G D P

全球发展展望

Global Development Perspective

中国社会科学院世界经济与政治研究所 经济发展研究中心

Working Paper No. 21.002

May 12th, 2021

Lei Hou, 侯蕾

houlei@cass.org.cn

Revisiting the location of FDI in China: A panel data approach with heterogeneous shocks[1]

Lei Hou, Kunpeng Li, Qi Li, Min Ouyang

Abstract Foreign Direct Investment (FDI) is viewed as a primary driving force in shaping the global economy and receives particular attention in empirical studies. In this paper, we argue that many of the existing studies ignore endogeneities that arise from shocks in source and destination countries. To address this endogeneity issue, we take the "controlling through estimating" idea from the econometric literature and propose using panel data models with heterogeneous shocks to deal with it. We consider the quasi maximum likelihood (QML) method to estimate our proposed model. We investigate the asymptotic properties of the QML estimator, including the consistency, the asymptotic representation, and the limiting distribution. We also propose new statistics to test the validity of the use of traditional dynamic and static panel data estimation methods. Applying it to the location determinants of inward FDI in China, we find that the endogeneity issue does exist, and that controlling for heterogeneous shocks helps to improve the estimation results.

Key words FDI, location determinants, Endogeneity issue, Panel data models, Quasi maximum likelihood estimation, Heterogeneous shocks

[1] Published in *Journal of Econometrics* 2021, Volume 221, Issue 2, Pages 483-509.

1 Introduction

In past decades, Foreign Direct Investment (FDI) has been a driving force in shaping the global economy. After the global economic and financial crisis of 2008-2009, the strong recovery of FDI was a principal factor in the global rebound (World Investment Report 2016, UNCTAD). The importance of FDI to the world economy has led to extensive research in multiple disciplines, including international business, economics, and management (e.g., Nielsen, Asmussen and Weatherall (2017)).

Following from the Ownership-Location-Internalization paradigm of FDI theory, a large stream of the FDI literature focuses on the location decision of FDI firms. Various location-specific characteristics, the fundamental factors in attracting FDI, have been investigated in empirical studies. Some typical economic factors – such as labor cost, market size, infrastructure, taxes, tariffs, and exchange rate – are well-established determinants for the FDI decision (e.g. Cheng and Kwan (2000), Blyde and Molina (2015)). Blonigen (2005), Nielsen, Asmussen and Weatherall (2017) provide excellent reviews of the important factors that determine FDI. Further, other local attributes – including the quality of institutions, corruptions and government efficiency, the access to market and supplier, and the proximity to alternative locations – also play important roles in determining FDI (Kang and Jiang (2012), Du, Lu and Tao (2008), Amiti and Javorcik (2008), Blanc-Brude et al. (2014), Cole, Elliott and Zhang (2009)).

In this paper, we argue that many of the existing empirical studies on FDI locations may suffer one source of endogeneity arising from the fact that FDI is a cross-border activity. To illustrate this, we note that FDI is an economic activity in its source countries. So, its volume is directly linked with economic conditions (or economic shocks) in source countries. Negative shocks in source countries would decrease the volume of FDI even although the destination region may possess many attractive features. On the other hand, the local characteristics of the destination region, which are used to explain the variations in FDI volume among different regions, are subject to economic shocks in its destination country. If shocks in source countries are correlated with those in the destination country, as we believe is typical in the real world, then endogeneity occurs.

We consider China's FDI inflow as an example to further explain this point. Figure 1 presents a picture of inward FDI on China from 1990 to 2010. As can be seen there, the total FDI inflow grows rapidly over the sample period, except for two episodes of dips that correspond to the 1998-2000 Asian Financial Crisis and the 2008-2009 Subprime Crisis. This basic fact lends strong support to the previous argument that FDI volume is linked to shocks in source countries. Moreover, as a response to the subprime crisis the central government of China launched an economic stimulus package of four trillion RMB, designed to offset the negative effects of the subprime crisis on the Chinese economy. This huge stimulus plan, which we view as a domestic shock to the host country, has a great impact on local FDI determinants, including wages, infrastructure, market demand, etc. Given these facts, we believe that one should take the endogeneity problem seriously when analyzing FDI data.

Insert Figure 1

In this paper we adopt the "controlling through estimating" idea in the econometric literature to address the endogeneity issue. We introduce domestic and foreign shocks into traditional fixed-effects panel data models to deal with this problem. Because the nature of the shocks is pervasive, we follow Stock and Watson (1998), modeling transitions of these shocks with a factor model. Specifically, we use factors to denote the shocks, and factor loadings to denote responses to shocks. The endogeneity is controlled for by estimating the regression coefficients simultaneously with factor loadings and variances of factors. A detailed description of the model is elaborated in Section 2.

China has experienced a growth miracle over the last four decades. To drive the economic recovery as fast as possible in its early stages, the central government of China authorized local governments to launch various policies to attract FDI. This makes China a unique subject for studying the location determinants of FDI. There have been many studies on this topic; see Amiti and Javorcik (2008), Cole, Elliott and Zhang (2009), Du, Lu and Tao (2008), and Fung, Iizaka and Parker (2002)), to name only a few. However, nearly all of these studies ignore the possible endogeneity issue resulting from the heterogeneous shocks mentioned above. Using our proposed new model, we revisit the location determinants of inward FDI on China. We find that regions with more FDI inflows in last period, larger local markets, broader infrastructure stock, lower labor costs, a higher level of openness, more government intervention and better human capital availability, attract more FDI. These results are consistent with economic theory. A comparison with traditional estimation results suggests that controlling for heterogeneous shocks helps to improve the estimation results.

Our paper contributes to the existing literature in several dimensions. First, we point out a potential endogeneity issue, which to a large extent is ignored in the existing empirical FDI studies. Because the source of the endogeneity stems from cross-border economic activities, the same endogeneity issue may exist in other areas of international economics. Second, we develop a panel data model with heterogeneous shocks to exclusively address this potential issue. Our proposed model is closely related to, but different from, a rapidly growing literature on panel data models with interactive effects. Models of this type have received much attention recently in applied studies; see, e.g., Gobillon and Magnac (2016) and Xu (2017) who estimate average treatment effects based on factor structure models. We use quasi maximum likelihood (QML) method to estimate the model. The QML method has been used often in factor analysis, e.g., Doz, Giannone and Reichlin (2012), Bai and Li (2012, 2016), etc. It has advantage in dealing with large dimensional cross-sectional heteroskedasticity parameters. We establish the asymptotic properties of the QML estimator. Third, we propose a new statistic to test the validity of the use of traditional panel data methods, which are popular in empirical FDI studies. We discuss the rotational indeterminacy issue associated with this test. The proposed statistic can be viewed as the sum of square of sample canonical correlations and is invariant to the rotational indeterminacy. We investigate the asymptotic properties of this new statistic. Fourth, we apply our new model to China's FDI location study. We find that after controlling for the endogeneity, the estimation results are much improved.

This paper is organized as follows. Section 2 gives a detailed description of our methodology. Section 3 gives the quasi likelihood function that is used in our theoretical analysis, and discusses the identification issue. Section 4 imposes some assumptions needed for the theoretical analysis. Section 5 examines the asymptotic properties of the QML estimator and finds that the QML estimator has five non-negligible bias terms. The estimation of the biases and the limiting variances is also discussed. Section 6 investigates hypothesis testing on the validity of the traditional panel data estimation methods. Section 7 discusses the method on obtaining the numerical value of QML estimator based on the expectation conditional maximization (ECM) algorithm. Section 8 runs Monte Carlo simulations to investigate the performance of the QML estimator and its competitors. Section 9 applies our method to the FDI location determinants in China. Section 10 concludes. In the appendix, we provide detailed proofs of the main theoretical results.

2 Methodology

When a regression model suffers from endogeneity, one widely-used method is to find instrumental variables (IV). But qualified IVs are difficult to find because they are required to be uncorrelated with the errors but correlated with the endogenous explanatory variables. It is very challenging for researchers to verify these two conditions, whether the proposed IVs are qualified (or not) remains largely agnostic. This is the typical case in empirical FDI location studies. In the previous section, we point out that the shocks in source countries may cause an endogeneity problem, but the nature and characteristics of those shocks was largely left unspecified. This makes it difficult for applied researchers to find the qualified IVs .

To address this concern, we adopt the "controlling through estimating" idea from the econometric literature. We propose the following dynamic panel data model with heterogeneous shocks

$$Y_{it} = \alpha_i + \rho Y_{it-1} + X_{it}'\beta + \kappa_i' g_t + e_{it}, \tag{2.1}$$

$$X_{it} = \nu_i + \gamma_i' h_t + v_{it}, \tag{2.2}$$

where Y_{it} is the dependent variable; X_{it} is a k-dimensional vector of explanatory variables. α_i and ν_i are the individual intercepts; and e_{it} and v_{it} are the idiosyncratic errors. g_t denotes r_1-dimensional outside shocks and h_t denotes r_2-dimensional domestic shocks. We allow g_t to be either correlated or uncorrelated with h_t, or allow g_t to contain partial factors which are also contained in h_t (i.e., g_t and h_t can have some common components). We note that equation (2.1) is capable to allow the lags of explanatory variables.

This model is related to the rapidly growing literature on panel data models with interactive effects or common shocks[①]. Overall, the existing literature can be classified into two branches. One branch assumes that the errors in the Y equation have a factor structure, but the inner structure of explanatory variables is unspecified. Representative studies include Bai (2009), Moon and Weidner (2017) and Li, Qian and Su (2016), among others. In terms of model specification, these studies are more general than the one in this paper, because we also assume that the explanatory variables have a factor structure, as seen in equation (2.2). However, one undesirable feature of these studies is that their estimations are all performed using the least-squares-method, under the assumption of cross-sectional homoskedasticity. If cross-sectional heteroskedasticity is heavy, the lease-squares estimator will have a $O(\frac{1}{N})$ bias and its limiting variance will be unduly large. The unsatisfactory performance of the least-squares estimator can be seen by simulations in Section 8. As shown in Figure 2, the FDI inflows into different provinces of China vary widely,

[①]For the applied studies on panel data models with interactive effects, see Eberhardt et al. (2013), Holly et al. (2010), Pesaran and Kapetanios (2005), Castagnetti and Rossi (2013), Moscone and Tosetti (2010), to name a few.

from 81 million dollars in Ningxia province to 28,500 million dollars in Jiangsu province in 2010. Because of this, we believe that the least-squares method is not the best way to estimate the model.

Insert Figure 2

The second branch assumes that both the errors of the Y equation and the explanatory variables have factor structures. Studies in this branch include Pesaran (2006), Pesaran and Kapetanios (2005), Bai and Li (2014), Greenaway-McGrevy, Han and Sul (2012), among others. Our study falls within this scope. Pesaran (2006) proposes the common correlated effect (CCE) estimation method: it seeks to approximate the space spanned by unknown factors with one spanned by cross-sectionally average observations. Once the space spanned by factors is well approximated, the correlated effects arising from the common factors can be controlled. Chudik and Pesaran (2015) extend the CCE method to the dynamic panel data models. Alternatively, Pesaran and Kapetanios (2005) propose a two-step principal component (PC) estimation method in which the factors are first estimated by the PC method and then the regression coefficients are estimated by replacing the factors with their PC estimates. Pesaran and Kapetanios (2005) use simulations to show the feasibility of their two-step estimation procedure. The asymptotic properties of the two-step PC estimators are investigated by Greenaway-McGrevy, Han and Sul (2012). The generalized method of moment (GMM) is considered by Ahn, Lee and Schmidt (2001, 2013) under the fixed-T setup. Again, we note that the above studies except Bai and Li (2014) do not take into account cross-sectional heteroskedasticity.

The current paper differs from Bai and Li (2014) in several aspects. First, Bai and Li's model is static but ours is dynamic. This seemingly-trivial difference actually poses new challenge on the analysis of consistency. In the QML estimation, the endogeneity is controlled by explicitly estimating the loadings and variance of factors $f_t = [g_t', h_t']'$ because these two parts are responsible for the endogenous correlation. However, the lagged dependent variable Y_{it-1} does not include g_t or h_t as the factors. So controlling the endogeneity arising from f_t does not guarantee the consistency of the QML estimator for ρ. New arguments are therefore needed to justify validity of the QML method. Second, Bai and Li (2014) make restrictive assumptions on the idiosyncratic errors, that is, the errors are assumed to be identically and independently distributed over time and uncorrelated over the cross section. In this paper we relax them and allow for general heteroskedasticiy and correlation to the idiosyncratic errors. Our theoretical analysis indicates that the QML estimator has five bias terms. Among these five terms, three terms can be found in Bai (2009) and Moon and Weidner (2017) in different formulas. The other two terms are new in the literature. We find that one bias term is related with the skewness of errors, and the other arises because the lagged dependent variable does not include h_t as factors. Furthermore, we discuss the estimation of the biases and the limiting variances under two interchangeable assumptions, which is also non-trivial because one may encounter a similar long-run variance estimation issue. Third, the endogeneity is inherited in Bai and Li (2014) through their factor model structure (i.e., endogeneity always exists). But our specification is more flexible on this. To illustrate this point, consider model (3.1) of Bai and Li (2014).

$$Y_{it} = \alpha_i + X_{it}'\beta + \kappa_i' g_t + e_{it},$$

$$X_{it} = \nu_i + \psi_i' g_t + \gamma_i' h_t + v_{it}.$$

Under Bai and Li's specification, both the X equation and the Y equation share the same factors (g_t). So the endogeneity is always present in their model. One consequence of their specification is that the relationship of g_t and h_t is no longer important. In our model specification, we only assume that the explanatory variables contain the factors h_t. The relationship of g_t and h_t is left unspecified. It is possible that g_t is independent with h_t, or it is possible that g_t and h_t are highly correlated. Our flexible specification on factors allows us to test an interesting issue: whether the classical estimation methods for panel data models is valid in the presence of domestic and foreign shocks. In Section 6, we perform this work to propose a statistic based on validity of the Anderson and Hsiao's (1981, 1982) instrumental variable, which lays the base for the popular Arellano and Bond's (1990) GMM method. As can be seen there, that test is a direct benefit from our flexible specification on the factors of the Y and X equations. In some particular applications, the relationship of these two types of shocks are also of interest.

3 Quasi likelihood function

Before formally presenting the model, it is necessary to define some notations. We use $\dot{a}_t$ to denote $a_t - \frac{1}{T}\sum_{t=1}^T a_t$ for any vector a_t, use M_{uv} to denote $\frac{1}{T}\sum_{t=1}^T \dot{u}_t \dot{v}_t'$, and use $\|\cdot\|$ to denote the Frobenius norm, i.e., for any matrix A, $\|A\| = \sqrt{\mathrm{tr}(A'A)}$.

We first write model (2.1)-(2.2) in a matrix form, to make it more convenient for the subsequent analysis. Let $f_t = (g_t', h_t')'$, $\mu_i = (\alpha_i, \nu_i')'$, $\epsilon_{it} = (e_{it}, v_{it}')'$ and

$$\Lambda_i = \begin{bmatrix} \kappa_i & 0_{r_1\times k} \\ 0_{r_2\times 1} & \gamma_i \end{bmatrix}, \quad \begin{bmatrix} Y_{it} - \rho Y_{it-1} - X_{it}'\beta \\ X_{it} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & -\beta' \\ 0_{k\times 1} & I_k \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} Y_{it} - \rho Y_{it-1} \\ X_{it} \end{bmatrix}}_{z_{it}(\rho)} = Bz_{it}(\rho),$$

with B and $z_{it}(\rho)$ implicitly defined above. Then model (2.1)-(2.2) can be written as

$$Bz_{it}(\rho) = \mu_i + \Lambda_i' f_t + \epsilon_{it}. \tag{3.1}$$

Let $z_t(\rho) = (z_{1t}(\rho)', z_{2t}(\rho)', \ldots, z_{Nt}(\rho)')'$, $\Lambda = (\Lambda_1, \Lambda_2, \ldots, \Lambda_N)'$, $\mu = (\mu_1', \mu_2', \ldots, \mu_N')'$, and ϵ_t is similarly defined. Then we can rewrite (3.1) as

$$\mathcal{B}z_t(\rho) = \mu + \Lambda f_t + \epsilon_t. \tag{3.2}$$

with $\mathcal{B} = I_N \otimes B$.

Suppose that (i) f_t is gaussian with mean zero and variance Σ_{ff}, and is independent with ϵ_{is} for all i, t and s; (ii) ϵ_{it} is independently and identically distributed over t and independent over i; (iii) e_{it} is independent with v_{it} for all i and t; (iv) the loadings Λ are nonrandom. With these assumptions, we have $\mathrm{var}(\Lambda f_t + \epsilon_t) = \Lambda \Sigma_{ff} \Lambda' + \Phi \triangleq \Sigma_{zz}$ with $\Phi = \mathrm{var}(\epsilon_t)$. We therefore can write out the likelihood function as

$$\mathcal{L}^*(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2NT}\sum_{t=1}^T \Big[\mathcal{B}z_t(\rho) - \mu\Big]' \Sigma_{zz}^{-1} \Big[\mathcal{B}z_t(\rho) - \mu\Big],$$

where $\theta = (\psi, \Lambda, \Phi, \Sigma_{ff})$ and $\psi = (\rho, \beta')'$. Given B and Σ_{zz}, it is easy to see that μ maximizes the above likelihood function at $\mu = \mathcal{B}\frac{1}{T}\sum_{t=1}^{T} z_t(\rho)$. Substituting the above formula into the likelihood function, we have

$$\mathcal{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}\Big[\mathcal{B}M_{zz}(\rho)\mathcal{B}'\Sigma_{zz}^{-1}\Big], \tag{3.3}$$

with $M_{zz}(\rho) = \frac{1}{T}\sum_{t=1}^{T}\dot{z}_t(\rho)\dot{z}_t(\rho)'$. Assumptions (i)-(iv) are restrictive in applications, and are relaxed to a large extent in the next section. But we still use (3.3) to motivate our estimation method. The QML estimator is therefore defined as

$$\widehat{\theta} = \underset{\theta^\dagger=(\psi^\dagger,\Lambda^\dagger,\Phi^\dagger,\Sigma_{ff}^\dagger)\in\Theta}{\operatorname{argmax}} \mathcal{L}(\theta^\dagger)$$

where $\theta^\dagger$ is the input arguments and Θ is the parameters space, which satisfy the normalization restrictions and Assumption E below. The first-order conditions with respect to free parameters are given in Appendix A.

We end this section with some discussions on the identification issue. Although the regression coefficient ψ is identifiable, the factors and factor loadings are not, which is well known in factor models, see, e.g., Bai and Li (2012). A notable feature of the current model is the presence of zero elements in Λ, which precludes some rotation possibilities. As a result, the needed number of restrictions for full identification is $r_1^2 + r_2^2$ instead of $(r_1 + r_2)^2$. In this paper, we only impose the following $(r_1^2 + r_1 + r_2^2 + r_2)/2$ normalization restrictions:

$$\widehat{\Sigma}_{gg} = M_{gg} = \frac{1}{T}\sum_{t=1}^{T}\dot{g}_t\dot{g}_t', \qquad \widehat{\Sigma}_{hh} = M_{hh} = \frac{1}{T}\sum_{t=1}^{T}\dot{h}_t\dot{h}_t'.$$

These restrictions would facilitate the theoretical analysis, and are not enough for full identification. In addition, these restrictions do not change the maximum value of objective function, so can be viewed as the loose ones. A direct implication is that these restrictions can be ignored in the computation method of Section 7.

4 Assumptions

We make the following assumptions. Let c and C be two generic constants, which are not necessarily the same at each appearance.

Assumption A. The r-dimensional factors $f_t = (g_t', h_t')'$ is a zero-mean covariance stationary process with absolute summable autocorrelation coefficients, $\sup_t E(\|f_t\|^4) < C$, and is independent with $\epsilon_{is} = (e_{is}, v_{is}')'$ for each i, t and s.

Assumption B. The factor loadings Λ_i are nonrandom, such that

(B.1) $\|\Lambda_j\| \le C$ for all $j = 1, \cdots, N$.

(B.2) The limit $\lim\limits_{N\to\infty} N^{-1}\Lambda'\Lambda$ is an $r \times r$ positive definite matrix.

Assumption C. The idiosyncratic error $\epsilon_{it} = (e_{it}, v_{it}')'$ satisfies $\sup_{i,t} E\|\epsilon_{it}\|^{16} \le C$, and the following conditions:

(C.1) $e_t = (e_{1t}, \ldots, e_{Nt})'$ is a martingale difference sequence with respect to the adaptive filtration $\mathscr{F}_{t-1} = \sigma(e_1, e_2, \ldots, e_{t-1})$. Let $\tau_{ij,t} = E(e_{it}e_{jt})$, there exists a sequence of fixed constants $\{\tau_{ij}\}_{j=1}^{\infty}$ for each given i such that $\sum_{j=1}^{N} \tau_{ij} < C$ and $\max_t |\tau_{ij,t}| \leq \tau_{ij}$.

(C.2) e_{it} is independent with v_{js} for all i, j, t and s.

(C.3) Let $\xi_{ij,t} = E(v_{it}v'_{jt})$ and $\zeta_{i,ts} = E(v_{it}v'_{is})$. For each i and t, there exist positive sequences $\{\xi_{ij}\}_{j=1}^{\infty}$ and $\{\zeta_{ts}\}_{s=1}^{\infty}$ such that $\sum_{j=1}^{N} \xi_{ij} < \infty$ and $\sum_{s=1}^{T} \|\zeta_{ts}\| < \infty$ with $\max_t \|\xi_{ij,t}\| \leq \xi_{ij}$ and $\max_i \|\zeta_{i,ts}\| \leq \zeta_{ts}$.

(C.4) $c \leq \tau_{min}(\Phi_j) \leq \tau_{max}(\Phi_j)$ for each j, where $\Phi_j = E(\epsilon_{jt}\epsilon'_{jt})$ with $\epsilon_{jt} = (e_{jt}, v'_{jt})'$, and $\tau_{min}(\Phi_j)$ and $\tau_{max}(\Phi_j)$ denote the smallest and largest eigenvalues of Φ_j, respectively.

(C.5) Moreover,

$$(a) \quad \frac{1}{NT} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \|E(\epsilon_{it}\epsilon'_{js})\| \leq C,$$

$$(b) \quad \frac{1}{T} \sum_{t,s=1}^{T} |\text{cov}(e_{it}e_{jt}, e_{is}e_{js})| \leq C, \quad \textit{for any } i \textit{ and } j,$$

$$(c) \quad \frac{1}{T} \sum_{t,s=1}^{T} \max_{1 \leq p,p',q,q' \leq k} \Big(|\text{cov}(v_{it,p}v_{jt,q}, v_{is,p'}v_{js,q'}| \Big) \leq C, \quad \textit{for any } i \textit{ and } j$$

$$(d) \quad \frac{1}{N^2T} \sum_{i,j,i',j'=1}^{N} \sum_{t,s=1}^{T} |\text{cov}(e_{it}e_{jt}, e_{i's}e_{j's})| \leq C,$$

$$(e) \quad \frac{1}{N^2T} \sum_{i,j,i',j'=1}^{N} \sum_{t,s=1}^{T} \max_{1 \leq p,p',q,q' \leq k} \Big(|\text{cov}(v_{it,p}v_{js,q}, v_{i's,p'}v_{j's,q'})| \Big) \leq C,$$

$$(f) \quad \frac{1}{NT^2} \sum_{i,j=1}^{N} \sum_{t,s,t',s'=1}^{T} \Big|\text{cov}\Big(e_{it}e_{it-l}(e_{is}^2 - \sigma_i^2), e_{jt'}e_{jt'-l}(e_{js'}^2 - \sigma_j^2)\Big)\Big| \leq C, \quad \text{for each } l.$$

Assumption D. For the regression coefficients, we assume that $|\rho| < 1$ and $\|\beta\| \leq C$.

Assumption E. The variances Φ_i for all i and $\Sigma_{ff} = E(f_t f'_t)$ are estimated in a compact set, i.e. all the eigenvalues of $\widehat{\Phi}_i$ and $\widehat{\Sigma}_{ff}$ are in an interval $[C^{-1}, C]$.

Assumption F. Let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$ and $\boldsymbol{K} = (\kappa_1, \kappa_2, \ldots, \kappa_N)'$. Define $\mathbf{B}_t$ and $\mathbf{U}_t$ as

$$\mathbf{B}_t = \frac{\boldsymbol{\alpha}}{1-\rho} + \boldsymbol{K} \sum_{\tau=0}^{\infty} \rho^{\tau} g_{t-\tau} + \sum_{\tau=0}^{\infty} \rho^{\tau} X_{t-\tau}, \quad \mathbf{U}_t = \sum_{\tau=0}^{\infty} \rho^{\tau} e_{t-\tau},$$

The following two weak convergences hold:

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \widetilde{\mathcal{Q}}'_t e_t \xrightarrow{d} N(0, V_a), \qquad \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \mathbf{U}'_{t-1}\Sigma_{ee}^{-1} e_t \xrightarrow{d} N(0, V_b),$$

with $V_a = \text{plim}_{N,T\to\infty} \frac{1}{NT} \sum_{t=1}^{T} \widetilde{\mathcal{Q}}'_t E(e_t e'_t) \widetilde{\mathcal{Q}}_t$ and $V_b = \text{plim}_{N,T\to\infty} \mathbf{U}'_{t-1}\Sigma_{ee}^{-1} E(e_t e'_t)\Sigma_{ee}^{-1}\mathbf{U}_{t-1}$, where $\widetilde{\mathcal{Q}}_t = M_K(\widetilde{Q}_t - \sum_{s=1}^{T} \pi_{st}\widetilde{Q}_s)$, $\pi_{st} = \dot{g}'_s(\dot{G}'\dot{G})^{-1}\dot{g}_t$, $M_{\boldsymbol{K}} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}\boldsymbol{K}(\boldsymbol{K}'\Sigma_{ee}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\Sigma_{ee}^{-1}$, and $\widetilde{Q}_t = (\dot{\mathbf{B}}_{t-1}, \dot{X}_t)$ with $\dot{X}_t = [\dot{X}_{1t}, \dot{X}_{2t}, \ldots, X_{Nt}]'$.

Assumption A is regarding factors. The randomness assumption on factors is different from the existing fixed-valued assumption in the QML studies such as Bai and Li (2012). The randomness is motivated by the dynamics of our model. Similar assumptions are made in a number of related studies, such as Pesaran (2006), Doz, Giannone and Reichlin (2012), etc. Stationarity assumption is largely for theoretical analysis and can be relaxed to some extent. Such an assumption precludes some practically interesting cases, for example, the factors are integrated process, see Bai, Kao and Ng (2009) and Kapetanios et al. (2011). The current paper confines the analysis on the stationary case and leaves the nonstationary case as a future work. Assumption B is about factor loadings. It assumes that the loadings are fixed values. Assumption B.2 precludes the possibility of degenerate expression of factor structure.

Assumption C is about the idiosyncratic errors, which extends the counterpart of Bai and Li (2014) to allow general weak correlations and heteroskedasticities over two dimensions. Due to the presence of the lagged dependent variable, the serial correlation of e_{it} is precluded by Assumption C.1. But it permits the cross-sectional correlation and heteroskedasticity. Assumption C.2 postulates that the idiosyncratic part of the regressors are exogenous. This assumption is necessary for the consistent estimation of parameters in the proposed model. Assumption C.3 allows the correlations and heteroskedasticities in the idiosyncratic errors of regressors. Assumption C.4 assumes that the variances of idiosyncratic errors are bonded away from zero. This assumption is standard. Note that with Assumptions C.1, C.2 and C.3, it is easy to verify that $\tau_{max}(\Phi_j)$ is uniformly bounded for each j. Assumption C.5 impose some moment conditions. These moment conditions implicitly further restrict the correlation strength of the errors, and are necessary for theoretical analysis.

Assumption D is regarding the regression coefficients. This assumption is standard in the econometric literature. Assumption E assumes that partial parameters are estimated in a compact set, which is often made when dealing with highly nonlinear objective functions. The likelihood function (3.3) is highly nonlinear. Assumption F provides two weak convergences of partial sum, which are used to derive the limiting distribution of the QML estimator. We implicitly use the martingale difference assumption to derive the limiting variances.

5 Asymptotic properties

This section presents the asymptotic results of the QML estimator. We first provide the consistency result in the following proposition.

Proposition 5.1 *Let $\widehat{\theta} = (\widehat{\rho}, \widehat{\beta}, \widehat{\Lambda}, \widehat{\Psi})$ be the solution obtained by maximizing (3.2). Under Assumptions A-E, we have*

$$\widehat{\rho} - \rho \xrightarrow{p} 0, \qquad \widehat{\beta} - \beta \xrightarrow{p} 0,$$
$$\frac{1}{N}\sum_{i=1}^{N}\left(\|\widehat{\kappa}_i - \mathbf{R}_1\kappa_i\|^2 + \|\widehat{\gamma}_i - \mathbf{R}_2\gamma_i\|^2\right) \xrightarrow{p} 0,$$
$$\frac{1}{N}\sum_{i=1}^{N}\left(|\widehat{\sigma}_i^2 - \sigma_i^2|^2 + \|\widehat{\Sigma}_{ii} - \Sigma_{ii}\|^2\right) \xrightarrow{p} 0.$$

where $\mathbf{R}_1 = M_{gg}^{-1}\mathbf{E}_1'\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda M_{ff}\mathbf{E}_1$ *and* $\mathbf{R}_2 = M_{hh}^{-1}\mathbf{E}_2'\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda M_{ff}\mathbf{E}_2$, $\mathbf{E}_1$ *and* $\mathbf{E}_2$ *are implicitly defined by* $I_r = [\mathbf{E}_1, \mathbf{E}_2]$, $\sigma_i^2 = \frac{1}{T}\sum_{t=1}^T E(e_{it}^2)$ *and* $\Sigma_{ii} = \frac{1}{T}\sum_{t=1}^T E(v_{it}v_{it}')$.

Let Ω_{NT} be defined as

$$\Omega_{NT} = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}(\dot{Y}_{-1}'M_{\boldsymbol{K}}\dot{Y}_{-1}M_{\dot{G}}) & \mathrm{tr}(\dot{Y}_{-1}'M_{\boldsymbol{K}}\dot{X}_1M_{\dot{G}}) & \cdots & \mathrm{tr}(\dot{Y}_{-1}'M_{\boldsymbol{K}}\dot{X}_kM_{\dot{G}}) \\ \mathrm{tr}(\dot{X}_1'M_{\boldsymbol{K}}\dot{Y}_{-1}M_{\dot{G}}) & \mathrm{tr}(\dot{X}_1'M_{\boldsymbol{K}}\dot{X}_1M_{\dot{G}}) & \cdots & \mathrm{tr}(\dot{Y}_{-1}'M_{\boldsymbol{K}}\dot{X}_kM_{\dot{G}}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\dot{X}_k'M_{\boldsymbol{K}}\dot{Y}_{-1}M_{\dot{G}}) & \mathrm{tr}(\dot{X}_k'M_{\boldsymbol{K}}\dot{X}_1M_{\dot{G}}) & \cdots & \mathrm{tr}(\dot{X}_k'M_{\boldsymbol{K}}\dot{X}_kM_{\dot{G}}) \end{bmatrix}$$

where $\dot{Y}_{-1} = [\dot{Y}_{it-1}]_{N\times T}$, $\dot{X}_p = [\dot{X}_{itp}]_{N\times T}$ for $p = 1, 2, \ldots, k$, $M_{\boldsymbol{K}} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}\boldsymbol{K}(\boldsymbol{K}'\Sigma_{ee}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\Sigma_{ee}^{-1}$ and $M_{\dot{G}} = I_T - \dot{G}(\dot{G}'\dot{G})^{-1}\dot{G}'$. Let $\Delta_{1,NT}$, $\Delta_{2,NT}$, $\Delta_{3,NT}$, $\Delta_{4,NT}$ and $\Delta_{5,NT}$ be all $(k+1)$-dimensional vectors with their pth element defined sequentially as

$$\Delta_{1,p,NT} = \begin{cases} \frac{1}{N}\mathrm{tr}\Big[(\dot{G}'\dot{G})^{-1}\dot{G}'E(e'\Sigma_{ee}^{-1}e)M_{\dot{G}}\dot{Y}_{-1}'\Sigma_{ee}^{-1}\boldsymbol{K}(\boldsymbol{K}'\Sigma_{ee}^{-1}\boldsymbol{K})^{-1}\Big] & \text{if } \ p = 1; \\ \frac{1}{N}\mathrm{tr}\Big[(\dot{G}'\dot{G})^{-1}\dot{G}'E(e'\Sigma_{ee}^{-1}e)M_{\dot{G}}\dot{X}_{p-1}'\Sigma_{ee}^{-1}\boldsymbol{K}(\boldsymbol{K}'\Sigma_{ee}^{-1}\boldsymbol{K})^{-1}\Big] & \text{if } \ p > 1; \end{cases}$$

$$\Delta_{2,p,NT} = \begin{cases} \frac{1}{T}\mathrm{tr}\Big[(\boldsymbol{K}'\Sigma_{ee}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\Sigma_{ee}^{-1}E(ee')M_{\boldsymbol{K}}\dot{Y}_{-1}\dot{G}(\dot{G}'\dot{G})^{-1}\Big] & \text{if } \ p = 1; \\ \frac{1}{T}\mathrm{tr}\Big[(\boldsymbol{K}'\Sigma_{ee}^{-1}\boldsymbol{K})^{-1}\boldsymbol{K}'\Sigma_{ee}^{-1}E(ee')M_{\boldsymbol{K}}\dot{X}_{p-1}\dot{G}(\dot{G}'\dot{G})^{-1}\Big] & \text{if } \ p > 1; \end{cases}$$

$$\Delta_{3,p,NT} = \begin{cases} \frac{1}{NT}\mathrm{tr}\Big[\dot{Y}_{-1}'M_{\boldsymbol{K}}\Sigma_{ee}^{-1}LM_{\dot{G}}\Big] & \text{if } \ p = 1; \\ \frac{1}{NT}\mathrm{tr}\Big[\dot{X}_{p-1}'M_{\boldsymbol{K}}\Sigma_{ee}^{-1}LM_{\dot{G}}\Big] & \text{if } \ p > 1; \end{cases}$$

$$\Delta_{4,p,NT} = \begin{cases} \mathrm{tr}\Big[(\dot{F}'\dot{F})^{-1}\dot{F}'\dot{Y}_{-1}'\Sigma_{ee}^{-1}\boldsymbol{K}\mathbf{E}_1'\mathcal{G}\Big] & \text{if } \ p = 1; \\ 0 & \text{otherwise} \end{cases}$$

$$\Delta_{5,p,NT} = \begin{cases} \frac{1}{N}\mathrm{tr}\Big[P_{G^\circ}E(\mathbf{U}_{-1}'\Sigma_{ee}^{-1}e)\Big] & \text{if } \ p = 1; \\ 0 & \text{otherwise} \end{cases}$$

where $L = [E(e_{it}^3)]_{N\times T}$ and $P_{G^\circ} = G^\circ(G^{\circ\prime}G^\circ)^{-1}G^{\circ\prime}$ with $G^\circ = [\mathbb{1}_T, G]$ and $\mathbb{1}_T$ is the T-dimensional vector with all its elements equal to 1.

With the above notations, we have the following theorem on the asymptotic representation of the QML estimator.

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

$$\Omega_{NT}(\widehat{\psi} - \psi) = \frac{1}{NT}\sum_{t=1}^T \begin{bmatrix} \dot{\mathbf{B}}_{t-1}'M_{\boldsymbol{K}}(e_t - \sum_{s=1}^T \pi_{st}e_s) + \mathbf{U}_{t-1}'\Sigma_{ee}^{-1}e_t \\ \dot{X}_t'M_{\boldsymbol{K}}(e_t - \sum_{s=1}^T \pi_{st}e_s) \end{bmatrix}$$
$$- \frac{1}{T}\Delta_{1,NT} - \frac{1}{N}\Delta_{2,NT} - \frac{1}{T}\Delta_{3,NT} + \frac{1}{N}\Delta_{4,NT} - \frac{1}{T}\Delta_{5,NT} + O_p(\delta_{NT}^{-3}) + o_p(\|\widehat{\psi} - \psi\|).$$

where $\mathbf{B}_t$, π_{st}, $M_{\boldsymbol{K}}$, $\mathbf{U}_t$, $\widetilde{Q}_t$, *and* $\widetilde{\mathcal{Q}}_t$ *are defined in Assumption F. The above expression can be alternatively written as*

$$\Omega_{NT}(\widehat{\psi} - \psi) = \frac{1}{NT}\sum_{t=1}^T \widetilde{\mathcal{Q}}_t'e_t + \frac{1}{NT}\sum_{t=1}^T \mathbf{U}_{t-1}'\Sigma_{ee}^{-1}e_t\ell_{k+1}$$

$$-\frac{1}{T}\Delta_{1,NT}-\frac{1}{N}\Delta_{2,NT}-\frac{1}{T}\Delta_{3,NT}+\frac{1}{N}\Delta_{4,NT}-\frac{1}{T}\Delta_{5,NT}+O_p(\delta_{NT}^{-3})+o_p(\|\widehat{\psi}-\psi\|),$$

where ℓ_{k+1} is the first column of the $(k+1)$-dimensional identity matrix.

Theorems 5.1 presents the asymptotic representation of the QML estimator. The QML estimator has five bias terms. The first bias term $\frac{1}{T}\Delta_{1,NT}$ arises from the temporal heteroskedasticity[②]. If the errors are homoskedastic over the time, the first term is gone. The second term $\frac{1}{N}\Delta_{2,NT}$ is related to the cross sectional correlation. If the errors are uncorrelated over cross section, the second term disappears. Since we estimate the cross sectional heteroskedasticity simultaneously with other parameters, the QML estimator is immune to the bias arising from this heteroskedasticity. The first two bias terms can be found in Bai (2009) and Moon and Weidner (2017). However, our bias formulas are different from theirs because our formulas are heteroscedasticity-adjusted. The third term $\frac{1}{T}\Delta_{3,NT}$ is $O_p(\frac{1}{T})$, which is due to the non-zero skewness of errors. Note that we have two sets of cross sectional incidental parameters, the factor loadings and cross sectional heteroskedasticity. The asymptotic representation of the former estimators is a linear expression of errors and the leading term of the latter estimators involves a quadratic form of errors. The interactions of the estimators of these two types give rise to the skewness of errors. If e_{it} is strictly stationary over time, $E(e_{it}^3)$ is a constant over t for each i and the matrix L can be written as $a\mathbb{1}_T'$ for some N-dimensional vector a. Since $\mathbb{1}_T'M_{\dot{G}}\dot{Z}'=0$ for $Z=Y_{-1}$ or X_p. So this term disappears in this special case. The fourth term $\frac{1}{N}\Delta_{4,NT}$ arises because $\dot{Y}_{-1}$ does not have h_t as the factors. Note that X_p has h_t as factors. It can be shown, due to the fact $\widehat{\mathcal{G}}_{12}=O_p(\frac{1}{N^2})$ where $\widehat{\mathcal{G}}_{12}$ is the $r_1\times r_2$ upper-right subblock of $\widehat{\mathcal{G}}=(\widehat{\Sigma}_{ff}^{-1}+\widehat{\Lambda}'\widehat{\Phi}^{-1}\widehat{\Lambda})^{-1}$, that $\Delta_{4,NT}=O(\frac{1}{N})$, which implies $\frac{1}{N}\Omega_{NT}^{-1}\Delta_{4,NT}=O(\frac{1}{N^2})$, a term negligible under $N/T\to c^*\in(0,\infty)$. So the regressor X_p does not cause this bias term. The last term is due to the presence of the predetermined regressor Y_{it-1} and the within-group transformation. If the interactive effects is absent, this term is equal to $\mathrm{tr}[P_TE(\mathbf{U}_{-1}'\Sigma_{ee}^{-1}e)]$ where $P_T=\frac{1}{T}\mathbb{1}_T\mathbb{1}_T'$, a result which has been found in classical dynamic panel data studies, see Alvarez and Arellano (2003). The same bias term can be found in Moon and Weidner (2017)

Theorem 5.2 *Under Assumptions A-F, as $N,T\to\infty$ and $N/T\to c^*\in(0,\infty)$, we have*

$$\sqrt{NT}\Big(\widehat{\psi}-\psi^*+\Omega_{NT}^{-1}\Big[\frac{1}{T}\Delta_{1,NT}+\frac{1}{N}\Delta_{2,NT}+\frac{1}{T}\Delta_{3,NT}-\frac{1}{N}\Delta_{4,NT}+\frac{1}{T}\Delta_{5,NT}\Big]\Big)\xrightarrow{d}\mathcal{N}\Big(0,\overline{\Omega}^{-1}\widetilde{\Omega}\overline{\Omega}^{-1}\Big),$$

where $\overline{\Omega}=\lim_{N,T\to\infty}\Omega_{NT}$ and $\widetilde{\Omega}=\operatorname{plim}_{N,T\to\infty}\frac{1}{NT}\sum_{t=1}^{T}\mathcal{Q}_t'E(e_te_t')\mathcal{Q}_t$ with $\mathcal{Q}_t=M_{\boldsymbol{K}}(W_t-\sum_{s=1}^{T}\pi_{st}W_s)$ and $W_t=[\dot{Y}_{t-1},\dot{X}_t]$.

Theorems 5.1 and 5.2 regard to our dynamic panel data model. If the model is static, we can show that the bias terms $\Delta_{4,NT}$ and $\Delta_{5,NT}$ would disappear from the asymptotic representation. This gives the following asymptotic results on the QML estimator for the static panel data model.

Corollary 5.1 *Suppose that the model is $Y_{it}=\alpha_i+X_{it}'\beta+\kappa_i'g_t+e_{it}$ in (2.1). Under Assumptions A-F, as $N,T\to\infty$, we have*

$$\Omega_{NT}^{\star}(\widehat{\beta}-\beta)=\frac{1}{NT}\sum_{t=1}^{T}\dot{X}_t'M_{\boldsymbol{K}}(e_t-\sum_{s=1}^{T}\pi_{st}e_s)-\frac{1}{T}\Delta_{1,NT}^{\star}-\frac{1}{N}\Delta_{2,NT}^{\star}-\frac{1}{T}\Delta_{3,NT}^{\star}+O_p(\delta_{NT}^{-3})+o_p(\|\widehat{\psi}-\psi\|),$$

[②] The serial correlation is precluded by Assumption C.1 due to the presence of lagged dependent variable.

where $\Omega^\star_{NT}$ is the lower right $k \times k$ submatrix of Ω_{NT} and $\Delta^\star_1, \Delta^\star_2$ and $\Delta^\star_3$ are the lower k-dimensional subvectors of Δ_1, Δ_2 and Δ_3, respectively. The above asymptotic representation implies

$$\sqrt{NT}\Big(\widehat{\beta} - \beta + \Omega^{\star -1}_{NT}\Big[\frac{1}{T}\Delta^\star_{1,NT} + \frac{1}{N}\Delta^\star_{2,NT} + \frac{1}{T}\Delta^\star_{3,NT}\Big]\Big) \xrightarrow{d} N\Big(0, \overline{\Omega}^{\star -1}\widetilde{\Omega}^\star\overline{\Omega}^{\star -1}\Big),$$

where $\widetilde{\Omega}^\star = \operatorname*{plim}_{N,T\to\infty} \frac{1}{NT}\sum_{t=1}^T \mathcal{X}_t' E(e_t e_t')\mathcal{X}_t$ with $\mathcal{X}_t = M_{\boldsymbol{K}}(\dot{X}_t - \sum_{s=1}^T \pi_{st}\dot{X}_s)$.

Theorems 5.1 and 5.2 and Corollary 5.1 are derived under the assumptions to allow general correlations and heteroskedasticities structure of the idiosyncratic errors. These assumptions are very appealing in practical applications, however, it generates a nontrivial issue on consistent estimation of the biases and the limiting variance matrices. For example, the second bias term Δ_2 involves $\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N \frac{1}{\sigma_i^2\sigma_j^2}\kappa_i\kappa_j \sum_{t=1}^T E(e_{it}e_{jt})$, which cannot be consistently estimated by the plug-in estimator $\frac{1}{NT}\sum_{i=1}^N\sum_{j=1}^N \frac{1}{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}\widehat{\kappa}_i\widehat{\kappa}_j\sum_{t=1}^T \widehat{e}_{it}\widehat{e}_{jt}$. As remarked in Bai and Ng (2006), this problem is analogous to inconsistency of the unweighted sum of sample autocovariances as a long run variance estimator in the time series context. The widely-used kernel method in time series, which puts small weight on distant correlations, is inapplicable to estimation of the cross-sectional correlations since there is no natural ordering on the cross sectional units. Similar issue also exists in estimation of the covariance matrix $\widetilde{\Omega}$. To deal with this issue, we further make interchangeable Assumptions G and G$'$. We emphasize that Assumptions G and G$'$ are only needed for the statistical inference.

Assumption G: e_{it} is uncorrelated with e_{jt} for $j \neq i$.

Assumption G$'$: e_t is a strictly stationary process.

Assumption G assumes no correlations over cross section. As pointed out by Bai and Ng (2006), this assumption is not especially restrictive in the presence of factor structure since much of the cross-correlations is presumably captured by the common factors. Under this assumption, we only need to estimate the heteroskedasticity parameters over cross section and time, as well as the skewness of errors. Alternatively, we can assume Assumption G$'$ and estimate the biases and the covariance matrices. Bai and Ng (2006) make this assumption and propose the cross-section heteroskedastic autocorrelation consistent (CS-HAC) estimator for the limiting covariance matrix in a factor-augmented regression model. The intuition is as follows. With the strictly stationarity assumption, the cross-sectional correlation for each (i, j) pair can be consistently estimated by the observations over time. The estimation precision is proportional to the time periods T. Suppose that we only choose partial cross sectional units to construct the estimator, say only n units. As long as n is carefully chosen to be divergent not so fast, the total estimation errors will be under control for large T.

Let $\widehat{\theta} = (\widehat{\psi}, \widehat{\Lambda}, \widehat{M}_{ff}, \widehat{\Phi})$ be the QML estimators, which are obtained via the ECM algorithm stated below. The factor $\dot{f}_t$ can be consistently estimated (up to a rotation) by $\widehat{f}_t = \widehat{\Sigma}_{ff}\widehat{\Lambda}'\widehat{\Sigma}_{zz}^{-1}\widehat{\mathcal{B}}\dot{z}_t(\widehat{\rho})$ for each t, where $\widehat{\Sigma}_{zz} = \widehat{\Lambda}\widehat{M}_{ff}\widehat{\Lambda}' + \widehat{\Phi}$ and $\widehat{\mathcal{B}}$ is the estimator for $\mathcal{B}$ by replacing β with $\widehat{\beta}$. Once $\widehat{F}$ is obtained, the estimator of $\dot{G}$ (i.e., $\widehat{G}$) can be easily constructed by picking out the first r_1 columns. Let $\widehat{M}_{\boldsymbol{K}} = \widehat{\Sigma}_{ee}^{-1} - \widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}}(\widehat{\boldsymbol{K}}'\widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}})^{-1}\widehat{\boldsymbol{K}}'\widehat{\Sigma}_{ee}^{-1}$, where $\widehat{\Sigma}_{ee} = \operatorname{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_N^2)$ and $\widehat{\boldsymbol{K}}$ is implicitly given in $\widehat{\Lambda}$ and $M_{\widehat{G}} = I_T - \widehat{G}(\widehat{G}'\widehat{G})^{-1}\widehat{G}'$. Furthermore, let $\widehat{e}_{it} = \dot{Y}_{it} - \widehat{\rho}\dot{Y}_{it-1} - \dot{X}_{it}'\widehat{\beta} - \widehat{\kappa}_i'\widehat{g}_t$. With the above estimators, we now discuss the estimation of $\Delta_{1,NT}, \ldots, \Delta_{5,NT}$. Basically, the estimators for these five vectors are based on the plug-in method except for some cases in which long-run variance estimation issue exists. We only elaborate the estimator of the first elements of these vectors. The estimators of the remaining elements either are zeros or simply replace Y_{-1} with X_{p-1} according to the bias formula given above. For $\Delta_{1,1,NT}$, it can be estimated by

$$\widehat{\Delta}_{1,1,NT} = \mathrm{tr}\Big[(\widehat{G}'\widehat{G})^{-1}\widehat{G}'\mathrm{diag}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i1}^2, \ldots, \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{iT}^2\Big)M_{\widehat{G}}\dot{Y}_{-1}'\widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}}(\widehat{\boldsymbol{K}}'\widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}})^{-1}\Big].$$

For $\Delta_{2,1,NT}$, it is equal to zero under Assumption G, and can be estimated under Assumption G′ by

$$\widehat{\Delta}_{2,1,NT} = \mathrm{tr}\Big[\Big(\frac{1}{N}\widehat{\boldsymbol{K}}'\widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}}\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{\widehat{\sigma}_i^2\widehat{\sigma}_j^2}\widehat{\kappa}_i\frac{1}{T}\sum_{t=1}^{T}\widehat{e}_{it}\widehat{e}_{jt}\Big(\dot{Y}_{j,-1}' - \widehat{\kappa}_j(\widehat{\boldsymbol{K}}'\widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}})^{-1}\widehat{\boldsymbol{K}}'\widehat{\Sigma}_{ee}^{-1}\dot{Y}_{-1}\Big)\widehat{G}(\widehat{G}'\widehat{G})^{-1}\Big]$$

For $\Delta_{3,1,NT}$, it can be estimated by $\frac{1}{NT}\mathrm{tr}\big[\dot{Y}_{-1}'M_{\widehat{\boldsymbol{K}}}\widehat{\Sigma}_{ee}^{-1}\widehat{L}M_{\widehat{G}}\big]$ with $\widehat{L} = [\widehat{e}_{it}^3]_{N\times T}$ under Assumption G, and is zero under Assumption G′. For $\Delta_{4,1,NT}$, it can be estimated by $\mathrm{tr}\big[(\widehat{F}'\widehat{F})^{-1}\widehat{F}'\dot{Y}_{-1}'\widehat{\Sigma}_{ee}^{-1}\widehat{\boldsymbol{K}}\mathbf{E}_1'\widehat{\mathcal{G}}\big]$. For $\Delta_{5,1,NT}$, it can be estimated by $\mathrm{tr}[P_{\widehat{G}^\circ}\mathrm{thr}_{n^*}(\widehat{J})]$, where $\mathrm{thr}_\gamma(\cdot)$ is the thresholding operation which sets the (t,s)th element of its arguments to zero if $|t-s| > n^*$, and $\widehat{G}^\circ = [\mathbb{1}_T, \widehat{G}]$ and $\widehat{J}$ defined as

$$\widehat{J} = \frac{1}{N}\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i1}^2 & 0 & 0 & \cdots & 0 \\ \widehat{\rho}\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i1}^2 & \sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i2}^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \widehat{\rho}^{T-2}\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i1}^2 & \widehat{\rho}^{T-3}\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i2}^2 & \widehat{\rho}^{T-4}\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{e}_{i3}^2 & \cdots & 0 \end{bmatrix}.$$

The matrix $\overline{\Omega}$ can be readily estimated by replacing $M_{\boldsymbol{K}}$ and $M_{\dot{G}}$ in Ω_{NT} with $M_{\widehat{\boldsymbol{K}}}$ and $M_{\widehat{G}}$, respectively. The matrix $\widetilde{\Omega}$ can be estimated by $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathcal{Q}}_{it}\widehat{\mathcal{Q}}_{it}'\widehat{e}_{it}^2$ under Assumption G, and $\frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\widehat{\mathcal{Q}}_{it}\widehat{\mathcal{Q}}_{jt}'\frac{1}{T}\sum_{s=1}^{T}\widehat{e}_{is}\widehat{e}_{js}$ under Assumption G′, where $\widehat{\mathcal{Q}}_{it}$ is the transpose of the i-th row of $\widehat{\mathcal{Q}}_t$, and $\widehat{\mathcal{Q}}_t$ is the matrix obtained by replacing $M_{\boldsymbol{K}}$ and π_{st} in $\mathcal{Q}_t$ with $M_{\widehat{\boldsymbol{K}}}$ and $\widehat{\pi}_{st} = \widehat{g}_t'(\widehat{G}'\widehat{G})^{-1}\widehat{g}_s$, respectively.

We have the following theorem shows consistency of the above estimators.

Theorem 5.3 *Under Assumptions A-F, together with Assumption G or G′, if $n \to \infty$ and $\frac{n^2}{\min(N,T)} \to 0$, we have $\widehat{\Delta}_{p,NT} - \Delta_{p,NT} = o_p(1)$ for $p = 1, 2, \ldots, 4$ and $\widehat{\Omega}_{NT} - \overline{\Omega} = o_p(1)$ and $\widehat{\widetilde{\Omega}}_{NT} - \widetilde{\Omega} = o_p(1)$. In addition, if $n^* \to \infty$ and $\frac{n^{*2}}{\min(N,T)} \to 0$, we have $\widehat{\Delta}_{5,NT} - \Delta_{5,NT} = o_p(1)$.*

6 Hypothesis testing

The FDI literature widely employs traditional dynamic panel data models to investigate the determinants of the FDI locations, see, for example, Cole, Elliott and Zhang (2009), Cheng and Kwan (2000), etc. It is well documented in the econometric literature that dynamic panel data models may suffer the celebrated incidental parameter issue of Neyman and Scott (1948). Anderson and Hsiao (1981, 1982) propose an IV method, in which one first differences the model over time and then use y_{it-2} as the instrument to the endogenous regressor. Arellano and Bond

(1991) generalize Anderson and Hasio's idea to the difference GMM. In the current framework, since we consider a large-T scenario, the GMM would entail the so-called many moments issue, see Alvarez and Arellano (2003) and Han and Phillips (2006). But one can still use Anderson and Hsiao's IV method in large-T setup. This evokes us to study the hypothesis testing on whether the IV proposed in Anderson and Hsiao's method is valid in our heterogeneous shocks model.

Consider the Y equation,

$$Y_{it} = \alpha_i + \rho Y_{it-1} + X_{it}'\beta + \kappa_i' g_t + e_{it}.$$

After first differencing, we obtain

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta X_{it}'\beta + \kappa_i' \Delta g_t + \Delta e_{it}.$$

For Anderson and Hsiao's method, one uses Y_{it-2} and ΔX_{it} as the instruments of ΔY_{it-1} and ΔX_{it}, respectively. In our heterogeneous shocks model, the errors in the Y equation have an additional term $\kappa_i' g_t$. However, if Δg_t is uncorrelated with Δh_t and $\mathbf{f}_{t-2}^{\rho}$, where $\mathbf{f}_t^{\rho}$ is the factors contained in Y_{it}, it is easy to verify that the orthogonal conditions in Anderson and Hsiao's method continue to hold and their method would deliver a consistent estimator.

Let $\mathcal{P}_t = (\Delta h_t', \mathbf{f}_{t-2}^{\rho\prime})'$. We are interested in testing $E(\Delta g_t \mathcal{P}_t') = 0$, which is equivalent to $E(\mathcal{P}_t \otimes \Delta g_t) = 0$. Note that the equation $E(\Delta g_t \mathcal{P}_t') = 0$ involves the unobserved factors, which can only be identified up to a rotation. But we emphasize that $E(\Delta g_t \mathcal{P}_t') = 0$ does not suffer the rotational indeterminacy and is well defined. This is due to the robustness of zero matrix to rotational transformation since $\mathcal{R}_1 0_{r_1 \times r_2} = 0_{r_1 \times r_2}$ and $0_{r_1 \times r_2}\mathcal{R}_2 = 0_{r_1 \times r_2}$, where $\mathcal{R}_1$ and $\mathcal{R}_2$ are two invertible $r_1 \times r_1$ and $r_2 \times r_2$ matrices.

To construct the statistic, a nature way is to replace Δg_t and $\mathcal{P}_t$ with their estimates, i.e., $\widehat{\Delta g_t}$ and $\widehat{\mathcal{P}}_t$. Unfortunately, since $\widehat{\Delta g_t}$ and $\widehat{\mathcal{P}}_t$ are non-zeros, they both suffer rotational indeterminacy. To address this issue, we consider the following sum of the square of canonical correlations between Δg_t and $\mathcal{P}_t$, which is defined as

$$\mathcal{C} = T\mathrm{tr}\Big[[E(\Delta g_t \Delta g_t')]^{-1} E(\Delta g_t \mathcal{P}_t')[E(\mathcal{P}_t \mathcal{P}_t')]^{-1} E(\mathcal{P}_t \Delta g_t')\Big].$$

An advantage of the above quantity is its immunity to the rotational indeterminacy. Our statistic is based on the sample analog of $\mathcal{C}$, i.e.,

$$\mathcal{C}_s = T\mathrm{tr}\Big[\Big(\sum_{t=3}^{T} \widehat{\Delta g}_t \widehat{\Delta g}_t'\Big)^{-1}\Big(\sum_{t=3}^{T} \widehat{\Delta g}_t \widehat{\mathcal{P}}_t'\Big)\Big(\sum_{t=3}^{T} \widehat{\mathcal{P}}_t \widehat{\mathcal{P}}_t'\Big)^{-1}\Big(\sum_{t=3}^{T} \widehat{\mathcal{P}}_t \widehat{\Delta g}_t'\Big)\Big], \tag{6.1}$$

where $\widehat{\Delta g}_t$ and $\widehat{\Delta h}_t$ can be obtained from the estimators $\widehat{g}_t$ and $\widehat{h}_t$. The latter two estimators can be directly obtained in the ECM algorithm stated in Section 7. The estimators of the factors $\mathbf{f}_t^{\rho}$, $\widehat{\mathbf{f}}_t^{\rho}$, can be obtained by invoking the ML estimation method proposed by Bai and Li (2012). Once the above estimators are obtained, the statistic $\mathcal{C}_s$ can be constructed by the formula (6.1). We have the following theorem on the asymptotic properties of $\mathcal{C}_s$.

Theorem 6.1 *Under Assumptions A-E, as $N, T \to \infty$ and $\sqrt{N}/T \to 0$, then*

- *under the null hypothesis* $H_0 : E(\Delta g_t \mathcal{P}_t') = 0$*, if* $f_t = [g_t', h_t']'$ *is an identical and independently distributed process, we have*

$$\mathcal{C}_s \xrightarrow{d} \frac{3}{2}\chi_a^2(r_1 r_2) + (1-\rho)\chi_b^2(r_1^2 + r_1 r_2);$$

where $\chi_a^2(r_1r_2)$ *and* $\chi_b^2(r_1^2 + r_1r_2)$ *are two independent chi-square distributed random variables with the degrees of freedom* r_1r_2 *and* $r_1^2 + r_1r_2$*, respectively.*

- *under the alternative hypothesis* $H_1 : E(\Delta g_t \mathcal{P}_t') \neq 0$*, we have* $\mathcal{C}_s \longrightarrow \infty$.

Theorem 6.1 implies that the statistic $\mathcal{C}_s$ converges in distribution to a weighted sum of chi-square distribution. Similar asymptotic results appear in likelihood ratio test on non-nested hypothesis, see Vuong (1989). The critical values for $\mathcal{C}_s$ can be obtained by the Monte Carlo method. More specifically, we generate $r_1^2 + 2r_1r_2$ standard normal variables. For each variable, we calculate its squared value. Next, we take a weighted sum of these squared values with the weight 1.5 for the first r_1r_2 values and the weight $1 - \widehat{\rho}$ for the remaining values, where $\widehat{\rho}$ is the QML estimator for ρ. We repeat these procedures with 1000 repetitions and sort the 1000 values with descending order. Then the critical values for 1%, 5% and 10% significance level are the 10th, 50th, 100th largest ones, respectively.

We note that the proposed test is only applicable under the dynamic scenario. For a static panel data model, one rarely uses the first differencing but within group transformation to eliminate the individual fixed effects. We can readily show that if $E(g_t h_t') = 0$, the traditional within group method would deliver the consistent estimation. This motivates us to test $E(g_t h_t') = 0$. Following the above idea, we propose the following statistic in static panel data model

$$\widetilde{\mathcal{C}}_s = T\mathrm{tr}\Big[\widehat{\Sigma}_{hg}\widehat{\Sigma}_{gg}^{-1}\widehat{\Sigma}_{gh}\widehat{\Sigma}_{hh}^{-1}\Big],$$

where $\widehat{\Sigma}_{gg}, \widehat{\Sigma}_{hh}, \widehat{\Sigma}_{gh}$ and $\widehat{\Sigma}_{hg}$ are implicitly defined by $\widehat{\Sigma}_{ff} = [\widehat{\Sigma}_{gg}, \widehat{\Sigma}_{gh} \mid \widehat{\Sigma}_{hg}, \widehat{\Sigma}_{hh}]$ and $\widehat{\Sigma}_{ff}$ is directly delivered by the ECM algorithm below. The following theorem provides the asymptotic results of $\widetilde{\mathcal{C}}_s$, whose proof is similar as (actually easier than) that of Theorem 5.2 and is therefore omitted.

Theorem 6.2 *Under Assumptions A-E, as* $N, T \to \infty$ *and* $\sqrt{N}/T \to 0$*, then*

- *under the null hypothesis* $H_0 : E(g_t h_t') = 0$*, if* $f_t = [g_t', h_t']'$ *is an identical and independently distributed process, we have*

$$\widetilde{\mathcal{C}}_s \xrightarrow{d} \chi^2(r_1 r_2);$$

- *under the alternative hypothesis* $H_1 : E(g_t h_t') \neq 0$*, we have* $\widetilde{\mathcal{C}}_s \longrightarrow \infty$.

7 Computing method

In this section we discuss the computation method. The QML estimators can be calculated via the ECM algorithm. Let $\theta^{(k)} = (\psi^{(k)}, \Lambda^{(k)}, \Phi^{(k)}, \Sigma_{ff}^{(k)})$ be the estimated value at the kth iteration, where Σ_{ff} is the variance of $f_t = (g_t', h_t')'$. In our ECM algorithm, we do not impose the identification conditions. One reason is that the identification conditions are imposed to uniquely determine the underlying parameters. Hence, it has no effects on the minimum value of the likelihood function. Another reason is that once we impose the restrictions such as $\Sigma_{gg} = I_{r_1}$ and $\Sigma_{hh} = I_{r_2}$, the elements of Σ_{gh} are implicitly restricted in $[-1, 1]$. But such restrictions cannot be guaranteed in the ECM iterating formulas.

Let $\mathcal{G}^{(k)} = \left[(\Sigma_{ff}^{(k)})^{-1} + \Lambda^{(k)\prime}(\Phi^{(k)})^{-1}\Lambda^{(k)}\right]^{-1}$. We first calculate

$$\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | Z, \theta^{(k)}) = \mathcal{G}^{(k)} + \mathcal{G}^{(k)}\Lambda^{(k)\prime}(\Phi^{(k)})^{-1}\mathcal{B}^{(k)}\frac{1}{T}\sum_{t=1}^{T} z_t(\rho^{(k)})z_t(\rho^{(k)})'\mathcal{B}^{(k)\prime}(\Phi^{(k)})^{-1}\Lambda^{(k)}\mathcal{G}^{(k)},$$

$$\frac{1}{T}\sum_{t=1}^{T} E(\mathcal{B}z_t(\rho) f_t' | Z, \theta^{(k)}) = \mathcal{B}^{(k)}\frac{1}{T}\sum_{t=1}^{T} z_t(\rho^{(k)})z_t(\rho^{(k)})'\mathcal{B}^{(k)\prime}(\Phi^{(k)})^{-1}\Lambda^{(k)}\mathcal{G}^{(k)},$$

where $\mathcal{B}^{(k)}$ is $\mathcal{B}$ when $\beta = \beta^{(k)}$. Once we obtain $\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | Z, \theta^{(k)})$, matrices $\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}(g_t g_t' | Z, \theta^{(k)})$, $\frac{1}{T}\sum_{t=1}^{T} E(g_t h_t' | Z, \theta^{(k)})$ and $\frac{1}{T}\sum_{t=1}^{T} E(h_t h_t' | Z, \theta^{(k)})$ are all known. $\boldsymbol{K}$ and Γ are updated according to

$$\boldsymbol{K}^{(k+1)} = \Upsilon_1^{(k)}\Big[\frac{1}{T}\sum_{t=1}^{T} E(g_t g_t' | Z, \theta^{(k)})\Big]^{-1}, \text{ and } \Gamma^{(k+1)} = \Upsilon_2^{(k)}\Big[\frac{1}{T}\sum_{t=1}^{T} E(h_t h_t' | Z, \theta^{(k)})\Big]^{-1},$$

where $\Upsilon_1^{(k)}$ is $N \times r_1$, whose ith row is the first r_1 subvector of the $[(i-1)(K+1)+1]$th row of $\frac{1}{T}\sum_{t=1}^{T} E(\mathcal{B}z_t(\rho) f_t' | Z, \theta^{(k)})$; and $\Upsilon_2^{(k)}$ is $NK \times r_2$, which is obtained by deleting the rows of $\frac{1}{T}\sum_{t=1}^{T} E(\mathcal{B}z_t(\rho) f_t' | Z, \theta^{(k)})$ corresponding to $\Upsilon_1^{(k)}$ and the first r_1 columns. Once $\boldsymbol{K}^{(k+1)}$ and $\Gamma^{(k+1)}$ are calculated, $\Lambda^{(k+1)}$ is obtained accordingly. We then calculate

$$\begin{aligned}\Phi^{(k+1)} = \mathrm{Dg}\Bigg\{&\mathcal{B}^{(k)}\frac{1}{T}\sum_{t=1}^{T} z_t(\rho^{(k)})z_t(\rho^{(k)})'\mathcal{B}^{(k)\prime} + \Lambda^{(k+1)}\Big[\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | Z, \theta^{(k)})\Big]\Lambda^{(k+1)\prime} \\ &- \Lambda^{(k+1)}\Big[\frac{1}{T}\sum_{t=1}^{T} E(f_t z_t(\rho)'\mathcal{B}' | Z, \theta^{(k)})\Big] - \Big[\frac{1}{T}\sum_{t=1}^{T} E(\mathcal{B}z_t(\rho) f_t' | Z, \theta^{(k)})\Big]\Lambda^{(k+1)\prime}\Bigg\},\end{aligned}$$

$$\psi^{(k+1)} = \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{(k+1)^2}}\dot{W}_{it}\dot{W}_{it}'\Big]^{-1}\Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^{(k+1)^2}}\dot{W}_{it}(\dot{Y}_{it} - \kappa_i^{(k+1)\prime}g_t^{(k)})\Big],$$

where $\mathrm{Dg}(\cdot)$ denotes the operator which sets the element of the input argument to zero if the counterpart of Φ is specified to zero; $(\sigma_i^{(k+1)})^2$ is the $(i-1)(K+1)+1$th diagonal element of Φ; and $\dot{W}_{it} = (\dot{Y}_{it-1}, \dot{X}_{it}')'$. $g_t^{(k)}$ is the first r_1 subvector of

$$f_t^{(k)} = \mathcal{G}^{(k)}\Lambda^{(k)\prime}(\Phi^{(k)})^{-1}\mathcal{B}^{(k)}\dot{z}_t(\rho^{(k)}).$$

Finally, the $\Sigma_{ff}^{(k+1)}$ is updated by

$$\Sigma_{ff}^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T} E(f_t f_t' | Z, \theta^{(k)}). \tag{7.1}$$

Putting everything together, we obtain $\theta^{(k+1)} = (\psi^{(k+1)}, \Lambda^{(k+1)}, \Phi^{(k+1)}, \Sigma_{ff}^{(k+1)})$. The iteration goes on until $\|\psi^{(k+1)} - \psi^{(k)}\|$ is smaller than some tolerance value. As regard to the initial values, we have three choices, that is, the within-group estimator, the GMM estimator, and the PC estimator of Moon and Weidner (2017). Theoretically, the PC estimator is the best choice because it is the consistent for the unknown coefficients. However, the main issue related with the PC initial value is the multiple local maximas (minimas) of the objective function. If the likelihood objective function has this issue, it is likely that the objective function of the PC method would have the same one. Hence, if the PC estimator is unfortunate to be the local minima, which is possible as pointed out by Moon and Weidner (2018), such an initial value would possibly make the QML estimator the local maxima. For this reason, only using the PC estimator as the initial value is not enough to guarantee the desirable estimator. We suggest using all the three estimators as initial values and choose the estimator, if different, that corresponds to the largest likelihood value, to be the final QML estimator.

8 Simulations

We conduct Monte Carlo simulations to investigate the finite sample performance of the ML estimators and the proposed $\mathcal{C}_s$ test. The underlying data generating process is

$$\begin{aligned} Y_{it} &= \alpha_i + \rho Y_{it-1} + X_{it}\beta + \kappa_i g_t + e_{it}, \\ X_{it} &= \nu_i + \gamma_{i1} h_{t1} + \gamma_{i2} h_{t2} + v_{it}. \end{aligned} \tag{8.1}$$

We set $\rho = 0.8$ and $\beta = 1$. The above specification indicates that $r_1 = 1$ and $r_2 = 2$. The factor loadings are generated by $\kappa_i = N(0,1)$, $\gamma_{i1} = \kappa_i + N(0,1)$ and $\gamma_{i2} = 2\kappa_i + N(0,1)$. The factors are generated by $h_{t1} = g_t + N(0,1)$ and $h_{t2} = g_t + N(0,1)$ with g_t drawn from $N(0,1)$. Other parameters, such as α_i and ν_i, are generated from $N(0,1)$. The idiosyncratic errors of X is generated by $\mathcal{A}_v^{1/2}\mathcal{E}_v\mathcal{B}_v^{1/2}$, where $\mathcal{E}_v$ is an $N \times T$ matrix with each element being independent $N(0,1)$. $\mathcal{A}_v$ is an $N \times N$ Toeplitz matrix whose the first diagonal elements are 1, the second diagonal elements are 0.48, and the remaining elements are zeros. $\mathcal{B}_v$ is defined similarly as $\mathcal{A}_v$ except that the second diagonal elements are -0.48. The idiosyncratic errors of the Y equation (i.e., e_{it}) is generated according to $\mathcal{A}_e\mathcal{E}_e\mathcal{B}_e$, where $\mathcal{E}_e$ is an $N \times T$ random matrix with each element being independent $(\chi^2(2) - 2)/2$. To investigate the performance of different data scenarios, we consider three choices of $\mathcal{A}_e$ and $\mathcal{B}_e$:

Case One: $\mathcal{A}_e$ and $\mathcal{B}_e$ are both identity matrices.

Case Two: $\mathcal{A}_e$ is an N-dimensional diagonal matrix with its i-th diagonal element equal to $0.1 + \frac{1}{1-u_i}\kappa_i^2 u_i$ where $u_i \sim U[0.3, 0.7]$, and $\mathcal{B}_e$ is an identity matrix.

Case Three: $\mathcal{A}_e$ is defined the same as above, and $\mathcal{B}_e$ is an T-dimensional diagonal matrix with its s-th diagonal element being $0.1 + \frac{1}{1-v_s}g_s^2 v_s$ where $v_s \sim U[0.3, 0.7]$.

The simulations of this section focus on the comparison of the ML and PC methods, where the PC estimator for the dynamic interactive-effects model is given in Moon and Weidner (2017). Chudik and Pesaran (2015) propose the CCE method to estimate a dynamic model. Their paper focuses much on the heterogeneous coefficients. Although it is possible to adapt their method to the setup of homogeneous coefficient, the asymptotic properties of the CCE estimator are much unknown. For this reason, we do not include their method for comparison. Throughout the simulations, we assume that r_1 and r_2 are known. Since both the QML and PC methods need to know the number of factors, such an assumption does not impair the fairness of comparison. In practice, this assumption is implausible, but the values r_1 and r_2 can be consistently estimated by a number of celebrated methods, e.g., Bai and Ng (2002), Onatski (2010), Ahn and Horenstein (2013). Moreover, as pointed out by Moon and Weidner (2015), the estimation of the regression coefficients is little affected by the number of factors as long as it is not underestimated.

Tables 1 - 3 present the simulation results on the ML and PC estimators, which are obtained from 1000 repetitions. There are several points worthy to emphasize. First, the bias issue in the dynamic models is nontrivial. If ignored, the t-test would suffer severe size distortion. Take ρ as the example to illustrate. In the sample size of $N = 150$ and $T = 50$, the actual coverage probabilities based on the ML estimator is 37.3% in case one, 70.0% in case two, and 64.4% in case three, which are far away from the nominal 95% coverage probability. The results based on the PC estimator exhibit the similar observations. Second, the bias correction formula can effectively remove the biases of the estimators. In all the combinations of N and T, we see that the magnitudes of the biases are appreciably reduced. This result is reflected in the performance of the t-test. As can be seen, the empirical coverage probabilities are much improved. In the sample when N and T are both large, the empirical probabilities are close to 95%, the nominal one. Third, the standard deviations of the ML estimator is much smaller than those of the PC in the presence of cross-sectional heteroskedasticity, which implies that the power of the t-test based on the ML estimator would be larger than the PC estimator. The result indicates that the ML estimator should be preferred when the data exhibits heavy heteroskedasticity.

Insert Tables 1 and 3

We next evaluate the empirical sizes and powers of the proposed test. All parameters except the factors are generated by the same way as the above. The factors are generated by the following way. First, we generate error $\varepsilon_t^f = [\varepsilon_t^g, \varepsilon_t^{h\prime}]'$, where ε_t^g is drawn independently from $N(0,1)$ and $\varepsilon_t^h = (\varepsilon_{t,1}^h, \varepsilon_{t,2}^h)$ with $\varepsilon_{t,1}^h$ and $\varepsilon_{t,2}^h$ being generated by $\varepsilon_{t,1}^h = \frac{c}{\sqrt{T}}\varepsilon_t^g + N(0,1)$ and $\varepsilon_{t,2}^h = \frac{c}{\sqrt{T}}\varepsilon_t^g + N(0,1)$. Once ε_t^f is obtained, the factors f_t are generated by $f_t = \frac{c}{\sqrt{T}} f_{t-1} + \varepsilon_t^f$. It is easy to see that when $c = 0$, the factor f_t satisfies the null hypothesis; if $c \neq 0$, the factor f_t represents the local Pitman alternative case and can be used to study the power of the test. According to Theorem 6.1, the final limiting distribution under the null is $1.5\chi_a^2(2) + 0.2\chi_b^2(3)$. We use simulation method to obtain the 90%, 95% and 99% quantiles of the distribution, which are 7.46, 9.62, and 10.07, respectively.

Table 4 presents the simulation results of the $\mathcal{C}_s$ test, which are obtained by 1000 repetitions. As seen, the $\mathcal{C}_s$ test performs well in all the combinations of N and T. When $c = 0$, the empirical sizes of the $\mathcal{C}_s$ test are close to the nominal levels. Only when N and T are both relatively small, say $N = 50, T = 50$, the $\mathcal{C}_s$ has a slight size distortion. However, when the sample size becomes larger, the size of test improves much. In addition, we see that the $\mathcal{C}_s$ test has good power. For the case $c = 3$, we see that one has a low chance to make type-*II* error under 95% nominal level.

Insert Table 4

9 The Location Determinants of China's FDI Inflow

9.1 Data and variables

Data

In this study, we employ a balanced panel data covering 29 provinces, autonomous regions and municipalities ("provinces" henceforth) over a 22-year period from 1993 to 2014. The 29 provinces included in our sample are AnHui, BeiJing, FuJian, GanSu, GuangDong, GuangXi, GuiZhou, HaiNan, HeBei, HeiLongJiang, HeNan, HuBei, HuNan, JiangSu, JiangXi, JiLin, LiaoNing, NeiMeng, NingXia, QingHai, ShanDong, ShangHai, ShanXi, ShaanXi, SiChuan, TianJin, XinJiang, YunNan, and ZheJiang, excluding Chongqing and Tibet. We confine our sample period from 1993 to 2014 since Chinese FDI inflow increased dramatically from 1993 after Deng Xiaoping's South Tour. Moreover, the Asian Financial Crisis and the Subprime Crisis are covered during this period that creates a typical context with heterogeneous shocks to match our model. All the data are obtained from the regional database of the National Bureau of Statistics of China. These data have also been used in previous FDI studies as aggregate level control variables (e.g. Kang and Lee (2007), Du, Lu and Tao (2008)).

Variable Constructions and Summary Statistics

Dependent variable. To measure the aggregate size of FDI inflow, we use the annual total investment of Foreign-invested enterprises (FIEs) at province level. All the investment are converted from US dollars into RMB with time-varying exchange rate and deflated with local Consumer Price Index (henceforth CPI) taking 2005 as the base year.

Explanatory variables.

Local GDP. According to traditional FDI theory, an incentive for firms seeking to produce abroad is to access local market of host country. Thus, the greater the local market size, the more likely it will attract foreign investment. Following Kang and Lee (2007), we include deflated provincial GDP to proxy for the size of local market, and we expect that it has a positive effect on FDI.

Infrastructure. The availability and quality of local infrastructure has a positive impact in attracting FDI, which is a particularly important factor impacting the location decision for firms engaging in cross-border production sharing (Blude and Molina (2015)). We take the sum of the length of road and railway over the area and calculate the density of road and railway to measure the availability of infrastructure in various provinces.

Wage. China is a labor-abundant country. Low labor costs make China an ideal destination for FDI, especially those taking China as an international fragmented production base and re-exporting platform. However, the rapid growth rate of wage would make the firms' cost management difficult, causing negative effects on attracting FDI. We use the growth rate of logarithm of the deflated average wage of urban employee to measure this wage cost.

Openness. In the early stage, China initiated the strategies to open up Special Economic Zones (SEZs), Open Coastal Cities (OCCs), and Economic and Technological Development Zones (ETDZs) and implement preferential policies towards FDI. Kang and Lee (2007) and Du, Lu and Tao (2008) find government preferential policies increase significantly the attractiveness of certain area for foreign investors. Here we add a variable Openness, which is measured by the ratio of total import and export over GDP, to control for the effect of these preferential policies.[③]

Government. As an economy during transition, China's various levels of governments still play substantial roles in economic activities. Hence we include Government measured with the ratio of local government fiscal expenditure over GDP to control for the possible impact. In existing literature, the expected sign of Government is mixed. A positive argument states that government spending contributes to aggregate demand (e.g. Blanc-Brude et al. (2014)) and state budgetary appropriation as a substitute for bank loans provides finance to credit constrained enterprises. In contrary, the negative argument emphasizes government's intervention in business operation and the induced corruption and unequal competition among different types of enterprises (e.g., Du, Lu and Tao (2008)).

Labor quality. FDI firms usually have technology and knowledge asset advantages, so they require better labor quality to adapt their technology and knowledge. Therefore, the abundance of skilled labor, as a specific resource, could be a local advantage for destination countries to attract FDI inflows (UNCTAD (1998), Zhang and Markusen (1999)). In the existing literature (e.g. Noorbakhsh and Paloni (2001), Gao (2005), Du, Lu and Tao (2008)), the percentage of educated persons is used to indicate the local human resources availability and labor quality. During our sample period, China also experienced a sharp increase in education level, which may contribute to its dramatic growth in FDI inflows. Thus, a positive sign of labor quality is expected.

We note that, exchange rate, tariff and taxes are also important FDI determinants affecting FDI location and magnitude (See Blonigen (2005) for a literature review). However, we study the FDI location choice across Chinese provinces, which are assumed to face national uniform exchange rate, tariffs and taxes. Further tariff or tax reduction or exemption in some special areas as preferential policies has been taken into account in constructing the variable Openness.

Variable definition and the descriptive statistics are summarized in Tables 5 and 6.

Insert Tables 5 and 6

9.2 Empirical results

As the first step, the number of factors r_1 and r_2, the numbers of the global shocks and the domestic shocks, need to be determined. On this matter, we have at hand a number of popular methods to use, e.g. Bai and Ng (2002), Onatski (2010) and Ahn and Horenstein (2013), to name a few. Existing studies, such as Boivin, Giannoni and Mihov (2009), suggest determining the number of factors based on economic analysis. In our FDI study, we find that the empirical results are sensitive to r_1, but not sensitive to r_2. So we choose the existing methods to estimate r_1. More concretely, we consider three popular methods, i.e., the information criterion (IC) method of Bai and Ng (2002), the edge distribution (ED) method of Onatski (2010) and the eigenvalue-ratio (ER) and growth-ratio (GR) methods of Ahn and Horenstein (2013). As regards to the IC method, as pointed out by Boivin and Ng (2006) and Onatski (2012), it tends to select a relatively

[③] In Kang and Lee (2007), Du, Lu and Tao (2008), Liu, Lovely and Ondrich (2010), they add dummy variables to control for the preferential policies. However, in our model, the effect of a time-invariant dummy variable will be included in the constant. Therefore, we use the openness of local economy, as a result of various preferential policies, to control for such an effect.

large number of factors in the limited sample size of real data applications. Li (2018) proposes a modified IC method by imposing a heavier penalty to avoid the overestimation issue. Since our sample size is $N = 29, T = 22$, which is a typical case that Li (2018) considers, we therefore use her method, instead of the original Bai and Ng's method. In addition, we note that ED, ER and GR methods are designed for the approximate factor models. They cannot be directly applied to the current model. To address this issue, we conduct the following procedure: first assume $\widehat{r}_1$ and estimate the regression coefficients, then calculate the corresponding residuals of the Y equation, and see whether this assumed number of factors in the residuals is supported by the ED, ER and GR methods. In all the methods, we set $r_{\max} = 4$.

The modified IC method and ED methods both suggest $\widehat{r}_1 = 1$. The modified IC values are -0.567 for $r_1 = 0$; -3.726 for $r_1 = 1$; -2.388 for $r_1 = 2$; -0.921 for $r_1 = 3$; and 0.541 for $r_1 = 4$, which suggest $\widehat{r}_1 = 1$ to be the estimated value. The ED method also suggests $r_1 = 1$. But the ER and GR methods fail to deliver the number of factors, i.e., the estimated number of factors are not the same as the value assumed a priori. Based on these results, we set $\widehat{r}_1 = 1$, which means one general global shock influencing the source of FDI. We note that one global shock is consistent with the specification of Kose et al. (2003) in which the authors study global economic circle. As for r_2, we experiment with different values and choose the smallest value r under which the estimation results are almost the same with those under $r + 1$[④].

Table 7 presents the empirical results for the five estimation methods, i.e., difference GMM, system GMM, Anderson and Hsiao's IV, PC, and QML methods. Retrospectively, the previous studies generally use the dynamic panel data model to study the FDI locations, so we calculate the results of difference GMM, system GMM, and Anderson and Hsiao's IV methods for the purpose of comparion[⑤]. In addition, the PC estimator is a primary competitor of our QML estimator, so we also include it for consideration. The sixth column to the ninth column of Table 7 are the results from our QML method. The sixth column is the result that we mainly rely on. The next three columns are used to check the robustness.

From table 7, we can draw the following conclusions. First, the QML method delivers impressive results. As seen in the sixth column, all the regression coefficients have correct signs and are statistically significant. More specifically, large inward FDI in last period, larger local market, broader infrastructure stock, lower growth rate of labor cost, a higher level of openness, more government intervention and better human capital availability attract more FDI. Second, the four other methods do not give very satisfactory results. For example, all the other methods suggest positive signs for the coefficient of WAGE. However, we emphasize that for these four methods, if ignoring those coefficients statistically insignificant, we find that the remaining ones are consistent with what the economic theory predicts. For example, the coefficient of WAGE, albeit positive in the columns of the second to the fifth, are insignificant. Third, our QML estimation results have their robustness. Since labor quality is closely related with wage, we delete LABOR in the eighth column, but find that the coefficients still have correct signs and statistically significant. In addition, since market, wage and openness are the three primary factors that drive the volumes of FDI, we consider a small regression with only these three regressors. Again, we find

[④]We also use the IC, ED, ER and GR methods to estimate r_2. But different methods give different values. The modified IC method estimates $\widehat{r}_2 = 1$. The ED method estimates $\widehat{r}_2 = 2$, and both ER and GR methods estimate $\widehat{r}_2 = 3$. However, as shown in Appendix C.2, the empirical results are little affected by this value.

[⑤]We use the Stata software to calculate the different GMM and system GMM results.

that the QML method gives the correct signs of the coefficients in the last column. Furthermore, we experiment different r_2 values and find that the r_2 value has very small effect on the final results. As seen, the results in the seventh column obtained under $r_2 = 2$ is almost the same as those under $r_2 = 1$. More empirical results under larger r_2 values can be found in Appendix C.2. Fourth, the endoneneity issue does exist in the data. The $\mathcal{C}_s$ statistic is 6.11, but the critical values are 4.87 for 10%, 5.92 for 5%, 8.28 for 1%. So our empirical example emphasizes the necessity to take account of correlated heterogeneous shocks in FDI study. Finally, the coefficient of GOVERN from the QML method is strongly positive. But the same coefficient from other methods is either insignificant (D-GMM, AH, PC), or strongly negative (S-GMM). As pointed out above, the sign of this coefficient is controversial in the literature. Over this controversy, our empirical results are inclined to support the positive effects of government activities. There are many reasons responsible for this estimated result, and the whole mechanism may be complicated. We left the delicate explanation as a future work.

Insert Table 7

In Appendix C.1, we use the Monte Carlo methods and the bootstrap methods to explore the effects of the limited sample size. We find that the the bias issue is not pronounced. But the t-statistics are much affected by the relatively large changes of the standard deviation. However, except for the regressor LABOR, the remaining coefficients are still significant under the 10% significance level. Given the results in the eighth column, the main conclusions found in Table 7 are not changed.

10 Conclusion

In this paper we emphasize one source of endogeneity in the existing empirical FDI location studies. With consideration that qualified IVs are hard to find in general, and they are unavailable in our application in particular, we take "controlling through estimating" idea in econometric literature and propose panel data models with heterogeneous shocks to address the endogeneity. We consider the QML method to estimate the proposed model. The QML method has a striking advantage that it explicitly takes into account of cross-sectional heteroskedasticity, which is a key feature in China's FDI pattern. We investigate the asymptotic properties including asymptotic representation and limiting distributions. We also propose a new statistic to conduct the hypothesis testing on the presence of endogeneity within the model. The asymptotic properties of the proposed test is further explored.

We apply our methodology to study the location determinants of China's FDI inflow, allowing the FDI inflow subject to a global shock and the local characteristics subject to various domestic shocks. Results from the conventional linear regression suggest the endogeneity problem does exist in this study, while results from the QML estimators show that our method, which explicitly models the endogenous shocks, generate estimation results consistent with the theory.

References

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

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

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

Alvarez, J., and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 1121-1159.

Anderson, T. W., and Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375), 598-606.

Anderson, T. W., and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18(1), 47-82.

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, 58(2), 277-297.

Amiti, M., and Javorcik, B. S. (2008). Trade costs and location of foreign firms in China. *Journal of Development Economics*, 85(1), 129-149.

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. (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension, *Review of Economics and Statistics*, 98(2), 298-309.

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

Bai, J., Li, K., and Lu, L. (2016). Estimation and inference of FAVAR models. *Journal of Business & Economic Statistics*, 34(4), 620-641.

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

Bai, J., and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4), 1133-1150.

Bai, J., Kao, C., and Ng, S. (2009). Panel cointegration with global stochastic trends. *Journal of Econometrics*, 149(1), 82-99.

Blanc-Brude, F., Cookson, G., Piesse, J., Strange, R.(2014). The FDI location decision: Distance and the effects of spatial dependence, *International Business Review*, 23, 797-810.

Blonigen, B. A.(2005). A review of the empirical literature on FDI determinants, *Atlantic Economic Journal*, 33, 383-403.

Blyde, J., Molina, D.(2015). Logistic infrastructure and the international location of fragmented production, *Journal of International Economics*, 95, 319-332.

Boivin, J., Giannoni, M. P., and Mihov, I. (2009). Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data. *American Economic Review*, 99(1), 350-84.

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

Castagnetti, C., and Rossi, E. (2013). Euro corporate bond risk factors. *Journal of Applied Econometrics*, 28(3), 372-391.

Chamberlain, G., and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51, 1305-1324.

Cheng, L. K., Kwan, Y. K.(2000). What are the determinants of the location of foreign direct investment? The Chinese experience, *Journal of International Economics*, 51, 379-400.

Chudik, A., and Pesaran, M. H. (2015). Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. *Journal of Econometrics*, 188(2), 393-420.

Cole, M. A., Elliott, R. J., and Zhang, J. (2009). Corruption, governance and FDI location in China: A province-level analysis. *The Journal of Development Studies*, 45(9), 1494-1512.

Doz, C., Giannone, D., and Reichlin, L. (2012). A quasi maximum likelihood approach for large, approximate dynamic factor models. *Review of economics and statistics*, 94(4), 1014-1024.

Du, J., Lu, Y., and Tao, Z. (2008). Economic institutions and FDI location choice: Evidence from US multinationals in China. *Journal of comparative Economics*, 36(3), 412-429.

Eberhardt, M., Helmers, C., and Strauss, H. (2013). Do spillovers matter when estimating private returns to R&D?. *Review of Economics and Statistics*, 95(2), 436-448.

Eberhardt, M., and Teal, F. (2011). Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys*, 25(1), 109-155.

Fung, K. C., Iizaka, H., and Parker, S. (2002). Determinants of US and Japanese direct investment in China. *Journal of Comparative Economics*, 30(3), 567-578.

Gao, T. (2005). Labor quality and the location of foreign direct investment: Evidence from China. *China Economic Review*, 16, 274-292.

Gobillon, L., and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3), 535-551.

Greenaway-McGrevy, R., Han, C., and Sul, D. (2012) Asymptotic distribution of factor augmented estimators for panel regression, *Journal of Econometrics*, 169(1), 48-53.

Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton university press.

Han, C., and Phillips, P. C. (2006). GMM with many moment conditions. *Econometrica*, 74(1), 147-192.

Holly, S., Pesaran, M. H., and Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics*, 158(1), 160-173.

Kang, Y., Jiang, F.(2012). FDI location choice of Chinese multinationals in East and Southeast Asia: Traditional economic factors and institutional perspective, *Journal of World Business*, 47, 45-53.

Kang, S. J., Lee, H. S.(2007). The determinants of location choice of South Korean FDI in China, *Japan and the World Economy*, 19, 441-460.

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

Kose, M. A., Otrok, C. and Whiteman, C. (2003). International Business Cycles: World, Region, and Country-Specific Factors, *American Economic Review*, 93(4), 1216-1239.

Li, D., Qian, J., and Su, L. (2016). Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516), 1804-1819.

Li, K. (2018). Inference for factor model based average treatment effects. *Available at SSRN 3112775*.

Li, K., Cui, G., and Lu, L. (2019). Efficient estimation of heterogeneous coefficients in panel data models with common shocks. Forthcoming in *Journal of Econometrics*.

Liu, X., Lovely, M. E., Ondrich, J.(2010). The location decisions of foreign investors in China: Untangling the effect of wages using a control function approach, *The Review of Economics and Statistics*, 92, 160-166.

Moon, H. R., and Weidner, M. (2015). Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica*, 83(4), 1543-1579.

Moon, H. R., and Weidner, M. (2017). Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory*, 1-38.

Moon, H. R., and Weidner, M. (2018). Nuclear norm regularized estimation of panel regression models. arXiv preprint arXiv:1810.10987.

Moscone, F., and Tosetti, E. (2010). Testing for error cross section independence with an application to US health expenditure. *Regional Science and Urban Economics*, 40(5), 283-291.

Neyman, J., and Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1), 1-32.

Nielsen, B. B., Asmussen, C. G., Weatherall, C. D.(2017). The location choice of foreign direct investments: Empirical evidence and methodological challenges, *Journal of World Business*, 52, 62-82.

Noorbakhsh, F., Paloni A. (2001). Human capital and FDI inflows to developing countries: New empirical evidence. *World Development*, 29(9), 1593-1610.

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

Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2), 244-258.

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

Pesaran, M., and Kapetanios, G. (2005). Alternative approaches to estimation and inference in large multifactor panels: Small sample results with an application to modelling of asset returns. *Manuscript*

Stock, J. H., and Watson, M. W. (1998). Diffusion indexes. *NBER working paper No. 6702*.

UNCTAD (1998). *World Investment Report 1998: Trends and Determinants*. ISBN 92-1-112426-3.

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2), 307-333.

Xu, Y. (2017). Generalized Synthetic Control method: Causal inference with interactive fixed effects models. *Political Analysis* 25, 57-76.

Zhang, K., Markusen, J. (1999). Vertical multinationals and host-country characteristics, *Journal of Development Economics*, 59, 233šC252.

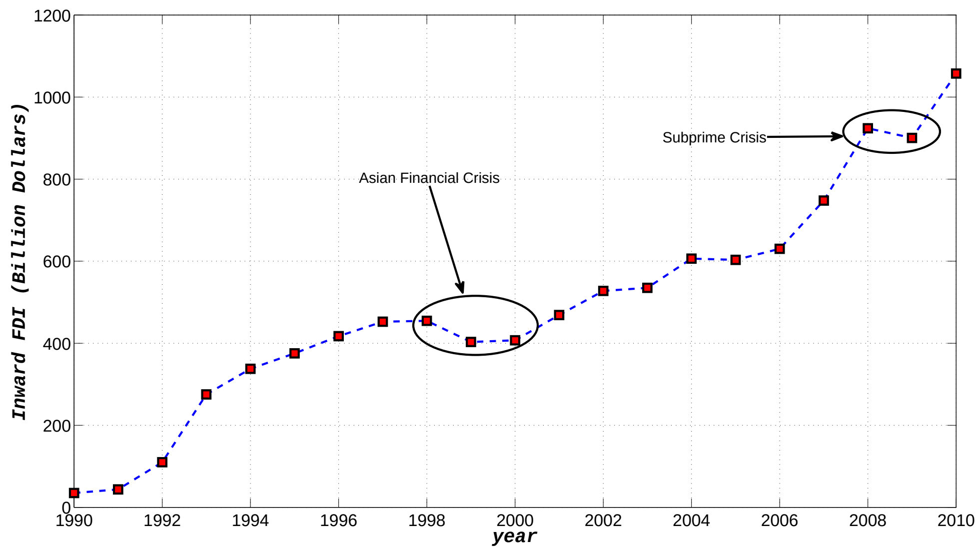

Figure 1: China's FDI Inflow over 1990-2010

Note: the figure presents the value of total inflow of foreign direct investment of China from 1990 to 2010. The data is from the China's Statistical Year Books. All numbers are in billion U.S. dollars.

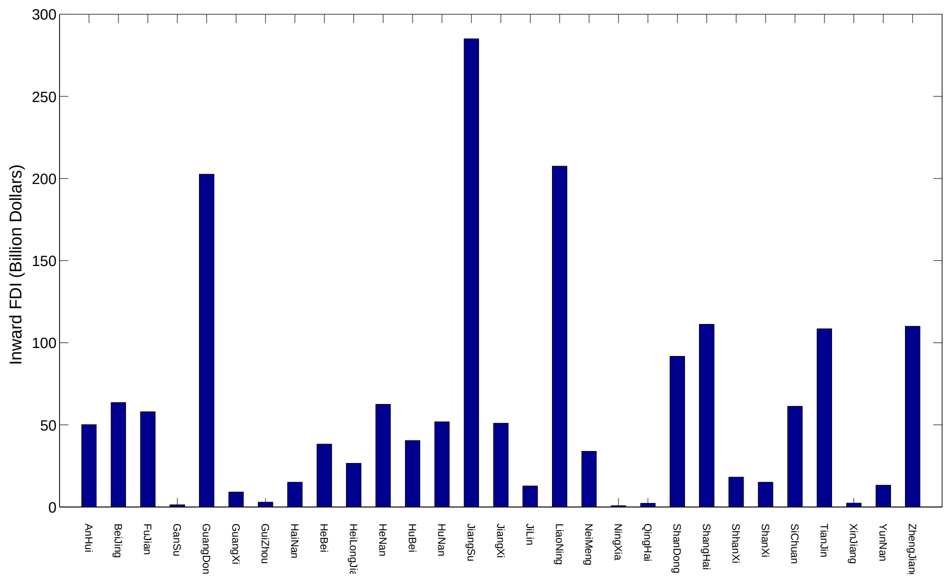

Figure 2: The 2010 FDI Inflow of 29 Provinces

Note: Figure 2 presents the FDI inflow received by 29 provinces in 2010: AnHui, BeiJing, FuJian, GanSu, GuangDong, GuangXi, GuiZhou, HaiNan, HeBei, HeiLongJiang, HeNan, HuBei, HuNan, JiangSu, JiangXi, JiLin, LiaoNing, NeiMeng, NingXia, QingHai, ShanDong, ShangHai, ShanXi, ShaanXi, SiChuan, TianJin, XinJiang, YunNan, and ZheJiang. The data is from the China's Statistical Year Books. All numbers are in billion U.S. dollars.

Table 1: Finite-sample performance of the QML and PC estimators under Case one

Size		ML						PC					
N	T	ρ			β			ρ			β		
		Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP
50	50	-0.0071	0.2768	75.5%	-0.0044	0.5500	90.3%	-0.0095	0.2989	64.7%	-0.0049	0.5847	91.2%
		-0.0027	-	88.5%	-0.0033	-	90.3%	-0.0024	-	91.0%	-0.0027	-	93.1%
100	50	-0.0074	0.2764	54.7%	-0.0046	0.5463	88.2%	-0.0088	0.2983	47.7%	-0.0051	0.5859	87.7%
		-0.0023	-	87.0%	-0.0033	-	88.7%	-0.0018	-	92.2%	-0.0030	-	91.9%
150	50	-0.0075	0.2754	37.3%	-0.0044	0.5462	85.2%	-0.0087	0.2969	32.7%	-0.0048	0.5873	86.4%
		-0.0022	-	85.5%	-0.0030	-	87.1%	-0.0017	-	91.0%	-0.0027	-	90.8%
50	100	-0.0033	0.2681	87.4%	-0.0015	0.5652	93.9%	-0.0042	0.2782	79.6%	-0.0017	0.5790	93.3%
		-0.0011	-	90.9%	-0.0011	-	93.7%	-0.0009	-	93.2%	-0.0009	-	94.2%
100	100	-0.0033	0.2678	77.4%	-0.0012	0.5623	93.6%	-0.0038	0.2781	72.0%	-0.0013	0.5808	94.4%
		-0.0007	-	91.0%	-0.0007	-	93.7%	-0.0005	-	93.2%	-0.0005	-	94.6%
150	100	-0.0035	0.2674	63.5%	-0.0014	0.5618	93.4%	-0.0038	0.2774	59.8%	-0.0015	0.5809	92.9%
		-0.0008	-	89.0%	-0.0009	-	93.7%	-0.0005	-	91.7%	-0.0008	-	93.8%
50	150	-0.0021	0.2662	91.7%	-0.0005	0.5693	94.0%	-0.0026	0.2728	85.8%	-0.0005	0.5760	93.9%
		-0.0005	-	91.5%	-0.0002	-	93.8%	-0.0004	-	92.3%	-0.0000	-	94.0%
100	150	-0.0021	0.2660	83.5%	-0.0006	0.5683	94.8%	-0.0024	0.2728	80.3%	-0.0006	0.5796	95.3%
		-0.0003	-	92.5%	-0.0002	-	94.9%	-0.0003	-	94.6%	-0.0001	-	95.2%
150	150	-0.0022	0.2655	76.9%	-0.0007	0.5670	94.7%	-0.0024	0.2724	73.1%	-0.0007	0.5792	94.7%
		-0.0004	-	93.1%	-0.0003	-	94.5%	-0.0003	-	95.1%	-0.0003	-	94.6%

Notes: In case one, the idiosyncratic errors e_{it} are both homoskedastic over cross section and time. SD denotes the standard deviation and CP the coverage probability. In each combination of N and T, the first row presents the results before bias correction, and the second row presents the results after bias correction.

Table 2: Finite-sample performance of the QML and PC estimators under Case two

Size		ML						PC					
N	T	ρ			β			ρ			β		
		Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP
50	50	-0.0029	0.1717	85.4%	-0.0027	0.3159	90.6%	-0.0093	0.3570	68.8%	0.0032	0.6219	82.1%
		-0.0008	-	91.0%	-0.0022	-	89.7%	-0.0021	-	89.9%	0.0026	-	85.4%
100	50	-0.0029	0.1712	78.2%	-0.0027	0.3140	84.4%	-0.0097	0.3910	58.2%	-0.0028	0.6582	87.8%
		-0.0007	-	91.1%	-0.0022	-	84.5%	-0.0020	-	91.4%	-0.0016	-	90.0%
150	50	-0.0029	0.1697	70.0%	-0.0024	0.3116	85.6%	-0.0100	0.3948	42.0%	-0.0035	0.6628	87.0%
		-0.0008	-	89.6%	-0.0019	-	85.5%	-0.0022	-	89.3%	-0.0017	-	90.4%
50	100	-0.0014	0.1671	90.3%	-0.0009	0.3234	92.8%	-0.0042	0.3375	81.9%	0.0040	0.6164	82.7%
		-0.0004	-	90.7%	-0.0008	-	92.5%	-0.0007	-	90.6%	0.0025	-	87.2%
100	100	-0.0013	0.1641	86.9%	-0.0008	0.3200	92.7%	-0.0043	0.3590	75.3%	0.0001	0.6409	90.5%
		-0.0002	-	92.4%	-0.0006	-	92.8%	-0.0006	-	92.7%	-0.0002	-	92.0%
150	100	-0.0013	0.1641	83.9%	-0.0006	0.3184	94.0%	-0.0046	0.3636	64.7%	-0.0004	0.6415	93.6%
		-0.0002	-	93.1%	-0.0004	-	93.9%	-0.0009	-	93.1%	-0.0001	-	94.5%
50	150	-0.0009	0.1642	93.1%	-0.0001	0.3258	93.4%	-0.0025	0.3289	86.9%	0.0050	0.6144	80.3%
		-0.0002	-	94.1%	0.0000	-	93.5%	-0.0003	-	92.8%	0.0031	-	87.3%
100	150	-0.0009	0.1634	89.6%	-0.0003	0.3234	94.2%	-0.0029	0.3473	81.7%	0.0008	0.6295	91.0%
		-0.0001	-	93.5%	-0.0002	-	93.9%	-0.0005	-	92.8%	0.0004	-	92.6%
150	150	-0.0009	0.1614	85.0%	-0.0004	0.3214	93.0%	-0.0028	0.3530	76.0%	0.0000	0.6368	93.1%
		-0.0001	-	93.9%	-0.0001	-	94.2%	-0.0004	-	93.3%	0.0000	-	94.6%

Notes: In case two, the idiosyncratic errors e_{it} are heteroskedastic over cross section, and homoskedastic over time. SD denotes the standard deviation and CP the coverage probability. In each combination of N and T, the first row presents the results before bias correction, and the second row presents the results after bias correction.

Table 3: Finite-sample performance of the QML and PC estimators under Case three

Size		ML						PC					
		ρ			β			ρ			β		
N	T	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP	Bias	$\sqrt{NT}$SD	CP
50	50	-0.0028	0.1679	85.3%	-0.0005	0.3161	87.8%	-0.0082	0.3502	69.2%	0.0227	0.6633	67.2%
		-0.0011	0	87.8%	-0.0006	0	86.0%	-0.0027	0	84.5%	0.0221	0	69.7%
100	50	-0.0027	0.1665	75.5%	-0.0018	0.3113	87.1%	-0.0086	0.3925	57.3%	0.0114	0.7231	71.1%
		-0.0008	0	85.9%	-0.0017	0	85.4%	-0.0027	0	79.7%	0.0108	0	76.5%
150	50	-0.0029	0.1658	64.4%	-0.0015	0.3093	84.4%	-0.0091	0.4016	46.5%	0.0093	0.7370	71.6%
		-0.0010	0	83.9%	-0.0014	0	83.0%	-0.0030	0	79.7%	0.0085	0	77.8%
50	100	-0.0013	0.1688	90.3%	-0.0004	0.3361	94.2%	-0.0038	0.3458	80.6%	0.0147	0.6973	73.5%
		-0.0003	0	89.3%	-0.0005	0	92.9%	-0.0007	0	87.9%	0.0127	0	77.7%
100	100	-0.0014	0.1658	84.7%	-0.0003	0.3305	92.5%	-0.0045	0.3714	72.6%	0.0070	0.7284	80.4%
		-0.0004	0	90.8%	-0.0003	0	92.0%	-0.0012	0	88.7%	0.0055	0	85.1%
150	100	-0.0013	0.1654	81.1%	-0.0005	0.3293	92.4%	-0.0043	0.3859	66.6%	0.0036	0.7456	83.4%
		-0.0003	0	90.6%	-0.0005	0	92.0%	-0.0010	0	90.2%	0.0024	0	88.4%
50	150	-0.0009	0.1704	92.5%	-0.0001	0.3476	94.2%	-0.0028	0.3449	85.7%	0.0121	0.7120	73.2%
		-0.0002	0	91.1%	-0.0001	0	93.6%	-0.0006	0	91.0%	0.0096	0	78.5%
100	150	-0.0009	0.1678	89.7%	-0.0002	0.3401	94.0%	-0.0029	0.3731	79.7%	0.0050	0.7375	80.2%
		-0.0002	0	92.0%	-0.0002	0	94.1%	-0.0005	0	90.8%	0.0031	0	88.0%
150	150	-0.0009	0.1673	85.9%	-0.0002	0.3378	93.4%	-0.0028	0.3803	75.8%	0.0031	0.7354	84.7%
		-0.0002	0	93.0%	-0.0002	0	93.9%	-0.0005	0	92.7%	0.0020	0	90.1%

Notes: In case three, the idiosyncratic errors e_{it} are heteroskedastic over cross section and time. SD denotes the standard deviation and CP the coverage probability. In each combination of N and T, the first row presents the results before bias correction, and the second row presents the results after bias correction.

Table 4: Rejecting frequency of the $\mathcal{C}_s$ test under 1%, 5% and 10% nominal level

N	T	$c=0$			$c=3$			$c=5$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
50	50	13.6%	6.6%	0.6%	90.3%	83.1%	54.0%	97.5%	97.1%	92.9%
75	50	10.6%	5.4%	0.6%	91.6%	83.3%	54.2%	98.9%	98.9%	96.7%
100	50	10.5%	4.7%	0.4%	91.6%	83.4%	55.1%	99.4%	99.0%	96.9%
125	50	11.0%	5.5%	0.6%	92.1%	85.3%	55.3%	99.8%	99.8%	98.5%
50	75	11.6%	5.7%	0.8%	91.5%	85.2%	61.7%	97.5%	97.4%	96.3%
75	75	11.5%	5.3%	0.7%	91.9%	86.4%	63.7%	99.5%	99.5%	99.2%
100	75	11.1%	5.4%	1.1%	92.8%	86.7%	63.5%	100.0%	100.0%	99.6%
125	75	11.2%	5.8%	1.1%	94.5%	88.1%	63.5%	99.6%	99.5%	98.9%
50	100	9.6%	4.6%	1.0%	92.9%	85.5%	66.3%	98.1%	98.1%	97.3%
75	100	10.5%	5.9%	1.2%	93.5%	88.4%	69.3%	99.2%	99.1%	99.0%
100	100	10.4%	5.3%	1.0%	93.7%	89.0%	69.3%	100.0%	100.0%	100.0%
125	100	10.7%	5.4%	1.1%	93.4%	88.4%	70.4%	99.8%	99.8%	99.4%

Table 5: Variable Descriptions

Variable	Description	Expected Sign
Dependent variable		
FDI	The logarithm of total investment of FDI firms (RMB in 2005)	
Explanatory variables		
market	The logarithm of GDP (RMB in 2005)	+
infra	The density of road and railway	+
wage	The growth rate of logarithm of the average wage of urban employee (RMB in 2005)	−
open	The total export and import over GDP	+
govern	Local government fiscal expenditure over GDP	+/-
labor	The high school graduates over population	+

Table 6: Summary statistics

Variable	Obs	Mean	Std	Max	Min
FDI	638	11.8072	1.5123	15.0340	6.5425
market	609	8.3833	1.1278	10.8819	5.2363
infra	609	0.5711	0.4455	2.3103	0.0187
wage	609	0.0102	0.0056	0.0441	-0.0272
open	609	3.1598	4.0134	22.0295	0.3204
govern	609	0.1632	0.0825	0.6121	0.0492
labor	609	4.1549	1.9710	8.9012	1.0705

Table 7: The estimation results of five different methods

Explanatory variables	D-GMM	S-GMM	AH	PC	QML			
					$r_2 = 1$	$r_2 = 2$	$r_2 = 1$	$r_2 = 1$
FDI(-1)	0.1035	0.5108***	0.4422	0.7062***	0.7093***	0.7093***	0.7066***	0.7124***
	(0.0896)	(0.1116)	(0.3701)	(0.0659)	(0.0220)	(0.0220)	(0.0220)	(0.0225)
market	0.3969**	0.3558**	0.2568	0.1213**	0.0439**	0.0439**	0.0644***	0.1147***
	(0.1665)	(0.1452)	(0.2726)	(0.0537)	(0.0208)	(0.0208)	(0.0181)	(0.0126)
infra	0.0371	0.1624	0.0916	-0.0209	0.0622**	0.0622**	0.0635**	
	(0.1661)	(0.1215)	(0.1824)	(0.0343)	(0.0257)	(0.0257)	(0.0261)	
wage	0.7596	1.9373	0.7196	0.0037	-2.7830***	-2.7830***	-2.4600***	-2.7355***
	(2.8363)	(2.6238)	(2.6351)	(0.1183)	(0.9581)	(0.9581)	(0.9094)	(0.8893)
open	0.0323	0.0671***	0.0136	0.0241***	0.0222***	0.0222***	0.0243***	0.0241***
	(0.0259)	(0.0179)	(0.0205)	(0.0060)	(0.0039)	(0.0039)	(0.0039)	(0.0038)
govern	-0.7355	-2.0484***	0.0713	-0.1269	0.4965***	0.4966***	0.5185***	
	(0.9460)	(0.5259)	(1.0830)	(0.2569)	(0.1705)	(0.1705)	(0.1708)	
labor	0.0609**	-0.0109	-0.0064	0.0172*	0.0088**	0.0088**		
	(0.0290)	(0.0168)	(0.0330)	(0.0089)	(0.0040)	(0.0040)		
Individual effects	yes	yes	yes	yes	yes	yes	yes	yes
Hetero-shock/ interacitve effects	no	no	no	yes	yes	yes	yes	yes
$\mathcal{C}_s$ test					6.11			

Notes: "D-GMM", "S-GMM", "AH", "PC" and "QML" denote the difference GMM method, the system GMM method, the Anderson and Hsiao's method, the PC method and the QML method, respectively. The numbers in the parentheses are standard errors. *, ** and *** denote the 10%, 5% and 1% statistical significance. The PC and QML methods are estimated under $r_1 = 1$. The biases and standard errors of the PC and QML methods are calculated under Assumption G. Critical values for $\mathcal{C}_s$ are 4.87 for 10%, 5.92 for 5%, 8.28 for 1%.

Appendix A: First order conditions

Let $\Phi_i = \text{diag}(\sigma_i^2, \Sigma_{ii})$ and $\Phi = \text{diag}(\Phi_1, \Phi_2, \ldots, \Phi_N)$, and $\widehat{\Phi}_i$ and $\widehat{\Phi}$ be the corresponding QML estimators. In this appendix, we presents the first order conditions for the objective function (3.3). For ease of exposition, we define $\widehat{\mathcal{H}} = (\widehat{\Lambda}'\widehat{\Phi}^{-1}\widehat{\Lambda})^{-1}$ and $\widehat{\mathcal{G}} = (\widehat{\Sigma}_{ff}^{-1} + \widehat{\Lambda}'\widehat{\Phi}^{-1}\widehat{\Lambda})^{-1}$ and partition it into $\widehat{\mathcal{G}} = [\mathcal{G}_1', \mathcal{G}_2']'$, where $\widehat{\mathcal{G}}_1$ is the first r_1 rows and $\widehat{\mathcal{G}}_2$ the remaining r_2 rows of $\widehat{\mathcal{G}}$, respectively. By $(A+B)^{-1} = B^{-1} - (A+B)^{-1}AB^{-1}$, we have $\widehat{\mathcal{G}} = \widehat{\mathcal{H}} - \widehat{\mathcal{G}}\widehat{\Sigma}_{ff}^{-1}\widehat{\mathcal{H}}$.

Let $W_{it} = (Y_{it-1}, X_{it}')'$ and $\psi = (\rho, \beta')'$. The first order condition with respect to ρ and β is

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\widehat{\sigma}_i^2}\dot{W}_{it}\left\{(\dot{Y}_{it} - \dot{W}_{it}'\widehat{\psi}) - \widehat{\kappa}_i'\widehat{\mathcal{G}}_1\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\begin{bmatrix}\dot{Y}_{jt} - \dot{W}_{jt}'\widehat{\psi}\\ \dot{X}_{jt}'\end{bmatrix}\right\} = 0. \tag{A.1}$$

The first order condition with respect to Σ_{ff} is

$$\widehat{\Lambda}'\widehat{\Sigma}_{zz}^{-1}\left[\widehat{\mathcal{B}}\frac{1}{T}\sum_{t=1}^{T}z_t(\widehat{\rho})z_t(\widehat{\rho})'\widehat{\mathcal{B}}' - \widehat{\Sigma}_{zz}\right]\widehat{\Sigma}_{zz}^{-1}\widehat{\Lambda} = \begin{bmatrix}\times & 0_{r_1\times r_2}\\ 0_{r_2\times r_1} & \times\end{bmatrix}, \tag{A.2}$$

where on the right-hand-side of (A.2), only zero components correspond to the first order conditions, while $\times$ denote components that do not subject to the first order (zero) conditions.

The first order condition for Λ leads to

$$\widehat{\Sigma}_{ff}\widehat{\Lambda}'\widehat{\Sigma}_{zz}^{-1}\left[\widehat{\mathcal{B}}\frac{1}{T}\sum_{t=1}^{T}z_t(\widehat{\rho})z_t(\widehat{\rho})'\widehat{\mathcal{B}}' - \widehat{\Sigma}_{zz}\right]\widehat{\Sigma}_{zz}^{-1} = \Upsilon', \tag{A.3}$$

where Υ is an $N(k+1)\times r$ matrix, whose entry is zero if the counterpart of Λ is not specified to be zero, otherwise undetermined. Pre-multiplying $\widehat{\Sigma}_{ff}^{-1}$ and post-multiplying $\widehat{\Lambda}$ to (A.3), we get

$$\widehat{\Lambda}'\widehat{\Sigma}_{zz}^{-1}\left[\widehat{\mathcal{B}}\frac{1}{T}\sum_{t=1}^{T}z_t(\widehat{\rho})z_t(\widehat{\rho})'\widehat{\mathcal{B}}' - \widehat{\Sigma}_{zz}\right]\widehat{\Sigma}_{zz}^{-1}\widehat{\Lambda} = \widehat{\Sigma}_{ff}^{-1}\Upsilon'\widehat{\Lambda}. \tag{A.4}$$

Equating the right hand sides of (A.2) and (A.4), together with special structure of Υ and $\widehat{\Lambda}$, we have $\Upsilon'\widehat{\Lambda} = 0$. Thus,

$$\widehat{\Lambda}'\widehat{\Phi}^{-1}\left[\widehat{\mathcal{B}}\frac{1}{T}\sum_{t=1}^{T}z_t(\widehat{\rho})z_t(\widehat{\rho})'\widehat{\mathcal{B}}' - \widehat{\Sigma}_{zz}\right]\widehat{\Phi}^{-1}\widehat{\Lambda} = 0, \tag{A.5}$$

where we use the fact $\widehat{\Lambda}'\widehat{\Sigma}_{zz}^{-1} = \widehat{\Sigma}_{ff}^{-1}\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}$. Let $\Sigma_{zz}^{ij} = \Lambda_i'\Sigma_{ff}\Lambda_j + \mathbb{1}(i=j)\Phi_i$ and

$$M_{zz}^{ij}(\rho,\beta) = \frac{1}{T}\sum_{t=1}^{T}\begin{bmatrix}\dot{Y}_{it} - \rho\dot{Y}_{it-1} - \dot{X}_{it}'\beta\\ \dot{X}_{it}\end{bmatrix}\left[\dot{Y}_{jt} - \rho\dot{Y}_{jt-1} - \dot{X}_{jt}'\beta, \dot{X}_{jt}'\right].$$

With the above notations, the first order condition for κ_i gives

$$\widehat{\mathcal{G}}_1\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\left[M_{zz}^{ji}(\widehat{\rho},\widehat{\beta}) - \widehat{\Sigma}_{zz}^{ji}\right]\widehat{\Phi}_i^{-1}\boldsymbol{e}_1 = 0, \tag{A.6}$$

where $\boldsymbol{e}_1$ is the first column of the $(k+1)$-dimensional identity matrix.

The first order condition for γ_i is

$$\widehat{\mathcal{G}}_2 \sum_{j=1}^{N} \widehat{\Lambda}_j \widehat{\Phi}_j^{-1} \left[M_{zz}^{ji}(\widehat{\rho}, \widehat{\beta}) - \widehat{\Sigma}_{zz}^{ji} \right] \widehat{\Phi}_i^{-1} \boldsymbol{e}^- = 0, \tag{A.7}$$

where $\boldsymbol{e}^-$ is the matrix which is obtained by deleting first column of the $(k+1)$-dimensional identity matrix. The first order condition for Σ_{jj} is

$$\begin{aligned} M_{zz}^{jj}(\widehat{\rho}, \widehat{\beta}) - \widehat{\Sigma}_{zz}^{jj} - \widehat{\Lambda}_j' \widehat{\mathcal{G}} \sum_{i=1}^{N} \widehat{\Lambda}_i \widehat{\Phi}_i^{-1} \left(M_{zz}^{ij}(\widehat{\rho}, \widehat{\beta}) - \widehat{\Sigma}_{zz}^{ij} \right) & \\ - \sum_{i=1}^{N} \left(M_{zz}^{ji}(\widehat{\rho}, \widehat{\beta}) - \widehat{\Sigma}_{zz}^{ji} \right) \widehat{\Phi}_i^{-1} \widehat{\Lambda}_i' \widehat{\mathcal{G}} \widehat{\Lambda}_j = J, & \end{aligned} \tag{A.8}$$

where J is a $(k+1) \times (k+1)$ matrix whose upper-left 1×1 and lower-right $k \times k$ submatrices are both zeros, the remaining elements are undetermined. The undetermined elements correspond to the zero elements of Σ_{ii}.

These first order conditions are used for the asymptotic analysis.

Appendix B: Proof of Proposition 5.1

For convenience of exposition, we further introduce the following notations. Let $\dot{\mathbf{W}}_{it} = (\dot{W}_{it}, 0_{(k+1)\times k})'$, $\widehat{\xi}_t = N^{-1} \sum_{i=1}^{N} \widehat{\Lambda}_i \widehat{\Phi}_i^{-1} \dot{\mathbf{W}}_{it}$ and $\widehat{\chi}_t = N^{-1} \sum_{i=1}^{N} \widehat{\Lambda}_i \widehat{\Phi}_i^{-1} \dot{\epsilon}_{it}$. $\mathbf{E}_1$ is the left r_1 columns of the $r \times r$ identity matrix and $\mathbf{E}_2$ is the remaining right r_2 columns, i.e., $[\mathbf{E}_1, \mathbf{E}_2] = I_r$.

Throughout the proof of consistency, to avoid the potential ambiguity, we use the symbols with asterisk to denote the underlying true values, the symbols with hat to denote the QML estimators, and the symbols themselves to denote the input argument of the likelihood functions. Once we have obtained the consistency, we drop asterisks from the symbols of the true values for notational simplicity.

The following two lemmas are useful in our theoretical analysis, whose proofs are given in the online supplement.

Lemma B.1 *Under Assumptions A-C,*

$$(a) \quad E\Big\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} f_t \epsilon_{it}' \Big\|^2 \le C, \qquad \text{for each } i,$$

$$(b) \quad E\Big\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\epsilon_{it}\epsilon_{jt}' - E(\epsilon_{it}\epsilon_{jt}')] \Big\|^2 \le C, \qquad \text{for each } i \text{ and } j$$

$$(c) \quad E\Big\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Lambda_i \Phi_i^{-1} \epsilon_{it} f_t' \Big\|^2 \le C,$$

$$(d) \quad E\Big\| \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Lambda_i \Phi_i^{-1} [\epsilon_{it}\epsilon_{jt}' - E(\epsilon_{it}\epsilon_{jt}')] \Big\|^2 \le C, \qquad \text{for each } j,$$

$$(e) \quad E\Big\| \frac{1}{N\sqrt{T}} \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_i \Phi_i^{-1} \sum_{t=1}^{T} [\epsilon_{it}\epsilon_{jt} - E(\epsilon_{it}\epsilon_{jt}')] \Phi_j^{-1} \Lambda_j' \Big\|^2 \le C,$$

$$(f)\quad E\Big\|\frac{1}{\sqrt{N}T}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^4}Y_{it-1}e_{it}(e_{is}^2-\sigma_i^2)\Big\|^2\le C.$$

where C is a generic constant.

Lemma B.2 *Let $\theta=(\rho,\beta,\Lambda,\Psi)$. Under Assumption A-E,*

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

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

$$(c)\quad \sup_{\theta\in\Theta}\mathrm{tr}\Big[\mathcal{B}^{*-1}\bar\epsilon\bar\epsilon'\mathcal{B}^{*-1\prime}\mathcal{B}'\Sigma_{zz}^{-1}\mathcal{B}\Big]=o_p(1)$$

$$(d)\quad \sup_{\theta\in\Theta}\Big|\mathrm{tr}\Big[\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot\epsilon_t'\mathcal{B}^{*-1\prime}\mathcal{B}'\Sigma_{zz}^{-1}\mathcal{B}\Big]\Big|=o_p(1)$$

where $\mathcal{B}^*=I_N\otimes B^*$ and $\mathcal{B}=I_N\otimes B$.

Proof of Proposition 5.1. Throughout the proof, we use the following centered objective function:

$$\mathcal{L}(\theta)=-\frac{1}{2N}\ln|\Sigma_{zz}|-\frac{1}{2N}\mathrm{tr}\Big[\mathcal{B}M_{zz}(\rho)\mathcal{B}'\Sigma_{zz}^{-1}\Big]+\frac{1}{2N}\ln|\Sigma_{zz}^*|+\frac{1}{2}(k+1),\tag{B.1}$$

where $\Sigma_{zz}^*=\Lambda^*\Sigma_{ff}^*\Lambda^{*\prime}+\Phi^*$ with $\Phi^*=\mathrm{Bdiag}(\overline{\Sigma}_{\epsilon\epsilon}^*)$. The above objective function differs by a constant from the original one, so it has the same maximizer. The model specification gives $B^*z_{it}(\rho^*)=\mu_i^*+\Lambda_i^{*\prime}f_t^*+\epsilon_{it}$. Applying the within-group transformation to this equation, we have $B^*\dot z_{it}(\rho^*)=\Lambda_i^{*\prime}\dot f_t^*+\dot\epsilon_{it}$, implying

$$\mathcal{B}^*\dot z_t(\rho^*)=\Lambda^*\dot f_t^*+\dot\epsilon_t\tag{B.2}$$

with $\dot z_t(\rho^*)=[z_{1t}(\rho^*)',z_{2t}(\rho^*)',\ldots,z_{Nt}(\rho^*)']'$.

Consider the second term on the right hand side of (B.1), which we use II_1 to denote. We have

$$II_1=\frac{1}{2N}\mathrm{tr}\Big[\mathcal{B}\frac{1}{T}\sum_{t=1}^{T}\dot z_t(\rho)\dot z_t(\rho)'\mathcal{B}'\Sigma_{zz}^{-1}\Big]=\frac{1}{2N}\mathrm{tr}\Big[\frac{1}{T}\sum_{t=1}^{T}\dot z_t(\rho)\dot z_t(\rho)'\mathcal{B}'\Sigma_{zz}^{-1}\mathcal{B}\Big].$$

where $\Sigma_{zz}=\Lambda\Sigma_{ff}\Lambda'+\Phi$. By the definition of $\dot z_{it}(\rho)$, for each i,

$$\dot z_{it}(\rho)=\begin{bmatrix}\dot Y_{it}-\rho\dot Y_{it-1}\\ \dot X_{it}\end{bmatrix}=-(\rho-\rho^*)\dot{\mathbf{Y}}_{it-1}+\dot z_{it}(\rho^*)=-(\rho-\rho^*)\dot{\mathbf{Y}}_{it-1}+B^{*-1}(\Lambda_i^{*\prime}\dot f_t^*+\dot\epsilon_{it}),$$

where $\dot{\mathbf{Y}}_{it-1}=(\dot Y_{it-1},0_{1\times k})'$. The above expression is equivalent to

$$\dot z_t(\rho)=-(\rho-\rho^*)\dot{\mathbf{Y}}_{t-1}+(I_N\otimes B^{*-1})(\Lambda^*\dot f_t^*+\dot\epsilon_t)=-(\rho-\rho^*)\dot{\mathbf{Y}}_{t-1}+\mathcal{B}^{*-1}(\Lambda^*\dot f_t^*+\dot\epsilon_t).$$

Now we introduce some symbols to simplify the notation. Define

$$\widetilde\Lambda^*=\mathcal{B}^{*-1}\Lambda^*,\qquad \widetilde\Sigma_{\epsilon\epsilon}^*=\mathcal{B}^{*-1}\overline\Sigma_{\epsilon\epsilon}^*\mathcal{B}^{*-1\prime},\qquad \overline\Sigma_{\epsilon\epsilon}^*=\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\epsilon_t'$$

$$\widetilde{\Lambda} = \mathcal{B}^{-1}\Lambda, \qquad \widetilde{\Phi} = \mathcal{B}^{-1}\Phi\mathcal{B}^{-1\prime}, \qquad \widetilde{\Sigma}_{zz} = \mathcal{B}^{-1}\Sigma_{zz}\mathcal{B}^{-1\prime}.$$

With the above symbols, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\dot{z}_t(\rho)\dot{z}_t(\rho)' = {} & \widetilde{\Sigma}^*_{zz} + (\rho-\rho^*)^2\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot{\mathbf{Y}}'_{t-1} - (\rho-\rho^*)\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot{f}_t^{*\prime}\widetilde{\Lambda}^{*\prime} \\
& - (\rho-\rho^*)\widetilde{\Lambda}^*\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*\dot{\mathbf{Y}}'_{t-1} - (\rho-\rho^*)\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot{\epsilon}'_t\mathcal{B}^{*-1\prime} \\
& - (\rho-\rho^*)\mathcal{B}^{*-1}\frac{1}{T}\sum_{t=1}^{T}\dot{\epsilon}_t\dot{\mathbf{Y}}'_{t-1} - \mathcal{B}^{*-1}\bar{\epsilon}\bar{\epsilon}'\mathcal{B}^{*-1\prime} \\
& + \mathcal{B}^{*-1}\Big(\Lambda^*\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*\epsilon'_t\Big)\mathcal{B}^{*-1\prime} + \mathcal{B}^{*-1}\Big(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\dot{f}_t^{*\prime}\Lambda^{*\prime}\Big)\mathcal{B}^{*-1\prime} \\
& + \mathcal{B}^{*-1}\Big(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\epsilon'_t - E(\epsilon_t\epsilon'_t)\Big)\mathcal{B}^{*-1\prime}.
\end{aligned}$$

where $\widetilde{\Sigma}^*_{zz} = \widetilde{\Lambda}^*(\frac{1}{T}\dot{F}^{*\prime}\dot{F}^*)\widetilde{\Lambda}^{*\prime} + \widetilde{\Sigma}^*_{\epsilon\epsilon}$. Let $\dot{\mathbf{Y}}_{-1} = (\dot{\mathbf{Y}}_0, \dot{\mathbf{Y}}_1, \ldots, \mathbf{Y}_{T-1})$. Notice that

$$\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot{\mathbf{Y}}'_{t-1} = \frac{1}{T}\dot{\mathbf{Y}}_{-1}\dot{\mathbf{Y}}'_{-1} = \frac{1}{T}\dot{\mathbf{Y}}_{-1}M_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1} + \frac{1}{T}\dot{\mathbf{Y}}_{-1}P_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1},$$

Thus,

$$\begin{aligned}
(\rho-\rho^*)^2\operatorname{tr}\Big[\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot{\mathbf{Y}}'_{t-1}\widetilde{\Sigma}^{-1}_{zz}\Big] = {} & (\rho-\rho^*)^2\operatorname{tr}\Big[\frac{1}{T}\dot{\mathbf{Y}}_{-1}M_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1}\widetilde{\Sigma}^{-1}_{zz}\Big] \\
& + (\rho-\rho^*)^2\operatorname{tr}\Big[\frac{1}{T}\dot{\mathbf{Y}}_{-1}P_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1}\widetilde{\Sigma}^{-1}_{zz}\Big].
\end{aligned}$$

Given the above expression, we can rewrite II_1 as

$$\begin{aligned}
II_1 = {} & (\rho-\rho^*)^2\frac{1}{2NT}\operatorname{tr}\Big[\dot{\mathbf{Y}}_{-1}M_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1}\widetilde{\Sigma}^{-1}_{zz}\Big] + \frac{1}{2N}\operatorname{tr}\Big[\widetilde{\Sigma}^*_{\epsilon\epsilon}\widetilde{\Sigma}^{-1}_{zz}\Big] \\
& + \frac{1}{2NT}\operatorname{tr}\Big[\Psi(\rho)'\widetilde{\Sigma}^{-1}_{zz}\Psi(\rho)\Big] + \mathcal{R}(\theta),
\end{aligned}$$

where $\Psi(\rho) = (\rho-\rho^*)\dot{\mathbf{Y}}_{-1}\dot{F}^*(\dot{F}^{*\prime}\dot{F}^*)^{-1/2} - \widetilde{\Lambda}^*(\dot{F}^{*\prime}\dot{F}^*)^{1/2}$ and

$$\begin{aligned}
\mathcal{R}(\theta) = {} & -(\rho-\rho^*)\frac{1}{N}\operatorname{tr}\Big[\frac{1}{T}\sum_{t=1}^{T}\dot{\mathbf{Y}}_{t-1}\dot{\epsilon}'_t\widetilde{\Sigma}^{-1}_{zz}\Big] - (\rho-\rho^*)\frac{1}{2N}\operatorname{tr}\Big[\mathcal{B}^{*-1}\bar{\epsilon}\bar{\epsilon}'\mathcal{B}^{*-1\prime}\widetilde{\Sigma}^{-1}_{zz}\Big] \\
& + \frac{1}{N}\operatorname{tr}\Big[\mathcal{B}^{*-1}\Big(\Lambda^*\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*\epsilon'_t\Big)\mathcal{B}^{*-1\prime}\widetilde{\Sigma}^{-1}_{zz}\Big] \\
& + \frac{1}{2N}\operatorname{tr}\Big[\mathcal{B}^{*-1}\Big(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t\epsilon'_t - \Sigma^*_{\epsilon\epsilon}\Big)\mathcal{B}^{*-1\prime}\widetilde{\Sigma}^{-1}_{zz}\Big].
\end{aligned}$$

Let $\widetilde{\Phi} = \mathcal{B}^{-1}\Phi\mathcal{B}^{-1\prime}$. By $\ln|\Sigma_{zz}| = \ln|\Phi| + \ln|I_r + \Sigma_{ff}^{1/2}\Lambda'\Phi^{-1}\Lambda\Sigma_{ff}^{1/2}|$ together with $\ln|\mathcal{B}| = \ln|I_N \otimes B| = 0$, we have

$$\ln|\Sigma_{zz}| = \ln|\Phi| + \ln|I_r + \Sigma_{ff}^{1/2}\Lambda'\Phi^{-1}\Lambda\Sigma_{ff}^{1/2}| = \ln|\widetilde{\Phi}| + \ln\Big|I_r + \Sigma_{ff}^{1/2}\Lambda'\Phi^{-1}\Lambda\Sigma_{ff}^{1/2}\Big|.$$

Let $\widetilde{\Phi}^* = \mathcal{B}^{*-1}\Phi^*\mathcal{B}^{*-1\prime}$. By the same arguments,

$$\ln|\Sigma_{zz}^*| = \ln|\widetilde{\Phi}^*| + \ln\left|I_r + \Sigma_{ff}^{*1/2}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*\Sigma_{ff}^{*1/2}\right|.$$

Given the preceding two expressions, together with the facts (i) $\widetilde{\Sigma}_{zz}^{-1} = \widetilde{\Phi}^{-1} - \widetilde{\Phi}^{-1}\widetilde{\Lambda}\widetilde{\mathcal{G}}\widetilde{\Lambda}'\widetilde{\Phi}^{-1}$ and (ii) $\mathrm{tr}(\widetilde{\Sigma}_{\epsilon\epsilon}^*\widetilde{\Phi}^{-1}) = \mathrm{tr}(\widetilde{\Phi}^*\widetilde{\Phi}^{-1})$, we can rewrite the objective function (B.1) as

$$\begin{aligned}
\mathcal{L}(\theta) = &-\frac{1}{2N}\Big[-\ln|\widetilde{\Phi}^*\widetilde{\Phi}^{-1}| + \mathrm{tr}(\widetilde{\Phi}^*\widetilde{\Phi}^{-1}) - N(k+1)\Big] \\
&- \frac{1}{2N}\ln\left|I_r + \Sigma_{ff}^{1/2}\Lambda'\Phi^{-1}\Lambda\Sigma_{ff}^{1/2}\right| \\
&+ \frac{1}{2N}\ln\left|I_r + \Sigma_{ff}^{*1/2}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*\Sigma_{ff}^{*1/2}\right| \\
&- \frac{1}{2NT}\mathrm{tr}\Big[\Psi(\rho)'\widetilde{\Sigma}_{zz}^{-1}\Psi(\rho)\Big] \\
&- (\rho-\rho^*)^2\frac{1}{2NT}\mathrm{tr}\Big[\dot{\mathbf{Y}}_{-1}M_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1}\widetilde{\Sigma}_{zz}^{-1}\Big] \\
&+ \frac{1}{2N}\mathrm{tr}\Big[\widetilde{\Sigma}_{\epsilon\epsilon}^*\widetilde{\Phi}^{-1}\widetilde{\Lambda}\widetilde{G}\widetilde{\Lambda}'\widetilde{\Phi}^{-1}\Big] - \mathcal{R}(\theta)
\end{aligned}$$

with $\widetilde{\mathcal{G}} = (\Sigma_{ff}^{-1} + \widetilde{\Lambda}'\widetilde{\Phi}^{-1}\widetilde{\Lambda})^{-1}$. The above expression can be further written as

$$\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \mathcal{L}_2(\theta),$$

where

$$\begin{aligned}
\mathcal{L}_1(\theta) = &-\frac{1}{2N}\Big[-\ln|\widetilde{\Phi}^*\widetilde{\Phi}^{-1}| + \mathrm{tr}(\widetilde{\Phi}^*\widetilde{\Phi}^{-1}) - N(k+1)\Big] - \frac{1}{2NT}\mathrm{tr}\Big[\Psi(\rho)'\widetilde{\Sigma}_{zz}^{-1}\Psi(\rho)\Big] \\
&- \frac{1}{2N}\ln\left|I_r + \Sigma_{ff}^{1/2}\Lambda'\Phi^{-1}\Lambda\Sigma_{ff}^{1/2}\right| - (\rho-\rho^*)^2\frac{1}{2NT}\mathrm{tr}\Big[\dot{\mathbf{Y}}_{-1}M_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1}\widetilde{\Sigma}_{zz}^{-1}\Big],
\end{aligned}$$

and

$$\mathcal{L}_2(\theta) = \frac{1}{2N}\ln\left|I_r + \Sigma_{ff}^{*1/2}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*\Sigma_{ff}^{*1/2}\right| + \frac{1}{2N}\mathrm{tr}\Big[\widetilde{\Sigma}_{\epsilon\epsilon}^*\widetilde{\Phi}^{-1}\widetilde{\Lambda}\widetilde{G}\widetilde{\Lambda}'\widetilde{\Phi}^{-1}\Big] - \mathcal{R}(\theta).$$

The first term of $\mathcal{L}_2(\theta)$ only involves the underlying true values, so it is easy to show that this term is $O(\ln N/N)$. The second term is $O_p(N^{-1})$ since $\widetilde{\Sigma}_{\epsilon\epsilon}^*\widetilde{\Phi}^{-1} \leq C\widetilde{\Sigma}_{\epsilon\epsilon}^*$ for some large constant C by the boundedenss of β and β^* and Assumption E, and the latter term is further bounded by $\tilde{C}I_N$ for some $\tilde{C}$ by Assumption C. The last term is $o_p(1)$ uniformly on Θ due to Lemma 1. Given these results, we have $\mathcal{L}_2(\theta) = o_p(1)$ uniformly on Θ. Since $\widehat{\theta}$ maximizes $\mathcal{L}(\theta)$, $\mathcal{L}(\widehat{\theta}) \geq \mathcal{L}(\theta^*)$, implying $\mathcal{L}_1(\widehat{\theta}) \geq \mathcal{L}_1(\theta^*) + \mathcal{L}_2(\theta^*) - \mathcal{L}_2(\widehat{\theta})$. But $\mathcal{L}_1(\theta^*) = 0$. Therefore, $\mathcal{L}_1(\widehat{\theta}) \geq -2\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = o_p(1)$. However, all the four terms in $\mathcal{L}_1(\theta)$ are non-positive. This leads to

$$(\widehat{\rho}-\rho^*)^2\frac{1}{2NT}\mathrm{tr}\Big[\dot{\mathbf{Y}}_{-1}M_{\dot{F}^*}\dot{\mathbf{Y}}'_{-1}\widetilde{\Sigma}_{zz}^{-1}\Big] = o_p(1). \tag{B.3}$$

$$\frac{1}{2N}\Big[-\ln|\widetilde{\Phi}^*\widetilde{\Phi}^{-1}| + \mathrm{tr}(\widetilde{\Phi}^*\widetilde{\Phi}^{-1}) - N(k+1)\Big] = o_p(1). \tag{B.4}$$

$$\frac{1}{2NT}\mathrm{tr}\Big[\Psi(\rho)'\widetilde{\Sigma}_{zz}^{-1}\Psi(\rho)\Big] = o_p(1). \tag{B.5}$$

Consider $Y_{it} = \alpha_i^* + \rho^* Y_{it-1} + X_{it}'\beta^* + \kappa_i^{*\prime} g_t^* + e_{it}$. Substitute $X_{it} = \nu_i^* + \gamma_i^{*\prime} h_t^* + v_{it}$ into the expression of Y_{it},

$$Y_{it} = (\alpha_i^* + \nu_i^{*\prime}\beta^*) + \rho^* Y_{it-1} + \kappa_i^{*\prime} g_t^* + \beta^{*\prime}\gamma_i^{*\prime} h_t^* + (e_{it} + v_{it}'\beta^*).$$

We can transform the above autoregressive expression to the following moving average one:

$$Y_{it-1} = \underbrace{\frac{\alpha_i^* + \nu_i^{*\prime}\beta^*}{1-\rho^*} + \kappa_i^{*\prime}\sum_{s=0}^{\infty}\rho^{*s} g_{t-s-1}^* + \beta^{*\prime}\gamma_i^{*\prime}\sum_{s=0}^{\infty}\rho^{*s} h_{t-s-1}^*}_{\widetilde{\mathbf{b}}_{it-1}} + \underbrace{\sum_{s=0}^{\infty}\rho^{*s}(e_{it-s-1} + v_{it-s-1}'\beta^*)}_{\widetilde{\mathbf{u}}_{it-1}}. \quad \text{(B.6)}$$

Let $\widetilde{\mathbf{B}}_{t-1}$ and $\widetilde{\mathbf{U}}_{t-1}$ be defined similarly as $\mathbf{Y}_{t-1}$, and $\widetilde{\mathbf{B}}_{-1}$ and $\widetilde{\mathbf{U}}_{-1}$ be defined similarly as $\mathbf{Y}_{-1}$. By definition, $\dot{\mathbf{Y}}_{-1} = \dot{\widetilde{\mathbf{B}}}_{-1} + \dot{\widetilde{\mathbf{U}}}_{-1}$. Consider $\frac{1}{2NT}\operatorname{tr}[\dot{\mathbf{Y}}_{-1} M_{\dot{F}^*} \dot{\mathbf{Y}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}]$, which can be written as

$$\begin{aligned}\frac{1}{2NT}\operatorname{tr}[\dot{\mathbf{Y}}_{-1} M_{\dot{F}^*} \dot{\mathbf{Y}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}] &= \frac{1}{2NT}\operatorname{tr}\Big[\dot{\widetilde{\mathbf{B}}}_{-1} M_{\dot{F}^*} \dot{\widetilde{\mathbf{B}}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}\Big] + \frac{1}{NT}\operatorname{tr}\Big[\dot{\widetilde{\mathbf{B}}}_{-1} M_{\dot{F}^*} \dot{\widetilde{\mathbf{U}}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}\Big] \\ &\quad + \frac{1}{2NT}\operatorname{tr}\Big[\dot{\widetilde{\mathbf{U}}}_{-1} M_{\dot{F}^*} \dot{\widetilde{\mathbf{U}}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}\Big].\end{aligned}$$

However, by the similar arguments in the proof of Lemma B.2, we can show

$$\sup_{\theta\in\Theta}\left|\frac{1}{NT}\operatorname{tr}\Big[\dot{\widetilde{\mathbf{B}}}_{-1} M_{\dot{F}^*} \dot{\widetilde{\mathbf{U}}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}\Big]\right| = o_p(1).$$

and

$$\sup_{\theta\in\Theta}\left|\frac{1}{NT}\operatorname{tr}\Big[\dot{\widetilde{\mathbf{U}}}_{-1} M_{\dot{F}^*} \dot{\widetilde{\mathbf{U}}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}\Big] - \frac{1}{NT}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}E(\widetilde{\mathbf{u}}_{it-1}^2)\right| = o_p(1).$$

Given the above results, we have

$$\frac{1}{2NT}\operatorname{tr}[\dot{\mathbf{Y}}_{-1} M_{\dot{F}^*} \dot{\mathbf{Y}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}] = \frac{1}{2NT}\operatorname{tr}\Big[\dot{\widetilde{\mathbf{B}}}_{-1} M_{\dot{F}^*} \dot{\widetilde{\mathbf{B}}}_{-1}' \widetilde{\Sigma}_{zz}^{-1}\Big] + \frac{1}{2NT}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}E(\widetilde{\mathbf{u}}_{it-1}^2) + o_p(1),$$

which is greater than zero since the first term is non-negative, the second is strictly positive by Assumptions C.4 and E. This implies $\widehat{\rho} \xrightarrow{p} \rho^*$ by (B.3).

Further consider (B.4). Some straightforward computations lead to

$$\frac{1}{N}\operatorname{tr}(\widetilde{\Phi}^*\widetilde{\Phi}^{-1}) = \frac{1}{N}\operatorname{tr}(\Phi^*\Phi^{-1}) + (\beta-\beta^*)'\Big[\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\sum_{t=1}^{T}E(v_{it}v_{it}')\Big](\beta-\beta^*).$$

The above result, together with $\ln|\widetilde{\Phi}^*\widetilde{\Phi}^{-1}| = \ln|\Phi^*\Phi^{-1}|$ and (B.4), gives

$$\begin{aligned}&\frac{1}{2N}\Big[-\ln|\Phi^*\Phi^{-1}| + \operatorname{tr}(\Phi^*\Phi^{-1}) - N(k+1)\Big] \\ &\quad + \frac{1}{2N}(\beta-\beta^*)'\Big[\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\sum_{t=1}^{T}E(v_{it}v_{it}')\Big](\beta-\beta^*) = o_p(1).\end{aligned}$$

Let λ be a genetic eigenvalue of the matrix $\Phi^{-1/2}\Phi^*\Phi^{-1/2}$. Apparently, it is real. Given the fact that $f(\lambda) = \lambda - \ln\lambda - 1 \geq 0$ for all real λ, we have that the first expression on the left hand side of the preceding equation is non-negative. The second is obviously non-negative too. Given this, we have

$$\frac{1}{2N}(\beta-\beta^*)'\Big[\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\sum_{t=1}^{T}E(v_{it}v_{it}')\Big](\beta-\beta^*)=o_p(1),$$

implying $\widehat{\beta}\xrightarrow{p}\beta^*$. In addition, we also have

$$\frac{1}{2N}\Big[-\ln|\Phi^*\Phi^{-1}|+\operatorname{tr}(\Phi^*\Phi^{-1})-N(k+1)\Big]=o_p(1),$$

which, by the same arguments in Bai and Li (2014), gives $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\Phi}_i-\Phi_i^*\|^2=o_p(1)$, which is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}\Big[|\widehat{\sigma}_i^2-\sigma_i^{*2}|^2+\|\widehat{\Sigma}_{ii}-\Sigma_{ii}^*\|^2\Big]=o_p(1).$$

Next consider (B.5). Given that $\widehat{\rho}\xrightarrow{p}\rho^*$, $\widehat{\beta}\xrightarrow{p}\beta^*$ and $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\Phi}_i-\Phi_i^*\|^2=o_p(1)$, we can readily show that

$$\frac{1}{NT}\operatorname{tr}\Big[(\dot{F}^{*\prime}\dot{F}^*)^{1/2}\Lambda^{*\prime}\widehat{\Sigma}_{zz}^{-1}\Lambda^*(\dot{F}^{*\prime}\dot{F}^*)^{1/2}\Big]=o_p(1),$$

which is equivalent to $\frac{1}{N}\Lambda^{*\prime}\widehat{\Sigma}_{zz}^{-1}\Lambda^*=o_p(1)$ since the matrix $\Lambda^{*\prime}\widehat{\Sigma}_{zz}^{-1}\Lambda^*$ is semi-positive definite. By the Woodbury formula $\widehat{\Sigma}_{zz}^{-1}=\widehat{\Phi}^{-1}-\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}$, this result can be alternatively written as

$$\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\Lambda^*-\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^*=o_p(1).$$

Note that $\widehat{\mathcal{G}}\geq\widehat{\mathcal{H}}$ by the definitions of $\widehat{\mathcal{G}}$ and $\widehat{\mathcal{H}}$, then

$$\begin{aligned}0\leq\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\Lambda^*&-\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{H}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^*\\&\leq\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\Lambda^*-\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^*=o_p(1).\end{aligned}$$

The above result gives

$$\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\Lambda^*-\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{H}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^*=o_p(1).\tag{B.7}$$

By $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\Phi}_i-\Phi_i^*\|^2=o_p(1)$, we have $\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\Lambda^*-\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*=o_p(1)$. However, by Assumption B.2,

$$\tau_{\min}\Big(\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*\Big)\geq\tau_{\max}(\Phi^*)^{-1}\tau_{\min}\Big(\frac{1}{N}\Lambda^{*\prime}\Lambda^*\Big)>0,$$

implying that $\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*$ is a positive definite matrix. This implies that $\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{H}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^*=\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*+o_p(1)$ and $\frac{1}{N}\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}\widehat{\mathcal{G}}\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^*=\frac{1}{N}\Lambda^{*\prime}\Phi^{*-1}\Lambda^*+o_p(1)$. Substituting $\widehat{\mathcal{G}}=\widehat{\mathcal{H}}-\widehat{\mathcal{G}}\widehat{\Sigma}_{ff}^{-1}\widehat{\mathcal{H}}$ into the last expression, with some algebra manipulations, we conclude $\widehat{\mathcal{G}}=o_p(1)$, which further implies $\widehat{\mathcal{H}}=o_p(1)$ due to $\widehat{\mathcal{G}}=\widehat{\mathcal{H}}-\widehat{\mathcal{G}}\widehat{\Sigma}_{ff}^{-1}\widehat{\mathcal{H}}$.

Define $\mathcal{A}\equiv\Lambda^{*\prime}\widehat{\Phi}^{-1}\widehat{\Lambda}(\widehat{\Lambda}'\widehat{\Phi}^{-1}\widehat{\Lambda})^{-1}=\operatorname{diag}(\mathcal{A}_{11},\mathcal{A}_{22})$ with

$$\mathcal{A}_{11}=\Big[\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\kappa_i^*\widehat{\kappa}_i'\Big]\Big[\sum_{i=1}^{N}\frac{1}{\widehat{\sigma}_i^2}\widehat{\kappa}_i\widehat{\kappa}_i'\Big]^{-1},\quad\mathcal{A}_{22}=\Big[\sum_{i=1}^{N}\gamma_i^*\widehat{\Sigma}_{ii}^{-1}\widehat{\gamma}_i'\Big]\Big[\sum_{i=1}^{N}\widehat{\gamma}_i\widehat{\Sigma}_{ii}^{-1}\widehat{\gamma}_i'\Big]^{-1}$$

where the second equation is due to the special structures of Λ and Φ. The first order condition (A.5) can be written as

$$\begin{aligned}\widehat{\Sigma}_{ff} &= \mathcal{A}'M_{ff}\mathcal{A} + \mathcal{A}'\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*\widehat{\chi}_t'\widehat{\mathcal{H}}_N + \widehat{\mathcal{H}}_N\frac{1}{T}\sum_{t=1}^{T}\widehat{\chi}_t\dot{f}_t^{*\prime}\mathcal{A} \\ &+ \widehat{\mathcal{H}}\sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\Lambda}_i\widehat{\Phi}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}[\epsilon_{it}\epsilon_{jt}' - E(\epsilon_{it}\epsilon_{jt}')]\widehat{\Phi}_j^{-1}\widehat{\Lambda}_j'\widehat{\mathcal{H}} \\ &- \mathcal{A}'\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*(\widehat{\psi}-\psi^*)'\widehat{\xi}_t'\widehat{\mathcal{H}}_N - \widehat{\mathcal{H}}_N\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)\dot{f}_t^{*\prime}\mathcal{A} \\ &- \widehat{\mathcal{H}}_N\frac{1}{T}\sum_{t=1}^{T}\widehat{\chi}_t(\widehat{\psi}-\psi^*)'\widehat{\xi}_t'\widehat{\mathcal{H}}_N - \widehat{\mathcal{H}}_N\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)\widehat{\chi}_t'\widehat{\mathcal{H}}_N \\ &+ \widehat{\mathcal{H}}_N\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)(\widehat{\psi}-\psi^*)'\widehat{\xi}_t'\widehat{\mathcal{H}}_N - \widehat{\mathcal{H}} \\ &- \mathcal{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\Lambda}_i\widehat{\Phi}_i^{-1}\bar{\epsilon}_i\bar{\epsilon}_j'\widehat{\Phi}_j^{-1}\widehat{\Lambda}_j'\widehat{\mathcal{H}} \\ &+ \widehat{\mathcal{H}}\sum_{i=1}^{N}\sum_{j=1}^{N}\widehat{\Lambda}_i\widehat{\Phi}_i^{-1}\frac{1}{T}\sum_{t=1}^{T}E(\epsilon_{it}\epsilon_{jt}')\widehat{\Phi}_j^{-1}\widehat{\Lambda}_j'\widehat{\mathcal{H}},\end{aligned} \tag{B.8}$$

where $\widehat{\mathcal{H}}_N = N\cdot\widehat{\mathcal{H}}$. By the arguments in the proof of Lemma A.2 in Li, Cui and Lu (2019), together with $\widehat{\mathcal{H}} = o_p(1)$ and $\widehat{\rho} = \rho^* + o_p(1)$ and $\widehat{\beta} = \beta^* + o_p(1)$, we can show

$$\widehat{\Sigma}_{ff} = \mathcal{A}'M_{ff}\mathcal{A} + \|\widehat{\mathcal{H}}_N\|^2\cdot\Big(O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N})\Big) + o_p(1). \tag{B.9}$$

Given (B.7) and (B.9), by the arguments on Page 460 of Bai and Li (2012), we have $\mathcal{A} = O_p(1)$ and $\mathcal{A}^{-1} = O_p(1)$. There results, together with (B.7), give $\widehat{\mathcal{H}} = O_p(\frac{1}{N})$ and $\widehat{\mathcal{G}} = O_p(\frac{1}{N})$.

Consider the first order conditions on κ_i and γ_i. By the basic fact that

$$M_{zz}^{ij}(\widehat{\rho},\widehat{\beta}) = \frac{1}{T}\sum_{t=1}^{T}\Big[\Lambda_i^{*\prime}\dot{f}_t^* + \dot{\epsilon}_{it} - \dot{\mathbf{W}}_{it}(\widehat{\psi}-\psi^*)\Big]\Big[\Lambda_j^{*\prime}\dot{f}_t^* + \dot{\epsilon}_{jt} - \dot{\mathbf{W}}_{jt}(\widehat{\psi}-\psi^*)\Big]',$$

together with some algebra manipulations, we have

$$\begin{aligned}\widehat{\kappa}_i - \mathbf{R}_1\kappa_i^* &= M_{gg}^{-1}\widehat{\mathcal{G}}_{1N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\chi}_t\dot{g}_t^{*\prime}\kappa_i^* + M_{gg}^{-1}\widehat{\mathcal{G}}_1\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\Lambda_j^*\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*\dot{e}_{it} \\ &+ M_{gg}^{-1}\widehat{\mathcal{G}}_1\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\frac{1}{T}\sum_{t=1}^{T}\dot{\epsilon}_{jt}\dot{e}_{it} - M_{gg}^{-1}\widehat{\mathcal{G}}_{1N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)\dot{e}_{it} \\ &- M_{gg}^{-1}\widehat{\mathcal{G}}_{1N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\chi}_t(\widehat{\psi}-\psi^*)'\dot{W}_{it} - M_{gg}^{-1}\widehat{\mathcal{G}}_{1N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)\dot{g}_t^{*\prime}\kappa_i^* \\ &- M_{gg}^{-1}\widehat{\mathcal{G}}_1\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\Lambda_j^{*\prime}\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*(\widehat{\psi}-\psi^*)'\dot{W}_{it}\end{aligned} \tag{B.10}$$

$$+ M_{gg}^{-1}\widehat{\mathcal{G}}_{1N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)(\widehat{\psi}-\psi^*)'\dot{W}_{it}$$

with $\widehat{\mathcal{G}}_N = N\cdot\widehat{\mathcal{G}}$, $\widehat{\mathcal{G}}_{1N} = \mathbf{E}_1'\widehat{\mathcal{G}}_N$, and

$$\begin{aligned}\mathbf{R}_1 &= M_{gg}^{-1}\widehat{\mathcal{G}}_1\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^* M_{ff}\mathbf{E}_1 = M_{gg}^{-1}\mathbf{E}_1'(\widehat{\mathcal{H}}-\widehat{\mathcal{G}}\widehat{\Sigma}_{ff}^{-1}\widehat{\mathcal{H}})\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^* M_{ff}\mathbf{E}_1\\ &= M_{gg}^{-1}\mathcal{A}_{11}'M_{gg} - M_{gg}^{-1}\mathbf{E}_1'\widehat{\mathcal{G}}\widehat{\Sigma}_{ff}^{-1}\mathcal{A}'M_{ff}\mathbf{E}_1.\end{aligned}$$

and

$$\begin{aligned}\widehat{\gamma}_i - \mathbf{R}_2\gamma_i^* &= M_{hh}^{-1}\widehat{\mathcal{G}}_{2N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\chi}_t\dot{h}_t^{*\prime}\gamma_i^* + M_{hh}^{-1}\widehat{\mathcal{G}}_2\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\Lambda_j^{*\prime}\frac{1}{T}\sum_{t=1}^{T}\dot{f}_t^*\dot{v}_{it}' \\ &\quad + M_{hh}^{-1}\widehat{\mathcal{G}}_2\sum_{j=1}^{N}\widehat{\Lambda}_j\widehat{\Phi}_j^{-1}\frac{1}{T}\sum_{t=1}^{T}\dot{\epsilon}_{jt}\dot{v}_{it}' - M_{hh}^{-1}\widehat{\mathcal{G}}_{2N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)\dot{v}_{it}' \\ &\quad - M_{hh}^{-1}\widehat{\mathcal{G}}_{2N}\frac{1}{T}\sum_{t=1}^{T}\widehat{\xi}_t(\widehat{\psi}-\psi^*)\dot{h}_t^{*\prime}\gamma_i^*,\end{aligned} \tag{B.11}$$

with $\widehat{\mathcal{G}}_{2N} = N\widehat{\mathcal{G}}_2$ and

$$\mathbf{R}_2 = M_{hh}^{-1}\widehat{\mathcal{G}}_2\widehat{\Lambda}'\widehat{\Phi}^{-1}\Lambda^* M_{ff}\mathbf{E}_2 = M_{hh}^{-1}\mathcal{A}_{22}'M_{hh} - M_{hh}^{-1}\mathbf{E}_2'\widehat{\mathcal{G}}\widehat{\Sigma}_{ff}^{-1}\mathcal{A}'M_{ff}\mathbf{E}_2.$$

Consider (B.10). We use $I_{i1}, I_{i2}, \ldots, I_{i8}$ to denote the eight terms on the right hand side of (B.10). By the Cauchy-Schwarz inequality,

$$\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\kappa}_i - \mathbf{R}_1\kappa_i^*\|^2 \le 8\frac{1}{N}\sum_{i=1}^{N}\Big(\|I_{i1}\|^2 + \|I_{i2}\|^2 + \cdots + \|I_{i8}\|^2\Big).$$

By $\widehat{\mathcal{H}}_N = O_p(1)$ and $\widehat{\mathcal{G}}_N = O_p(1)$, Lemma B.3 in the online supplement indicates that $\frac{1}{N}\sum_{i=1}^{N}\|I_{il}\|^2 = o_p(1)$ for each $l = 1, 2, \ldots, 8$. Therefore, $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\kappa}_i - \mathbf{R}_1\kappa_i^*\|^2 = o_p(1)$. Similarly, we can show that $\frac{1}{N}\sum_{i=1}^{N}\|\widehat{\gamma}_i - \mathbf{R}_2\gamma_i^*\|^2 = o_p(1)$. Given this, we have

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\|\widehat{\kappa}_i - \mathbf{R}_1\kappa_i^*\|^2 + \|\widehat{\gamma}_i - \mathbf{R}_2\gamma_i^*\|^2\Big) \xrightarrow{p} 0.$$

This completes the proof. □

The proofs of Theorems 5.1, 5.2, 5.3 and 6.1 are given in the online supplement.

Appendix C: Simulation results based on the mimic data

This appendix consist of two parts. The first part adds the Monte Carlo simulation that mimics the real data in Section 9. The second part provides the empirical results under different r_2 values.

Appendix C.1: Monte Carlo simulation based on the real data

Since the explanatory variables are observed, we use these observed data in simulations to mimic the real data to the largest extent. Given the data of explanatory variables, we only need to generate the pseudo data of the dependent variables. More specifically, the Y data are generated according to

$$\widetilde{Y}_{it} = \widehat{\rho}\widetilde{Y}_{it-1} + X_{it}'\widehat{\beta} + \widehat{\kappa}_i'\widehat{g}_t + \varepsilon_{it}.$$

where $\widehat{\rho}, \widehat{\beta}, \widehat{\kappa}_i$ and $\widehat{g}_t$ are the QML estimators. Here we use the symbol with tilde to denote the pseudo dependent variables. We consider four ways to generated ε_{it}.

1. $\varepsilon_{it} = \widehat{\sigma}_i \varsigma_{it}$, where $\widehat{\sigma}_i^2$ is the QML estimator and ς_{it} is a standard normal variable;
2. $\varepsilon_{it} = \widehat{e}_{it}\varsigma_{it}$, where $\widehat{e}_{it}$ is the residual of e_{it} and ς_{it} is defined the same as above;
3. Let $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$. ε_t is the bootstrapped value, which is drawn from the pool $\{\widehat{e}_1, \widehat{e}_2, \dots, \widehat{e}_T\}$ with $\widehat{e}_t = (\widehat{e}_{1t}, \dots, \widehat{e}_{Nt})'$.
4. Let b be the size of block and $\widehat{e}_t^\star = (\widehat{e}_t, \widehat{e}_{t+1}, \dots, \widehat{e}_{t+b-1})$. For each t, if it satisfies $t = mb + 1$ with m being an integer, the next b-period errors $(\varepsilon_t, \dots, \varepsilon_{t+b-1})$ are drawn from the pool $\{\widehat{e}_1^\star, \dots, \widehat{e}_{T-b+1}^\star\}$. We set $b = 3$ in simulations.

The first two ways are the Monte Carlo methods and the next two are the bootstrap methods. For the first set, only the information of cross sectional heteroskedasticity is used. For the second set, it is possible to additionally keep the serial heteroskedasticity in error. However, the first two sets discard the cross sectional and serial correlations among the residuals $\widehat{e}_{it}$. For the third set, note that the bootstrap procedure is based on the N-dimensional vector, so the cross sectional correlation is maintained. For the fourth set, since we use the block bootstrap procedure, the cross sectional correlation and partial serial correlation are maintained.

Our simulations are conducted with 1000 repetitions. To evaluate the finite sample performance, we calculate the estimated coefficients, which are the mean of the 1000 estimated coefficients, the estimated SDs, which are the mean of the estimated standard deviations, and the simulated SDs, which are the standard deviations of the 1000 estimated coefficient. The distances of the estimated means and the QML estimators indicate the magnitude of the bias. In addition, the simulated SD is the precise SD values because it is obtained from the 1000 estimated coefficients. The estimated SD is the nominal one because it is calculated based on the asymptotic theory. So the ratios of the simulated SDs relative to the estimated SDs denote magnitudes of distortion in the limited sample size.

Table C.1: Estimation results based on the mimic data

		Lag	Market	Infra	Wage	Open	Govern	Labor
	True values	0.7093	0.0439	0.0622	-2.7830	0.0222	0.4965	0.0088
	SD	0.0220	0.0208	0.0257	0.9581	0.0039	0.1705	0.0040
Set One	Estimated values	0.7093***	0.0445*	0.0625**	-2.7354**	0.0221***	0.4867**	0.0087*
	Estimated SD	0.0023	0.0189	0.0229	0.7268	0.0027	0.1487	0.0038
	Simulated SD	0.0027	0.0232	0.0283	0.8850	0.0032	0.1816	0.0046
Set Two	Estimated values	0.7093***	0.0435*	0.0628**	-2.6868**	0.0222***	0.5222**	0.0084*
	Estimated SD	0.0031	0.0184	0.0216	0.8190	0.0028	0.1801	0.0035
	Simulated SD	0.0050	0.0242	0.0285	1.1847	0.0041	0.2554	0.0047
Set Three	Estimated values	0.7084***	0.0427*	0.0614*	-2.7553**	0.0224***	0.5027*	0.0091
	Estimated SD	0.0023	0.0190	0.0222	0.7102	0.0026	0.1543	0.0039
	Simulated SD	0.0094	0.0265	0.0327	1.1910	0.0042	0.2330	0.0087
Set Four	Estimated Values	0.7098***	0.0428*	0.0628*	-2.5227**	0.0221***	0.4922*	0.0089
	Estimated SD	0.0020	0.0206	0.0231	0.6490	0.0024	0.1364	0.0035
	Simulated SD	0.0068	0.0271	0.0344	1.0658	0.0038	0.2277	0.0085

Table C.1 presents the simulation results based on the mimic data. Given the above values, we compute the adjusted t-statistic as follows

$$\text{Adjusted t} = \frac{\text{Estimated SD}}{\text{Simulalted SD}} \frac{(2 * \text{Ture Value} - \text{Estimated Value})}{\text{SD}}.$$

Given the adjusted t-statistics, we relabel the significance levels on the coefficients, as shown in the above table.

From Table C.1, we can draw the following conclusions. First, the bias issue is not pronounced. Given the QML estimator as the underlying values, the estimated coefficients based on the simulated data in all the design types are around underlying true values, and the biases are too small to change the signs of estimated coefficients. Second, the adjusted t-statistics have changed a lot, which lead to the changes of the significance level. However, except for the regressor LABOR, the remaining coefficients are still significant under the 10% significance level. However, as shown in the main text, if we delete LABOR, the remaining coefficients still have correct signs and are statistically significant. Given these, we believe that although the limited sample size indeed has large effects on the empirical results, the main conclusions of the paper are not changed.

Appendix C.2: Empirical results under different r_2 values

We present the empirical results under different r_2 values. As point out in the main text, the data-driven methods fail to give the same estimated value. However, we find that the empirical results are little affected by the r_2 value, as seen in the following table. Due to this reason, we think that the conclusions drawn from Table 7 are reliable.

Table C.2: Additional empirical results of the QML method

Explanatory variables	$r_1 = 1$ $r_2 = 1$	$r_1 = 1$ $r_2 = 2$	$r_1 = 1$ $r_2 = 3$	$r_1 = 1$ $r_2 = 4$
FDI(-1)	0.7093***	0.7093***	0.7092***	0.7092***
	(0.0220)	(0.0220)	(0.0220)	(0.0220)
market	0.0439**	0.0439**	0.0438**	0.0438**
	(0.0208)	(0.0208)	(0.0208)	(0.0208)
infra	0.0622**	0.0622**	0.0624***	0.0623***
	(0.0257)	(0.0257)	(0.0257)	(0.0257)
wage	-2.7830***	-2.7830***	-2.7872***	-2.7861***
	(0.9581)	(0.9581)	(0.9577)	(0.9578)
open	0.0222***	0.0222***	0.0222***	0.0222***
	(0.0039)	(0.0039)	(0.0039)	(0.0039)
govern	0.4965***	0.4966***	0.4960***	0.4963***
	(0.1705)	(0.1705)	(0.1705)	(0.1705)
labor	0.0088**	0.0088**	0.0088**	0.0088**
	(0.0040)	(0.0040)	(0.0040)	(0.0040)

联系邮箱：xuqiy@163.com