsilon}_c = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}e_{it}v_{it}.$$

It is easy to check that $E(\boldsymbol{\varepsilon}_a) = E(\boldsymbol{\varepsilon}_b) = 0$ and $E(\boldsymbol{\varepsilon}_c) = \mathbf{0}$. By the well-known result that

$$E[(\boldsymbol{v}_t' A \boldsymbol{v}_t)^2] = [\operatorname{tr}(A)]^2 + \operatorname{tr}(A^2) + \operatorname{tr}(A'A) + \kappa \operatorname{tr}(A \circ A) \tag{4.2}$$

where "$\circ$" denotes the Hadamard product and $\boldsymbol{v}_t$ are iid over t with zero mean and identity variance matrix, and the elements of v_t are also iid with the fourth moment $3+\kappa$, we have

$$\operatorname{var}(\boldsymbol{\varepsilon}_a) = \operatorname{tr}(S_N^{\circ 2}) + \operatorname{tr}(\Sigma_{ee} S_N^{\circ\prime} \Sigma_{ee}^{-1} S_N^{\circ}).$$

This follows from $\operatorname{tr}(\Sigma_{ee}^{1/2} S_N^{\circ\prime} \Sigma_{ee}^{-1/2}) = 0$ and $\operatorname{tr}[(\Sigma_{ee}^{1/2} S_N^{\circ\prime} \Sigma_{ee}^{-1/2}) \circ (\Sigma_{ee}^{1/2} S_N^{\circ\prime} \Sigma_{ee}^{-1/2})] = 0$. From $S_N^{\circ} = S_N - S_N^d$, and from S_N^d being a diagonal matrix, the above result can be alternatively written as

$$\operatorname{var}(\boldsymbol{\varepsilon}_a) = \frac{1}{N}\Big(\operatorname{tr}[S_N^2] + \operatorname{tr}[\Sigma_{ee}^{-1} S_N \Sigma_{ee} S_N'] - 2\sum_{i=1}^N S_{ii,N}^2\Big).$$

In addition, it is relatively easy to show that

$$\operatorname{var}(\boldsymbol{\varepsilon}_b) = \frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \frac{1}{\sigma_i^2} S_{ij,N}^2 (\beta' \Sigma_{jjv} \beta).$$

Combining results, we have

$$\operatorname{var}(\boldsymbol{\varepsilon}_a + \boldsymbol{\varepsilon}_b) = \operatorname{var}(\boldsymbol{\varepsilon}_a) + \operatorname{var}(\boldsymbol{\varepsilon}_b) = \frac{1}{N}\Big[\operatorname{tr}[S_N^2] + \operatorname{tr}[\Sigma_{ee}^{-1} S_N \Psi S_N'] - 2\sum_{i=1}^N S_{ii,N}^2\Big]$$

by the definition of Ψ. In addition,

$$\operatorname{cov}(\boldsymbol{\varepsilon}_a + \boldsymbol{\varepsilon}_b, \boldsymbol{\varepsilon}_c) = \operatorname{cov}(\boldsymbol{\varepsilon}_b, \boldsymbol{\varepsilon}_c) = \frac{1}{N}\Big[\beta' \sum_{i=1}^N \frac{1}{\sigma_i^2} S_{ii,N} \Sigma_{iiv}\Big], \quad \operatorname{var}(\boldsymbol{\varepsilon}_c) = \frac{1}{N}\Big[\sum_{i=1}^N \frac{1}{\sigma_i^2} \Sigma_{iiv}\Big].$$

These results imply that the variance of (4.1) is Ω.

Corollary 4.1 *Under the assumptions of Theorem 4.2, if $\sqrt{N}/T \to 0$, we have*

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

where $\overline{\Omega} = \lim_{N\to\infty} \Omega$.

Remark 4.4 Ω can be consistently estimated by

$$\hat{\Omega} = \frac{1}{N}\begin{bmatrix} \operatorname{tr}[\hat{S}_N^2] + \operatorname{tr}[\hat{\Sigma}_{ee}^{-1} \hat{S}_N \hat{\Psi} \hat{S}_N'] - 2\sum_{i=1}^N \hat{S}_{ii,N}^2 & \hat{\beta}'(\sum_{i=1}^N \hat{S}_{ii,N} \hat{\Sigma}_{iiv}/\hat{\sigma}_i^2) \\ (\sum_{i=1}^N \hat{S}_{ii,N} \hat{\Sigma}_{iiv}/\hat{\sigma}_i^2)\hat{\beta} & \sum_{i=1}^N \hat{\Sigma}_{iiv}/\hat{\sigma}_i^2 \end{bmatrix}.$$

where $\hat{S}_N = W_N(I_N - \hat{\rho} W_N)^{-1}$ and $\hat{\Psi} = \operatorname{diag}(\hat{\sigma}_1^2 + \hat{\beta}' \hat{\Sigma}_{11v} \hat{\beta}, \ldots, \hat{\sigma}_N^2 + \hat{\beta}' \hat{\Sigma}_{NNv} \hat{\beta})$.

Remark 4.5 To gain an intuition of the asymptotic results in Theorem 4.2, consider the following spatial panel data model without common shocks

$$y_{it} = \alpha_i + \rho \sum_{j=1}^N w_{ij,N} y_{jt} + v_{it}' \beta + e_{it} \tag{4.3}$$

where e_{it} and v_{it} satisfies the conditions listed in Assumption B but v_{it} is assumed to be observable. Conditional on v_{it}, the likelihood function by concentrating out α_i is

$$\mathcal{L}'(\theta) = -\frac{1}{2N}\sum_{i=1}^{N}\ln\sigma_i^2 + \frac{1}{N}\ln|I_N - \rho W_N| - \frac{1}{2NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\dot{y}_{it} - \rho\ddot{y}_{it} - \dot{v}_{it}'\beta)^2$$

where, again, $\ddot{y}_{it} = \sum_{j=1}^{N} w_{ij,N}\dot{y}_{jt}$. Let $\tilde{\theta} = (\tilde{\rho}, \tilde{\beta}, \tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_N^2)$ be the MLE of the above likelihood function. It can be shown that $\tilde{\omega} - \omega$ has the same asymptotic expression as in Theorem 4.2. This means that the likelihood approach of this paper eliminates the endogenous part of x_{it} (common factors). Theorem 4.2 also provides the asymptotic theory of MLE under heteroskedasticity for the usual spatial models.

Remark 4.6 From Corollary 4.1, we see that the MLE achieves asymptotic efficiency for heteroskedastic spatial models in the sense that the limiting variance is not a sandwich form. This result contrasts with the existing results in the literature such as Yu et al. (2008) and Lee and Yu (2010), in which the limiting variance of the MLE has a sandwich formula. The reason for the difference is the heteroskedasticity estimation. In the present paper we allow cross-sectional heteroskedasticity, while Yu et al. (2008) assume homoskedasticity. Under heteroskedasticity, the asymptotic expression does not involve e_{it}^2, as shown in Theorem 4.2. But under the homoskedasticity, the situation is different. Still consider model (4.3). If homoskedasticity is assumed and is imposed in estimation, the asymptotic expression for the MLE is

$$\tilde{\omega} - \omega = \tilde{\Omega}^{-1}\frac{1}{\sqrt{NT}\sigma^2}\begin{bmatrix}\sum_{t=1}^{T} e_t' S_N^{\circ\prime} e_t + \sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}(\sum_{j=1}^{N} S_{ij,N}v_{jt}')\beta + \upsilon \\ \sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}v_{it}\end{bmatrix} + o_p(1),$$

where $\upsilon = \sum_{i=1}^{N}\sum_{t=1}^{T}[S_{ii,N} - \frac{1}{N}\mathrm{tr}(S_N)](e_{it}^2 - \sigma^2)$ and

$$\tilde{\Omega} = \frac{1}{N}\begin{bmatrix}\frac{1}{\sigma^2}\mathrm{tr}[S_N\tilde{\Psi}S_N'] + \mathrm{tr}(S_N S_N') + \mathrm{tr}(S_N^2) - \frac{2}{N}[\mathrm{tr}(S_N)]^2 & \frac{1}{\sigma^2}\beta'\sum_{i=1}^{N} S_{ii,N}\Sigma_{iiv} \\ \frac{1}{\sigma^2}\sum_{i=1}^{N} S_{ii,N}\Sigma_{iiv}\beta & \frac{1}{\sigma^2}\sum_{i=1}^{N}\Sigma_{iiv}\end{bmatrix},$$

here $\tilde{\Psi}$ is a diagonal matrix with its ith diagonal element being $\beta'\Sigma_{iiv}\beta$. From the above, we can see that the asymptotic expression under the homoskedasticity involves e_{it}^2. So the limiting variance of $\tilde{\omega} - \omega$ will depend on the kurtosis of e_{it}. Because $\tilde{\Omega}$ does not depend on the kurtosis, the limiting variance of $\tilde{\omega} - \omega$ has a sandwich formula. In contrast, the MLE under heteroskedasticity has a limiting variance not of a sandwich form, regardless of normality. This is an interesting result. Thus estimating heteroskedasticity is desirable from two considerations: the limiting distribution is robust to the underlying distributions; it avoids potential inconsistency when homoskedasticity is incorrectly imposed.

5 Computation and algorithm

We show how spatial panel data models with common shocks can be easily estimated by the EM algorithm. Lee (2004a) uses the usual maximization procedures (e.g. the Newton-Raphson method) to estimate the spatial models; Bai and Li (2014) use the EM algorithm to estimate panel data models with common shocks. Neither these methods are suitable for models with both spatial interactions and common shocks without modification.

An obstacle to using the standard EM algorithm is that there is no closed form solution for the maximizer $\hat{\rho}$ in (3.1) even other parameters are all known. However, this problem can be easily overcome because ρ is a scalar and actual maximization (a low dimension maximization) can be directly carried out. Thus the algorithm combines the usual maximization procedures with the EM algorithm. Let $\theta^{(s)} = (\rho^{(s)}, \beta^{(s)}, \Phi^{(s)}, \Sigma_{\epsilon\epsilon}^{(s)})$ denote the estimated value at the sth iteration. Our updating procedures consist of two steps. In the first step, we update $\Phi, \Sigma_{\epsilon\epsilon}$ and β according to the EM algorithm

$$\Phi^{(s+1)} = \Big[\frac{1}{T}\sum_{t=1}^{T} E(Dz_t f_t' | \theta^{(s)})\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | \theta^{(s)})\Big]^{-1}, \tag{5.1}$$

$$\Sigma_{\epsilon\epsilon}^{(s+1)} = \mathrm{Dg}\Big[D^{(s)} M_{zz} D^{(s)\prime} - \Phi^{(s+1)}\Phi^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)} M_{zz} D^{(s)\prime}\Big] \tag{5.2}$$

and

$$\begin{aligned} \beta^{(s+1)} = & \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(\sigma_i^{(s+1)})^2}\dot{x}_{it}\dot{x}_{it}'\Big]^{-1} \\ & \times \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(\sigma_i^{(s+1)})^2}\dot{x}_{it}\Big(\dot{y}_{it} - \rho^{(s)}\sum_{j=1}^{N} w_{ij}\dot{y}_{jt} - \lambda_i^{(s+1)\prime} f_t^{(s)}\Big)\Big], \end{aligned} \tag{5.3}$$

where Dg is the operator which sets the entries of its argument to zeros if the counterparts of $E(\epsilon_t \epsilon_t')$ are zeros; $(\sigma_i^{(s+1)})^2$ is the $[(i-1)(k+1)+1]$th diagonal element of $\Sigma_{\epsilon\epsilon}^{(s+1)}$ and $\lambda_i^{(s+1)}$ is the transpose of the $[(i-1)(k+1)+1]$th row of $\Phi^{(s+1)}$. In addition,

$$\frac{1}{T}\sum_{t=1}^{T} E(Dz_t f_t' | \theta^{(s)}) = D^{(s)} M_{zz} D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}\Phi^{(s)},$$

$$\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | \theta^{(s)}) = I_r - \Phi^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}\Phi^{(s)} + \Phi^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)} M_{zz} D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}\Phi^{(s)},$$

and

$$f_t^{(s)} = \Phi^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)}\dot{z}_t. \tag{5.4}$$

In the second step, we update ρ by maximizing (3.1) with respect to ρ at $\beta = \beta^{(s+1)}, \Phi = \Phi^{(s+1)}$ and $\Sigma_{\epsilon\epsilon} = \Sigma_{\epsilon\epsilon}^{(s+1)}$ with an initial value of ρ at $\rho^{(s)}$. The suggested procedure is a version of the ECME procedure of Liu and Rubin (1994). Putting together, we obtain $\theta^{(s+1)} = (\rho^{(s+1)}, \beta^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)})$.

This procedure guarantees that the value of likelihood function in each iteration does not decrease. This is because

$$\mathcal{L}(\rho^{(s)}, \beta^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \geq \mathcal{L}(\rho^{(s)}, \beta^{(s)}, \Phi^{(s)}, \Sigma_{\epsilon\epsilon}^{(s)}), \tag{5.5}$$

$$\mathcal{L}(\rho^{(s+1)}, \beta^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \geq \mathcal{L}(\rho^{(s)}, \beta^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}). \tag{5.6}$$

Letting $\rho = \rho^{(s)}$ be fixed, inequality (5.5) can be verified by the standard theory of the EM algorithm, see Dempster et al. (1977) and McLachlan and Krishnan (1997). Inequality (5.6) is due to the definition of $\rho^{(s+1)}$. In Appendix E, we show that the limit of the iterated solution satisfies the first order conditions (3.2)-(3.5) and therefore possesses the local optimality property.

For the initial value $\theta^{(1)} = (\rho^{(1)}, \beta^{(1)}, \Phi^{(1)}, \Sigma_{\epsilon\epsilon}^{(1)})$, $\rho^{(1)}$ and $\beta^{(1)}$ can be set to the within group estimator, ignoring the endogeneity problem. And $\Phi^{(1)}$ and $\Sigma_{\epsilon\epsilon}^{(1)}$ is then the maximizer of (3.1) at $\rho = \rho^{(1)}$ and $\beta = \beta^{(1)}$.

6 Finite sample properties

In this section, we run simulations to investigate the finite sample properties of the MLE. The data are generated according to

$$
\begin{aligned}
y_{it} &= \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + x_{it1}\beta_1 + x_{it2}\beta_2 + \lambda_i' f_t + e_{it} \\
x_{itp} &= \nu_{ip} + \gamma_{ip}' f_t + v_{itp}, \qquad \text{for} \quad p = 1, 2.
\end{aligned}
$$

The dimension of f_t is fixed to 2. We set $\beta_1 = 1$ and $\beta_2 = 2$. All the elements of α_i, ν_{ip}, λ_i and f_t are generated from $N(0,1)$ and $\gamma_{ip} = \lambda_i + u_{ip}$ with both elements of u_{ip} being independent $N(0,1)$ for $p = 1, 2$. This allows correlations between λ_i and γ_{ip}. To generate the errors and heteroscedasticity, we use the method of Bai and Li (2014) to set $\epsilon_t = \sqrt{\operatorname{diag}(\Xi)}\Upsilon\varepsilon_t$, where ε_t is an $N(k+1)$ dimensional vector with all the elements being $(\chi_2^2 - 2)/2$, where χ_2^2 denotes the chi-squared distribution with two degrees of freedom, which is normalized to zero mean and unit variance. In addition, Ξ is also $N(k+1)$ dimensional, whose ith element is set to

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

where η_i is drawn from $U[0.1, 0.9]$ and ι_i' is the ith row of Φ; Υ is defined as $\operatorname{diag}(\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_N)$ with $\Upsilon_i = \operatorname{diag}\{1, (M_i'M_i)^{-1/2}M_i\}$ where M_i is a $k \times k$ standard normal random matrix for each i.

The generated data exhibit heteroskedasticity. The generated x_{it} is correlated with the factors and factor loadings in the y_{it} equation, and the two regressors x_{it1} and x_{it2} are also correlated; the errors are non-normal and skewed. The simulation results under the normal and student's t distributions are given in Appendix F.

The spatial weights matrices generated in the simulation are similar to Kelejian and Prucha (1999) and Kapoor et al. (2007). More specifically, all the units are arranged in a circle and each unit is affected only by the q units immediately before it and immediately after it with equal weight. Following Kelejian and Prucha (1999), we normalize the spatial weights matrix by letting the sum of each row be equal to 1 (so the weight is $\frac{1}{2q}$) and call this specification of spatial weights matrix "q ahead and q behind."

Adapting a criterion in Bai and Li (2014), the number of factors is determined by

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

with

$$IC(m) = \frac{1}{2N\bar{k}} \ln \left| \hat{\Phi}^m \hat{\Phi}^{m\prime} + \hat{\Sigma}_{\epsilon\epsilon}^m \right| - \frac{1}{N\bar{k}} \ln |I_N - \hat{\rho}^m W_N| + m \frac{N\bar{k} + T}{2N\bar{k}T} \ln[\min(N\bar{k}, T)].$$

where $\hat{\rho}^m$, $\hat{\Phi}^m$ and $\hat{\Sigma}_{\epsilon\epsilon}^m$ are the respective estimators of ρ, Φ and $\Sigma_{\epsilon\epsilon}$ when the factor number is set to m; $\bar{k} = k + 1$. We set $r_{\max} = 4$.

The following four tables present the simulation results which are obtained based on 1000 repetitions. Biases and root mean square errors (RMSE) are both computed. The percentage that the factor number is correctly estimated is given in the third column of each table. Different values of ρ and different spatial weights matrices are considered. The tables show that the MLE has good finite sample properties. The number of factors can be correctly estimated with high probability. The biases are small. The RMSE of the estimators decreases as the sample becomes larger, indicating that they are consistent.

Table 1: The performance of the MLE under $\rho = 0.2$
with "1 ahead and 1 behind" spatial weights matrix

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.7	0.0004	0.0039	0.0001	0.0095	0.0003	0.0090
50	50	99.9	0.0000	0.0025	-0.0001	0.0058	-0.0003	0.0059
100	50	100.0	0.0000	0.0016	0.0000	0.0037	-0.0002	0.0036
25	100	99.7	-0.0000	0.0027	0.0000	0.0068	0.0001	0.0063
50	100	100.0	0.0000	0.0018	0.0000	0.0041	-0.0001	0.0042
100	100	100.0	0.0001	0.0011	0.0001	0.0026	-0.0001	0.0026

Table 2: The performance of the MLE under $\rho = 0.9$
with "1 ahead and 1 behind" spatial weights matrix

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.6	-0.0000	0.0011	0.0003	0.0097	0.0004	0.0099
50	50	100.0	0.0000	0.0006	0.0001	0.0058	-0.0002	0.0059
100	50	100.0	-0.0001	0.0004	0.0000	0.0036	0.0000	0.0038
10	100	99.8	0.0000	0.0007	-0.0001	0.0067	-0.0001	0.0068
50	100	100.0	-0.0001	0.0005	0.0001	0.0040	0.0004	0.0042
100	100	100.0	-0.0001	0.0003	0.0001	0.0026	0.0003	0.0027

Table 3: The performance of the MLE under $\rho = 0.2$
with "3 ahead and 3 behind" spatial weights matrix

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.3	-0.0002	0.0062	0.0004	0.0092	0.0004	0.0095
50	50	100.0	-0.0002	0.0037	-0.0001	0.0058	-0.0003	0.0057
100	50	100.0	0.0001	0.0025	-0.0001	0.0037	-0.0001	0.0037
25	100	100.0	-0.0002	0.0045	0.0001	0.0069	-0.0000	0.0067
50	100	100.0	0.0000	0.0028	0.0002	0.0040	0.0001	0.0040
100	100	100.0	0.0000	0.0017	-0.0001	0.0024	0.0000	0.0025

Table 4: The performance of the MLE under $\rho = 0.9$
with "3 ahead and 3 behind" spatial weights matrix

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.4	0.0000	0.0013	0.0005	0.0093	0.0000	0.0094
50	50	99.9	-0.0000	0.0008	-0.0002	0.0065	-0.0001	0.0060
100	50	100.0	0.0000	0.0005	-0.0001	0.0037	-0.0001	0.0038
25	100	99.9	0.0000	0.0009	-0.0003	0.0063	-0.0002	0.0067
50	100	100.0	0.0000	0.0006	0.0000	0.0039	-0.0001	0.0041
100	100	100.0	0.0000	0.0004	-0.0001	0.0027	0.0000	0.0026

7 Extensions

This section discusses two extensions: one allows time-invariant and common regressors, and the other allows a spatial autoregressive (SAR) specification for the errors. Both extensions are of practical relevance, and both can be studied within the ML framework.

7.1 Models with time-invariant and common regressors

Consider the following extended spatial panel data models with common shocks

$$
\begin{aligned}
y_{it} &= \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \sum_{p=1}^{k} x_{itp}\beta_p + r_i' h_t + \tau_i' p_t + \lambda_i' f_t + e_{it}; \\
x_{itp} &= r_i' s_{tp} + \eta_{ip}' p_t + \gamma_{ip}' f_t + v_{itp}; \qquad \text{for } \ p = 1, 2, \dots, k.
\end{aligned}
\tag{7.1}
$$

where r_i represents a vector of observable time-invariant variables such as race, gender, and education; p_t represents a vector of observable common variables (not varying with i) such as aggregate prices, unemployment rates, and other macroeconomic policy variables. Note that we allow the regression coefficients of time-invariant regressors to be time-varying, and allow the coefficients of common regressors to be individual dependent. This is a sensible way to include time-invariant and common regressors. For example, returns to schooling are likely to be time varying, and individual responses to policy variables are likely to be individual dependent. Imposing constant coefficients for these variables are also easy (Bai, 2009). Also note that we allow x_{it} to be correlated with the time-invariant regressors r_i and with the common regressors p_t, as shown in the x equation.

Model (7.1) falls within the framework of commons shocks. Let $f_t^\dagger = (h_t', p_t', s_{t1}', \dots, s_{tk}', f_t')'$, and let $\Phi_i^\dagger$ be defined as

$$\Phi_i^\dagger = \begin{bmatrix} r_i' & \tau_i' & 0 & \lambda_i' \\ 0 & \eta_i' & I_k \otimes r_i' & \gamma_i' \end{bmatrix}$$

where $\eta_i = (\eta_{i1}, \dots, \eta_{ik})$ and $\gamma_i = (\gamma_{i1}, \dots, \gamma_{ik})$. We can rewrite model (7.1) as

$$\begin{bmatrix} y_{it} - \rho \sum_{j=1}^N w_{ij,N} y_{jt} - \sum_{p=1}^k x_{itp}\beta_p \\ x_{it} \end{bmatrix} = \Phi_i^\dagger f_t^\dagger + \epsilon_{it},$$

which is similar to the model in Section 2. The difference here is that some components of the common factors $f_t^\dagger$ are observable, and some components of the factor loadings $\Phi_i^\dagger$ are observable. The maximum likelihood method is good at imposing restrictions. The observed components of $f_t^\dagger$ and of $\Phi_i^\dagger$ are not estimated but are restricted to their observed values.

We will not pursue the asymptotic analysis for this model to conserve space. A related investigation is given in Bai and Li (2014) in the absence of spatial effects. Instead we run a small simulation to demonstrate the performance of the MLE. The data are generated according to (7.1). The way to generate the factors, factor loadings, errors and heteroscedasticity is similar to Section 6. Other prespecified parameters are also the same except $\rho = 0.5$. The spatial weights matrix is that of "3 ahead and 3 behind." The dimensions of f_t, h_t and p_t are all one. We do not report the estimated coefficients for the time-invariant regressor and the common regressor (r_i, p_t) (there are many of them). For simplicity, we assume that the number of factors is known (it can also be estimated easily). Table 5 reports the simulation results based on 1000 repetitions. It is seen that the MLE performs quite well in the presence of time-invariant and common regressors.

Table 5: The performance of the MLE
in the presence of time-invariant and common regressors

N	T	ρ		β_1		β_2	
		Bias	RMSE	Bias	RMSE	Bias	RMSE
25	50	0.0000	0.0065	0.0001	0.0138	0.0009	0.0133
50	50	0.0002	0.0042	-0.0002	0.0084	0.0004	0.0088
100	50	-0.0001	0.0028	-0.0001	0.0056	-0.0001	0.0056
25	100	0.0001	0.0049	0.0007	0.0108	0.0004	0.0109
50	100	0.0000	0.0030	-0.0002	0.0064	-0.0004	0.0060
100	100	0.0000	0.0020	-0.0002	0.0039	-0.0002	0.0041

7.2 SAR disturbances

Spatial autoregressive (SAR) disturbances have been an important part of spatial modeling, and to which recent panel literature pays further attention; for example, Baltagi et al.

(2003), Baltagi et al. (2007), Kapoor et al. (2007), and Lee and Yu (2010). Here we consider spatial effects both in the dependent variable and in the errors, together with common shocks. Consider the following model

$$Y_t = \alpha + \rho W_N Y_t + X_t\beta + \Lambda f_t + U_t \tag{7.2}$$

with $U_t = \pi M_N U_t + e_t$, where M_N is another spatial weights matrix with its diagonal elements being zeros; Y_t, α, and U_t are $N \times 1$ vectors, X_t is $N \times k$, and $\Lambda = (\lambda_1, ..., \lambda_N)'$. If the common shock term Λf_t does not exist, the preceding model reduces to that of Lee and Yu (2010); but again we allow cross sectional heteroskedasticity.

To render an expression consistent with model (1.1), premultiply $I_N - \pi M_N$ on both sides. Then

$$Y_t = (\alpha - \pi M_N\alpha) + \rho W_N Y_t + \pi M_N Y_t - \rho\pi M_N W_N Y_t + X_t\beta - \pi M_N X_t\beta + (\Lambda - \pi M_N\Lambda) f_t + e_t.$$

We can treat $\alpha - \pi M_N\alpha$ as a new α and $\Lambda - \pi M_N\Lambda$ as a new Λ since they are free parameters. Now the above equation can be written as

$$Y_t = \alpha + \rho W_N Y_t + \pi M_N Y_t - \rho\pi M_N W_N Y_t + X_t\beta - \pi M_N X_t\beta + \Lambda f_t + e_t,$$

which can be alternatively written as

$$\begin{aligned} y_{it} = \alpha_i &+ \rho\Big(\sum_{j=1}^N w_{ij,N} y_{jt}\Big) + \pi\Big(\sum_{j=1}^N m_{ij,N} y_{jt}\Big) - \rho\pi\Big(\sum_{j=1}^N\sum_{l=1}^N m_{ij,N} w_{jl,N} y_{lt}\Big) \\ &+ x_{it}'\beta - \pi\Big(\sum_{j=1}^N m_{ij,N} x_{jt}'\Big)\beta + \lambda_i' f_t + e_{it} \end{aligned} \tag{7.3}$$

where $w_{ij,N}$ and $m_{ij,N}$ are the elements of W_N and M_N. Similar to model (1.1), we allow the regressors also to be affected by the common shocks,

$$x_{it} = \nu_i + \gamma_i' f_t + v_{it}. \tag{7.4}$$

Combining (7.3) and (7.4), by the same method in Section 2, we can rewrite the model as

$$D(\rho, \beta, \pi) z_t = \mu + \Phi f_t + \epsilon_t \tag{7.5}$$

where μ, Φ and ϵ_t are defined in the same way as in Section 2; $D(\rho, \beta, \pi)$ is an $N(k+1) \times N(k+1)$ matrix, whose (i, j) subblock, denoted by $D_{ij}(\rho, \beta, \pi)$, is equal to

$$D_{ij}(\rho, \beta, \pi) = \begin{cases} \begin{bmatrix} 1 + \rho\pi m_{i*,N} w_{*i,N} & -\beta' \\ 0 & I_k \end{bmatrix} & \text{if } \ i = j \\ \begin{bmatrix} -\rho w_{ij,N} - \pi m_{ij,N} + \rho\pi m_{i*,N} w_{*j,N} & \pi m_{ij,N}\beta' \\ 0 & 0 \end{bmatrix} & \text{if } \ i \neq j \end{cases}$$

where $m_{i*,N}$ is the ith row of M_N and $w_{*j,N}$ is the jth column of W_N.

Model (7.5) is similar to (2.2) except that the transformation matrix D is more complicated. Nevertheless, the inverse matrix of $D(\rho, \beta, \pi)$, denoted by $V(\rho, \beta, \pi)$, still has a closed form. Let $V_{ij}(\rho, \beta, \pi)$ be the (i, j)th subblocks of $V(\rho, \beta, \pi)$, then we have

$$V_{ij}(\rho, \beta, \pi) = \begin{cases} \begin{bmatrix} [(I_N - \pi M_N)(I_N - \rho W_N)]^{ii} & (I_N - \rho W_N)^{ii}\beta' \\ 0 & I_k \end{bmatrix} & \text{if } \; i = j \\ \begin{bmatrix} [(I_N - \pi M_N)(I_N - \rho W_N)]^{ij} & (I_N - \rho W_N)^{ij}\beta' \\ 0 & 0 \end{bmatrix} & \text{if } \; i \neq j \end{cases}$$

where $[(I_N - \pi M_N)(I_N - \rho W_N)]^{ij}$ and $(I_N - \rho W_N)^{ij}$ are the respective (i, j)th elements of $[(I_N - \pi M_N)(I_N - \rho W_N)]^{-1}$ and $(I_N - \rho W_N)^{-1}$. Using the above result, the analysis of the MLE is similar as model (1.1).

We use simulations to illustrate the performance of the MLE. The data are generated according to (7.5). The factors, factor loadings, errors and heteroskedasticity are generated in the same way as in Section 6. Other prespecified parameters such as the number of factors, the number of regressors and the true values of β are also the same; we set $\rho = 0.5$ and $\pi = 0.4$; W_N and M_N are set equal to each other and to be the "3 ahead and 3 behind" weights matrix. For simplicity, the number of factors is assumed to be known. Table 6 reports the simulation results based on 1000 repetitions.

Table 6 shows that the maximum likelihood method continue to perform well. The RMSE decreases as the sample size becomes larger, implying that the MLE is consistent.

Table 6: The performance of the MLE under SAR disturbances

N	T	ρ		β_1		β_2		π	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
25	50	-0.0002	0.0069	-0.0002	0.0092	-0.0002	0.0091	-0.0014	0.0185
50	50	0.0001	0.0042	0.0003	0.0056	0.0000	0.0055	-0.0004	0.0113
100	50	0.0002	0.0026	0.0001	0.0035	0.0000	0.0036	-0.0003	0.0073
25	100	0.0002	0.0049	-0.0001	0.0062	0.0000	0.0066	-0.0008	0.0126
50	100	0.0001	0.0031	0.0001	0.0041	-0.0001	0.0039	-0.0004	0.0079
100	100	0.0000	0.0019	0.0000	0.0025	0.0000	0.0026	-0.0002	0.0049

A further extension is to consider spatial weights on the lag Y_{t-1}. The idea of joint modeling of spatial effects and common shocks is similar and the MLE is still applicable.

8 Conclusion

This paper considers spatial panel data models with common shocks, in which the spatial lag term is endogenous and the explanatory variables are correlated with the unobservable common factors and factor loadings. The proposed maximum likelihood estimator is capable of handling of both types of cross sectional dependence. The results make it possible to determine which type of cross-section dependence or both are present. Heteroskedasticity is explicitly allowed. It is found that when heteroskedasticity is estimated, the limiting variance of MLE is no longer of a sandwich form regardless of normality. We provide a rigorous analysis for the asymptotic theory of the MLE, demonstrating its desirable properties. We also show that a version of the EM algorithm is very effective in estimating the model. The Monte Carlo simulations show that the MLE can be easily computed and has good finite sample properties. We also discuss some extensions of the model.

References

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

Anderson, T.W. (2003). *Introduction to Multivariate Statistical Analysis.* Hoboken, New Jersey: John Wiley & Sons.

Anselin, L. (1988). Spatial econometrics: methods and models. The Netherlands: Kluwer Academic Publishers.

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

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

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

Baltagi, B., Song, S.H., and Koh, W. (2003) Testing panel data regression models with spatial error correlation, *Journal of Econometrics*, 117, 123-150.

Baltagi, B., Song, S.H., Jung, B.C., and Koh, W. (2007) Testing for serial correlation, spatial autocorrelation and random effects using panel data, *Journal of Econometrics*, 140, 5-51.

Case, A. C. (1991) Spatial patterns in household demand. *Econometrica*, 59(4), 953–965.

Case, A. C., Rosen, H. S., and Hines, J. R. (1993) Budget spillovers and fiscal policy interdependence: Evidence from the states. *Journal of Public Economics*, 52(3), 285-307.

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

Cliff, A. D., and Ord, J. K. (1973) *Spatial autocorrelation,* London: Pion Ltd.

Conley, T. G. (1999). GMM Estimation with Cross Sectional Dependence, Journal of Econometrics, 92, 1-45.

Conley, T.G., and Dupor, B. (2003) “A Spatial Analysis of Sectoral Complementarity,” Journal of Political Economy, 111, 311-352.

Cressie, N. (1993): Statistics for Spatial Data. New York: John Wiley & Sons.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1-38.

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996) Crime and social interactions. *The Quarterly Journal of Economics*, 111(2), 507-548.

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

Kapoor, M., Kelejian, H. H., and Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97-130.

Kelejian, H. H., and Prucha, I. R. (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1), 99-121.

Kelejian, H. H., and Prucha, I. R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2), 509-533.

Kelejian, H. H., and Prucha, I. R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1), 53-67.

Lee, L. (2004a) Asymptotic distribution of quasi-maximum likelihood estiators for spatial augoregressive models, *Econometrica*, 72(6), 1899–1925.

Lee, L. (2004b) A supplement to "Asymptotic distribution of quasi-maximum likelihood estiators for spatial augoregressive models", *Manuscipt.*

Lee, L. and Yu, J. (2010) Estimation of spatial autoregressive panel data models with fixed effects, *Journal of Econometrics*, 154(2), 165-185.

Lin, X., and Lee, L. F. (2010) GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1), 34-52.

Liu, C., and Rubin, D. B. (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81(4), 633-648.

McLachlan, G. J., and Krishnan, T. *The EM Algorithm and Extensions.* 1997. New York: Wiley

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

Newey, W. and McFadden, D. (1994): Large Sample Estimation and Hypothesis Testing, in Engle, R.F. and D. McFadden (eds.) *Handbook of Econometrics,* North Holland.

Ord, K. (1975). Estimation methods for models of spatial interaction. Journal of the American Statistical Association, 70(349), 120-126.

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

Pesaran, M. H., and Tosetti, E. (2011). Large panels with spatial correlations and common factors, *Journal of Econometrics*, 161, 182–202.

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

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

Topa, G. (2001). Social Interactions, Local Spillovers and Unemployment, *Review of Economic Studies,* 68, 261-295.

Yu, J., de Jong, R., and Lee, L. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics*, 146(1), 118-134.

Appendix: Proofs for the theorems in the main text

In the appendix, we provide the detailed proofs for the theorems in the main text. We first define some notations which will be used throughout the appendix.

$$\hat{H} = (\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi})^{-1}; \qquad \hat{H}_N = N \cdot \hat{H};$$
$$\hat{G} = (I_r + \hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi})^{-1}; \qquad \hat{G}_N = N \cdot \hat{G}.$$

Appendix A: Proof for consistency

While in the main text, we use $(\rho, \beta, \Phi, \Sigma_{\epsilon\epsilon})$ to denote the true value of the coefficients. For proving consistency, we shall use a superscript "*" to denote the true values of parameters; the variables without "*" denote the input variables of the likelihood function. This notation is only used in Appendix A. Once consistency is established, we will drop "*" in Appendices B to F. The following lemmas are useful for the proof of consistency.

Lemma A.1 *Let $V(\rho,\beta)$ be the inverse matrix of $D(\rho,\beta)$, then*

$$V_{ij}(\rho,\beta) = \begin{cases} \begin{bmatrix} (I-\rho W_N)^{ii} & (I-\rho W_N)^{ii}\beta' \\ 0 & I_k \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} (I-\rho W_N)^{ij} & (I-\rho W_N)^{ij}\beta' \\ 0 & 0 \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{A.1}$$

where $V_{ij}(\rho,\beta)$ is the (i,j)th subblock of V and $(I-\rho W_N)^{ij}$ is the (i,j)th element of $(I-\rho W_N)^{-1}$. Furthermore, let $R = (I_N - \rho W_N)(I_N - \rho^* W_N)^{-1}$ and $\mathbb{D} = DD^{*-1}$ with $D^* = D(\rho^*, \beta^*)$. We have

$$\mathbb{D}_{ij} = \begin{cases} \begin{bmatrix} R_{ii} & R_{ii}\beta^{*\prime} - \beta' \\ 0 & I_K \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} R_{ij} & R_{ij}\beta^{*\prime} \\ 0 & 0 \end{bmatrix} & \text{if } i \neq j \end{cases}$$

where $\mathbb{D}_{ij}$ is the (i,j)th subblock of $\mathbb{D}$ and R_{ij} is the (i,j)th element of R.

Proof of Lemma A.1. This follows from direct verification.

Lemma A.2 *Let $(\rho,\beta) \in \mathcal{A}\times\mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are both compact sets. Under Assumptions A-F, uniformly on $\mathcal{A}\times\mathcal{B}$,*

$$(a) \quad \Big\| \sum_{j=1}^{N} R_{ij}(\gamma_j^*\beta^* + \lambda_j^*) - \gamma_i^{*\prime}\beta \Big\| \le C, \quad \text{for all } i;$$

$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{j=1}^{N}R_{ij}(e_{jt}+\beta^{*\prime}v_{jt})-\beta' v_{it}\Big)f_t\Big\|^2=O_p(T^{-1}),$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}^2-E(\tilde{e}_{it}^2)]\Big|^2=O_p(T^{-1}),$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}e_{it}-E(\tilde{e}_{it}e_{it})]\Big|^2=O_p(T^{-1}),$$

$$(e)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}v_{it}-E(\tilde{e}_{it}v_{it})]\Big\|^2=O_p(T^{-1}),$$

$$(f)\quad \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}\tilde{e}_{jt}-E(\tilde{e}_{it}\tilde{e}_{jt})]\Big|^2=O_p(T^{-1}),$$

$$(g)\quad \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}e_{jt}-E(\tilde{e}_{it}e_{jt})]\Big|^2=O_p(T^{-1}),$$

$$(h)\quad \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}v_{jt}-E(\tilde{e}_{it}v_{jt})]\Big\|^2=O_p(T^{-1}),$$

where $\tilde{e}_{it}=\sum_{j=1}^{N}S_{ij,N}(e_{jt}+\beta^{*\prime}v_{jt})$ *with* $S_{ij,N}$ *being the* (i,j)*th element of* $S_N=W_N(I_N-\rho^* W_N)^{-1}$*;* R *and* R_{ij} *are defined in Lemma A.1.*

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

$$(a)\quad \sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\Phi^*\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\epsilon_t'\Big)\mathbb{D}'\Sigma_{zz}^{-1}\Big]\Big|=o_p(1)$$

$$(b)\quad \sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t'-\Sigma_{\epsilon\epsilon}^*)\mathbb{D}'\Sigma_{zz}^{-1}\Big]\Big|=o_p(1)$$

$$(c)\quad \sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\bar{\epsilon}\bar{\epsilon}'\mathbb{D}'\Sigma_{zz}^{-1}\Big]\Big|=o_p(1)$$

where $\mathbb{D}=DD^{*-1}$ *and* $\Sigma_{zz}=\Phi\Phi'+\Sigma_{\epsilon\epsilon}$*.*

The proofs of the preceding two lemmas are given in Appendix D.

Proof of Proposition 4.1: Throughout the proof, we use the following centered objective function:

$$\begin{aligned}\mathcal{L}(\theta)=&-\frac{1}{2N}\ln|\Sigma_{zz}|+\frac{1}{N}\ln|D|-\frac{1}{2N}\mathrm{tr}[DM_{zz}D'\Sigma_{zz}^{-1}]\\&+\frac{1}{2N}\ln|\Sigma_{zz}^*|-\frac{1}{N}\ln|D^*|+\frac{k+1}{2}.\end{aligned}\tag{A.2}$$

The above objective function differs by a constant from the original one. By $D^*z_t=\mu^*+\Phi^*f_t+\epsilon_t$, we have

$$D^*M_{zz}D^{*\prime}=\Sigma_{zz}^*+\Phi^*\Big(\frac{1}{T}\sum_{t=1}^{T}f_t\epsilon_t'\Big)+\Big(\frac{1}{T}\sum_{t=1}^{T}\epsilon_tf_t'\Big)\Phi^{*\prime}+\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t'-\Sigma_{\epsilon\epsilon}^*)-\bar{\epsilon}\bar{\epsilon}',$$

where $\Sigma_{zz}^* = \Phi^*\Phi^{*\prime} + \Sigma_{\epsilon\epsilon}^*$ and $\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^T \epsilon_t$. The above equation uses the fact that $\dot{f}_t = f_t$ for $\bar{f} = 0$. Thus,

$$M_{zz} = D^{*-1}\Sigma_{zz}^* D^{*\prime -1} + \mathbb{S}, \tag{A.3}$$

where

$$\begin{aligned}\mathbb{S} = & D^{*-1}\Phi^*\Big(\frac{1}{T}\sum_{t=1}^T f_t\epsilon_t'\Big)D^{*\prime-1} + D^{*-1}\Big(\frac{1}{T}\sum_{t=1}^T \epsilon_t f_t'\Big)\Phi^{*\prime}D^{*\prime-1} \\ & + D^{*-1}\frac{1}{T}\sum_{t=1}^T(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*)D^{*\prime-1} - D^{*-1}\bar{\epsilon}\bar{\epsilon}'D^{*\prime-1}.\end{aligned}$$

Substituting (A.3) into (A.2),

$$\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \mathcal{L}_2(\theta), \tag{A.4}$$

where

$$\begin{aligned}\mathcal{L}_1(\theta) = & -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\mathrm{tr}[DD^{*-1}\Sigma_{zz}^*D^{*-1\prime}D'\Sigma_{zz}^{-1}] \\ & + \frac{1}{2N}\ln|\Sigma_{zz}^*| - \frac{1}{N}\ln|D^*| + \frac{k+1}{2}\end{aligned}$$

and

$$\mathcal{L}_2(\theta) = -\frac{1}{2N}\mathrm{tr}[D\mathbb{S}D'\Sigma_{zz}^{-1}]. \tag{A.5}$$

By Lemma A.3, we have $\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = o_p(1)$. Since $\hat{\theta}$ maximizes $\mathcal{L}(\theta)$, we have $\mathcal{L}(\hat{\theta}) \geq \mathcal{L}(\theta^*)$, implying $\mathcal{L}_1(\hat{\theta}) \geq \mathcal{L}_1(\theta^*) + \mathcal{L}_2(\theta^*) - \mathcal{L}_2(\hat{\theta})$. By Lemma A.3, $|\mathcal{L}_2(\theta^*) - \mathcal{L}_2(\hat{\theta})| \geq -2\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = -|o_p(1)|$. Given this result, together with $\mathcal{L}_1(\theta^*) = 0$, we have

$$\mathcal{L}_1(\hat{\theta}) \geq -|o_p(1)|. \tag{A.6}$$

Letting $\hat{\mathbb{D}} = D(\hat{\rho},\hat{\beta})D^{*-1}$, we rewrite $\mathcal{L}_1(\hat{\theta})$ as

$$\begin{aligned}\mathcal{L}_1(\hat{\theta}) = & -\frac{1}{2N}\ln|\hat{\Sigma}_{zz}| + \frac{1}{N}\ln|\hat{D}| - \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] \\ & + \frac{1}{2N}\ln|\Sigma_{zz}^*| - \frac{1}{N}\ln|D^*| + \frac{k+1}{2}\end{aligned}$$

First consider $\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]$, which can be written as, in view of $\Sigma_{zz}^* = \Phi^*\Phi^{*\prime} + \Sigma_{\epsilon\epsilon}^*$,

$$\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] = \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] + \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Phi^*\Phi^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] \triangleq i_1 + i_2, \quad \text{say.}$$

By the Woodbery formula $\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{\epsilon\epsilon}^{-1} - \hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}\hat{G}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}$, where $\hat{G} = (I_r + \hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi})^{-1}$, i_1 can be written as

$$i_1 = \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}] - \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}\hat{G}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}] \triangleq i_3 - i_4, \quad \text{say}$$

Let

$$\hat{R} = (I_N - \hat{\rho}W_N)(I_N - \rho^*W_N)^{-1} = I_N - (\hat{\rho} - \rho^*)S_N.$$

Algebra shows that

$$
\begin{aligned}
i_3 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{R}_{ij}^2\frac{\sigma_j^2}{\hat{\sigma}_i^2} + \frac{1}{2N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}(\hat{R}_{ii}\beta^* - \hat{\beta})'\Sigma_{iiv}^*(\hat{R}_{ii}\beta^* - \hat{\beta}) \\
+ \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{R}_{ij}^2\beta^{*\prime}\Sigma_{jjv}^*\beta^* + \frac{1}{2N}\mathrm{tr}\Big[\sum_{i=1}^{N}\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}\Big]
\end{aligned}
$$

where $\hat{R}_{ij}$ is the (i,j)-th element of $\hat{R}$. In addition, we also have $i_4 = o_p(1)$ uniformly on Θ. To see this, by the boundedness of $\hat{\Sigma}_{ii}$ and Σ_{ii}^*, $\hat{\mathbb{D}}\Sigma_{\epsilon\epsilon}^*\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}$ is less than[2] $C\hat{\mathbb{D}}\hat{\mathbb{D}}'$ for some C, which is further less than $CC'I_{N(k+1)}$ for some constant C', as shown in the proof of Lemma A.3(c). This result leads to $i_4 \leq C_1C_2\frac{1}{2N}\mathrm{tr}[\hat{\Phi}'\hat{\Sigma}_{\varepsilon\varepsilon}\hat{\Phi}\hat{G}] = O_p(N^{-1})$. Given the results on i_1, i_2, i_3 and i_4, together with

$$
\ln|\Sigma_{zz}| = \ln|\Sigma_{\epsilon\epsilon}| + \ln|I_r + \Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi| = \sum_{i=1}^{N}(\ln\sigma_i^2 + \ln|\Sigma_{iiv}|) + \ln|I_r + \Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi|,
$$

we have

$$
\begin{aligned}
\mathcal{L}_1(\hat{\theta}) = -\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{R}_{ij}^2\frac{\sigma_j^{*2}}{\hat{\sigma}_i^2} - \frac{1}{2N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}(\hat{R}_{ii}\beta^* - \hat{\beta})'\Sigma_{iiv}^*(\hat{R}_{ii}\beta^* - \hat{\beta}) \\
-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{R}_{ij}^2\beta^{*\prime}\Sigma_{jjv}^*\beta^* - \frac{1}{2N}\mathrm{tr}\Big[\sum_{i=1}^{N}\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}\Big] - \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Phi^*\Phi^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] \\
-\frac{1}{2N}\sum_{i=1}^{N}(\ln\hat{\sigma}_i^2 + \ln|\hat{\Sigma}_{iiv}|) - \frac{1}{2N}\ln|I_r + \hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}| + \frac{1}{2N}\sum_{i=1}^{N}(\ln\sigma_i^{*2} + \ln|\Sigma_{iiv}^*|) \\
+\frac{1}{2N}\ln|I_r + \Phi^{*\prime}\Sigma_{\epsilon\epsilon}^{*-1}\Phi^*| + \frac{1}{N}\ln|\hat{R}| + \frac{k+1}{2} \geq -|o_p(1)| \qquad \text{(A.7)}
\end{aligned}
$$

where we use the fact that $\ln|\hat{D}| - \ln|D^*| = \ln|\hat{D}D^{*-1}| = \ln|\hat{\mathbb{D}}|$ and $\det(\hat{\mathbb{D}}) = \det(\hat{R})$. Consider

$$
-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{R}_{ij}^2\frac{\sigma_j^{*2}}{\hat{\sigma}_i^2} + \frac{1}{N}\ln|\hat{R}|
$$

which is equivalent to

$$
-\frac{1}{2N}\mathrm{tr}[\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}] + \frac{1}{2N}\ln|\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}| + \frac{1}{2N}\sum_{i=1}^{N}(\ln\hat{\sigma}_i^2 - \ln\sigma_i^{*2}) \qquad \text{(A.8)}
$$

Substituting (A.8) into (A.7), we can rewrite $\mathcal{L}_1(\hat{\theta})$ as

$$
\begin{aligned}
\mathcal{L}_1(\hat{\theta}) = -\Big\{\frac{1}{2N}\mathrm{tr}[\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}] - \frac{1}{2N}\ln|\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}| - \frac{1}{2}\Big\} - \Big\{\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Phi^*\Phi^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]\Big\} \\
-\Big\{\frac{1}{2N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}(\hat{R}_{ii}\beta^* - \hat{\beta})'\Sigma_{iiv}^*(\hat{R}_{ii}\beta^* - \hat{\beta})\Big\} - \Big\{\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{R}_{ij}^2\beta^{*\prime}\Sigma_{jjv}^*\beta^*\Big\}
\end{aligned}
$$

[2]For matrices A and B, we say $A \leq B$ if $B - A$ is a semi-definite positive matrix.

$$-\Big\{\frac{1}{2N}\sum_{i=1}^{N}\Big(\ln|\hat{\Sigma}_{iiv}\Sigma_{iiv}^{*-1}|+\mathrm{tr}[\Sigma_{iiv}^{*}\hat{\Sigma}_{iiv}^{-1}]-k\Big)\Big\}-\Big\{\frac{1}{2N}\ln|I_r+\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}|\Big\}\geq -|o_p(1)|$$

where we use the fact that $\frac{1}{N}\ln|I_r+\Phi^{*\prime}\Sigma_{\epsilon\epsilon}^{*-1}\Phi^*| = o(1)$. In the above equation, all the expressions in the braces are nonnegative. The non-negativity can be easily verified for the expressions in the 2nd, 3rd, 4th and 6th braces. For the first expression, let $\iota_1,\iota_2,\dots,\iota_N$ be the eigenvalues of matrix $\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}$. Then the first expression is equal to $\frac{1}{2N}\sum_{i=1}^N(\iota_i-\ln\iota_i-1)$. Consider the function $y=x-\ln x-1$, which achieves its minimum value 0 at $x=1$. The same arguments also work for the 5th expression. Given that all the expressions in the braces are nonnegative, and $\mathcal{L}_1(\hat{\theta})\geq -|o_p(1)|$, each of the five expressions must be $o_p(1)$, that is,

$$\frac{1}{2N}\mathrm{tr}[\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}]-\frac{1}{2N}\ln|\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}|-\frac{1}{2}=o_p(1) \tag{A.9}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}(\hat{R}_{ii}\beta^*-\hat{\beta})'\Sigma_{iiv}^*(\hat{R}_{ii}\beta^*-\hat{\beta})=o_p(1) \tag{A.10}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{R}_{ij}^2\beta^{*\prime}\Sigma_{jjv}^*\beta^*=o_p(1) \tag{A.11}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\Big(\mathrm{tr}[\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}]-\ln|\Sigma_{iiv}\hat{\Sigma}_{iiv}^{*-1}|-k\Big)=o_p(1) \tag{A.12}$$

$$\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Phi^*\Phi^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]=o_p(1) \tag{A.13}$$

Now we first prove $\hat{\rho}\xrightarrow{p}\rho^*$ under the local identification condition (2.5). By $\beta^*\neq 0$, together with the boundedness of $\Sigma_{iiv}^*,\hat{\sigma}_i^2$, there exists a positive constant c such that $\frac{1}{\hat{\sigma}_i^2}\beta^{*\prime}\Sigma_{jjv}^*\beta^*>c$ for all i and j. This result, together with (A.11), implies

$$o_p(1)=\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{R}_{ij}^2\beta^{*\prime}\Sigma_{jjv}^*\beta^*>c\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\hat{R}_{ij}^2>0,$$

implying $\frac{1}{2N}\sum_{i=1}^N\sum_{j=1,j\neq i}^N\hat{R}_{ij}^2=o_p(1)$. By $\hat{R}=I_N-(\hat{\rho}-\rho^*)S_N$, the above result can be written as

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}S_{ij,N}^2\Big)(\hat{\rho}-\rho^*)^2=o_p(1),$$

implying $\hat{\rho}\xrightarrow{p}\rho^*$ by (2.7).

We next prove $\hat{\rho}\xrightarrow{p}\rho^*$ under the global identification condition (2.6). Consider (A.9), which can be written as

$$\frac{1}{2N}\mathrm{tr}[\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1/2}]-\frac{1}{2N}\ln|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1/2}|-\frac{1}{2}=o_p(1) \tag{A.14}$$

We use ι_i ($i=1,2,\dots,N$) to denote the eigenvalues of $\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1/2}$ temporarily. By the boundedness of $\hat{\rho},\hat{\sigma}_i^2$, it is easy to see $\iota_i\in[C^{-1},C]$ for all i for some large constant C. In addition, there exists a constant b (for example $b = \frac{1}{4C^2}$), such that $x - \ln x - 1 \geq b(x-1)^2$ for all $x \in [C^{-1}, C]$. Given this result, we have

$$\frac{1}{2N}\mathrm{tr}[\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}] - \frac{1}{2N}\ln|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2}| - \frac{1}{2} = \frac{1}{2N}\sum_{i=1}^{N}(\iota_i - \ln \iota_i - 1)$$

$$\geq b\frac{1}{2N}\sum_{i=1}^{N}(\iota_i - 1)^2 = b\frac{1}{2N}\left\|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_N\right\|^2,$$

implying

$$\frac{1}{N}\left\|\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_N\right\|^2 = o_p(1).$$

The above result is equivalent to

$$\frac{1}{N}\mathrm{tr}\Big[(\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_N)'(\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_N)\Big] = o_p(1),$$

which can be written as

$$\frac{1}{N}\mathrm{tr}\Big[\hat{\Sigma}_{ee}^{-1/2}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})'\hat{\Sigma}_{ee}^{-1/2}\Big] = o_p(1).$$

However, by the boundedness of $\hat{\sigma}_i^2$, there exists some constant c such that $\hat{\Sigma}_{ee}^{-1/2} \geq cI_N$. Thus

$$o_p(1) = \frac{1}{N}\mathrm{tr}\Big[\hat{\Sigma}_{ee}^{-1/2}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})'\hat{\Sigma}_{ee}^{-1/2}\Big]$$

$$\geq c^4\frac{1}{N}\mathrm{tr}\Big[(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})'\Big] = c^4\frac{1}{N}\left\|\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee}\right\|^2 > 0.$$

So we have

$$\frac{1}{N}\left\|\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee}\right\|^2 = o_p(1).$$

By $\hat{R} = I_N - (\hat{\rho} - \rho^*)S_N$, the above results is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\sigma_i^{*2} - \hat{\sigma}_i^2 - 2(\hat{\rho} - \rho^*)S_{ii,N}\sigma_i^{*2} + (\hat{\rho} - \rho^*)^2\sum_{j=1}^{N}S_{ij,N}^2\sigma_j^{*2}\Big)^2$$

$$+(\hat{\rho} - \rho^*)^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\Big(S_{ij,N}\sigma_j^{*2} + S_{ji,N}\sigma_i^{*2} - (\hat{\rho} - \rho^*)\sum_{p=1}^{N}S_{ip,N}S_{jp,N}\sigma_p^{*2}\Big)^2 = o_p(1).$$

The two expressions on the left hand side are both nonnegative, so we have

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\sigma_i^{*2} - \hat{\sigma}_i^2 - 2(\hat{\rho} - \rho^*)S_{ii,N}\sigma_i^{*2} + (\hat{\rho} - \rho^*)^2\sum_{j=1}^{N}S_{ij,N}^2\sigma_j^{*2}\Big)^2 = o_p(1), \tag{A.15}$$

$$(\hat{\rho} - \rho^*)^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\Big(S_{ij,N}\sigma_j^{*2} + S_{ji,N}\sigma_i^{*2} - (\hat{\rho} - \rho^*)\sum_{p=1}^{N}S_{ip,N}S_{jp,N}\sigma_p^{*2}\Big)^2 = o_p(1). \tag{A.16}$$

Result (A.16) implies $\hat{\rho} \xrightarrow{p} \rho^*$ in view of (2.6). So we have proved the consistency of $\hat{\rho}$ under both the local and global identification conditions.

By $\hat{\rho} \xrightarrow{p} \rho^*$, we have $\hat{R}_{ii} \xrightarrow{p} 1$ for all i. This result, together with (A.10), leads to $\hat{\beta} \xrightarrow{p} \beta^*$. Given $\hat{\rho} \xrightarrow{p} \rho^*$, (A.15) implies that

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\sigma}_i^2 - \sigma_i^{*2}\right)^2 = o_p(1)$$

From (A.12), using the arguments of Bai and Li (2014), we can show

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iiv} - \Sigma_{iiv}^*\|^2 = o_p(1).$$

Combining the above two results, we have

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

The last claim of Proposition 4.1 can be proved from (A.13) in a similar way as in Bai and Li (2014). The details are omitted. This completes the proof. $\square$

Appendix B: Detailed proofs for the convergence rates

Given consistency, we now drop the superscript "*" from the true parameters for notational simplicity. We also drop the terms involving $\bar{\epsilon}$ because they are negligible. Let $\omega = (\rho, \beta')'$ and $\hat{\omega} = (\hat{\rho}, \hat{\beta}')'$. The following lemmas are useful for the subsequent analysis.

Lemma B.1 *Under Assumptions A-F, together with IC, we have*

$$(a)\quad \left\|\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\boldsymbol{b}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\right\| = O_p(\|\hat{\omega}-\omega\|^2)$$

$$(b)\quad \left\|(\hat{\rho}-\rho)^2\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}\boldsymbol{a}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\right\| = O_p(\|\hat{\omega}-\omega\|^2)$$

$$(c)\quad \left\|(\hat{\rho}-\rho)\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}(\hat{\beta}-\beta)'\boldsymbol{b}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\right\| = O_p(\|\hat{\omega}-\omega\|^2)$$

$$(d)\quad \left\|(\hat{\rho}-\rho)\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}f_t'\right\| = O_p(\|\hat{\omega}-\omega\|)$$

$$(e)\quad \left\|(\hat{\rho}-\rho)\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}\epsilon_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\right\| = O_p(\|\hat{\omega}-\omega\|)$$

$$(f)\quad \left\|\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)f_t'\right\| = O_p(\|\hat{\omega}-\omega\|)$$

$$(g)\quad \left\|\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)\epsilon_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\right\| = O_p(\|\hat{\omega}-\omega\|)$$

where $\boldsymbol{a}_{it} = (\ddot{y}_{it}, 0_{1\times k})'$*;* $\boldsymbol{b}_{it} = (x_{it}, 0_{k\times k})'$ *with* $\ddot{y}_{it} = \sum_{j=1}^{N} w_{ij,N}\dot{y}_{jt}$.

Lemma B.2 *Under Assumptions A-F, together with IC,*

$$(a)\quad \Big\|\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\Big\| = O_p(T^{-1/2})$$

$$(b)\quad \Big\|\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t\epsilon_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\Big\| = O_p(T^{-1/2})$$

$$(c)\quad \Big\|\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Phi}_i'\hat{H}\Big\| = o_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big]^{1/2}\Big)$$

The proof of Lemma B.1 is given in Appendix D. The proof of Lemma B.2 is similar to that of Lemma A.5 in the supplement of Bai and Li (2014) and hence are omitted.

Proposition B.1 *Under Assumptions A-F, together with IC, we have*

$$A = O_p(T^{-1/2})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2\Big)+O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big]^{\frac{1}{2}}\Big)+O_p(\|\hat{\omega}-\omega\|),$$

where $A=(\hat{\Phi}-\Phi)'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}\hat{H}$ *with* $\hat{H}=(\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi})^{-1}$.

Proof of Proposition B.1. By (3.2), we have

$$\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{D}M_{zz}\hat{D}'-\hat{\Sigma}_{zz})\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}=0, \tag{B.1}$$

which is equivalent to

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{l=1}^{N}\hat{D}_{il}z_{lt}\Big)\Big(\sum_{l=1}^{N}\hat{D}_{jl}z_{lt}\Big)'-\hat{\Phi}_i\hat{\Phi}_j'-\hat{\Sigma}_{ii}1(i=j)\Big]\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j=0. \tag{B.2}$$

Let $\boldsymbol{a}_{it}=(\ddot{y}_{it},0_{1\times k})'$; $\boldsymbol{b}_{it}=(x_{it},0_{k\times k})'$, then

$$\sum_{l=1}^{N}\hat{D}_{il}z_{lt}=-\boldsymbol{a}_{it}(\hat{\rho}-\rho)-\boldsymbol{b}_{it}(\hat{\beta}-\beta)+\Phi_i'f_t+\epsilon_{it}. \tag{B.3}$$

Substituting (B.3) into (B.2), with some algebra, we have

$$\begin{aligned} A+A' &= AA'+\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\boldsymbol{b}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\\ &+(\hat{\rho}-\rho)^2\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}\boldsymbol{a}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}+\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\\ &+(\hat{\rho}-\rho)\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}(\hat{\beta}-\beta)'\boldsymbol{b}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}+(I-A)'\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t\epsilon_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\\ &+\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)\boldsymbol{a}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}(\hat{\rho}-\rho)+\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}f_t'(I-A)\end{aligned}$$

$$-(\hat{\rho}-\rho)\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}f_t'(I-A)-(I-A)'\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t\boldsymbol{a}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}(\hat{\rho}-\rho)$$

$$-(\hat{\rho}-\rho)\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{it}\epsilon_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}-\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)f_t'(I-A)$$

$$-\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}\boldsymbol{a}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}(\hat{\rho}-\rho)-(I-A)'\frac{1}{T}\sum_{j=1}^{N}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\boldsymbol{b}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}$$

$$-\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{it}(\hat{\beta}-\beta)\epsilon_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}-\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Phi}_i'\hat{H}$$

$$-\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}(\hat{\beta}-\beta)'\boldsymbol{b}_{jt}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}. \tag{B.4}$$

where $\varepsilon_{ij,t}=\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{it}\epsilon_{jt}'-E(\epsilon_{it}\epsilon_{jt}')]$. Consider the right hand side of (B.4). Term AA' is of smaller order term than A and hence negligible. The magnitudes of the remaining 16 terms are given in Lemmas B.1 and B.2. Then (B.4) can be written as

$$A+A'=O_p(T^{-1/2})+o_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big]^{1/2}\Big)+O_p(\|\hat{\omega}-\omega\|). \tag{B.5}$$

Furthermore, $\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}$ and $\frac{1}{N}\Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi$ are both diagonal matrices, thus

$$\text{Nondiag}\Big\{\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{\Phi}-\frac{1}{N}\Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi\Big\}=0$$

where *Nondiag* denotes non-diagonal elements. Adding and subtracting terms, the above equation can be written as

$$\text{Nondiag}\Big\{A\hat{Q}+\hat{Q}A'\Big\}=\text{Nondiag}\Big\{\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Phi}-\Phi)-\frac{1}{N}\Phi'(\hat{\Sigma}_{\epsilon\epsilon}^{-1}-\Sigma_{\epsilon\epsilon}^{-1})\Phi\Big\}$$

$$=\text{Nondiag}\Big\{O_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2\Big)+O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big]^{1/2}\Big)\Big\} \tag{B.6}$$

Given (B.5) and (B.6), together with $\hat{Q}\xrightarrow{p}Q$, we solve for A as

$$A=O_p(T^{-1/2})+O_p\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2\Big)+O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big]^{\frac{1}{2}}\Big)+O_p(\|\hat{\omega}-\omega\|).$$

This completes the proof. □

Lemma B.3 *Under the assumptions of Proposition B.1,*

(a) $$\frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)^2\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}\boldsymbol{a}_{it}'\Big\|^2=o_p(\|\hat{\omega}-\omega\|^2),$$

(b) $$\frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}\epsilon_{it}'\Big\|^2=O_p(\|\hat{\omega}-\omega\|^2),$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}\boldsymbol{a}_{it}'\Big\|^2=O_p(\|\hat{\omega}-\omega\|^2),$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'\Big\|^2=o_p(\|\hat{\omega}-\omega\|^2),$$

$$(e)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)\boldsymbol{a}_{it}'\Big\|^2=o_p(\|\hat{\omega}-\omega\|^2),$$

$$(f)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}f_t'\Phi_i\Big\|^2=O_p(\|\hat{\omega}-\omega\|^2),$$

$$(g)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t\boldsymbol{a}_{it}'\Big\|^2=O_p(\|\hat{\omega}-\omega\|^2).$$

Lemma B.4 *Under the assumptions of Proposition B.1,*

$$(a)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'\Big\|=o_p(\|\hat{\omega}-\omega\|^2),$$

$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)f_t'\Phi_i\Big\|^2=O_p(\|\hat{\omega}-\omega\|^2),$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'\Big\|^2=O_p(\|\hat{\omega}-\omega\|^2),$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)\epsilon_{it}'\Big\|^2=o_p(\|\hat{\omega}-\omega\|^2),$$

$$(e)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'\Big\|^2=o_p(\|\hat{\omega}-\omega\|^2).$$

Lemma B.5 *Under the assumptions of Proposition B.1,*

$$(a)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{jt}\epsilon_{it}'-E(\epsilon_{jt}\epsilon_{it}')]\Big\|^2=O_p(T^{-1}),$$

$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t\epsilon_{it}'\Big\|^2=O_p(T^{-1}),$$

$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}f_t'\Phi_i\Big\|^2=O_p(T^{-1}),$$

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\hat{H}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\Big\|^2=o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2).$$

The proofs of Lemmas B.3 and B.4 are similar to that of Lemma B.1. The proof of Lemma B.5 is similar to that of Lemma A.7 in the supplement of Bai and Li (2014). So the proofs of the three lemmas are omitted.

Proposition B.2 *Under the assumptions of Proposition B.1, we have*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2=O_p(T^{-1})+O_p(\|\hat{\omega}-\omega\|^2),$$

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

Proof of Proposition B.2. Consider the first order condition (3.2). Using (B.3), we can rewrite (3.2) as

$$\begin{aligned}
\hat{\Phi}_i-\Phi_i=&-A'\Phi_i+(\hat{\rho}-\rho)^2\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}\boldsymbol{a}_{it}'+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\varepsilon_{ji,t}\\
&+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}\epsilon_{it}'\\
&-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}\boldsymbol{a}_{it}'+(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'\\
&+(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)\boldsymbol{a}_{it}'-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{a}_{jt}f_t'\Phi_i\\
&-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t\boldsymbol{a}_{it}'-\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)f_t'\Phi_i\\
&-\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'-\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{b}_{jt}(\hat{\beta}-\beta)\epsilon_{it}'\\
&-\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}(\hat{\beta}-\beta)'\boldsymbol{b}_{it}'+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t\epsilon_{it}'\\
&+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}f_t'\Phi_i-\hat{H}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii}).
\end{aligned}\tag{B.7}$$

for each i, where $\varepsilon_{ij,t}=T^{-1}\sum_{t=1}^{T}[\epsilon_{it}\epsilon_{jt}'-E(\epsilon_{it}\epsilon_{jt}')]$ and $\boldsymbol{a}_{it}$ and $\boldsymbol{b}_{it}$ are defined in the proof of Proposition B.1. There are 17 terms on the right hand side of (B.7), which we use $\boldsymbol{i}_{i,1},\boldsymbol{i}_{i,2},\ldots,\boldsymbol{i}_{i,17}$ to denote. Then

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2\le C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Phi}_i-\Phi_i\|^2\le C\frac{1}{N}\sum_{i=1}^{N}(\|\boldsymbol{i}_{i,1}\|+\cdots+\|\boldsymbol{i}_{i,17}\|)^2\\
\le 17C\frac{1}{N}\sum_{i=1}^{N}(\|\boldsymbol{i}_{i,1}\|^2+\cdots+\|\boldsymbol{i}_{i,17}\|^2)
\end{aligned}\tag{B.8}$$

By Proposition B.1, we have

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

The remaining 16 terms are governed by Lemmas B.3, B.4 and B.5. Given these results, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2=O_p(T^{-1})+O_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2)+O_p(\|\hat{\omega}-\omega\|^2). \quad \text{(B.9)}$$

The first order condition (3.3) gives

$$\begin{aligned}\hat{\sigma}_i^2-\sigma_i^2&=\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)+(\hat{\rho}-\rho)^2\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}^2+\frac{1}{T}\sum_{t=1}^{T}\left[x_{it}'(\hat{\beta}-\beta)\right]^2\\&+2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}x_{it}'(\hat{\beta}-\beta)-2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}f_t'\lambda_i-2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}e_{it}\\&-2\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\hat{\beta}-\beta)f_t'\lambda_i-2\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\hat{\beta}-\beta)\epsilon_{it}+2\lambda_i'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\\&-(\hat{\lambda}_i-\lambda_i)'(\hat{\lambda}_i-\lambda_i)-2\lambda_i'(\hat{\lambda}_i-\lambda_i). \quad \text{(B.10)}\end{aligned}$$

and

$$\begin{aligned}\hat{\Sigma}_{iiv}-\Sigma_{iiv}&=\frac{1}{T}\sum_{t=1}^{T}(v_{it}v_{it}'-\Sigma_{iiv})+\gamma_i'\frac{1}{T}\sum_{t=1}^{T}f_tv_{it}'+\frac{1}{T}\sum_{t=1}^{T}v_{it}f_t'\gamma_i \quad \text{(B.11)}\\&-(\hat{\gamma}_i-\gamma_i)'\gamma_i-\gamma_i'(\hat{\gamma}_i-\gamma_i)-(\hat{\gamma}_i-\gamma_i)'(\hat{\gamma}_i-\gamma_i)\end{aligned}$$

However, the first column of $\hat{\Phi}_i-\Phi_i$ is $\hat{\lambda}_i-\lambda_i$. This gives

$$\begin{aligned}\hat{\lambda}_i-\lambda_i&=-A'\lambda_i+(\hat{\rho}-\rho)^2\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}\ddot{y}_{it}-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}e_{it}\\&+\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_{it}+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{jt}e_{it}-E(\epsilon_{jt}e_{it})]\\&-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}\ddot{y}_{it}+(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}(\hat{\beta}-\beta)'x_{it}\\&+(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)\ddot{y}_{it}-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}f_t'\lambda_i\\&-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t\ddot{y}_{it}-\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)f_t'\lambda_i \quad \text{(B.12)}\\&-\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'x_{it}-\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)e_{it}\\&-\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}(\hat{\beta}-\beta)'x_{it}+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\\&+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}f_t\lambda_i'-\hat{H}\hat{\lambda}_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}\end{aligned}$$

Similarly we have

$$\hat{\gamma}_i - \gamma_i = -A'\gamma_i - (\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}f_t'\gamma_i \tag{B.13}$$

$$-(\hat{\rho}-\rho)\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}v_{it}' + \hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{jt}v_{it}' - E(\epsilon_{jt}v_{it}')]$$

$$-\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)f_t'\gamma_i - \hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)v_{it}'$$

$$+\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_tv_{it}' + \hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}f_t'\gamma_i - \hat{H}\hat{\gamma}_i\hat{\Sigma}_{iiv}^{-1}(\hat{\Sigma}_{iiv}-\Sigma_{iiv})$$

Substituting (B.12) into (B.10), we have

$$\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2) + (\hat{\rho}-\rho)^2\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}^2 + \frac{1}{T}\sum_{t=1}^{T}\left[x_{it}'(\hat{\beta}-\beta)\right]^2$$

$$+2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}x_{it}'(\hat{\beta}-\beta) - 2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}f_t'\lambda_i - 2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}e_{it}$$

$$-2\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\hat{\beta}-\beta)f_t'\lambda_i - 2\frac{1}{T}\sum_{t=1}^{T}x_{it}'(\hat{\beta}-\beta)e_{it} + 2\lambda_i'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2\lambda_i'A'\lambda_i \tag{B.14}$$

$$-2(\hat{\rho}-\rho)^2\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}\ddot{y}_{it} - 2\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{jt}e_{it} - E(\epsilon_{jt}e_{it})]$$

$$-2\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_{it} + 2(\hat{\rho}-\rho)\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}e_{it}$$

$$+2(\hat{\rho}-\rho)\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}\ddot{y}_{it} - 2(\hat{\rho}-\rho)\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}(\hat{\beta}-\beta)'x_{it}$$

$$-2(\hat{\rho}-\rho)\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)\ddot{y}_{it} + 2(\hat{\rho}-\rho)\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{jt}f_t'\lambda_i$$

$$+2(\hat{\rho}-\rho)\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t\ddot{y}_{it} + 2\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)f_t'\lambda_i$$

$$+2\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_t(\hat{\beta}-\beta)'x_{it} + 2\lambda_i'\hat{H}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j\frac{1}{T}\sum_{t=1}^{T}x_{jt}'(\hat{\beta}-\beta)e_{it}$$

$$+2\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}(\hat{\beta}-\beta)'x_{it} - 2\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\Phi_j'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}$$

$$-2\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}\epsilon_{jt}f_t\lambda_i' + 2\lambda_i'\hat{H}\hat{\lambda}_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2} - (\hat{\lambda}_i-\lambda_i)'(\hat{\lambda}_i-\lambda_i)$$

There are 27 terms on the right hand side of (B.14), which are denoted by $\boldsymbol{ii}_{i,1}, \dots, \boldsymbol{ii}_{i,27}$. Then

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\sigma}_i^2-\sigma_i^2\|^2 \le 27\frac{1}{N}\sum_{i=1}^{N}(\|\boldsymbol{ii}_{i,1}\|^2+\cdots+\|\boldsymbol{ii}_{i,27}\|^2).$$

Checking these 27 terms one by one, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\sigma}_i^2-\sigma_i^2\|^2 = O_p(T^{-1})+o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2)+O_p(\|\hat{\omega}-\omega\|^2).$$

Applying the similar method to (B.11) and (B.13), we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iiv}-\Sigma_{iiv}\|^2 = O_p(T^{-1})+o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2)+O_p(\|\hat{\omega}-\omega\|^2).$$

Using the preceding two results, together with

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\hat{\sigma}_i^2-\sigma_i^2\|^2+\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iiv}-\Sigma_{iiv}\|^2,$$

we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 = O_p(T^{-1})+o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2)+O_p(\|\hat{\omega}-\omega\|^2). \quad \text{(B.15)}$$

Substituting (B.9) into (B.15), we have

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

Substituting the above result into (B.9), we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i-\Phi_i\|^2 = O_p(T^{-1})+O_p(\|\hat{\omega}-\omega\|^2).$$

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

Remark. Notice that the 10th term on the right hand side of (B.14) is $2\lambda_i'A'\lambda_i = \lambda_i'(A+A')\lambda_i$. However, the expression of $A+A'$ is given in (B.4). Using the expression on the right hand side of (B.4) to replace $A+A'$, we have the following alternative expression of $\hat{\sigma}_i^2-\sigma_i^2$, which is useful to prove Proposition B.3 and Lemma C.4 below.

$$\begin{aligned}\hat{\sigma}_i^2-\sigma_i^2 =& \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)-(\hat{\lambda}_i-\lambda_i)'(\hat{\lambda}_i-\lambda_i)-2(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}e_{it}\\ &-2\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{jt}e_{it}-E(\epsilon_{jt}e_{it})]+\mathcal{S}_{i,\sigma^2}+\mathcal{T}_{i,\sigma^2} \end{aligned} \quad \text{(B.16)}$$

where

$$\begin{aligned}
\mathcal{S}_{i,\sigma^2} &= (\hat\rho-\rho)^2\frac{1}{T}\sum_{t=1}^T \ddot y_{it}^2 + \frac{1}{T}\sum_{t=1}^T [x_{it}'(\hat\beta-\beta)]^2 + 2(\hat\rho-\rho)\frac{1}{T}\sum_{t=1}^T \ddot y_{it}x_{it}'(\hat\beta-\beta) \qquad \text{(B.17)}\\
&-2\frac{1}{T}\sum_{t=1}^T x_{it}'(\hat\beta-\beta)e_{it} - 2\lambda_i'A'\frac{1}{T}\sum_{t=1}^T f_t x_{it}'(\hat\beta-\beta) + 2\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T x_{jt}'(\hat\beta-\beta)e_{it}\\
&-2(\hat\rho-\rho)^2\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}\ddot y_{it} - 2\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T x_{jt}'(\hat\beta-\beta)(\hat\beta-\beta)'x_{it}\\
&-2(\hat\rho-\rho)\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}(\hat\beta-\beta)'x_{it} - 2(\hat\rho-\rho)\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T x_{jt}'(\hat\beta-\beta)\ddot y_{it}\\
&-2(\hat\rho-\rho)\lambda_i'A'\frac{1}{T}\sum_{t=1}^T f_t\ddot y_{it} + \lambda_i'\Big[\hat H\sum_{j=1}^N\sum_{l=1}^N \frac{1}{\hat\sigma_j^2\hat\sigma_l^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T x_{jt}'(\hat\beta-\beta)(\hat\beta-\beta)'x_{lt}\hat\lambda_l'\hat H\Big]\lambda_i\\
&+(\hat\rho-\rho)^2\lambda_i'\Big[\hat H\sum_{j=1}^N\sum_{l=1}^N \frac{1}{\hat\sigma_j^2\hat\sigma_l^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}\ddot y_{lt}\hat\lambda_l'\hat H\Big]\lambda_i + 2(\hat\rho-\rho)\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}f_t'A\lambda_i\\
&+2(\hat\rho-\rho)\lambda_i'\hat H\sum_{j=1}^N \hat\Phi_j\hat\Sigma_{jj}^{-1}\frac{1}{T}\sum_{t=1}^T \epsilon_{jt}\ddot y_{it} - 2(\hat\rho-\rho)\lambda_i'\Big[\hat H\sum_{j=1}^N\sum_{l=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}\epsilon_{lt}\hat\Sigma_{ll}^{-1}\hat\Phi_l'\hat H\Big]\lambda_i\\
&-2\lambda_i'\Big[\hat H\sum_{j=1}^N\sum_{l=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T x_{jt}'(\hat\beta-\beta)\epsilon_{lt}\hat\Sigma_{ll}^{-1}\hat\Phi_l'\hat H\Big]\lambda_i + 2(\hat\rho-\rho)\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}e_{it}\\
&+2\lambda_i'\hat H\sum_{j=1}^N \frac{1}{\hat\sigma_j^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T x_{jt}'(\hat\beta-\beta)f_t'A\lambda_i + 2\lambda_i'\hat H\sum_{j=1}^N \hat\Phi_j\hat\Sigma_{jj}^{-1}\frac{1}{T}\sum_{t=1}^T \epsilon_{jt}(\hat\beta-\beta)'x_{it}\\
&+2(\hat\rho-\rho)\lambda_i'\Big[\hat H\sum_{j=1}^N\sum_{l=1}^N \frac{1}{\hat\sigma_j^2\hat\sigma_l^2}\hat\lambda_j\frac{1}{T}\sum_{t=1}^T \ddot y_{jt}(\hat\beta-\beta)'x_{lt}\hat\lambda_l'\hat H\Big]\lambda_i
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{T}_{i,\sigma^2} &= 2\lambda_i'A'\frac{1}{T}\sum_{t=1}^T f_te_{it} - 2\lambda_i'A'\frac{1}{T}\sum_{t=1}^T f_t\epsilon_{jt}'\hat\Sigma_{jj}^{-1}\hat\Phi_j'\hat H\lambda_i + 2\lambda_i'\hat H\hat\lambda_i\frac{\hat\sigma_i^2-\sigma_i^2}{\hat\sigma_i^2} + \lambda_i'AA'\lambda_i\\
&+\lambda_i'\Big[\hat H\sum_{j=1}^N\sum_{l=1}^N \hat\Phi_j\hat\Sigma_{jj}^{-1}\varepsilon_{jl,t}\hat\Sigma_{ll}^{-1}\hat\Phi_l'\hat H\Big]\lambda_i - \lambda_i'\hat H\sum_{j=1}^N \hat\Phi_j\hat\Sigma_{jj}^{-1}(\hat\Sigma_{jj}-\Sigma_{jj})\hat\Sigma_{jj}^{-1}\hat\Phi_j'\hat H\lambda_i \qquad \text{(B.18)}
\end{aligned}$$

Using the results in Proposition B.2, we can strengthen Proposition B.1, which is given in Proposition B.3 below. We need the following lemmas.

Lemma B.6 *Under Assumptions A-F, together with IC,*

(a) $\Big\|\hat H\sum_{i=1}^N\sum_{j=1}^N \hat\Phi_i\hat\Sigma_{ii}^{-1}\varepsilon_{ij,t}\hat\Sigma_{jj}^{-1}\hat\Phi_j'\hat H\Big\| = O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat\omega-\omega\|)$;

(b) $\Big\|\frac{1}{T}\sum_{j=1}^N\sum_{t=1}^T f_t\epsilon_{jt}'\hat\Sigma_{jj}^{-1}\hat\Phi_j'\hat H\Big\| = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + o_p(\|\hat\omega-\omega\|)$;

(c) $\left\|\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Phi}_i'\hat{H}\right\| = O_p(N^{-1}T^{-1/2}) + o_p(\|\hat{\omega}-\omega\|)$.

Lemma B.7 *Under Assumptions A-F, together with IC,*

(a) $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'\mathcal{S}_{i,\sigma^2} = o_p(\|\hat{\omega}-\omega\|)$;

(b) $$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'\mathcal{T}_{i,\sigma^2} = o_p(A) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\omega}-\omega\|);$$

(c) $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2) = O_p(N^{-1/2}T^{-1/2})$;

(d) $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'(\hat{\lambda}_i-\lambda_i)'(\hat{\lambda}_i-\lambda_i) = O_p(T^{-1}) + o_p(\|\hat{\omega}-\omega\|)$;

(e) $(\hat{\rho}-\rho)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'\frac{1}{T}\sum_{t=1}^{T}\ddot{y}_{it}e_{it} = O_p(\|\hat{\omega}-\omega\|)$

(f) $$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'\lambda_i'\hat{H}\sum_{j=1}^{N}\hat{\Phi}_j\hat{\Sigma}_{jj}^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{jt}e_{it}-E(\epsilon_{jt}e_{it})] = O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\omega}-\omega\|);$$

The proofs of Lemmas B.6 and B.7 are given in Appendix D.

Proposition B.3 *Under the assumptions of Proposition B.1, together with IC, we have*

$$A = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + O_p(\|\hat{\omega}-\omega\|).$$

Proof of Proposition B.3. Consider equation (B.4). Using the results in Lemmas B.1 and B.6, we can be rewritten equation (B.4) as

$$A + A' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + O_p(\|\hat{\omega}-\omega\|) \tag{B.19}$$

By equation (B.6), we have

$$\text{Nondiag}\Big\{A\hat{Q}+\hat{Q}A'\Big\} = \text{Nondiag}\Big\{\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Phi}-\Phi) - \frac{1}{N}\Phi'(\hat{\Sigma}_{\epsilon\epsilon}^{-1}-\Sigma_{\epsilon\epsilon}^{-1})\Phi\Big\}$$

The first term on the right hand side is $O_p(T^{-1})+o_p(\|\hat{\omega}-\omega\|)$ by Proposition B.2. Consider the second term. From

$$-(\hat{\Sigma}_{\epsilon\epsilon}^{-1}-\Sigma_{\epsilon\epsilon}^{-1}) = \hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})\Sigma_{\epsilon\epsilon}^{-1} = \Sigma_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})\Sigma_{\epsilon\epsilon}^{-1} - \hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})\Sigma_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})\Sigma_{\epsilon\epsilon}^{-1}$$

we have

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

$$-\frac{1}{N}\Phi'\big[\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})\Sigma_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})\Sigma_{\epsilon\epsilon}^{-1}\big]\Phi = ii_1 - ii_2, \quad \text{say}$$

By the boundedness of $\Sigma_{ii}, \hat{\Sigma}_{ii}$ and Φ_i, we have

$$\begin{aligned}\|ii_2\| &= \Big\|\frac{1}{N}\sum_{i=1}^{N}\Phi_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\Sigma_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\Sigma_{ii}^{-1}\Phi_i'\Big\| \\ &\le C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 = O_p(T^{-1}) + o_p(\|\hat{\omega}-\omega\|)\end{aligned}$$

by Proposition B.2.

Now consider ii_1, which is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'(\hat{\sigma}_i^2-\sigma_i^2) + \frac{1}{N}\sum_{i=1}^{N}\gamma_i\Sigma_{iiv}^{-1}(\hat{\Sigma}_{iiv}-\Sigma_{iiv})\Sigma_{iiv}^{-1}\gamma_i'.$$

Substitute (B.16) into the first expression. By Lemma B.7, we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\lambda_i\lambda_i'(\hat{\sigma}_i^2-\sigma_i^2) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + O_p(\|\hat{\omega}-\omega\|).$$

Similarly, we also have

$$\frac{1}{N}\sum_{i=1}^{N}\gamma_i\Sigma_{iiv}^{-1}(\hat{\Sigma}_{iiv}-\Sigma_{iiv})\Sigma_{iiv}^{-1}\gamma_i' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + O_p(\|\hat{\omega}-\omega\|).$$

Then it follows

$$\text{Nondiag}\Big\{A\hat{Q}+\hat{Q}A'\Big\} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + O_p(\|\hat{\omega}-\omega\|). \tag{B.20}$$

By (B.19) and (B.20), together with $\hat{Q}\xrightarrow{p}Q$, we have, by solving for A,

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

as stated in Proposition B.3. $\square$

Appendix C: Proof for Theorem 4.2

We define the following notations to simplify expressions.

$\hat{\Pi}_i^{\ddot{y}\ddot{y}\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\ddot{y}_{is}\ddot{y}_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$	$\hat{\Pi}_i^{e\ddot{y}\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}e_{is}\ddot{y}_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$
$\hat{\Pi}_i^{\ddot{y}\epsilon\Phi} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\ddot{y}_{is}\epsilon_{js}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'$	$\hat{\Pi}_i^{x\ddot{y}\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}x_{is}'(\hat{\beta}-\beta)\ddot{y}_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$
$\hat{\Pi}^{f\ddot{y}\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}f_s\ddot{y}_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$	$\hat{\Pi}^{fx\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}f_s(\hat{\beta}-\beta)'x_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$
$\hat{\Pi}_i^{\ddot{y}f} = \frac{1}{T}\sum_{s=1}^{T}\ddot{y}_{is}f_s'$	$\hat{\Pi}_i^{\ddot{y}x\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}\ddot{y}_{is}(\hat{\beta}-\beta)'x_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$
$\hat{\Pi}_i^{xf} = \frac{1}{T}\sum_{s=1}^{T}x_{is}'(\hat{\beta}-\beta)f_s'$	$\hat{\Pi}_i^{ex\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}e_{is}(\hat{\beta}-\beta)'x_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$
$\Pi_i^{ef} = \frac{1}{T}\sum_{s=1}^{T}e_{is}f_s'$	$\hat{\Pi}_i^{x\epsilon\Phi} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}x_{is}'(\hat{\beta}-\beta)\epsilon_{js}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'$
$\hat{\Pi}^{f\epsilon\Phi} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}f_s\epsilon_{js}'\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'$	$\hat{\Pi}^{fx_p\lambda} = \frac{1}{NT}\sum_{j=1}^{N}\sum_{s=1}^{T}f_sx_{jsp}\hat{\lambda}_j'/\hat{\sigma}_j^2$
$\hat{\Pi}_i^{xx\lambda} = \frac{1}{NT}\sum_{j=1}^{T}\sum_{s=1}^{T}x_{is}'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_{js}\hat{\lambda}_j'/\hat{\sigma}_j^2$	

The proofs for the following four lemmas are given in Appendix D.

Lemma C.1 *Under the assumptions of Theorem 4.2,*

$$\mathcal{S}_\rho = o_p(\|\hat\omega - \omega\|),$$
$$\mathcal{S}_{\beta,p} = o_p(\|\hat\omega - \omega\|),$$

where $\mathcal{S}_\rho$ and $\mathcal{S}_{\beta,p}$ are defined in (C.7) and (C.10) below.

Lemma C.2 *Under the assumptions of Theorem 4.2*

$$\begin{aligned}\mathcal{T}_\rho = & -\frac{1}{N}\sum_{i=1}^N \frac{1}{\hat\sigma_i^2} S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big]'\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big] \\ & + O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat\omega - \omega\|). \\ \mathcal{T}_{\beta,p} = & \, O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat\omega - \omega\|).\end{aligned}$$

where $\mathcal{T}_\rho$ and $\mathcal{T}_{\beta,p}$ are defined in (C.8) and (C.11) below.

Remark. The first term of $\mathcal{T}_\rho$ in Lemma C.2 is $O_p(T^{-1})$ which comes from the 12th term of $\mathcal{T}_\rho$ in (C.8).

Lemma C.3 *Let $\varrho = \frac{1}{N}\Xi' S_N' \Sigma_{ee}^{-1}\Lambda$, $\varphi_p = \frac{1}{N}\Xi' S_N' \Sigma_{ee}\Gamma_p$, $\vartheta_p = \frac{1}{N}\Gamma_p'\Sigma_{ee}^{-1}\Lambda$ and $\varsigma_{pq} = \frac{1}{N}\Gamma_p'\Sigma_{ee}^{-1}\Gamma_q$, where $\Gamma_p = (\gamma_{1p}, \gamma_{2p}, \dots, \gamma_{Np})'$ and $\Xi = \Lambda + \sum_{p=1}^k \beta_p\Gamma_p$. Under the assumptions of Theorem 4.2 we have*

$$(a) \quad \hat\Pi^{f\ddot y\lambda} = \frac{1}{N}\Xi' S_N'\Sigma_{ee}^{-1}\Lambda + o_p(1) = \varrho + o_p(1)$$

$$(b) \quad \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat\sigma_i^2}\ddot y_{it}^2 = \frac{1}{N}\mathrm{tr}[\Sigma_{ee}^{-1}S_n(\Xi\Xi' + \Psi)S_n'] + o_p(1)$$

$$(c) \quad \frac{1}{T}\sum_{t=1}^T\Big(\frac{1}{N}\sum_{i=1}^N\frac{1}{\hat\sigma_i^2}\ddot y_{it}\hat\lambda_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^N\frac{1}{\hat\sigma_i^2}\hat\lambda_i\ddot y_{it}\Big) = \mathrm{tr}[\varrho Q^{-1}\varrho'] + o_p(1)$$

$$(d) \quad \frac{1}{N}\mathrm{tr}[S_N W_N(I_N - \hat\rho W_N)^{-1}] = \frac{1}{N}\mathrm{tr}(S_N^2) + o_p(1)$$

$$(e) \quad \mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\hat\sigma_i^2}f_t\ddot y_{it}\hat\Pi_i^{\ddot y f}\Big)(I-A)\Big] = \frac{1}{N}\mathrm{tr}(\Xi' S_N'\Sigma_{ee}^{-1}S_N\Xi) + o_p(1)$$

$$(f) \quad \mathrm{tr}\Big[\hat\Pi^{f\ddot y\lambda}\hat H_N\hat\Pi^{f\ddot y\lambda\prime}(I-A)\Big] = \mathrm{tr}(\varrho Q^{-1}\varrho') + o_p(1)$$

$$(g) \quad \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\hat\sigma_i^2}\ddot y_{it}x_{itp} = \mathrm{tr}(\varphi_p) + \frac{1}{N}\Big[\beta'\sum_{i=1}^N\frac{1}{\sigma_i^2}S_{ii,N}\Sigma_{iiv}I_k^p\Big] + o_p(1)$$

$$(h) \quad \frac{1}{T}\sum_{t=1}^T\Big(\frac{1}{N}\sum_{i=1}^N\frac{1}{\hat\sigma_i^2}\ddot y_{it}\hat\lambda_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^N\frac{1}{\hat\sigma_i^2}\hat\lambda_i x_{itp}\Big) = \mathrm{tr}(\varrho Q^{-1}\vartheta_p') + o_p(1)$$

$$(i) \quad \mathrm{tr}\Big[\hat\Pi^{f\ddot y\lambda}H_N\hat\Pi^{fx\lambda\prime}(I-A)\Big] = \sum_{p=1}^k\mathrm{tr}(\varrho Q^{-1}\vartheta_p')(\hat\beta_p - \beta) + o_p(\|\hat\omega - \omega\|)$$

$$(j)\quad \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{xf}(I-A)\Big]=\sum_{p=1}^{k}\mathrm{tr}(\varphi_p)(\hat{\beta}_p-\beta_p)+o_p(\|\hat{\omega}-\omega\|)$$

$$(k)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}x_{itq}=\mathrm{tr}(\varsigma_{pq})+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\Sigma_{iiv}^{(p,q)}+o_p(1)$$

$$(l)\quad \frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}x_{itp}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i x_{itq}\Big)=\mathrm{tr}(\vartheta_pQ^{-1}\vartheta_q')+o_p(1)$$

$$(m)\quad \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_tx_{itp}\hat{\Pi}_i^{xf}(I-A)\Big]=\sum_{q=1}^{k}\mathrm{tr}(\varsigma_{pq})(\hat{\beta}_q-\beta_q)+o_p(\|\hat{\omega}-\omega\|)$$

$$(n)\quad \mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\hat{\Pi}^{f\ddot{y}\lambda\prime}(I-A)\Big]=\mathrm{tr}(\varrho Q^{-1}\vartheta_p')+o_p(1)$$

$$(o)\quad \mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\hat{\Pi}^{fx\lambda\prime}(I-A)\Big]=\sum_{q=1}^{k}\mathrm{tr}[\vartheta_pQ^{-1}\vartheta_q'](\hat{\beta}_q-\beta_q)+o_p(\|\hat{\omega}-\omega\|)$$

$$(p)\quad \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_tx_{itp}\hat{\Pi}_i^{\ddot{y}f}(I-A)\Big]=\mathrm{tr}(\varphi_p)+o_p(1)$$

where I_k^p is the pth column of the k dimensional identity matrix and $\Sigma_{iiv}^{(p,q)}$ is the (p,q)th element of Σ_{iiv}. In addition, $\hat{\beta}_p$ and β_p are the pth element of $\hat{\beta}$ and β, respectively.

Lemma C.4 *Under the assumptions of Theorem 4.2,*

$$\begin{aligned}(a)\quad \frac{1}{NT}\sum_{t=1}^{T}\beta'V_t'S_N'\hat{\Sigma}_{ee}^{-1}e_t &= \frac{1}{NT}\sum_{t=1}^{T}\beta'V_t'S_N'\Sigma_{ee}^{-1}e_t+O_p(N^{-1/2}T^{-1})\\ &\quad+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|),\\ (b)\quad \frac{1}{NT}\sum_{t=1}^{T}e_t'S_N'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})e_t &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\\ &\quad-\frac{1}{N}\sum_{i=1}^{N}\frac{S_{ii,N}}{\sigma_i^2}\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)+\Big[\frac{2}{N}\sum_{i=1}^{N}S_{ii,N}^2\Big](\hat{\rho}-\rho)\\ &\quad+O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|),\\ (c)\quad \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}v_{itp}e_{it} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}v_{itp}e_{it}+O_p(N^{-1/2}T^{-1})\\ &\quad+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|),\end{aligned}$$

where $V_t=(v_{1t},v_{2t},\dots,v_{Nt})'$.

Remark. To prove (b), we need to substitute (B.16) into (b). The first term on the right hand side of (b) comes from the second term of (B.16) and the third term comes from the third term of (B.16). The proofs for Lemmas C.1-C.4 are in Appendix D.

Proof of Theorem 4.2. Consider (3.4), which is equivalent to

$$-\frac{1}{N}\mathrm{tr}((I-\hat{\rho}W)^{-1}W)-(\hat{\rho}-\rho)\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}^2\Big)$$

$$
\begin{aligned}
&-(\hat{\beta}-\beta)'\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{it}\ddot{y}_{it}\Big)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\lambda_i' f_t\ddot{y}_{it}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}e_{it}\ddot{y}_{it}\\
&\qquad+(\hat{\rho}-\rho)\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big] \qquad \text{(C.1)}\\
&\qquad+(\hat{\beta}-\beta)'\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}x_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big]\\
&-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}f_t\hat{\lambda}_i'\hat{G}\hat{\Phi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\Phi\Big]-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{G}\hat{\Phi}'\hat{\Sigma}_{\varepsilon\varepsilon}^{-1}\epsilon_t\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big]=0
\end{aligned}
$$

Let $\ddot{Y}_t=(\ddot{y}_{1t},\ddot{y}_{2t},\ldots,\ddot{y}_{Nt})'$, then

$$
\ddot{Y}_t=W_NY_t=S_N(X_t\beta+\Lambda f_t+e_t)=S_N(I_N\otimes\bar{\beta}')(\Phi f_t+\epsilon_t)=S_N\Xi f_t+S_N\zeta_t. \qquad \text{(C.2)}
$$

where $\bar{\beta}=(1,\beta')'$, $\xi_i=\lambda_i+\sum_{p=1}^{k}\beta_p\gamma_{ip}$, $\zeta_{it}=e_{it}+\sum_{p=1}^{k}\beta_p v_{itp}$, $\Xi=(\xi_1,\xi_2,\ldots,\xi_N)'$ and $\zeta_t=(\zeta_{1t},\zeta_{2t},\ldots,\zeta_{Nt})'$. Substituting (B.2) into (B.1), we have

$$
\begin{aligned}
&\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}^2\Big](\hat{\rho}-\rho)-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big](\hat{\rho}-\rho)\\
&\qquad+\frac{1}{N}\mathrm{tr}\Big[S_NW_N(I_N-\hat{\rho}W_N)^{-1}\Big](\hat{\rho}-\rho)+\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}x_{it}'\Big)(\hat{\beta}-\beta)\\
&\qquad\qquad-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i x_{it}'\Big)\Big](\hat{\beta}-\beta) \qquad \text{(C.3)}\\
&=\frac{1}{NT}\sum_{t=1}^{T}f_t'\Lambda'S_N'\hat{\Sigma}_{ee}^{-1}e_t+\frac{1}{NT}\sum_{t=1}^{T}\beta'X_t'S_N'\hat{\Sigma}_{ee}^{-1}e_t+\frac{1}{NT}\sum_{t=1}^{T}e_t'S_N'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})e_t\\
&+\frac{1}{NT}\sum_{t=1}^{T}\mathrm{tr}[S_N'\Sigma_{ee}^{-1}(e_te_t'-\Sigma_{ee})]-\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}[\epsilon_t\zeta_t'-E(\epsilon_t\zeta_t')]S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]\\
&\quad-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}(\hat{\lambda}_i-\lambda_i)'f_t\ddot{y}_{it}+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}f_t\hat{\lambda}_i'\hat{G}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Phi}-\Phi)\Big]\\
&\quad+\mathrm{tr}\Big[\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})(I_N\otimes\bar{\beta})S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]-\mathrm{tr}\Big[\frac{1}{N}(\hat{\Xi}-\Xi)'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]\\
&\qquad-\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\epsilon_tf_t'\Xi'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]+\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}f_t\zeta_t'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big].
\end{aligned}
$$

By the similar method, we can rewrite the first order condition (3.5) as

$$
\begin{aligned}
&\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}\ddot{y}_{it}\Big)(\hat{\rho}-\rho)-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}x_{itp}\hat{\lambda}_i'\Big)\hat{G}_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big](\hat{\rho}-\rho)\\
&+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}x_{it}'(\hat{\beta}-\beta)-\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}x_{itp}\hat{\lambda}_i'\Big)\hat{G}_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i x_{it}'\Big)(\hat{\beta}-\beta)\\
&\quad=-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}(\hat{\lambda}_i-\lambda_i)'f_t+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}\hat{\lambda}_i'\hat{G}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Phi}-\Phi)f_t \qquad \text{(C.4)}
\end{aligned}
$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}v_{itp}\hat{\lambda}_i'\hat{G}f_t-\mathrm{tr}\Big[\frac{1}{NT}\hat{G}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}[\epsilon_{it}v_{jtp}-E(\epsilon_{it}v_{jtp})\frac{1}{\hat{\sigma}_j^2}]\hat{\lambda}_j'\Big]$$
$$-\mathrm{tr}\Big[\frac{1}{NT}\hat{G}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\epsilon_t f_t'\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\gamma_{ip}\hat{\lambda}_i'\Big]+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}e_{it}$$
$$+\mathrm{tr}\Big[\frac{1}{N}\hat{G}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})I_{k+1}^{p+1}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i'\Big]-\mathrm{tr}\Big[\frac{1}{N}\hat{G}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}(\hat{\gamma}_{ip}-\gamma_{ip})\hat{\lambda}_i'\Big]$$

The fifth term on the right hand side of (C.3) involves $\hat{\lambda}_i-\lambda_i$, whose expression is given by (B.12). Substituting (B.12) into (C.3), we have

$$\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}^2\Big](\hat{\rho}-\rho)-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big](\hat{\rho}-\rho)$$
$$+\frac{1}{N}\mathrm{tr}\Big[S_NW_N(I_N-\hat{\rho}W_N)^{-1}\Big](\hat{\rho}-\rho)+\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}x_{it}'\Big)(\hat{\beta}-\beta)$$
$$-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i x_{it}'\Big)\Big](\hat{\beta}-\beta)$$
$$=\frac{1}{NT}\sum_{t=1}^{T}f_t'\Lambda'S_N'\hat{\Sigma}_{ee}^{-1}e_t+\frac{1}{NT}\sum_{t=1}^{T}\beta'X_t'S_N'\hat{\Sigma}_{ee}^{-1}e_t \tag{C.5}$$
$$+\frac{1}{NT}\sum_{t=1}^{T}e_t'S_N'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})e_t+\frac{1}{NT}\sum_{t=1}^{T}\mathrm{tr}[S_N'\Sigma_{ee}^{-1}(e_te_t'-\Sigma_{ee})]$$
$$+\mathrm{tr}[\hat{\Pi}^{f\ddot{y}\lambda}(A+A')]-tr[\hat{\Pi}^{f\ddot{y}\lambda}\hat{G}A']-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}(\hat{\lambda}_i-\lambda_i)'A\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})(I_N\otimes\bar{\beta})S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]-\mathrm{tr}\Big[\frac{1}{N}(\hat{\Xi}-\Xi)'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]$$
$$-\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\epsilon_tf_t'\Xi'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]+\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}f_t\zeta_t'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^4}\hat{\lambda}_i'\hat{H}f_t\ddot{y}_{it}-(\hat{\rho}-\rho)^2\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}f_t\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}\ddot{y}\lambda}\hat{H}\Big]$$
$$-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\sum_{s=1}^{T}[e_{is}\epsilon_{js}'-E(e_{is}\epsilon_{js}')]\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}f_t\ddot{y}_{it}-\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{xx\lambda}\hat{H}\Big]$$
$$+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{e\ddot{y}\lambda}\hat{H}\Big]+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}\epsilon\Phi}\hat{H}\Big]$$
$$-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{x\ddot{y}\lambda}\hat{H}\Big]-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}x\lambda}\hat{H}\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\lambda_i'\hat{\Pi}^{fx\lambda}\hat{H}\Big]+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}f}(I-A)\Big]$$
$$+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\lambda_i'\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\Big]+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{xf}(I-A)\Big]$$

$$
+\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{ex\lambda}\hat{H}\Big]+\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{x\epsilon\Phi}\hat{H}\Big]
$$

$$
-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{ef}(I-A)\Big]-\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\lambda_i'\hat{\Pi}^{f\epsilon\Phi}\hat{H}\Big]
$$

$$
-\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}[\epsilon_t\zeta_t'-E(\epsilon_t\zeta_t')]S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]
$$

The fifth term on the right hand side of (C.5) involves $A+A'$, which is given in (B.4). Substituting (B.4) into (C.5), we have

$$
\begin{aligned}
&\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}^2\Big](\hat{\rho}-\rho)-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big](\hat{\rho}-\rho)\\
&\quad+\frac{1}{N}\mathrm{tr}\Big[S_NW_N(I_N-\hat{\rho}W_N)^{-1}\Big](\hat{\rho}-\rho)+\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}x_{it}'\Big)(\hat{\beta}-\beta)\\
&\quad-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{y}_{it}\hat{\lambda}_i'\Big)G_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_ix_{it}'\Big)\Big](\hat{\beta}-\beta) \qquad \text{(C.6)}\\
&=\frac{1}{NT}\sum_{t=1}^{T}\beta'V_t'S_N'\hat{\Sigma}_{ee}^{-1}e_t+\frac{1}{NT}\sum_{t=1}^{T}e_t'S_N'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})e_t+\frac{1}{NT}\sum_{t=1}^{T}\mathrm{tr}[S_N'\Sigma_{ee}^{-1}(e_te_t'-\Sigma_{ee})]\\
&\quad-\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}_N\hat{\Pi}^{fx\lambda\prime}(I-A)\Big]+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}f}(I-A)\Big]\\
&\quad+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{xf}(I-A)\Big]-(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}_N\hat{\Pi}^{f\ddot{y}\lambda\prime}(I-A)\Big]+\mathcal{S}_\rho+\mathcal{T}_\rho.
\end{aligned}
$$

where V_t is defined in Lemma C.4 and

$$
\begin{aligned}
\mathcal{S}_\rho&=-(\hat{\rho}-\rho)^2\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}\ddot{y}\lambda}\hat{H}\Big]-\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{xx\lambda}\hat{H}\Big]\\
&\quad+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{e\ddot{y}\lambda}\hat{H}\Big]+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}\epsilon\Phi}\hat{H}\Big]\\
&\quad-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{x\ddot{y}\lambda}\hat{H}\Big]-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{\ddot{y}x\lambda}\hat{H}\Big]\\
&\quad+(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\hat{\Pi}_i^{\ddot{y}x\lambda}\hat{H}_N\Big]+(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\hat{\Pi}_i^{x\ddot{y}\lambda}\hat{H}_N\Big]\\
&\quad-(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\hat{\Pi}_i^{\ddot{y}\epsilon\Phi}\hat{H}_N\Big]-(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}_N\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\Pi}_i^{\ddot{y}\epsilon\Phi\prime}\hat{\lambda}_i'\hat{H}\Big] \qquad \text{(C.7)}\\
&\quad-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}(\hat{\lambda}_i-\lambda_i)'\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\Big]+(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}A'\hat{\Pi}^{f\ddot{y}\lambda}H_N\Big]\\
&\quad+\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\hat{\Pi}_i^{xx\lambda}H_N\Big]+(\hat{\rho}-\rho)^2\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\hat{\Pi}_i^{\ddot{y}\ddot{y}\lambda}\hat{H}_N\Big]
\end{aligned}
$$

$$-\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}(\hat{\lambda}_i-\lambda_i)'\hat{\Pi}^{fx\lambda}\hat{H}\Big]+\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}A'\hat{\Pi}^{fx\lambda}H_N\Big]$$
$$-\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\hat{\Pi}_i^{x\epsilon\Phi}\hat{H}_N\Big]-\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}_N\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\Phi}_i^{x\epsilon\Phi\prime}\hat{\lambda}_i'\hat{H}\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{ex\lambda}\hat{H}\Big]+\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{x\epsilon\Phi}\hat{H}\Big]$$

and

$$\mathcal{T}_\rho=\mathrm{tr}\Big[\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}(\hat{\Sigma}_{\epsilon\epsilon}-\Sigma_{\epsilon\epsilon})(I_N\otimes\bar{\beta})S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]-\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}A'\hat{\Pi}^{f\epsilon\Phi}\hat{H}_N\Big]$$
$$-\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}(\hat{\lambda}_i-\lambda_i)'A\Big]-\mathrm{tr}\Big[\frac{1}{N}(\hat{\Xi}-\Xi)'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]-\mathrm{tr}[\hat{\Pi}^{f\ddot{y}\lambda}\hat{G}A']$$
$$-\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}_N\hat{\Pi}^{f\epsilon\Phi\prime}A\Big]-\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}\sum_{s=1}^{T}[e_{is}\epsilon_{js}'-E(e_{is}\epsilon_{js}')]\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}f_t\ddot{y}_{it}$$
$$+\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}f_t\zeta_t'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big]-\mathrm{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\hat{\Phi}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}[\epsilon_t\zeta_t'-E(\epsilon_t\zeta_t')]S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big] \quad (C.8)$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^4}\hat{\lambda}_i'\hat{H}f_t\ddot{y}_{it}+\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\Big]$$
$$-\mathrm{tr}\Big[\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\sum_{i=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Phi}_i'\hat{H}\Big]-\mathrm{tr}\Big[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t\zeta_t'S_N'\hat{\Sigma}_{ee}^{-1}e_sf_s'\Big]$$
$$+\mathrm{tr}[\hat{\Pi}^{f\ddot{y}\lambda}AA']+\mathrm{tr}\Big[\Big(\hat{\Pi}^{f\ddot{y}\lambda}-\frac{1}{N}\Xi'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\Big)\hat{H}_N\hat{\Pi}^{f\epsilon\Phi\prime}\Big]+\mathrm{tr}\Big[\Xi'S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{H}\hat{G}\hat{\Pi}^{f\epsilon\Phi\prime}\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}(\hat{\lambda}_i-\lambda_i)'\hat{\Pi}^{f\epsilon\Phi}\hat{H}_N\Big]+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t\ddot{y}_{it}\hat{\Pi}_i^{ef}A\Big]$$

Substituting the expression of $\hat{\lambda}_i-\lambda_i$ into the first term on the right hand side of (C.4), we can rewrite (C.4) as

$$\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}\ddot{y}_{it}\Big](\hat{\rho}-\rho)+\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}x_{itp}x_{it}'\Big](\hat{\beta}-\beta)$$
$$-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}x_{itp}\hat{\lambda}_i'\Big)\hat{G}_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i\ddot{y}_{it}\Big)\Big](\hat{\rho}-\rho)$$
$$-\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}x_{itp}\hat{\lambda}_i'\Big)\hat{G}_N\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_ix_{it}'\Big)\Big](\hat{\beta}-\beta)$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}v_{itp}e_{it}+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_tx_{itp}\hat{\Pi}_i^{xf}(I-A)\Big] \quad (C.9)$$
$$-(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\hat{\Pi}^{f\ddot{y}\lambda\prime}(I-A)\Big]-\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\hat{\Pi}^{fx\lambda\prime}(I-A)\Big]$$
$$+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_tx_{itp}\hat{\Pi}_i^{\ddot{y}f}(I-A)\Big]+\mathcal{S}_{\beta,p}+\mathcal{T}_{\beta,p}$$

where

$$\mathcal{S}_{\beta,p} = -(\hat{\rho}-\rho)^2 \mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{\ddot{y}\ddot{y}\lambda}\hat{H}\Big] - \mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{xx\lambda}\hat{H}\Big]$$
$$+(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{e\ddot{y}\lambda}\hat{H}\Big] + (\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{\ddot{y}\epsilon\Phi}\hat{H}\Big]$$
$$-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{x\ddot{y}\lambda}\hat{H}\Big] - (\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{\ddot{y}x\lambda}\hat{H}\Big]$$
$$+(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i \hat{\Pi}_i^{\ddot{y}x\lambda}\hat{H}_N\Big] + (\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i \hat{\Pi}_i^{x\ddot{y}\lambda}\hat{H}_N\Big]$$
$$-(\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i \hat{\Pi}_i^{\ddot{y}\epsilon\Phi}\hat{H}_N\Big] - (\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\Pi}_i^{\ddot{y}\epsilon\Phi\prime}\hat{\lambda}_i'\hat{H}\Big]$$
$$-(\hat{\rho}-\rho)\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} (\hat{\lambda}_i-\lambda_i)'\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}\Big] + (\hat{\rho}-\rho)\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}A'\hat{\Pi}^{f\ddot{y}\lambda}\hat{H}_N\Big]$$
$$+\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i \hat{\Pi}_i^{xx\lambda}H_N\Big] + (\hat{\rho}-\rho)^2\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i \hat{\Pi}_i^{\ddot{y}\ddot{y}\lambda}\hat{H}_N\Big] \qquad \text{(C.10)}$$
$$-\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} (\hat{\lambda}_i-\lambda_i)'\hat{\Pi}^{fx\lambda}\hat{H}\Big] + \mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}A'\hat{\Pi}^{fx\lambda}H_N\Big]$$
$$-\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i \hat{\Pi}_i^{x\epsilon\Phi}\hat{H}_N\Big] - \mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\hat{\Phi}_i^{x\epsilon\Phi\prime}\hat{\lambda}_i'\hat{H}\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{ex\lambda}\hat{H}\Big] + \mathrm{tr}\Big[\frac{1}{T}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp} \hat{\Pi}_i^{x\epsilon\Phi}\hat{H}\Big]$$

and

$$\mathcal{T}_{\beta,p} = -\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp}(\hat{\lambda}_i-\lambda_i)'A\Big] + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} v_{itp}\hat{\lambda}_i'\hat{G}f_t$$
$$-\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{G}A'\Big] - \mathrm{tr}\Big[\frac{1}{NT}\hat{G}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}[\epsilon_{it}v_{jtp} - E(\epsilon_{it}v_{jtp})]\frac{1}{\hat{\sigma}_j^2}\hat{\lambda}_j'\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{N}\hat{G}\sum_{i=1}^N \hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})I_{k+1}^{p+1}\frac{1}{\hat{\sigma}_i^2}\hat{\lambda}_i'\Big] - \mathrm{tr}\Big[\frac{1}{N}\hat{G}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}(\hat{\gamma}_{ip}-\gamma_{ip})\hat{\lambda}_i'\Big]$$
$$-\frac{1}{NT^2}\sum_{i=1}^N\sum_{j=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2}\sum_{s=1}^T [e_{is}\epsilon_{js}' - E(e_{is}\epsilon_{js}')]\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}f_t x_{itp} + \mathrm{tr}[\hat{\Pi}^{fx_p\lambda}AA']$$
$$+\mathrm{tr}\Big[\frac{1}{NT^2}\sum_{i=1}^N\sum_{t=1}^T\sum_{s=1}^T \frac{1}{\hat{\sigma}_i^2} f_t f_s' v_{itp} e_{is}\Big] + \mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp}\hat{\Pi}_i^{ef}A\Big]$$
$$-\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^N \hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Phi}_i'\hat{H}\Big] - \mathrm{tr}\Big[A'\hat{\Pi}^{f\epsilon\Phi}\hat{H}_N\hat{\Pi}^{fx_p\lambda}\Big]$$
$$+\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\hat{\sigma}_i^2} f_t x_{itp}(\hat{\lambda}_i-\lambda_i)'\hat{\Pi}^{f\epsilon\Phi}H_N\Big] - \mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}_N\hat{\Pi}^{f\epsilon\Phi\prime}A\Big]$$

$$+\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\gamma_{ip}\hat{\lambda}_i'\hat{G}\hat{H}_N\hat{\Pi}^{f\epsilon\Phi\prime}\Big]+\mathrm{tr}\Big[\Big(\hat{\Pi}^{fx_p\lambda}-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\gamma_{ip}\hat{\lambda}_i'\Big)\hat{H}_N\hat{\Pi}^{f\epsilon\Phi\prime}\Big]$$
$$+\mathrm{tr}\Big[\hat{\Pi}^{fx_p\lambda}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\varepsilon_{ij,t}\hat{\Sigma}_{jj}^{-1}\hat{\Phi}_j'\hat{H}\Big]+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^4}x_{itp}f_t'\hat{H}\hat{\lambda}_i \qquad (C.11)$$

Consider (C.6). The five terms on the left hand side of (C.6) are given in Lemma C.3. The fourth-seventh terms on the right hand side are also given in Lemma C.3. The first two terms on the right hand side are given in Lemma C.4(a) and (b). The last two terms are given in Lemmas C.1 and C.2. Given these results, we have

$$\frac{1}{N}\Big[\mathrm{tr}(S_N^2)+\mathrm{tr}(\Sigma_{ee}^{-1}S_N\Psi S_N')-2\sum_{i=1}^{N}S_{ii,N}^2\Big](\hat{\rho}-\rho)+\frac{1}{N}\Big[\beta'(\sum_{i=1}^{N}\frac{1}{\sigma_i^2}S_{ii,N}\Sigma_{iiv})\Big](\hat{\beta}-\beta)$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}e_{it}\Big[\sum_{j=1,j\neq i}^{N}S_{ij,N}e_{jt}\Big]+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}e_{it}\Big[\sum_{j=1}^{N}S_{ij,N}v_{jt}'\Big]\beta$$
$$+O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2}). \qquad (C.12)$$

Further consider (C.9). Except for the fist term and the last two terms on the right hand side, the remaining terms are given in Lemma C.3. The first term is dealt with by Lemma C.4(c). The last two terms are given in Lemmas C.1 and C.2. Given these results, we have

$$\frac{1}{N}\Big[\beta'\sum_{i=1}^{N}\frac{1}{\sigma_i^2}S_{ii,N}\Sigma_{iiv}I_k^p\Big](\hat{\rho}-\rho)+\sum_{q=1}^{k}\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\Sigma_{iiv}^{(p,q)}\Big](\hat{\beta}_q-\beta_q)$$
$$=\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}e_{it}v_{itp}+O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2}). \qquad (C.13)$$

for all $p=1,2,\ldots,k$. Combining (C.12) and (C.13), we have, under $\sqrt{N}/T\to 0$,

$$\hat{\omega}-\omega=\Omega^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}\begin{bmatrix}\sum_{j=1}^{N}e_{it}\eta_{ij,t}S_{ij,N}\\ e_{it}v_{it}\end{bmatrix}+o_p(\frac{1}{\sqrt{NT}}).$$

where $\eta_{ij,t}=v_{jt}'\beta+1(i\neq j)e_{jt}$. This completes the proof of Theorem 4.2. $\square$

Appendix D: Detailed proofs for some lemmas

In this appendix, we provide the detailed proofs for some lemmas appearing in the earlier sections.

Proof of Lemma A.2. Notice

$$R=(I_N-\rho W_N)(I_N-\rho^*W_N)^{-1}=I_N-(\rho-\rho^*)W_N(I_N-\rho^*W_N)^{-1}=I_N-(\rho-\rho^*)S_N.$$

where S_N is defined in Assumption F. By Assumption E, we have

$$\|S_N\|_\infty\le\|W_N\|_\infty\cdot\|(I_N-\rho^*W_N)^{-1}\|_\infty\le C,$$

for some constant C. So R is uniformly bounded in row sums since ρ is in a compact set, i.e., $\sum_{j=1}^{N}|R_{ij}| \leq C$ for all i. Given this result, together with β in a compact set, (a) follows.

Consider (b). Notice

$$\begin{aligned}\sum_{j=1}^{N} R_{ij}(e_{jt}+\beta^{*\prime}v_{jt}) - \beta' v_{it} &= e_{it} - (\beta-\beta^*)' v_{it} - (\rho-\rho^*)\sum_{j=1}^{N} S_{ij,N}(e_{jt}+\beta^{*\prime}v_{jt}) \\ &= e_{it} - (\beta-\beta^*)' v_{it} - (\rho-\rho^*)\tilde{e}_{it}.\end{aligned}$$

where $\tilde{e}_{it} = \sum_{j=1}^{N} S_{ij,N}(e_{jt}+\beta^{*\prime}v_{jt})$. Now the left hand side of (b) is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}\Big\|\Big(\frac{1}{T}\sum_{t=1}^{T} e_{it}f_t\Big) - (\beta-\beta^*)'\Big(\frac{1}{T}\sum_{t=1}^{T} v_{it}f_t\Big) - (\rho-\rho^*)\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}f_t\Big)\Big\|^2$$

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

$$3\Big\{\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} e_{it}f_t\Big\|^2 + \|\beta-\beta^*\|^2\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} v_{it}f_t'\Big\|^2 + (\rho-\rho^*)^2\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}f_t\Big\|^2\Big\}$$

The first two terms are both $O_p(T^{-1})$. To show the third term is also $O_p(T^{-1})$, it suffices to prove

$$E\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} f_t\tilde{e}_{it}\Big\|^2\Big] = E\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} f_t' f_s\tilde{e}_{it}\tilde{e}_{is}\Big] = O(T^{-1}).$$

Since $\tilde{e}_{it}$ is independent with $\tilde{e}_{is}$ for $t \neq s$ by Assumption B, we have

$$E\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} f_t' f_s\tilde{e}_{it}\tilde{e}_{is}\Big] = E\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T} f_t' f_t\tilde{e}_{it}^2\Big] \leq CE\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{e}_{it}^2\Big].$$

By the definition of $\tilde{e}_{it}$,

$$E\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T^2}\sum_{t=1}^{T}\tilde{e}_{it}^2\Big] = \frac{1}{T}\Big[\frac{1}{N}\sum_{i=1}^{N}(\sigma_i^{*2}+\beta^{*\prime}\Sigma_{iiv}^*\beta^*)\Big(\sum_{j=1}^{N} S_{ij,N}^2\Big)\Big] = O(T^{-1}),$$

where the last equation is due to

$$\sum_{j=1}^{N} S_{ij,N}^2 \leq \Big(\max_{1\leq i,j\leq N}|S_{ij,N}|\Big)\sum_{j=1}^{N}|S_{ij,N}| \leq \big(\max_{1\leq i,j\leq N}|S_{ij,N}|\big)\|S_N\|_\infty \leq C. \tag{D.1}$$

Thus (b) follows.

Consider (c). It suffices to prove

$$E\Big[\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}^2 - E(\tilde{e}_{it}^2)]\Big|^2\Big] = O(T^{-1}).$$

The left hand side of the above is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\big(E(\tilde{e}_{it}^4) - [E(\tilde{e}_{it}^2)]^2\big) \leq \frac{1}{NT}\sum_{i=1}^{N} E(\tilde{e}_{it}^4)$$

By Assumption B, we have $E(e_{jt}+v_{jt}'\beta^*)^4 \le C'$ for some constant C'. Given this result, we have

$$E(\tilde{e}_{it}^4) = E\big(\sum_{j=1}^N S_{ij,N}(e_{jt}+v_{jt}'\beta^*)\big)^4 \le C'\Big(\sum_{j=1}^N |S_{ij,N}|\Big)^4 \triangleq C.$$

Then (c) follows.

Consider (d). Similarly as (c), it suffices to show

$$E\Big[\frac{1}{N}\sum_{i=1}^N \Big|\frac{1}{T}\sum_{t=1}^T [\tilde{e}_{it}e_{it} - E(\tilde{e}_{it}e_{it})]\Big|^2\Big] = O(T^{-1}).$$

The left hand side of the above expression is equal to

$$\frac{1}{NT}\sum_{i=1}^N \big(E(\tilde{e}_{it}^2 e_{it}^2) - [E(\tilde{e}_{it}e_{it})]^2\big) \le \frac{1}{NT}\sum_{i=1}^N E(\tilde{e}_{it}^2 e_{it}^2) \le \frac{1}{NT}\sum_{i=1}^N [E(\tilde{e}_{it}^4)]^{1/2}[E(e_{it}^4)]^{1/2}.$$

which is $O(T^{-1})$ by $E(\tilde{e}_{it}^4) \le C$ and $E(e_{it}^4) \le C$ for all i and t. Thus we have (d).

Consider (e). Notice, for each $p = 1, 2, \ldots, k$,

$$\frac{1}{N}\sum_{i=1}^N \Big|\frac{1}{T}\sum_{t=1}^T [\tilde{e}_{it}v_{itp} - E(\tilde{e}_{it}v_{itp})]\Big|^2 = O_p(T^{-1}),$$

which can be proved in the same way as (d). Then (e) follows since v_{it} is a finite dimensional vector.

Consider (f). Again, it suffices to prove

$$E\Big[\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \Big|\frac{1}{T}\sum_{t=1}^T [\tilde{e}_{it}\tilde{e}_{jt} - E(\tilde{e}_{it}\tilde{e}_{jt})]\Big|^2\Big] = O(T^{-1}). \tag{D.2}$$

The left hand side of the above expression is equal to

$$\frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N \big[E(\tilde{e}_{it}^2\tilde{e}_{jt}^2) - [E(\tilde{e}_{it}\tilde{e}_{jt})]^2\big] \le \frac{1}{N^2T}\sum_{i=1}^N\sum_{j=1}^N E(\tilde{e}_{it}^2\tilde{e}_{jt}^2) = O(T^{-1}).$$

where the last equation uses the fact that

$$E(\tilde{e}_{it}^2\tilde{e}_{jt}^2) \le [E(\tilde{e}_{it}^4)]^{1/2}[E(\tilde{e}_{jt}^4)]^{1/2}.$$

Then (f) follows. The proofs of (g) and (h) are similar to those of (d) and (e) and the details are hence omitted. This completes the proof. $\square$

Proof of Lemma A.3. Consider (a). By

$$\Sigma_{zz}^{-1} = \Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}\Phi G\Phi'\Sigma_{\epsilon\epsilon}^{-1} \tag{D.3}$$

we have that the left hand side of (a) is bounded by

$$\sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\Phi^*\Big(\frac{1}{T}\sum_{t=1}^T f_t\epsilon_t'\Big)\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\Big]\Big| + \sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\Phi^*\Big(\frac{1}{T}\sum_{t=1}^T f_t\epsilon_t'\Big)\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\Phi G\Phi'\Sigma_{\epsilon\epsilon}^{-1}\Big]\Big|.$$

We use ii_1 and ii_2 to denote the above expression, respectively. For ease of exposition, let $\ddot{\lambda}_i^*(\rho,\beta) = \sum_{j=1}^N R_{ij}(\gamma_j^*\beta^* + \lambda_j^*) - \gamma_i^{*\prime}\beta$ and $\varsigma_{it}(\rho,\beta) = \sum_{j=1}^N R_{ij}(e_{jt} + \beta^{*\prime} v_{jt}) - \beta' v_{it}$. We shall suppress ρ,β from the symbols for notational simplicity. Consider ii_1. By the expression of $\mathbb{D}$ in Lemma A.1, ii_1 is bounded by

$$ii_1 \le \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}\ddot{\lambda}_i^*\varsigma_{it}f_t'\Big]\Big| + \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \gamma_i^*\Sigma_{iiv}^{-1}v_{it}f_t'\Big]\Big| \triangleq ii_3 + ii_4.$$

Term ii_3 is $O_p(T^{-1/2})$ uniformly on Θ since

$$\Big|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}\ddot{\lambda}_i^*\varsigma_{it}f_t'\Big| \le C\Big(\frac{1}{N}\sum_{i=1}^N\|\ddot{\lambda}_i^*\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T\varsigma_{it}f_t\Big\|^2\Big)^{1/2} = O_p(T^{-1/2})$$

by Lemma A.2(a) and (b). Term ii_4 is also $O_p(T^{-1/2})$ uniformly on Θ since

$$\Big\|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \gamma_i^*\Sigma_{iiv}^{-1}v_{it}f_t'\Big\| \le C\Big(\frac{1}{N}\sum_{i=1}^N\|\gamma_i^*\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T f_t v_{it}'\Big\|^2\Big)^{1/2} = O_p(T^{-1/2}).$$

Thus $ii_1 = o_p(1)$ uniformly on Θ. Now consider ii_2, which is bounded by

$$\begin{aligned}
ii_2 \le\ & \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}f_t\varsigma_{it}\lambda_i'\Big)G\Big(\sum_{j=1}^N\frac{1}{\sigma_j^2}\lambda_j\ddot{\lambda}_j^{*\prime}\Big)\Big]\Big| \\
&+ \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}f_t\varsigma_{it}\lambda_i'\Big)G\Big(\sum_{j=1}^N\gamma_j\Sigma_{jjv}^{-1}\gamma_j^{*\prime}\Big)\Big]\Big| \\
&+ \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T f_t v_{it}'\Sigma_{iiv}^{-1}\gamma_i'\Big)G\Big(\sum_{j=1}^N\frac{1}{\sigma_j^2}\lambda_j\ddot{\lambda}_j^{*\prime}\Big)\Big]\Big| \\
&+ \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T f_t v_{it}'\Sigma_{iiv}^{-1}\gamma_i'\Big)G\Big(\sum_{j=1}^N\gamma_j\Sigma_{jjv}^{-1}\gamma_j^{*\prime}\Big)\Big]\Big| \\
=\ & ii_5 + ii_6 + ii_7 + ii_8, \qquad \text{say.}
\end{aligned}$$

Consider ii_5. Since

$$\Big(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}f_t\varsigma_{it}\lambda_i'\Big)G\Big(\sum_{j=1}^N\frac{1}{\sigma_j^2}\lambda_j\ddot{\lambda}_i^{*\prime}\Big)$$

is bounded in norm by

$$C\Big[\frac{1}{N}\sum_{i=1}^N\Big\|\frac{1}{T}\sum_{t=1}^T\varsigma_{it}f_t\Big\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^N\|\ddot{\lambda}_i^*\|^2\Big]^{1/2}\Big[\sum_{i=1}^N\frac{1}{\sigma_i^2}\|\lambda_i'H^{1/2}\|^2\Big]\cdot\|\mathscr{U}\|, \tag{D.4}$$

where $\mathscr{U} = (I + H^{-1})^{-1}$ and $H = \Phi'\Sigma_{\epsilon\epsilon}^{-1}\Phi$. Notice

$$\sum_{i=1}^N\|\Sigma_{ii}^{-1/2}\Phi_i'H^{1/2}\|^2 = \mathrm{tr}\Big[H^{1/2}\Big(\sum_{i=1}^N\Phi_i\Sigma_{ii}^{-1}\Phi_i'\Big)H^{1/2}\Big] = \mathrm{tr}[H^{1/2}H^{-1}H^{1/2}] = r \tag{D.5}$$

This implies $\sum_{i=1}^N\frac{1}{\sigma_i^2}\|\lambda_i'H^{1/2}\|^2 \le r$ since

$$\sum_{i=1}^N\frac{1}{\sigma_i^2}\|\lambda_i'H^{1/2}\|^2 + \sum_{i=1}^N\|\Sigma_{iiv}^{-1/2}\gamma_i'H^{1/2}\|^2 = \sum_{i=1}^N\|\Sigma_{ii}^{-1/2}\Phi_i'H^{1/2}\|^2 = r.$$

Given the above result and Lemma A.2(b), we have that the expression of (D.4) is $O_p(T^{-1/2})$, which implies $ii_5 = O_p(T^{-1/2})$ uniformly on Θ. Terms ii_6, ii_7 and ii_8 can be proved to be $O_p(T^{-1/2})$ uniformly on Θ in a similar way as the proof of ii_5. Thus $ii_2 = o_p(1)$ uniformly on Θ. This result, combined with the result of ii_1, leads to (a).

Consider (b). By (D.3), the left hand side of (b) is bounded by

$$\sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*)\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\Big]\Big| + \sup_{\theta\in\Theta}\Big|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*)\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\Phi G\Phi'\Sigma_{\epsilon\epsilon}^{-1}\Big]\Big|.$$

We use ii_9 and ii_{10} to denote the above two expressions. Consider ii_9. By Lemma A.1, ii_9 is bounded by

$$ii_9 = \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}v_{it}' - \Sigma_{iiv}^*)\Sigma_{iiv}^{-1}\Big]\Big| + \sup_{\theta\in\Theta}\Big|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}\ddot{\boldsymbol{u}}_{it}^2\Big| \triangleq ii_{11} + ii_{12}.$$

where

$$\ddot{\boldsymbol{u}}_{it}^2 = \varsigma_{it}^2 - \Big(\sum_{j=1}^{N}R_{ij}^2\sigma_j^{*2} + \sum_{j=1,j\neq i}^{N}R_{ij}^2\beta^{*\prime}\Sigma_{jjv}^*\beta^* + (R_{ii}\beta^* - \beta)'\Sigma_{iiv}^*(R_{ii}\beta^* - \beta)\Big).$$

By $R = I_N - (\rho - \rho^*)S_N$, we can rewrite $\ddot{\boldsymbol{u}}_{it}^2$ as

$$\begin{aligned}\ddot{\boldsymbol{u}}_{it}^2 = \varsigma_{it}^2 - \sigma_i^{*2} - \sum_{j=1}^{N}S_{ij,N}^2(\sigma_j^{*2} + \beta^{*\prime}\Sigma_{jjv}^*\beta^*) - (\beta - \beta^*)'\Sigma_{iiv}^*(\beta - \beta^*)\\ - 2(\rho - \rho^*)\beta^{*\prime}S_{ii,N}\Sigma_{iiv}^*(\beta - \beta^*) - 2(\rho - \rho^*)S_{ii,N}\sigma_i^{*2}.\end{aligned} \tag{D.6}$$

Consider ii_{11}. Notice

$$\Big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}v_{it}' - \Sigma_{iiv}^*)\Sigma_{iiv}^{-1}\Big\| \leq \Big(\frac{1}{N}\sum_{i=1}^{N}\|\Sigma_{iiv}^{-1}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}(v_{it}v_{it}' - \Sigma_{iiv}^*)\Big\|^2\Big)^{1/2}.$$

By the boundedness of Σ_{iiv}, the above expression is $O_p(T^{-1/2})$ uniformly on Θ, implying $ii_{11} = O_p(T^{-1/2})$ uniformly on Θ. Consider ii_{12}. By

$$\varsigma_{it} = \sum_{j=1}^{N}R_{ij}(e_{jt} + \beta^{*\prime}v_{jt}) - \beta'v_{it} = e_{it} - (\beta - \beta^*)'v_{it} - (\rho - \rho^*)\tilde{e}_{it}$$

where $\tilde{e}_{it} = \sum_{j=1}^{N}S_{ij,N}(e_{jt} + \beta^{*\prime}v_{jt})$. Substituting the above result into (D.6), we have

$$\begin{aligned}\ddot{\boldsymbol{u}}_{it}^2 = (e_{it}^2 - \sigma_i^{*2}) + (\beta - \beta^*)'(v_{it}v_{it}' - \Sigma_{iiv}^*)(\beta - \beta^*) + (\rho - \rho^*)^2[\tilde{e}_{it}^2 - E(\tilde{e}_{it}^2)]\\ -2(\beta - \beta^*)'v_{it}e_{it} - 2(\rho - \rho^*)[\tilde{e}_{it}e_{it} - E(\tilde{e}_{it}e_{it})] + 2(\rho - \rho^*)(\beta - \beta^*)'[v_{it}\tilde{e}_{it} - E(v_{it}\tilde{e}_{it})].\end{aligned}$$

By the boundedness of σ_i^2, ii_{12} is bounded by $C\sup_{\theta\in\Theta}\big|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\boldsymbol{u}}_{it}^2\big|$. Ignore C, the latter expression, by the preceding equation, is bounded by

$$\sup_{\theta\in\Theta}\Big|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^{*2})\Big| + \sup_{\theta\in\Theta}\Big|(\beta - \beta^*)'\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(v_{it}v_{it}' - \Sigma_{iiv}^*)\Big)(\beta - \beta^*)\Big|$$

$$+\sup_{\theta\in\Theta}\Big|(\rho-\rho^*)^2\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\tilde{e}_{it}^2-E(\tilde{e}_{it}^2)]\Big)\Big|+2\sup_{\theta\in\Theta}\Big|(\beta-\beta^*)'\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}v_{it}e_{it}\Big)\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|(\rho-\rho^*)\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\tilde{e}_{it}e_{it}-E(\tilde{e}_{it}e_{it})]\Big)\Big| \tag{D.7}$$

$$+2\sup_{\theta\in\Theta}\Big|(\beta-\beta^*)\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[v_{it}\tilde{e}_{it}-E(v_{it}\tilde{e}_{it})]\Big)(\rho-\rho^*)\Big|.$$

Consider the last term on the right hand side of (D.7), which is bounded by

$$2\|\beta-\beta^*\|\cdot|\rho-\rho^*|\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[v_{it}\tilde{e}_{it}-E(v_{it}\tilde{e}_{it})]\Big\|^2\Big)^{1/2}.$$

By the boundedness of β,ρ,β^* and ρ^*, this expression is $O_p(T^{-1/2})$ uniformly on Θ by Lemma A.2(e). The remaining 5 terms on the right hand side of (D.7) can be proved to be $O_p(T^{-1/2})$ uniformly on Θ in a similar way. So $ii_{12}=O_p(T^{-1/2})$ uniformly on Θ. Given the results of ii_{11} and ii_{12}, we have $ii_9=O_p(T^{-1/2})$ uniformly on Θ.

We then proceed to consider ii_{10}. Using the same method in analyzing ii_9, we have

$$ii_{10}\le\sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}H^{1/2}\lambda_i\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\lambda_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+\sup_{\theta\in\Theta}\Big|\sum_{p=1}^{k}\sum_{q=1}^{k}(\beta_p-\beta_p^*)(\beta_q-\beta_q^*)\mathrm{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}H^{1/2}\lambda_i\mathscr{V}_{ij,pq}\lambda_j'H^{1/2}\mathscr{U}\Big)\Big]\Big|$$

$$+\sup_{\theta\in\Theta}\Big|(\rho-\rho^*)^2\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}H^{1/2}\lambda_i\frac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{it}\tilde{e}_{jt}-E(\tilde{e}_{it}\tilde{e}_{jt})]\lambda_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|\sum_{p=1}^{k}(\beta_p-\beta_p^*)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}H^{1/2}\lambda_i\mathscr{G}_{ij,p}\lambda_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|(\rho-\rho^*)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}H^{1/2}\lambda_i\mathscr{J}_{ij}\lambda_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|(\rho-\rho^*)\Big\{\sum_{p=1}^{k}(\beta_p-\beta_p^*)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}H^{1/2}\lambda_i\mathscr{L}_{ij,p}\lambda_j'H^{1/2}\mathscr{U}\Big]\Big\}\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2}H^{1/2}\lambda_i\frac{1}{T}\sum_{t=1}^{T}e_{it}v_{jt}'\Sigma_{jjv}^{-1}\gamma_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|\sum_{p=1}^{k}(\beta_p-\beta_p^*)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2}H^{1/2}\lambda_i\mathscr{V}_{ij,p}\Sigma_{jjv}^{-1}\gamma_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+2\sup_{\theta\in\Theta}\Big|(\rho-\rho^*)\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2}H^{1/2}\lambda_i\mathscr{L}_{ij}\Sigma_{jjv}^{-1}\gamma_j'H^{1/2}\mathscr{U}\Big]\Big|$$

$$+\sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}H^{1/2}\gamma_i\Sigma_{iiv}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{it}v_{jt}'-E(v_{it}v_{jt}')]\Sigma_{jjv}^{-1}\gamma_j'H^{1/2}\mathscr{U}\Big]\Big|.$$

where $\mathscr{V}_{ij,pq} = \frac{1}{T}\sum_{t=1}^T [v_{itp}v_{jtq} - E(v_{itp}v_{jtq})]$; $\mathscr{V}_{ij,p} = \frac{1}{T}\sum_{t=1}^T [v_{itp}v_{jt} - E(v_{itp}v_{jt})]$; $\mathscr{G}_{ij,p} = \frac{1}{T}\sum_{t=1}^T [e_{it}v_{jtp} - E(e_{it}v_{jtp})]$; $\mathscr{J}_{ij} = \frac{1}{T}\sum_{t=1}^T [\tilde{e}_{it}e_{jt} - E(\tilde{e}_{it}e_{jt})]$; $\mathscr{L}_{ij,p} = \frac{1}{T}\sum_{t=1}^T [\tilde{e}_{it}v_{jtp} - E(\tilde{e}_{it}v_{jtp})]$; $\mathscr{L}_{ij} = \frac{1}{T}\sum_{t=1}^T [\tilde{e}_{it}v_{jt} - E(\tilde{e}_{it}v_{jt})]$ and $\mathscr{U} = (I + H^{-1})^{-1}$. Using the results in Lemma A.2, we can show that the ten terms on the right hand side are all $O_p(T^{-1/2})$. Since the proofs are similar, we only choose the second term as an illustration. The expression in $|\cdot|$ is bounded by

$$C\Big(\sum_{i=1}^N \frac{1}{\sigma_i^2}\|\lambda_i' H^{1/2}\|^2\Big)\Big\{\sum_{p=1}^k\sum_{q=1}^k |\beta_p - \beta_p^*|\cdot|\beta_q - \beta_q^*| \\ \times\Big[\frac{1}{N^2}\sum_{i=1}^N\sum_{t=1}^T\Big|\frac{1}{T}\sum_{t=1}^T [v_{itp}v_{jtq} - E(v_{itp}v_{jtq})]\Big|^2\Big]^{1/2}\Big\}\|\mathscr{U}\|,$$

which is $O_p(T^{-1/2})$ uniformly on Θ given the boundedness of β, β^* and $\sum_{i=1}^N \frac{1}{\sigma_i^2}\|\lambda_i' H^{1/2}\|^2 \le r$. So we have $ii_{10} = o_p(1)$. Given the results on ii_9 and ii_{10}, we have (b).

Consider (c). By

$$0 < \frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\bar{\epsilon}\bar{\epsilon}'\mathbb{D}'\Sigma_{zz}^{-1}\Big] < \frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\bar{\epsilon}\bar{\epsilon}'\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\Big],$$

it suffices to prove that $\sup\limits_{\theta\in\Theta}\frac{1}{N}\mathrm{tr}[\mathbb{D}\bar{\epsilon}\bar{\epsilon}'\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}] = o_p(1)$. By the boundedness of Σ_{ii}, $\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\mathbb{D}$ is bounded by $C\mathbb{D}'\mathbb{D}$ for some C. Since $\|S_N\|_1$ and $\|S_N\|_\infty$ are both uniformly bounded by Assumption E, we have that $\|R\|_\infty$ and $\|R\|_1$ are both uniformly bounded for $R = I_N - (\rho - \rho^*)S_N$ and for ρ in a compact set. This result means that $\mathbb{D}$ is uniformly bounded in both row and column sums by the definition of $\mathbb{D}$, which in turn implies that $\|\mathbb{D}\|_1$ and $\|\mathbb{D}\|_\infty$ are both bounded. By $\|\mathbb{D}\|_2 \le \sqrt{\|\mathbb{D}\|_1\cdot\|\mathbb{D}\|_\infty}$, we have $\tau_{max}(\mathbb{D}\mathbb{D}') = \tau_{max}(\mathbb{D}'\mathbb{D})$ is uniformly bounded, where $\tau(\cdot)$ denotes the largest eigenvalue of the argument. Then $\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\mathbb{D} \le CI_{N(k+1)}$ for some constant C. Given this result, we have

$$\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\bar{\epsilon}\bar{\epsilon}'\mathbb{D}'\Sigma_{\epsilon\epsilon}^{-1}\Big] \le C\frac{1}{N}\mathrm{tr}[\bar{\epsilon}\bar{\epsilon}'] = o_p(1).$$

So we have (c). This completes the proof of Lemma A.3. □

Proof of Lemma B.1. The proofs of the seven results in this lemma are similar and we only choose (a) to illustrate. By the Cauchy-Schwarz inequality, the left hand side of (a) is bounded by

$$C\cdot\|\hat{\beta} - \beta\|^2\cdot\|\hat{H}_N^{1/2}\|^2\cdot\Big(\sum_{i=1}^N\|\hat{H}^{1/2}\hat{\Phi}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\frac{1}{T}\sum_{t=1}^T\|\boldsymbol{b}_{it}\|\|\boldsymbol{b}_{jt}\|\Big)^{1/2}.$$

which is $O_p(\|\hat{\omega} - \omega\|^2)$ by (D.5), $\hat{H}_N \xrightarrow{p} Q$ and $\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\frac{1}{T}\sum_{t=1}^T\|\boldsymbol{b}_{it}\|\|\boldsymbol{b}_{jt}\| = O_p(1)$. □

Proof of Lemma B.6. The proof of (a) is similar to that of Lemma A.12(b) in the supplement of Bai and Li (2014). Result (b) can be proved in a similar way as Lemma

C.1(e) in the supplement of Bai and Li (2012). Consider (c). By the boundedness of $\hat{\Sigma}_{ii}$, the left hand side is bounded by

$$C \cdot \|\hat{H}\|^{1/2} \cdot \Big(\sum_{i=1}^{N} \|\hat{H}^{1/2} \hat{\Phi} \hat{\Sigma}_{ii}^{-1/2}\| \cdot \|(\hat{\Sigma}_{ii} - \Sigma_{ii}) \hat{\Sigma}_{ii}^{-1} \hat{\Phi}_i'\| \Big) \cdot \|\hat{H}\|,$$

which, by the Cauchy-Schwarz inequality, is further bounded by

$$C \cdot \|\hat{H}\|^{1/2} \cdot \Big(\sum_{i=1}^{N} \|\hat{H}^{1/2} \hat{\Phi} \hat{\Sigma}_{ii}^{-1/2}\|^2 \Big)^{1/2} \Big(\sum_{i=1}^{N} \|(\hat{\Sigma}_{ii} - \Sigma_{ii}) \hat{\Sigma}_{ii}^{-1} \hat{\Phi}_i'\|^2 \Big)^{1/2} \cdot \|\hat{H}\|. \tag{D.8}$$

Notice

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^{N} \|(\hat{\Sigma}_{ii} - \Sigma_{ii}) \hat{\Sigma}_{ii}^{-1} \hat{\Phi}_i'\|^2 &\le C \frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \cdot \|\hat{\Sigma}_{ii}^{-1/2} \hat{\Phi}_i'\|^2 \\ &\le 2C \frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \Big(\|\hat{\Sigma}_{ii}^{-1/2} (\hat{\Phi}_i - \Phi_i)\|^2 + \|\hat{\Sigma}_{ii}^{-1/2} \Phi_i\|^2 \Big) \\ = 2C \frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \cdot \|\hat{\Sigma}_{ii}^{-1/2} (\hat{\Phi}_i - \Phi_i)\|^2 &+ \frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \cdot \|\hat{\Sigma}_{ii}^{-1/2} \Phi_i\|^2. \end{aligned}$$

By the boundedness of $\hat{\Sigma}_{ii}$ and Σ_{ii}, the first term is bounded by $C \frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii}^{-1/2} (\hat{\Phi}_i - \Phi_i)\|^2 = O_p(T^{-1}) + O_p(\|\hat{\omega} - \omega\|^2)$ by Proposition B.2. Furthermore, by the boundedness of $\hat{\Sigma}_{ii}$ and Φ_i, the second term is bounded by $C \frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1}) + O_p(\|\hat{\omega} - \omega\|^2)$ by Proposition B.2. So we have

$$\frac{1}{N} \sum_{i=1}^{N} \|(\hat{\Sigma}_{ii} - \Sigma_{ii}) \hat{\Sigma}_{ii}^{-1} \hat{\Phi}_i'\|^2 = O_p(T^{-1}) + O_p(\|\hat{\omega} - \omega\|^2).$$

Given the above result, together with (D.8) and (D.5) as well as $\hat{H} = O_p(N^{-1})$, we have (c). $\square$

Proof of Lemma B.7. Result (f) can be proved similarly (and more easily) as Lemma B.6(a). Results (b), (c), (d) and (e) can be shown easily and the details are omitted. We only prove (a). By the boundedness of λ, σ_i^2, the left hand side of (a) is bounded in norm by $C \frac{1}{N} \sum_{i=1}^{N} \|\mathcal{S}_{i,\sigma^2}\|$ for some constant C. By (B.17), $\mathcal{S}_{i,\sigma^2}$ consists of 21 terms, which we use $ii_{i,1}, \dots, ii_{i,21}$ to denote. Then $\frac{1}{N} \sum_{i=1}^{N} \|\mathcal{S}_{i,\sigma^2}\| \le \sum_{p=1}^{21} \frac{1}{N} \sum_{i=1}^{N} \|ii_{i,p}\|$. Checking these 21 terms one by one, we obtain (a). $\square$

Proof of Lemma C.1. By (C.7), the expression of $\mathcal{S}_\rho$ consists of 20 terms, which we use $ii_1, \dots, ii_{20}$ to denote temporarily. These 20 terms can be classified into three groups. The first group, including $ii_1, ii_2, ii_5, ii_6, ii_7, ii_8, ii_{13}$ and ii_{14}, consists of the terms involving either $(\hat{\rho} - \rho)^2$, or $\|\hat{\beta} - \beta\|^2$, or $|\hat{\rho} - \rho| \cdot \|\hat{\beta} - \beta\|$. The proof of the first group is similar to that of Lemma B.1. The second group, including $ii_{15}, ii_{16}, ii_{17}, ii_{18}, ii_{19}$ and ii_{20}, only involves $\hat{\beta} - \beta$. The proof of the second group is similar to that of Lemma A.10(b) in the supplement of Bai and Li (2014). The last group, including $ii_3, ii_4, ii_9, ii_{10}, ii_{11}$ and ii_{12}, are the terms only involving $\hat{\rho}-\rho$. The proofs of these six terms are similar and we choose ii_3 to illustrate. By (C.2),

$$\ddot{Y}_t = S_N \Xi f_t + S_N \zeta_t = \ddot{\Xi} f_t + \ddot{\zeta}_t \tag{D.9}$$

with $\ddot{\Xi} = S_N \Xi$ and $\ddot{\zeta}_t = S_N \zeta_t$. Let $\ddot{\xi}_i$ be the transpose of the ith row of $\ddot{\Xi}$ and let $\ddot{\zeta}_{it}$ be the ith element of $\ddot{\zeta}_t$. Then it follows that

$$\ddot{y}_{it} = \ddot{\xi}_i' f_t + \ddot{\zeta}_{it}. \tag{D.10}$$

Now consider ii_3. By (D.10), the expression in the trace operator is equal to

$$\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2} f_t \ddot{\zeta}_{it}\hat{\Pi}_i^{e\ddot{y}\lambda}\hat{H} + \sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\ddot{\xi}_i\hat{\Pi}_i^{e\ddot{y}\lambda}\hat{H} = ii_{3,1} + ii_{3,2}, \quad \text{say.}$$

By the definition of $\hat{\Pi}_i^{e\ddot{y}\lambda}$, together with boundedness of $\hat{\sigma}_i^2$, $ii_{3,1}$ is bounded in norm by

$$C\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{s=1}^{T} e_{is}\ddot{y}_{js}\Big)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\|\hat{\lambda}_j\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} f_t\ddot{\zeta}_{it}\Big\|^2\Big]^{1/2}\|\hat{H}_N\|$$

which is $O_p(T^{-1/2})$. Proceed to consider $ii_{3,2}$, which is equivalent to

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{1}{\hat{\sigma}_i^2}\ddot{\xi}_i e_{is} f_s'\Big)\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\ddot{\xi}_j\hat{\lambda}_j'\Big)\hat{H}_N + \Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\sigma_i^2 S_{ji,N}}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\ddot{\xi}_i\hat{\lambda}_j'\Big)\hat{H}_N$$
$$+\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\ddot{\xi}_i\hat{\lambda}_j'\frac{1}{T}\sum_{s=1}^{T}[e_{is}\ddot{\zeta}_{js} - E(e_{is}\ddot{\zeta}_{js})]\Big)\hat{H}_N \tag{D.11}$$

The first expression of (D.11) is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\ddot{\xi}_i\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{s=1}^{T} f_s e_{is}\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\|\hat{\lambda}_j\|^2\Big)^{1/2}\|\hat{H}_N\|,$$

which is $O_p(T^{-1/2})$. The second expression is equal to

$$\frac{1}{N^2}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\Big(\sum_{i=1}^{N} S_{ji,N}\ddot{\xi}_i\Big)\hat{\lambda}_j'\hat{H}_N - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\hat{\sigma}_i^2-\sigma_i^2)S_{ji,N}}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\ddot{\xi}_i\hat{\lambda}_j'\hat{H}_N,$$

which is bounded in norm by

$$C\frac{1}{N}\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\sum_{i=1}^{N} S_{ji,N}\ddot{\xi}_i\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\|\hat{\lambda}_j\|^2\Big)^{1/2}\|\hat{H}_N\|$$
$$+C\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]\Big[\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\|\hat{\lambda}_j\|^2\Big]\|\hat{H}_N\| = O_p(N^{-1}) + O_p(T^{-1/2}).$$

The last expression of (D.11) is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\ddot{\xi}_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\hat{\sigma}_j^2}\|\hat{\lambda}_j\|^2\Big)^{1/2}\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{s=1}^{T}[e_{is}\ddot{\zeta}_{js}-E(e_{is}\ddot{\zeta}_{js})]\Big|^2\Big)^{1/2}\|\hat{H}_N\|,$$

which is also $O_p(T^{-1/2})$. Given this result, we have $ii_3 = o_p(\|\hat{\omega}-\omega\|)$. This proves the first part of the lemma.

The proof for the second part is similar, and is omitted. □

Proof of Lemma C.2. By (C.8), the expression of $\mathcal{T}_\rho$ consists of 18 terms, which we use $ii_1, ii_2, \dots, ii_{18}$ to denote temporarily. By Lemma B.6(b), we have $\hat{\Pi}^{f\epsilon\Phi} = O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) + o_p(\|\hat{\omega}-\omega\|)$. Given this result, together with Proposition B.3, it is relatively easy to show that $ii_2, ii_3, ii_5, ii_6, ii_{14}, ii_{15}, ii_{16}, ii_{17}$ and ii_{18} are all $O_p(N^{-1/2}T^{-1}) + O_p(N^{-1}T^{-1/2}) + O_p(T^{-3/2}) + o_p(\|\hat{\omega}-\omega\|)$. In addition, ii_{12} is shown to be $O_p(N^{-1}T^{-1/2}) + o_p(\|\hat{\omega}-\omega\|)$ in Lemma B.6(c) and ii_1 can be proved similarly as ii_{12}. Also, ii_{11} is shown to be $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\omega}-\omega\|)$ in Lemma B.6(a) and ii_7 and ii_9 can be proved similarly as ii_{11}; ii_4 is $O_p(N^{-1}T^{-1/2})$ by $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{\Phi}_i - \Phi_i\|^2 = O_p(T^{-1}) + o_p(\|\hat{\omega}-\omega\|)$; ii_{10} is $O_p(N^{-1}T^{-1/2}) + o_p(\|\hat{\omega}-\omega\|)$ by $\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2 = O_p(T^{-1}) + O_p(\|\hat{\omega}-\omega\|^2)$; ii_8 is $O_p(N^{-1}T^{-1/2})$ by

$$\Big\|\frac{1}{NT}\sum_{t=1}^{T}f_t\zeta_t' S_N'\hat{\Sigma}_{ee}^{-1}\hat{\Lambda}\hat{G}\Big\| \le C\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\|\hat{\lambda}_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t\ddot{\zeta}_{it}\Big\|^2\Big)^{1/2}\|\hat{G}\|.$$

Finally, it is relatively easy to see

$$ii_{13} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big] + O_p(T^{-3/2}) + o_p(\|\hat{\omega}-\omega\|).$$

Given the above analysis, we have the first part of Lemma C.2. The proof of the second part is similar and in fact is easier, thus omitted. This completes the proof of Lemma C.2. □

Proof of Lemma C.3. By (D.10), the proof of Lemma C.3 is similar to that of Lemma A.9 in the supplement of Bai and Li (2014) and the details are omitted. □

Proof of Lemma C.4. The proof of result (c) is similar to that of Lemma A.12(a) in the supplement of Bai and Li (2014) and hence are omitted. We only choose result (b) to prove since the proof of result (a) is similar and actually easier. Consider (b). The left hand side of (b) is equal to

$$\frac{1}{NT}\sum_{t=1}^{T}\mathrm{tr}[\Sigma_{ee}S_N'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})] + \frac{1}{NT}\sum_{t=1}^{T}\mathrm{tr}[(e_te_t'-\Sigma_{ee})S_N'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})]$$
$$= -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}S_{ii,N} - \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^2}\frac{1}{T}\sum_{t=1}^{T}[e_{it}\ddot{e}_{it}-E(e_{it}\ddot{e}_{it})] = ii_1 + ii_2, \quad \text{say}.$$

where $\ddot{e}_t = S_N e_t$ and $\ddot{e}_{it}$ is its ith element. Further consider ii_1. By (B.16), ii_1 is equal to

$$ii_1 = -\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{S_{ii,N}}{\hat{\sigma}_i^2}(e_{it}^2 - \sigma_i^2) + \frac{1}{N}\sum_{i=1}^N \frac{S_{ii,N}}{\hat{\sigma}_i^2}(\hat{\lambda}_i - \lambda_i)'(\hat{\lambda}_i - \lambda_i)$$

$$+2(\hat{\rho} - \rho)\frac{1}{N}\sum_{i=1}^N \frac{S_{ii,N}}{\hat{\sigma}_i^2}\Big(\frac{1}{T}\sum_{t=1}^T \ddot{y}_{it}e_{it}\Big) - \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}S_{ii,N}\mathcal{S}_{i,\sigma^2} - \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}S_{ii,N}\mathcal{T}_{i,\sigma^2}$$

$$+2\mathrm{tr}\Big[\hat{H}_N\Big(\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N \hat{\Phi}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^T[\epsilon_{it}e_{jt} - E(\epsilon_{it}e_{jt})]S_{jj,N}\frac{1}{\hat{\sigma}_j^2}\lambda_j'\Big)\Big]$$

$$= iii_1 + iii_2 + \cdots + iii_6, \quad \text{say.}$$

First consider iii_2. By (B.12), we have

$$\hat{\lambda}_i - \lambda_i = \frac{1}{T}\sum_{t=1}^T f_t e_{it} + \mathrm{Rem}_i$$

where Rem_i denotes the remainder terms. So we have

$$iii_2 = \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}\mathrm{Rem}_i'\mathrm{Rem}_i + 2\frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big]'\mathrm{Rem}_i$$

$$+ \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big]'\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big]$$

The first term is bounded in norm by $C\frac{1}{N}\sum_{i=1}^N \|\mathrm{Rem}_i\|^2$. The second term, by the boundedness of $\hat{\sigma}_i^2$ and $S_{ii,N}$, is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \|\mathrm{Rem}_i\|^2\Big)^{1/2}.$$

Some calculation shows that $\frac{1}{N}\sum_{i=1}^N \|\mathrm{Rem}_i\|^2 = O_p(N^{-1}T^{-1}) + O_p(T^{-2}) + O_p(\|\hat{\omega} - \omega\|^2)$. Given this result, we have

$$iii_2 = \frac{1}{N}\sum_{i=1}^N \frac{1}{\hat{\sigma}_i^2}S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big]'\Big[\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big] + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}) + o_p(\|\hat{\omega} - \omega\|).$$

Consider iii_3. By (D.10), we have

$$\frac{1}{N}\sum_{i=1}^N \frac{S_{ii,N}}{\hat{\sigma}_i^2}\Big(\frac{1}{T}\sum_{t=1}^T \ddot{y}_{it}e_{it}\Big) = \frac{1}{N}\sum_{i=1}^N \frac{S_{ii,N}}{\hat{\sigma}_i^2}\ddot{\xi}_i\frac{1}{T}\sum_{t=1}^T f_t e_{it} + \frac{2}{N}\sum_{i=1}^N \frac{\sigma_i^2}{\hat{\sigma}_i^2}S_{ii,N}^2$$

$$+ \frac{1}{N}\sum_{i=1}^N \frac{S_{ii,N}}{\hat{\sigma}_i^2}\frac{1}{T}\sum_{t=1}^T[\ddot{\zeta}_{it}e_{it} - E(\ddot{\zeta}_{it}e_{it})].$$

The first term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^N \|S_{ii,N}\ddot{\xi}_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N \Big\|\frac{1}{T}\sum_{t=1}^T f_t e_{it}\Big\|^2\Big)^{1/2} = O_p(T^{-1/2}).$$

The third term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}S_{ii,N}^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[\ddot{\zeta}_{it}e_{it}-E(\ddot{\zeta}_{it}e_{it})]\Big|^2\Big)^{1/2}=O_p(T^{-1/2}).$$

The second term is equal to

$$\frac{2}{N}\sum_{i=1}^{N}S_{ii,N}^2-\frac{2}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}S_{ii,N}^2.$$

Since

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}S_{ii,N}^2\leq C\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big)^{1/2}=o_p(1)$$

we have that the second term is $\frac{2}{N}\sum_{i=1}^{N}S_{ii,N}^2+o_p(1)$. Given these results, we have $iii_3=(\hat{\rho}-\rho)\frac{2}{N}\sum_{i=1}^{N}S_{ii,N}^2+o_p(\|\hat{\omega}-\omega\|)$. Substituting (B.17) into iii_4 and checking the terms one by one, we have $iii_4=o_p(\|\hat{\omega}-\omega\|)$. Substituting (B.18) into iii_5 and using the results of Proposition B.3 and Lemma B.6, we can show that $iii_5=O_p(N^{-1/2}T^{-1})+O_p(N^{-1}T^{-1/2})+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|)$. Next, iii_6 can be shown to be $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|)$ similarly as Lemma B.6(a). Summarizing the above results, we have

$$\begin{aligned}ii_1=iii_1&+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}S_{ii,N}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]+(\hat{\rho}-\rho)\frac{2}{N}\sum_{i=1}^{N}S_{ii,N}^2\\&+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|).\end{aligned}$$

We then consider iii_1+ii_2, which is equal to

$$-\frac{1}{N}\sum_{i=1}^{N}\frac{S_{ii,N}}{\sigma_i^2}\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)-\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^2}\frac{1}{T}\sum_{t=1}^{T}e_{it}\Big(\sum_{j=1,j\neq i}^{N}S_{ij,N}e_{jt}\Big)\triangleq iii_8+iii_7,\ \text{say}$$

Consider iii_7. By (B.16), we have

$$\hat{\sigma}_i^2-\sigma_i^2=\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)+\mathcal{R}\mathrm{em}_i.$$

where $\mathcal{R}\mathrm{em}_i$ denotes the remainder terms. Substituting the above result into iii_7, we have

$$\begin{aligned}iii_7=&\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2\sigma_i^2}\Big[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}e_{it}\Big(\sum_{j=1,j\neq i}^{N}S_{ij,N}e_{jt}\Big)(e_{is}^2-\sigma_i^2)\Big]\\&+\frac{1}{N}\sum_{i=1}^{N}\frac{\mathcal{R}\mathrm{em}_i}{\hat{\sigma}_i^2\sigma_i^2}\frac{1}{T}\sum_{t=1}^{T}e_{it}\Big(\sum_{j=1,j\neq i}^{N}S_{ij,N}e_{jt}\Big)=iii_a+iii_b,\quad\text{say.}\end{aligned}$$

The second term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\mathcal{R}\mathrm{em}_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}e_{it}\Big(\sum_{j=1,j\neq i}^{N}S_{ij,N}e_{jt}\Big)\Big|^2\Big)^{1/2},$$

which is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|)$ by $\frac{1}{N}\sum_{i=1}^N\|\mathcal{R}\mathrm{em}_i\|^2=O_p(N^{-1}T^{-1})+O_p(T^{-2})+O_p(\|\hat{\omega}-\omega\|^2)$. Term iii_a is equal to

$$\frac{1}{N}\sum_{i=1}^N\frac{1}{\sigma_i^4}\Big[\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T e_{it}\Big(\sum_{j=1,j\neq i}^N S_{ij,N}e_{jt}\Big)(e_{is}^2-\sigma_i^2)\Big]$$
$$-\frac{1}{N}\sum_{i=1}^N\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^4}\Big[\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T e_{it}\Big(\sum_{j=1,j\neq i}^N S_{ij,N}e_{jt}\Big)(e_{is}^2-\sigma_i^2)\Big].$$

The first term of the above expression is $O_p(N^{-1/2}T^{-1})$ and the second term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^N\Big|\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T e_{it}\Big(\sum_{j=1,j\neq i}^N S_{ij,N}e_{jt}\Big)(e_{is}^2-\sigma_i^2)\Big|^2\Big)^{1/2},$$

which is $O_p(T^{-3/2})+o_p(\|\hat{\omega}-\omega\|)$. Summarizing all the results, we have (b). $\square$

Appendix E: Local optimality of the iterated solution

In this appendix, we show that the iterated solution from the suggested estimation procedure in Section 5 possesses the local optimality property. That is, the solution satisfies the first order conditions. Let $\theta^\infty=(\rho^\infty,\beta^\infty,\Phi^\infty,\Sigma_{\epsilon\epsilon}^\infty)$ be the converged solution, i.e. $\lim_{s\to\infty}\rho^{(s)}=\rho^\infty$, $\lim_{s\to\infty}\beta^{(s)}=\beta^\infty$, $\lim_{s\to\infty}\Phi^{(s)}=\Phi^\infty$ and $\lim_{s\to\infty}\Sigma_{\epsilon\epsilon}^{(s)}=\Sigma_{\epsilon\epsilon}^\infty$. By (5.1),

$$\Phi^{(s+1)}\Big[\frac{1}{T}\sum_{t=1}^T E(f_tf_t'|\theta^{(s)})\Big]=\frac{1}{T}\sum_{t=1}^T E(Dz_tf_t'|\theta^{(s)}).$$

By the expressions of $\frac{1}{T}\sum_{t=1}^T E(f_tf_t'|\theta^{(s)})$ and $\frac{1}{T}\sum_{t=1}^T E(Dz_tf_t'|\theta^{(s)})$, then

$$\begin{aligned}&D^{(s)}M_{zz}D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}\Phi^{(s)}\\&=\Phi^{(s+1)}\Big[I_r-\Phi^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}\Phi^{(s)}+\Phi^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}D^{(s)}M_{zz}D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1}\Phi^{(s)}\Big].\end{aligned}$$

Letting $s\to\infty$, we have

$$\begin{aligned}\Phi^\infty-\Phi^\infty\Phi^{\infty\prime}(\Sigma_{zz}^\infty)^{-1}\Phi^\infty-D^\infty M_{zz}D^{\infty\prime}(\Sigma_{zz}^\infty)^{-1}\Phi^\infty&\\+\Phi^\infty\Phi^{\infty\prime}(\Sigma_{zz}^\infty)^{-1}D^\infty M_{zz}D^{\infty\prime}(\Sigma_{zz}^\infty)^{-1}\Phi^\infty&=0.\end{aligned}$$

By $\Sigma_{zz}^\infty=\Phi^\infty\Phi^{\infty\prime}+\Sigma_{\epsilon\epsilon}^\infty$, the above equation can be written as

$$\Sigma_{\epsilon\epsilon}^\infty(\Sigma_{zz}^\infty)^{-1}\Phi^\infty-\Sigma_{\epsilon\epsilon}^\infty(\Sigma_{zz}^\infty)^{-1}D^\infty M_{zz}D^{\infty\prime}(\Sigma_{zz}^\infty)^{-1}\Phi^\infty=0.$$

pre-multiplying $\Sigma_{zz}^\infty(\Sigma_{\epsilon\epsilon}^\infty)^{-1}$, we have

$$\Phi^{\infty\prime}(\Sigma_{zz}^\infty)^{-1}(D^\infty M_{zz}D^{\infty\prime}-\Sigma_{zz}^\infty)=0. \tag{E.1}$$

By $\Phi^{\infty\prime}(\Sigma_{zz}^{\infty})^{-1} = G^{\infty}\Phi^{\infty\prime}(\Sigma_{\epsilon\epsilon}^{\infty})^{-1}$ and G^{∞} is invertible,

$$\Phi^{\infty\prime}(\Sigma_{\epsilon\epsilon}^{\infty})^{-1}(D^{\infty}M_{zz}D^{\infty\prime} - \Sigma_{zz}^{\infty}) = 0,$$

which is equivalent to (3.2).

We next consider (5.2). Letting $s \to \infty$, (5.2) is equal to

$$\Sigma_{\epsilon\epsilon}^{\infty} = \mathrm{Dg}\Big[D^{\infty}M_{zz}D^{\infty\prime} - \Phi^{\infty}\Phi^{\infty\prime}(\Sigma_{zz}^{\infty})^{-1}D^{\infty}M_{zz}D^{\infty\prime}\Big].$$

By (E.1), we can rewrite the above result as

$$\Sigma_{\epsilon\epsilon}^{\infty} = \mathrm{Dg}\Big[D^{\infty}M_{zz}D^{\infty\prime} - \Phi^{\infty}\Phi^{\infty\prime}\Big],$$

which is equivalent to

$$0 = \mathrm{Dg}\Big[D^{\infty}M_{zz}D^{\infty\prime} - \Sigma_{zz}^{\infty}\Big],$$

the same as (3.3).

We then consider (5.3). Letting $s \to \infty$, (5.3) is equal to

$$\Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{\infty 2}}\dot{x}_{it}\dot{x}_{it}'\Big]\beta^{\infty} = \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{\infty 2}}\dot{x}_{it}\Big(\dot{y}_{it} - \rho^{\infty}\ddot{y}_{it} - \lambda_i^{\infty\prime}f_t^{\infty}\Big).$$

The above equation can be written as

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{\infty 2}}\dot{x}_{it}(\dot{y}_{it} - \rho^{\infty}\ddot{y}_{it} - \dot{x}_{it}'\beta^{\infty}) - \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{\infty 2}}\dot{x}_{it}\lambda_i^{\infty\prime}f_t^{\infty} = 0$$

By (5.4),

$$f_t^{\infty} = \Phi^{\infty\prime}(\Sigma_{zz}^{\infty})^{-1}D^{\infty}\dot{z}_t = G^{\infty}\Phi^{\infty\prime}(\Sigma_{\epsilon\epsilon}^{\infty})^{-1}D^{\infty}\dot{z}_t.$$

By the preceding two equations, we have

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{\infty 2}}\dot{x}_{it}(\dot{y}_{it} - \rho^{\infty}\ddot{y}_{it} - \dot{x}_{it}'\beta^{\infty}) - \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{\infty 2}}\dot{x}_{it}\lambda_i^{\infty\prime}G^{\infty}\Phi^{\infty\prime}(\Sigma_{\epsilon\epsilon}^{\infty})^{-1}D^{\infty}\dot{z}_t = 0$$

Dividing NT on both sides, we obtain the same equation as (3.5). The iterating formula for ρ at each step satisfies the first order condition for ρ by way of computation. So ρ^{∞} satisfies (3.4).

In summary, we show that the converged EM solutions of Section 5 satisfy the first order conditions given in Section 3 and hence possess the local optimality property. We have also verified that the numerical solutions indeed satisfy the first order conditions in our simulations.

Appendix F: Additional simulation results

The simulation results reported in the main text are based on an asymmetric error distribution (χ^2). This section presents additional simulation results when the idiosyncratic errors ϵ_{it} follow a normal distribution or student's t-distribution (t_5) (standardized to have a unit variance). The simulation results show that the MLE is not sensitive to the error distributions, as predicted by the theory.

Table F1: The performance of the MLE under $\rho = 0.2$
with "1 ahead and 1 behind" spatial weights matrix, normal distribution

N	T	%	ρ		β_1		β_2	
		$\hat{r} = r$	Bias	RMSE	Bias	RMSE	Bias	RMSE
25	50	99.7	-0.0002	0.0044	-0.0003	0.0092	0.0005	0.0096
50	50	100.0	0.0000	0.0025	-0.0002	0.0062	0.0003	0.0059
100	50	100.0	0.0000	0.0017	0.0001	0.0038	-0.0001	0.0038
25	100	99.8	0.0000	0.0030	-0.0001	0.0069	0.0000	0.0069
50	100	100.0	0.0000	0.0018	-0.0003	0.0041	0.0001	0.0042
100	100	100.0	0.0000	0.0011	0.0000	0.0026	-0.0001	0.0027

Table F2: The performance of the MLE under $\rho = 0.9$
with "1 ahead and 1 behind" spatial weights matrix, normal distribution

N	T	%	ρ		β_1		β_2	
		$\hat{r} = r$	Bias	RMSE	Bias	RMSE	Bias	RMSE
25	50	99.8	0.0000	0.0011	0.0002	0.0099	0.0002	0.0106
50	50	100.0	0.0001	0.0006	0.0002	0.0061	0.0004	0.0061
100	50	100.0	0.0001	0.0004	-0.0001	0.0037	-0.0004	0.0039
25	100	99.6	0.0000	0.0007	0.0002	0.0068	0.0001	0.0072
50	100	100.0	0.0000	0.0005	-0.0001	0.0041	0.0000	0.0043
100	100	100.0	0.0000	0.0003	-0.0001	0.0026	-0.0001	0.0027

Table F3: The performance of the MLE under $\rho = 0.2$
with "3 ahead and 3 behind" spatial weights matrix, normal distribution

N	T	%	ρ		β_1		β_2	
		$\hat{r} = r$	Bias	RMSE	Bias	RMSE	Bias	RMSE
25	50	99.8	0.0003	0.0065	0.0001	0.0094	0.0007	0.0102
50	50	100.0	-0.0001	0.0039	-0.0003	0.0058	0.0002	0.0061
100	50	100.0	0.0001	0.0025	0.0000	0.0037	0.0002	0.0039
25	100	99.7	-0.0004	0.0043	-0.0002	0.0064	0.0001	0.0068
50	100	100.0	0.0000	0.0028	0.0000	0.0041	-0.0001	0.0041
100	100	100.0	0.0001	0.0017	0.0001	0.0026	-0.0001	0.0026

Table F4: The performance of the MLE under $\rho = 0.9$
with "3 ahead and 3 behind" spatial weights matrix, normal distribution

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	100.0	0.0000	0.0013	0.0005	0.0100	-0.0005	0.0097
50	50	100.0	0.0000	0.0009	-0.0001	0.0061	0.0000	0.0060
100	50	100.0	0.0000	0.0006	-0.0001	0.0038	0.0001	0.0038
25	100	100.0	0.0000	0.0009	-0.0002	0.0064	0.0001	0.0070
50	100	100.0	0.0000	0.0006	0.0000	0.0040	-0.0003	0.0043
100	100	100.0	0.0000	0.0004	0.0000	0.0027	-0.0002	0.0027

Table F5: The performance of the MLE under $\rho = 0.2$
with "1 ahead and 1 behind" spatial weights matrix, student's t-distribution

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.7	-0.0002	0.0041	0.0001	0.0090	-0.0002	0.0097
50	50	99.9	0.0000	0.0026	-0.0003	0.0060	0.0000	0.0058
100	50	100.0	0.0000	0.0016	-0.0001	0.0037	0.0000	0.0036
25	100	99.8	0.0001	0.0029	-0.0002	0.0068	-0.0001	0.0065
50	100	100.0	0.0001	0.0018	-0.0003	0.0042	-0.0002	0.0039
100	100	100.0	0.0000	0.0011	-0.0001	0.0026	-0.0001	0.0026

Table F6: The performance of the MLE under $\rho = 0.9$
with "1 ahead and 1 behind" spatial weights matrix, student's t-distribution

N	T	% $\hat{r} = r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.7	0.0001	0.0010	-0.0003	0.0096	0.0000	0.0099
50	50	99.9	0.0000	0.0006	-0.0004	0.0061	0.0000	0.0060
100	50	100.0	0.0001	0.0004	-0.0001	0.0038	-0.0002	0.0038
25	100	99.9	0.0000	0.0007	0.0004	0.0065	-0.0002	0.0067
50	100	100.0	0.0000	0.0004	0.0000	0.0041	-0.0002	0.0042
100	100	100.0	0.0000	0.0003	-0.0001	0.0025	-0.0001	0.0027

Table F7: The performance of the MLE under $\rho = 0.2$
with "3 ahead and 3 behind" spatial weights matrix, student's t-distribution

N	T	% $\hat{r}=r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.6	-0.0006	0.0067	0.0001	0.0093	0.0000	0.0097
50	50	100.0	-0.0001	0.0038	-0.0002	0.0060	-0.0001	0.0060
100	50	100.0	-0.0001	0.0025	-0.0001	0.0037	0.0001	0.0037
25	100	99.7	-0.0001	0.0043	-0.0001	0.0069	0.0000	0.0069
50	100	100.0	-0.0001	0.0029	-0.0001	0.0039	-0.0002	0.0040
100	100	100.0	-0.0001	0.0017	0.0000	0.0025	0.0001	0.0026

Table F8: The performance of the MLE under $\rho = 0.9$
with "3 ahead and 3 behind" spatial weights matrix, student's t-distribution

N	T	% $\hat{r}=r$	ρ Bias	ρ RMSE	β_1 Bias	β_1 RMSE	β_2 Bias	β_2 RMSE
25	50	99.7	0.0000	0.0013	0.0001	0.0099	0.0005	0.0100
50	50	100.0	0.0000	0.0009	0.0003	0.0062	-0.0001	0.0059
100	50	100.0	0.0000	0.0005	0.0000	0.0038	-0.0001	0.0036
25	100	99.8	0.0000	0.0010	0.0000	0.0064	0.0001	0.0066
50	100	100.0	0.0000	0.0006	-0.0002	0.0041	-0.0002	0.0041
100	100	100.0	0.0000	0.0004	-0.0001	0.0027	0.0000	0.0026