Spatial dynamic panel data models with correlated random effects*

Liyao Li, Zhenlin Yang

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

Published in Journal of Econometrics (2021) 221 (2), 424-454. DOI: 10.1016/j.jeconom.2020.05.016

Abstract

In this paper, M-estimation and inference methods are developed for spatial dynamic panel data models with correlated random effects, based on short panels. The unobserved individual-specific effects are assumed to be correlated with the observed time-varying regressors linearly or in a linearizable way, giving the so-called correlated random effects model, which allows the estimation of effects of time-invariant regressors. The unbiased estimating functions are obtained by adjusting the conditional quasi-scores given the initial observations, leading to M-estimators that are consistent, asymptotically normal, and free from the initial conditions except the process starting time. By decomposing the estimating functions into sums of terms uncorrelated given idiosyncratic errors, a hybrid method is developed for consistently estimating the variance–covariance matrix of the M-estimators, which again depends only on the process starting time. Monte Carlo results demonstrate that the proposed methods perform well in finite sample. An empirical application on the political competition in China is presented.

Keywords: Adjusted quasi score, Dynamic panels, Correlated random effects, Initial-conditions, Martingale difference, Spatial effects, Short panels

* We are very much indebted to the two anonymous referees for their insightful and constructive comments that have led to significant improvements in the paper. We thank Editor, Oliver Linton, and the associate editor, and Ingmar Prucha, Badi Baltagi, Anil Bera, Lung-Fei Lee, Xiaodong Liu, and the participants of the XI World Conference of the Spatial Econometrics Association, June 13–15, 2017, Singapore, and the 8th Shanghai Econometrics Workshop, June 18–19, 2018, for their helpful comments. Zhenlin Yang gratefully acknowledges the financial support from Singapore Management University under Grant C244/MSS16E003.

INTRODUCTION

Consider the spatial dynamic panel data (SDPD) model where the spatial effects appear in the model in the forms of spatial lag (SL), space–time lag (STL), and spatial error (SE):

$$\begin{aligned} y_t &= \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + Z\gamma + \mu + \alpha_t 1_n + u_t, \\ u_t &= \lambda_3 W_3 u_t + v_t, \quad t = 1, 2, \ldots, T, \end{aligned} \tag{1.1}$$

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ and $v_t = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ are $n \times 1$ vectors of response values and idiosyncratic errors at time t, and $\{v_{it}\}$ are independent and identically distributed (iid) across i and t with mean zero and variance σ_v^2; the scalar parameter ρ characterizes the dynamic effect, λ_1 the SL effect, λ_2 the STL effect, and λ_3 the SE effect; $\{X_t\}$ are

$n \times p$ matrices containing values of p time-varying exogenous variables, Z is an $n \times q$ matrix containing the values of q time-invariant exogenous variables that may include the intercept, dummy variables (e.g., individuals' gender and race), etc.; β and γ are the usual regression coefficients; W_r, $r = 1, 2, 3$, are the given $n \times n$ spatial weight matrices; and μ is an $n \times 1$ vector of unobserved individual-specific effects, $\alpha = \{\alpha_t\}_{t=1}^T$ is a $T \times 1$ vector of unobserved time-specific effects, and 1_n is an $n \times 1$ vector of ones.

According to the way (μ, α) relate to $\{X_t\}$, the model is classified as: (*i*) *fixed effects* (FE) model if (μ, α) are correlated with X_t arbitrarily; (*ii*) *random effects* (RE) model if (μ, α) are uncorrelated with X_t; and (*iii*) *correlated random effects* (CRE) if (μ, α) are correlated with X_t linearly or in a *linearizable* way (see Footnote 1). M.-J. Lee (2002) called FE the *related effects*, and RE the *unrelated effects*. So, naturally the CRE can be called the *linearly related effects*. The term CRE is a tribute to Mundlak (1978), and Chamberlain (1982, 1984). In this work, we adopt the more popular terms: FE, RE and CRE, so that the SDPD models specified in (1.1) can be: FE-SDPD model, RE-SDPD model, or CRE-SDPD model.

Extensive discussions have appeared in the panel model literature, see, e.g. Cameron and Trivedi (2005), Wooldridge (2010), Baltagi (2013), and Hsiao (2014). The FE model has weaknesses (Cameron and Trivedi, 2005, p. 715–716): (*i*) it does not allow the estimation of the effects of time-invariant regressors, e.g., gender, race; (*ii*) while coefficients of time-varying regressors are estimable, these estimates may be very imprecise if most of the variation in a regressor is cross sectional rather than over time; (*iii*) prediction of the conditional mean is impossible, instead only changes in conditional mean caused by the changes in time-varying regressors can be predicted; and (*iv*) even coefficients of time-varying regressors may be difficult or theoretically impossible to identify in nonlinear models. The RE model overcomes these difficulties, but causal interpretation may then be unwarranted (Cameron and Trivedi, 2005, p. 715–716). The CRE model makes a compromise between the two: overcomes the weaknesses of the FE model and at the same time captures the linear or linearizable correlation between the 'effects' and the time-varying regressors.

The literature on spatial dynamic panels is fast expanding in recent years. However, most of the research on spatial dynamic panel data models focused on the long panels (with large n and large T), see, e.g., Yang et al. (2006), Mutl (2006), Yu et al. (2008), Yu and Lee (2010), Lee and Yu (2010a, 2012, 2014); Bai and Li (2015), Shi and Lee (2017), with relatively fewer works on the short panels, e.g., Elhorst (2010), Su and Yang (2015), Qu et al. (2016), Kuersteiner and Prucha (2018), and Yang (2018). Most of the works on short panels are on the FE-SDPD model, except Su and Yang (2015) who considered both FE- and RE-SDPD models but with only the SE effect built in the model. The general RE-SDPD model of the form (1.1) has not been formally considered, and the more general CRE-SDPD model specification has not even appeared in the literature. See Anselin et al. (2008), and Lee and Yu (2010b, 2015) for nice surveys on spatial panel data models. In this paper, we give a full treatment on the estimation and inference for the CRE-SDPD model, which includes the RE-SDPD model as a special case. We focus on the large-n and small-T setting, i.e., the *short panels*.

The CRE assumption renders a linear model for μ based on the observed X_t. We adopt the approach of Mundlak (1978) and specify that μ is linearly related to $\{X_t\}$ as,

$$\mu = \bar{X}\pi + \varepsilon, \tag{1.2}$$

where $\bar{X} = \frac{1}{T+1}\sum_{t=0}^T X_t$ and ε is an n-vector of $iid(0, \sigma_\varepsilon^2)$ errors, independent of v_t for all t. This can be extended to $\mu = X_0\pi_0 + X_1\pi_1 + \cdots + X_T\pi_T + \varepsilon$, as in Chamberlain (1982, 1984), a spatial Durbin form as in Debarsy (2012), or any *linearizable* relationship.[1]

Clearly, the advantages of the CRE-SDPD model over the FE-SDPD model are (*i*) it captures the typical correlation between μ and X_t and at the same time allows the effects of time-invariant variables Z, such as gender and race, be estimated, (*ii*) it may be more robust against possible existence of measurement errors and random coefficients, (*iii*) it makes the prediction of conditional mean possible as it works with levels rather than on differences series as in FE-approach, and (*iv*) it avoids the *incidental parameters* problem caused by the individual fixed effects, and hence may increase the estimation efficiency greatly.[2] Therefore, it is highly desirable to carry out a formal study on the CRE-SDPD model to provide a set of easy-to-use estimation and inference methods for applied researchers.

However, the CRE induces another set of errors ε, associated with the model for the individual-specific effects μ, besides the original set of idiosyncratic errors $\{v_t\}$, which further complicates the *initial conditions* in the model estimation and posts a much greater challenge in the estimation of the variance–covariance (VC) matrix of parameter estimates, compared with the FE-approach. The key problem is that in short panels, the error components in the disturbance cannot be separately estimated, rendering the *outer-product-of-martingale-difference* (OPMD) method of Yang (2018) for the FE-SDPD model unapplicable. The full quasi maximum likelihood (QML) approach of Su and Yang (2015) is also unapplicable as the usual way of modeling the initial observations based on a linear model may not be valid in the existence of spatial lag terms, as discussed in Yang (2018).

This paper contributes to the literature of dynamic short panel data models with spatial dependence by (*i*) providing an M-estimation method for the CRE-SDPD model, and (*ii*) introducing a new method for estimating the VC matrix of the

---

1 The intercept of Model (1.2) is absorbed into that of Model (1.1) for parameter identifiability (see Section 2.1 for details). By 'linearizable' we mean any CRE relationship that can be written as or approximated by a model linear in a finite number of parameters. To keep our exposition simple enough, we work with (1.2). For issue on parameter identification, see, e.g., Anselin et al. (2008, p.647), Elhorst (2012), and Lee and Yu (2016).

2 The FE-approach treats μ as unknown parameters, directly estimated or removed by some transformation. Hence, one period of the data is 'lost' which may consist of one third or one quarter of the 'usable' data if $T = 3$ or 4, making a significant difference in estimation efficiency.

M-estimators, of which both are **free from the initial conditions** except the process starting time $(-m)$. Our M-estimation strategy provides a complement to Yang (2018) for FE-SDPD model. It starts by adjusting the conditional quasi score function given the initial observations, to give a set of unbiased *estimation functions* or moment conditions that are free from the specification of the distribution of the y_0 (the *initial conditions*) apart from the process starting time $(-m)$. The vector of estimating functions is then written as a sum with the n summands being martingale differences with respect to individual-specific errors given idiosyncratic errors, so that a hybrid method that combines analytical derivations and the feasible sample analogues is proposed for estimating the VC matrix of the M-estimators. The resulting VC matrix estimator is also free from the initial conditions except the process starting time. The consistency and asymptotic normality of the M-estimators are established, and the consistency of the VC matrix estimator is also proved. Extensive Monte Carlo results show that, in finite samples, (*i*) proposed M-estimators perform very well, much superior to the conditional QML estimators (QMLE), (*ii*) proposed VC matrix estimator also performs well, and (*iii*) in case of the simple RE-SDPD model with only SE effect, the proposed M-estimator performs equally well as the full QMLE of Su and Yang (2015), but is numerically much more efficient. Without time-specific effects and if T goes large with n as in Yu et al. (2008), the proposed M-estimation method remains valid, and in this case, the usual method for estimating the VC matrix applies.

The CRE-SDPD model given in (1.1) is fairly general, embedding several important submodels obtained by dropping one or two spatial effects, none of which has been formally treated in the literature except Su and Yang (2015).[3] The proposed estimation and inference methods can easily be simplified to suit each special model of interest for a particular applied problem. Very interestingly, in a simple static panel data model, i.e., setting $\rho, \lambda_1, \lambda_2$ and λ_3 in Model (1.1) to zero, one can show that the CRE-estimators of β under Mundlak's and Chamberlain's specifications reduce to the usual FE-estimator (Cameron and Trivedi, 2005, Sec. 21.4.4, Krishnakumar, 2006; Hsiao, 2014, Sec. 3.4.2.1). However, we show that such an equivalence fails to hold once we move away from these formulations (e.g., a subset of X_t is correlated with μ), add dynamic terms, add spatial terms, etc.[4] These reinforce the need of a new set of estimation and inference methods for the general CRE-SDPD model.

The rest of the paper goes as follows. Section 2 introduces the M-estimator for the CRE-SDPD model and presents its asymptotic properties. Section 3 introduces the new method of estimating the VC matrix of the M-estimator. Section 4 presents Monte Carlo results. Section 5 presents an empirical application. Section 6 concludes the paper and offers some further discussions. All the technical proofs are relegated to the appendices.

2. Estimation of SDPD model with CRE

2.1. Conditional QML estimation of CRE-SDPD model

Let $B_r \equiv B_r(\lambda_r) = I_n - \lambda_r W_r$, $r = 1, 3$, and $B_2 \equiv B_2(\rho, \lambda_2) = \rho I_n + \lambda_2 W_2$. The CRE-SDPD model specified by (1.1)–(1.2) has reduced form, for $t = 1, \ldots, T$:

$$y_t = B_1^{-1}B_2 y_{t-1} + B_1^{-1}(X_t\beta + Z\gamma + \bar{X}\pi + \alpha_t 1_n) + B_1^{-1}\varepsilon + B_1^{-1}B_3^{-1}v_t. \tag{2.1}$$

Let $Y = (y_1', \ldots, y_T')'$, $Y_{-1} = (y_0', \ldots, y_{T-1}')'$, $X = (X_1', \ldots, X_T')'$, $D = (I_{T-1}\otimes 1_n',\ 0_{(T-1)}0_n')'$, and $\mathbf{X} = (1_{nT}, D, X, 1_T\otimes Z, 1_T\otimes\bar{X})$, where $\otimes$ denotes the Kronecker product, 1_k denotes a $k\times 1$ vector of ones, 0_k a $k\times 1$ vector of zeros, and I_k a $k\times k$ identity matrix. Further, let $\boldsymbol{\varepsilon} = 1_T\otimes\varepsilon$, $\mathbf{v} = (v_1', \ldots, v_T')'$, $\mathbf{W}_r = I_T\otimes W_r$, and $\mathbf{B}_r = I_T\otimes B_r$, $r = 1, 2, 3$. The reduced form (2.1) can be written compactly in matrix form:

$$Y = \mathbf{B}_1^{-1}\mathbf{B}_2 Y_{-1} + \mathbf{B}_1^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{B}_1^{-1}\boldsymbol{\varepsilon} + \mathbf{B}_1^{-1}\mathbf{B}_3^{-1}\mathbf{v}. \tag{2.2}$$

where $\boldsymbol{\beta} = (\breve{\alpha}', \beta', \gamma', \pi')'$ with $\mathtt{dim}(\boldsymbol{\beta}) = 2p + q + T$, and $\breve{\alpha} = (\alpha_T, \alpha_1 - \alpha_T, \ldots, \alpha_{T-1} - \alpha_T)'$.

Let $\mathbf{e} = \boldsymbol{\varepsilon} + \mathbf{B}_3^{-1}\mathbf{v}$ be the *composite error* vector. As $\{\varepsilon_i\}$ are $iid(0, \sigma_\varepsilon^2)$, $\{v_{it}\}$ are $iid(0, \sigma_v^2)$, and $\boldsymbol{\varepsilon}$ and $\mathbf{v}$ are independent, the variance–covariance (VC) matrix of $\mathbf{e}$ is:

$$\mathrm{Var}(\mathbf{e}) = \sigma_\varepsilon^2(J_T\otimes I_n) + \sigma_v^2(\mathbf{B}_3'\mathbf{B}_3)^{-1} = \sigma_v^2[\phi(J_T\otimes I_n) + (\mathbf{B}_3'\mathbf{B}_3)^{-1}] \equiv \sigma_v^2\Omega, \tag{2.3}$$

where $\phi = \sigma_\varepsilon^2/\sigma_v^2$. Let $\lambda = (\lambda_1, \lambda_2, \lambda_3)'$, $\theta = (\boldsymbol{\beta}', \rho, \lambda_1, \lambda_2)$ and $\psi = (\boldsymbol{\beta}', \sigma_v^2, \phi, \rho, \lambda')'$. Assume $\mathbf{X}$ is exogenously given. The **quasi** Gaussian loglikelihood, treating $\boldsymbol{\varepsilon}$ and $\mathbf{v}$ as normally distributed and y_0 as exogenously generated (conditioning on y_0), is

$$\ell_{\mathrm{SDPD}}(\psi) = -\frac{nT}{2}\log(2\pi\sigma_v^2) - \frac{1}{2}\log|\Omega(\phi, \lambda_3)| + \log|\mathbf{B}_1(\lambda_1)| - \frac{1}{2\sigma_v^2}e'(\theta)\Omega^{-1}(\phi, \lambda_3)e(\theta), \tag{2.4}$$

where $e(\theta) = \mathbf{B}_1 Y - \mathbf{B}_2 Y_{-1} - \mathbf{X}\boldsymbol{\beta}$, and $|\cdot|$ denotes the determinant of a square matrix.

Maximizing $\ell_{\mathrm{SDPD}}(\psi)$ gives the conditional QML estimator (QMLE) $\hat{\psi}_c$ of ψ. However, y_0 is not exogenous unless $m = 0$ (data collection starts when process starts) and ε and/or $\mathbf{v}$ may not be normal. Thus, $\ell_{\mathrm{SDPD}}(\psi)$ may not be a

[3] They considered an SDPD model with RE and spatial error (i.e., setting λ_1 and λ_2 to zero in Model (1.1), and setting π to zero in Model (1.2)), and a full QMLE by modeling the initial observations.

[4] We thank a referee for pointing out this simple connection and for raising the issue on its possible existence in general. See `Supplementary Appendix` at http://www.mysmu.edu/faculty/zlyang/ for details.

true loglikelihood function and maximizing it may not give a consistent estimate of ψ, in particular when $m > 0$ so that y_0 is endogenously generated. When T is also large, consistency may be achieved as ignoring the endogeneity in y_0 is asymptotically negligible. However, it may still suffer from the asymptotic bias problem. To solve these problems, we adopt the fundamental idea of Yang (2018) to 'correct' the quasi score functions to give a set of unbiased estimating functions or moment conditions.

2.2. M-Estimation of CRE-SDPD model

The quasi-score function, $S_{\text{SDPD}}(\psi) = \frac{\partial}{\partial\psi}\ell_{\text{SDPD}}(\psi)$, has the form:

$$S_{\text{SDPD}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\Omega^{-1}e(\theta), \\ \frac{1}{2\sigma_v^4}e'(\theta)\Omega^{-1}e(\theta) - \frac{nT}{2\sigma_v^2}, \\ \frac{1}{2\sigma_v^2}e'(\theta)\Omega^{-1}(J_T \otimes I_n)\Omega^{-1}e(\theta) - \frac{1}{2}\text{tr}[\Omega^{-1}(J_T \otimes I_n)], \\ \frac{1}{\sigma_v^2}e'(\theta)\Omega^{-1}Y_{-1}, \\ \frac{1}{\sigma_v^2}e'(\theta)\Omega^{-1}\mathbf{W}_1 Y - \text{tr}(\mathbf{B}_1^{-1}\mathbf{W}_1), \\ \frac{1}{\sigma_v^2}e'(\theta)\Omega^{-1}\mathbf{W}_2 Y_{-1}, \\ \frac{1}{2\sigma_v^2}e'(\theta)\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}e(\theta) - \frac{1}{2}\text{tr}(\Omega^{-1}\dot{\Omega}_{\lambda_3}), \end{cases} \tag{2.5}$$

where $\dot{\Omega}_{\lambda_3} = (\mathbf{B}_3'\mathbf{B}_3)^{-1}(\mathbf{B}_3'\mathbf{W}_3 + \mathbf{W}_3'\mathbf{B}_3)(\mathbf{B}_3'\mathbf{B}_3)^{-1}$, and $\text{tr}(\cdot)$ is the trace of a square matrix.

Let ψ_0 be the true value of ψ. A parametric quantity evaluated at the true parameters is denoted by adding a subscript '$_0$', e.g., B_{10}, Ω_0. The usual expectation and variance operators $\text{E}(\cdot)$ and $\text{Var}(\cdot)$ correspond to the true parameters. We derive $\text{E}[S_{\text{SDPD}}(\psi_0)]$, and show that the $(\rho, \lambda_1, \lambda_2)$-components of $\text{E}[S_{\text{SDPD}}(\psi_0)]$ are generally not zero, and that the same components of $\text{plim}_{n\to\infty}\frac{1}{nT}S_{\text{SDPD}}(\psi_0)$ are not zero. Thus, the conditional QMLE $\hat{\psi}_c$ cannot be consistent.

Assumption A. Assume (i) the processes started $m(\geq 0)$ periods before the start of data collection (0th period), and then evolve according to Models (1.1) and (1.2), (ii) y_{-m} and z_i are exogenous, and (iii) the individual specific effects μ are related to X_t linearly or in a linearizable way with additive errors ε independent of $v_t, t = -m+1, \ldots, T$.

It is easy to show that the $(\beta, \sigma_v^2, \phi, \lambda_3)$-components of $\text{E}[S_{\text{SDPD}}(\psi_0)]$ are all zero. The derivations of the other components are complicated by the additional time-invariant error component ε (induced by the CRE-formulation), which generates cumulative impact on y_t, $t = 0, 1, \ldots, T$. Recursive substitutions on (2.1) lead to the following important lemma.

Lemma 2.1. *Suppose Assumption A holds. Assume further that the errors $\{v_{it}\}$ in Model (1.1) are iid$(0, \sigma_{v0}^2)$ across i and t, the errors $\{\varepsilon_i\}$ in Model (1.2) are iid$(0, \sigma_{\varepsilon 0}^2)$, and $\{v_{it}\}$ and $\{\varepsilon_i\}$ are independent. If both B_{10}^{-1} and B_{30}^{-1} exist, then we have for $m \geq 1$,*

$$\text{E}(Y_{-1}\mathbf{e}') = \sigma_{v0}^2(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10}), \tag{2.6}$$

$$\text{E}(Y\mathbf{e}') = \sigma_{v0}^2(\phi_0\mathbf{C}_0 + \mathbf{D}_0), \tag{2.7}$$

where $\mathbf{C} \equiv \mathbf{C}(\rho, \lambda_1, \lambda_2, m)$, $\mathbf{C}_{-1} \equiv \mathbf{C}_{-1}(\rho, \lambda_1, \lambda_2, m)$, $\mathbf{D} \equiv \mathbf{D}(\rho, \lambda_1, \lambda_2, \lambda_3)$, and $\mathbf{D}_{-1} \equiv \mathbf{D}_{-1}(\rho, \lambda_1, \lambda_2, \lambda_3)$ are $nT \times nT$ matrices, defined as follows: $\mathbf{C} = [1_T \otimes (C_1', C_2', \ldots, C_T')]'$ and $\mathbf{C}_{-1} = [1_T \otimes (C_0', C_1', \ldots, C_{T-1}')]'$, where $C_t = (\sum_{i=0}^{t+m-1}\mathcal{B}^i)B_1^{-1}$ and $\mathcal{B} = B_1^{-1}B_2$;

$$\mathbf{D} = \begin{pmatrix} D_0 & 0 & \ldots & 0 & 0 \\ D_1 & D_0 & \ldots & 0 & 0 \\ D_2 & D_1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ D_{T-1} & D_{T-2} & \ldots & D_1 & D_0 \end{pmatrix} \quad \textit{and} \quad \mathbf{D}_{-1} = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ D_0 & 0 & \ldots & 0 & 0 \\ D_1 & D_0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ D_{T-2} & D_{T-3} & \ldots & D_0 & 0 \end{pmatrix},$$

where $D_t = \mathcal{B}^t B_1^{-1}(B_3'B_3)^{-1}$.

The results of Lemma 2.1 lead immediately to

$$\text{E}(\mathbf{e}'\Omega_0^{-1}Y_{-1}) = \text{tr}[(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10})\Omega_0^{-1}], \tag{2.8}$$

$$\text{E}(\mathbf{e}'\Omega_0^{-1}\mathbf{W}_1 Y) = \text{tr}[(\phi_0\mathbf{C}_0 + \mathbf{D}_0)\Omega_0^{-1}\mathbf{W}_1], \tag{2.9}$$

$$\text{E}(\mathbf{e}'\Omega_0^{-1}\mathbf{W}_2 Y_{-1}) = \text{tr}[(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10})\Omega_0^{-1}\mathbf{W}_2], \tag{2.10}$$

showing that the $(\rho, \lambda_1, \lambda_2)$-components of $\text{E}[S_{\text{SDPD}}(\psi_0)]$ are generally not zero, and more importantly, the $(\rho, \lambda_1, \lambda_2)$-components of $\text{plim}_{n\to\infty}\frac{1}{nT}S_{\text{SDPD}}(\psi_0)$ are not zero. Therefore, the conditional QMLE $\hat{\psi}_c$ cannot be consistent in general.

It is very interesting to note that these quantities are free from the specification of the distribution of the initial observations y_0, except the process starting time $(-m)$ embedded in the matrices $\mathbf{C}$ and $\mathbf{C}_{-1}$. Thus, these results provide a simple way to adjust the conditional quasi-scores, $S_{\text{SDPD}}(\psi_0)$, so as to give a set of unbiased estimating functions or moment conditions free from the initial conditions except m.[5] Unlike in the FE-approach of Yang (2018), where ε is differenced away, we need to account for its presence which is not trivial.

The adjusted quasi-score (AQS) functions are:

$$S^*_{\text{SDPD}}(\psi) = \begin{cases} \frac{1}{\sigma_v^2}\mathbf{X}'\Omega^{-1}e(\theta), \\ \frac{1}{2\sigma_v^4}e'(\theta)\Omega^{-1}e(\theta) - \frac{nT}{2\sigma_v^2}, \\ \frac{1}{2\sigma_v^2}e'(\theta)\Omega^{-1}(J_T\otimes I_n)\Omega^{-1}e(\theta) - \frac{1}{2}\text{tr}[\Omega^{-1}(J_T\otimes I_n)], \\ \frac{1}{\sigma_v^2}e'(\theta)\Omega^{-1}Y_{-1} - \text{tr}[(\phi\mathbf{C}_{-1}+\mathbf{D}_{-1})\Omega^{-1}], \\ \frac{1}{\sigma_v^2}e'(\theta)\Omega^{-1}\mathbf{W}_1 Y - \text{tr}[(\phi\mathbf{C}+\mathbf{D})\Omega^{-1}\mathbf{W}_1], \\ \frac{1}{\sigma_v^2}e'(\theta)\Omega^{-1}\mathbf{W}_2 Y_{-1} - \text{tr}[(\phi\mathbf{C}_{-1}+\mathbf{D}_{-1})\Omega^{-1}\mathbf{W}_2], \\ \frac{1}{2\sigma_v^2}e'(\theta)\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}e(\theta) - \frac{1}{2}\text{tr}(\Omega^{-1}\dot{\Omega}_{\lambda_3}). \end{cases} \tag{2.11}$$

It is easy to show that $\text{E}[S^*_{\text{SDPD}}(\psi_0)] = 0$, and that $\text{plim}_{n\to\infty}\frac{1}{nT}S^*_{\text{SDPD}}(\psi_0) = 0$. Solving the estimating equations $S^*_{\text{SDPD}}(\psi) = 0$ gives M-estimator $\hat{\psi}_{\text{M}}$, which is shown to be consistent and asymptotically normal under some regularity conditions in Theorems 2.1 and 2.2.

The equation solving process can be simplified by first solving the equations for $\boldsymbol{\beta}$ and σ_v^2 given $\delta = (\phi, \rho, \lambda')'$ to obtain the constrained M-estimators of $\boldsymbol{\beta}$ and σ_v^2 as

$$\hat{\boldsymbol{\beta}}(\delta) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}(\mathbf{B}_1 Y - \mathbf{B}_2 Y_{-1}), \tag{2.12}$$

$$\hat{\sigma}_v^2(\delta) = \frac{1}{nT}\hat{\mathbf{e}}'(\delta)\Omega^{-1}\hat{\mathbf{e}}(\delta), \tag{2.13}$$

where $\hat{\mathbf{e}}(\delta) = \mathbf{B}_1 Y - \mathbf{B}_2 Y_{-1} - \mathbf{X}\hat{\boldsymbol{\beta}}(\delta)$. Substituting them back into the last five components of the AQS functions gives the concentrated AQS functions $S^{*c}_{\text{SDPD}}(\delta)$ (see (B.1), Appendix B). Solving $S^{*c}_{\text{SDPD}}(\delta) = 0$, we obtain the unconstrained M-estimator $\hat{\delta}_{\text{M}}$ of δ, and the unconstrained M-estimators $\hat{\boldsymbol{\beta}}_{\text{M}} \equiv \hat{\boldsymbol{\beta}}(\hat{\delta}_{\text{M}})$ and $\hat{\sigma}^2_{v,\text{M}} \equiv \hat{\sigma}_v^2(\hat{\delta}_{\text{M}})$ of $\boldsymbol{\beta}$ and σ_v^2. Thus, $\hat{\psi}_{\text{M}} = (\hat{\boldsymbol{\beta}}'_{\text{M}}, \hat{\sigma}^2_{v,\text{M}}, \hat{\delta}'_{\text{M}})'$.

Remark 2.1. From the way that the AQS function is defined in (2.11), we see that the M-estimator $\hat{\psi}_{\text{M}}$ for the CRE-SDPD model specified by (1.1) and (1.2) is free from the specification of the distribution of y_0, except the value m that is unknown.

However, this does not pose a serious problem as (i) in practice one is often able to 'tell' roughly the value of m from the data, and (ii) $\hat{\psi}_{\text{M}}$ is quite robust against the changes in the value of m. See Elhorst (2010) and Su and Yang (2015) for similar remarks.[6]

2.3. Asymptotic properties of M-Estimator

To proceed with a formal study on the asymptotic properties of the proposed M-estimator, some generic notations are helpful: `blkdiag(···)` forms a block-diagonal matrix based on the given matrices, $\gamma_{\min}(\cdot)$ and $\gamma_{\max}(\cdot)$ denote the smallest and largest eigenvalues of a real symmetric matrix, and $\|\cdot\|$ denotes the Frobenius norm of a matrix.

Assumption B. The innovations v_{it} are iid for all i and t with $\text{E}(v_{it}) = 0$, $\text{Var}(v_{it}) = \sigma^2_{v0}$, and $\text{E}|v_{it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$. The innovations ε_i are iid for all i with $\text{E}(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2_{\varepsilon 0}$, and $\text{E}|\varepsilon_i|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$.

Assumption C. The parameter space $\boldsymbol{\Delta}$ of δ is compact, and the true parameter vector δ_0 lies in its interior.

Assumption D. The elements of (y_{-m}, Z, X_t), $t = -m+1, \ldots, 0, \ldots, T$, are uniformly bounded, and the $\lim_{n\to\infty}\frac{1}{nT}\mathbf{X}'\mathbf{X}$ exists and is nonsingular.

5 The full QMLEs of the regular dynamic panel data model of Hsiao et al. (2002) and the SE-SDPD model of Su and Yang (2015), where the initial observations are modeled, also depend on m.

6 Under simpler models with full QML estimation, Hsiao et al. (2002) recommended to estimate m together with the other common parameters; Su and Yang (2015) pointed out that m may not be separately identified unless the 'lag' parameter $\rho \neq 0$. In our general model, m may be identifiable unless the 'lag' parameters $(\rho, \lambda_1, \lambda_2)$ in $\mathcal{B}$ are all zero. We choose this practical approach to avoid additional numerical complications.

Assumption E. (i) For $r = 1, 2, 3$, the elements $w_{r,ij}$ of W_r are at most of order h_n^{-1}, uniformly in all i and j, and $w_{r,ii} = 0$ for all i; (ii) $h_n/n \to 0$ as $n \to \infty$; (iii) $\{W_r, r = 1, 2, 3\}$ and $\{B_{r0}^{-1}, r = 1, 3\}$ are uniformly bounded in both row and column sums; (iv) For $r = 1, 3$, $\{B_r^{-1}\}$ are uniformly bounded in either row or column sums, uniformly in λ_r in a compact parameter space $\boldsymbol{\Lambda}_r$, and $0 < \underline{c}_r \le \inf_{\lambda_r \in \boldsymbol{\Lambda}_r} \gamma_{\min}(B_r' B_r) \le \sup_{\lambda_r \in \boldsymbol{\Lambda}_r} \gamma_{\max}(B_r' B_r) \le \bar{c}_r < \infty$.

Assumption F. (i) $\sum_{i=0}^{\infty} \mathcal{B}_0^i$ exists and is uniformly bounded in both row and column sums, and (ii) $\sum_{i=0}^{\infty} \mathcal{B}^i$ is invertible for (λ_1, λ_2) in a neighborhood of $(\lambda_{10}, \lambda_{20})$.

Assumption B is standard in spatial panel data models with error components (see, e.g., Su and Yang, 2015). Assumption C is needed in establishing the consistency of $\hat{\delta}_{\mathrm{M}}$. Assumptions D and E($iv$) guarantee the existence and nonsingularity of $\lim_{n\to\infty} \frac{1}{nT}\mathbf{X}'\Omega^{-1}\mathbf{X}$, so that once δ is identified, the identifications of β and σ_v^2 follow. Assumption E parallels Assumption E of Yang (2018) and relates to Lee (2004). Allowing h_n to grow with n but at a slower rate is useful as it corresponds a spatial layout where the *degree of spatial dependence* increases with n. See Lee (2004) and Yang (2015) for related discussions. Due to the cumulative impact of ε_n from the past, we need Assumption F(i) to ensure that the initial observations y_0 have a proper stochastic behavior when $m = \infty$, e.g., $\frac{h_n}{n}[y_0'\Phi y_0 - \mathrm{E}(y_0'\Phi y_0)] = o_p(1)$ for Φ uniformly bounded in either row or column sums with elements of uniform order h_n^{-1}. Clearly, it is satisfied when $\|\mathcal{B}_0\| < 1$, giving $\sum_{i=0}^{\infty} \mathcal{B}_0^i = (I_n - \mathcal{B}_0)^{-1}$. A similar assumption is made in Yu et al. (2008). Assumption F(ii) is needed for feasible VC matrix estimation.

To establish the consistency of $\hat{\delta}_{\mathrm{M}}$, define $\bar{S}^*_{\mathrm{SDPD}}(\psi) = \mathrm{E}[S^*_{\mathrm{SDPD}}(\psi)]$, the population counter part of the AQS function. Given δ, the population AQS equations $\bar{S}^*_{\mathrm{SDPD}}(\psi) = 0$ are partially solved at $\bar{\boldsymbol{\beta}}(\delta) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}(\mathbf{B}_1\mathrm{E}Y - \mathbf{B}_2\mathrm{E}Y_{-1})$ and $\bar{\sigma}_v^2(\delta) = \frac{1}{n(T-1)}\mathrm{E}[\bar{\mathbf{e}}(\delta)'\Omega^{-1}\bar{\mathbf{e}}(\delta)]$, where $\bar{\mathbf{e}}(\delta) = \mathbf{e}(\theta)|_{\beta=\bar{\beta}(\delta)} = \mathbf{B}_1 Y - \mathbf{B}_2 Y_{-1} - \mathbf{X}\bar{\boldsymbol{\beta}}(\delta)$. Substituting $\bar{\boldsymbol{\beta}}(\delta)$ and $\bar{\sigma}_v^2(\delta)$ back into the last five components of $\bar{S}^*_{\mathrm{SDPD}}(\psi)$ leads to the population counter part of the concentrated AQS functions, which is denoted by $\bar{S}^{*c}_{\mathrm{SDPD}}(\delta)$ (see (B.2), Appendix B). It is easy to see that the M-estimator $\hat{\delta}_{\mathrm{M}}$ of δ_0 is a zero of $S^{*c}_{\mathrm{SDPD}}(\delta)$, and δ_0 is a zero of $\bar{S}^{*c}_{\mathrm{SDPD}}(\delta)$. Thus, by Theorem 5.9 of van der Vaart (1998), $\hat{\delta}_{\mathrm{M}}$ will be consistent for δ_0 if $\sup_{\delta\in\boldsymbol{\Delta}} \frac{1}{nT}\left\|S^{*c}_{\mathrm{SDPD}}(\delta) - \bar{S}^{*c}_{\mathrm{SDPD}}(\delta)\right\| \xrightarrow{p} 0$, and the following identification condition holds.

Assumption G. $\inf_{\delta:\, d(\delta,\delta_0)\ge\varepsilon} \left\|\bar{S}^{*c}_{\mathrm{SDPD}}(\delta)\right\| > 0$ *for every* $\varepsilon > 0$*, where* $d(\delta, \delta_0)$ *is a measure of distance between* δ_0 *and* δ.

Theorem 2.1. *Suppose Assumptions A–G hold. Assume further that* (i) $\gamma_{\max}[\mathrm{Var}(Y)]$ *and* $\gamma_{\max}[\mathrm{Var}(Y_{-1})]$ *are bounded, and* (ii) $\inf_{\delta\in\Delta}\gamma_{\min}\left[\mathrm{Var}(\mathbf{B}_1 Y - \mathbf{B}_2 Y_{-1})\right] \ge \underline{c}_y > 0$. *We have, as* $n \to \infty$, $\hat{\delta}_{\mathrm{M}} \xrightarrow{p} \delta_0$. *It follows that* $\hat{\psi}_{\mathrm{M}} \xrightarrow{p} \psi_0$.

To establish asymptotic normality of the proposed M-estimator $\hat{\psi}_{\mathrm{M}}$, the following representations of Y and Y_{-1} in terms of $\mathbf{y}_0 = 1_T \otimes y_0$ and $\mathbf{e}$ are very useful.

$$Y = \mathbb{Q}\mathbf{y}_0 + \boldsymbol{\eta} + \mathbb{S}\mathbf{e} \quad \text{and} \quad Y_{-1} = \mathbb{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1}\mathbf{e}, \tag{2.14}$$

where $\mathbb{Q} = \texttt{blkdiag}(\mathcal{B}_0, \mathcal{B}_0^2, \ldots, \mathcal{B}_0^T)$, $\mathbb{Q}_{-1} = \texttt{blkdiag}(I_n, \mathcal{B}_0^1, \ldots, \mathcal{B}_0^{T-1})$, $\mathbb{S} = \mathbb{R}\mathbf{B}_{10}^{-1}$, $\mathbb{S}_{-1} = \mathbb{R}_{-1}\mathbf{B}_{10}^{-1}$, $\boldsymbol{\eta} = \mathbb{S}\mathbf{X}\boldsymbol{\beta}_0$, $\boldsymbol{\eta}_{-1} = \mathbb{S}_{-1}\mathbf{X}\boldsymbol{\beta}_0$,

$$\mathbb{R} = \begin{pmatrix} I_n & 0 & 0 & \ldots & 0 \\ \mathcal{B}_0 & I_n & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{B}_0^{T-1} & \mathcal{B}_0^{T-2} & \mathcal{B}_0^{T-3} & \ldots & I_n \end{pmatrix} \quad \textit{and} \quad \mathbb{R}_{-1} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ I_n & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{B}_0^{T-2} & \mathcal{B}_0^{T-3} & \mathcal{B}_0^{T-4} & \ldots & 0 \end{pmatrix}.$$

By the representations given in (2.14) the AQS vector at ψ_0 is written as

$$S^*_{\mathrm{SDPD}}(\psi_0) = \begin{cases} \Pi_{10}'\mathbf{e}, \\ \mathbf{e}'\Phi_{10}\mathbf{e} - \frac{nT}{2\sigma_{v0}^2}, \\ \mathbf{e}'\Phi_{20}\mathbf{e} - \frac{1}{2}\mathrm{tr}[\Omega_0^{-1}(J_T \otimes I_n)], \\ \mathbf{e}'\Psi_{10}\mathbf{y}_0 + \Pi_{20}'\mathbf{e} + \mathbf{e}'\Phi_{30}\mathbf{e} - \mathrm{tr}[(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10})\Omega_0^{-1}], \\ \mathbf{e}'\Psi_{20}\mathbf{y}_0 + \Pi_{30}'\mathbf{e} + \mathbf{e}'\Phi_{40}\mathbf{e} - \mathrm{tr}[(\phi_0\mathbf{C}_0 + \mathbf{D}_0)\Omega_0^{-1}\mathbf{W}_1], \\ \mathbf{e}'\Psi_{30}\mathbf{y}_0 + \Pi_{40}'\mathbf{e} + \mathbf{e}'\Phi_{50}\mathbf{e} - \mathrm{tr}[(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10})\Omega_0^{-1}\mathbf{W}_2], \\ \mathbf{e}'\Phi_{60}\mathbf{e} - \frac{1}{2}\mathrm{tr}(\Omega_0^{-1}\dot{\Omega}_{\lambda_3 0}), \end{cases} \tag{2.15}$$

where $\Pi_1 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{X}$, $\Pi_2 = \frac{1}{\sigma_v^2}\Omega^{-1}\boldsymbol{\eta}_{-1}$, $\Pi_3 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{W}_1\boldsymbol{\eta}$, and $\Pi_4 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{W}_2\boldsymbol{\eta}_{-1}$; $\Phi_1 = \frac{1}{2\sigma_v^4}\Omega^{-1}$, $\Phi_2 = \frac{1}{2\sigma_v^2}\Omega^{-1}(J_T \otimes I_n)\Omega^{-1}$, $\Phi_3 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbb{S}_{-1}$, $\Phi_4 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{W}_1\mathbb{S}$, $\Phi_5 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{W}_2\mathbb{S}_{-1}$, and $\Phi_6 = \frac{1}{2\sigma_v^2}\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}$; $\Psi_1 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbb{Q}_{-1}$, $\Psi_2 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{W}_1\mathbb{Q}$, and $\Psi_3 = \frac{1}{\sigma_v^2}\Omega^{-1}\mathbf{W}_2\mathbb{Q}_{-1}$.

As $\mathbf{e} = \boldsymbol{\varepsilon} + \mathbf{B}_{30}^{-1}\mathbf{v}$, Y and Y_{-1} are further represented as

$$Y = \mathbb{Q}\mathbf{y}_0 + \boldsymbol{\eta} + \mathbb{S}\boldsymbol{\varepsilon} + \mathbb{B}\mathbf{v} \quad \text{and} \quad Y_{-1} = \mathbb{Q}_{-1}\mathbf{y}_0 + \boldsymbol{\eta}_{-1} + \mathbb{S}_{-1}\boldsymbol{\varepsilon} + \mathbb{B}_{-1}\mathbf{v}, \tag{2.16}$$

where $\mathbb{B} = \mathbb{S}\mathbf{B}_{30}^{-1}$ and $\mathbb{B}_{-1} = \mathbb{S}_{-1}\mathbf{B}_{30}^{-1}$. Thus, $S^*_{\text{SDPD}}(\psi_0)$ are further expressed in terms of $\mathbf{v}$, ε and y_0. Using backward substitution on Eq. (2.1), we have, for $m \geq 1$:

$$
\begin{aligned}
y_0 &= \mathcal{B}^m y_{-m} + \sum_{k=0}^{m-1} \mathcal{B}^k B_1^{-1} \mathbf{X}_{-k}\boldsymbol{\beta} + \sum_{k=0}^{m-1} \mathcal{B}^k B_1^{-1}\varepsilon + \sum_{k=0}^{m-1} \mathcal{B}^k B_1^{-1} B_3^{-1} v_{-k} \\
&\equiv \eta_m + K_m\varepsilon + V_m,
\end{aligned} \tag{2.17}
$$

where $\eta_m = \mathcal{B}^m y_{-m} + \sum_{k=0}^{m-1} \mathcal{B}^k B_1^{-1}\mathbf{X}_{-k}\boldsymbol{\beta}$, being the mean of y_0 given Y_{-m} and $\mathbf{X}_{-k}$, $k = 0, 1, \ldots, m$ and thus exogenous; $K_m = \sum_{k=0}^{m-1} \mathcal{B}^k B_1^{-1}$; $V_m = \sum_{k=0}^{m-1} \mathcal{B}^k B_1^{-1} B_3^{-1} v_{-k}$; and $\mathbf{X}_{-k}$ collects all the regressors' values at the $(-k)$th period. Obviously, V_m is independent of ε and v_t, $t = 1, 2, \ldots, T$. Therefore, the components of $S^*_{\text{SDPD}}(\psi_0)$ are linear combinations of terms linear-quadratic in $\mathbf{v}$, linear-quadratic in ε, and bilinear in ε and $\mathbf{v}$, in ε and V_m, and in $\mathbf{v}$ and V_m. These lead to a simple way for establishing the asymptotic normality of the AQS vector $S^*_{\text{SDPD}}(\psi_0)$, and thus the asymptotic normality of the proposed M-estimator.

Theorem 2.2. *Under assumptions of Theorem 2.1, we have, as $n \to \infty$,*

$$
\sqrt{nT}(\hat{\psi}_{\mathtt{M}} - \psi_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty} \Sigma^{*-1}_{\text{SDPD}}(\psi_0)\Gamma^*_{\text{SDPD}}(\psi_0)\Sigma^{*-1}_{\text{SDPD}}(\psi_0)\big],
$$

where $\Sigma^*_{\text{SDPD}}(\psi_0) = -\frac{1}{nT}\mathrm{E}[\frac{\partial}{\partial\psi'}S^*_{\text{SDPD}}(\psi_0)]$ and $\Gamma^*_{\text{SDPD}}(\psi_0) = \frac{1}{nT}\mathrm{Var}[S^*_{\text{SDPD}}(\psi_0)]$, both assumed to exist and $\Sigma^*_{\text{SDPD}}(\psi_0)$ to be positive definite, for sufficiently large n.

3. Robust VC matrix estimation of M-Estimators

The expected negative Hessian matrix $\Sigma^*_{\text{SDPD}}(\psi)$ can be consistently estimated by its observed counter part $\widehat{\Sigma}^*_{\text{SDPD}} = -\frac{1}{nT}\frac{\partial}{\partial\psi'}S^*_{\text{SDPD}}(\psi)|_{\psi=\hat{\psi}_{\mathtt{M}}}$. The detailed expression of $\frac{\partial}{\partial\psi'}S^*_{\text{SDPD}}(\psi)$ is given in Appendix B. Unfortunately, none of the existing methods can be used to estimate $\Gamma^*_{\text{SDPD}}(\psi_0)$. The traditional plug-in method requires the unconditional distribution of y_0 or a valid model for y_0 when T is fixed, of which neither is plausible as the unconditional distribution involves unobservables and a valid model seems very difficult (if not impossible) to formulate, in particular when the model contains spatial lag terms (Yang, 2018). To overcome these difficulties in estimating the VC matrix for the FE-SDPD model, Yang (2018) proposed an *outer-product-of-martingale-difference* (OPMD) method, where the AQS function of the FE-SDPD model is decomposed into a sum of vector martingale difference (MD) sequences so that the average of the outer products of the MDs gives a consistent estimate of the VC matrix of that AQS function. However, this OPMD method does not apply to our CRE-SDPD model due to the existence of two error components ε and v_t.

A new method of feasible and consistent VC matrix estimation is needed. We see that the representations given in (2.15) are crucial in obtaining such an estimate. From (2.15) we see that the AQS function contains three types of elements:

$$
\Pi'\mathbf{e}, \quad \mathbf{e}'\Phi\mathbf{e}, \quad \text{and} \quad \mathbf{e}'\Psi\mathbf{y}_0,
$$

where Π, Φ, and Ψ are nonstochastic matrices depending on ψ_0 with Π being $nT \times \dim(\boldsymbol{\beta})$ or $nT \times 1$, and Φ and Ψ being $nT \times nT$. The closed form expressions for the variances of $\Pi'e$ and $e'\Phi e$ can be derived but the plug-in method cannot be applied as their analytical expressions involve the 3rd and 4th moments of both ε_i and v_{it}, which cannot be consistently estimated simultaneously with a fixed T. Furthermore, the closed-form expressions for the variance of $\mathbf{e}'\Psi\mathbf{y}_0$ and its covariances with $\Pi'e$ and $e'\Phi e$ depend on the past values of the regressors and the process starting positions, which are unobserved. Thus, the plug-in method based on the full analytical expression of Γ^*_{SDPD} does not work either in this case.

As neither the traditional plug-in method nor the OPMD method works for estimating Γ^*_{SDPD}, an alternative method must be developed. To fix the idea, we again, as in Yang (2018), endeavor to decompose $S^*_{\text{SDPD}}(\psi_0)$ into a sum $\sum_{i=1}^n \mathbf{g}_i$ such that $\{\mathbf{g}_i\}$ possess some 'desirable properties' and a feasible estimator for Γ^*_{SDPD} can thus be developed. Difficulty lies in the fact that the composite error, $e_t = \varepsilon + B_3^{-1}v_t$, consists of two components v_t and ε, which cannot be 'consistently' estimated simultaneously due to the fixed T nature. Thus, although $\{\mathbf{g}_i\}$ can be written as MD sequences separately in terms of ε and v_t, it cannot be estimated this way as only the estimates $\hat{e}_t$ are available. However, if the decomposition $\sum_{i=1}^n \mathbf{g}_i$ is such that $\mathbf{g}_i$ and $\mathbf{g}_j$, $j \neq i$, are uncorrelated with respect to ε for given $\{v_t\}$, then a *hybrid method*, i.e., combining sample analogue and the analytical expressions, can be developed for estimating Γ^*_{SDPD}. Note that based on $S^*_{\text{SDPD}}(\psi_0) = \sum_{i=1}^n \mathbf{g}_i$,

$$
\Gamma^*_{\text{SDPD}} = \frac{1}{nT}\mathrm{E}[S^*_{\text{SDPD}}(\psi_0)S^{*\prime}_{\text{SDPD}}(\psi_0)] = \frac{1}{nT}\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}_i') + \frac{1}{nT}\sum_{i=1}^n\sum_{j=1, j\neq i}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}_j'). \tag{3.1}
$$

The single-sum term $\sum_{i=1}^n \mathrm{E}(\mathbf{g}_i\mathbf{g}_i')$ may be estimated by its sample analogue $\sum_{i=1}^n \hat{\mathbf{g}}_i\hat{\mathbf{g}}_i'$, where $\hat{\mathbf{g}}_i$ is the plug-in estimate of $\mathbf{g}_i$ by plugging $\hat{\psi}_{\mathtt{M}}$ and $\hat{e}_{it}$ in $\mathbf{g}_i$. For the double-sum term, we derive 'semi-analytical' expressions in terms of ψ_0, $\mu^{(3)}_{v0}$ and

$\mu_{v0}^{(4)}$ (the 3rd and 4th moments of the idiosyncratic error v_{it}), and the initial values y_0, so that a mixture of the plug-in and sample analogue methods can be applied. We choose $\hat{\mathbf{g}}_i$ in such a way that this method is free from the specifications of the distributions of the initial observations, and that it involves only $\mu_{v0}^{(3)}$ and $\mu_{v0}^{(4)}$, of which estimates are readily available. The latter is achieved by transforming y_0 so that the transformed y_0 has an error structure similar to e_t:

$$y_0^* = K_m^{-1} y_0 = \varepsilon + K_m^{-1}\eta_m + K_m^{-1} V_m \equiv \varepsilon + \eta_m^* + V_m^*, \tag{3.2}$$

see (2.17). Clearly, making ε 'stand out' in the above expression as in $\mathbf{e}$ is to take a full advantage of the MD structure in ε so that, in the double-sum part of (3.1), the 3rd and 4th moments of ε_i do not appear and some complicated terms disappear. This is important as the 3rd and 4th moments of ε_i cannot be consistently estimated together with these of v_{it}. The invertibility of K_m, $m \geq 1$, is ensured by Assumption F(ii).

To proceed, for a square matrix A, let A^u, A^l and A^d be, respectively, its upper-triangular, lower-triangular, and diagonal matrix such that $A = A^u + A^l + A^d$. Denote by Π_t, Φ_{ts} and Ψ_{ts} the submatrices of Π, Φ and Ψ partitioned according to $t, s = 2, \ldots, T$. Denote the partial sum of time-indexed quantities using the '$_+$' notation: e.g., $\Psi_{t+} = \sum_{s=1}^T \Psi_{ts}$, $\Psi_{+s} = \sum_{t=1}^T \Psi_{ts}$, $\Psi_{++} = \sum_{t=1}^T \sum_{s=1}^T \Psi_{ts}$, and similarly for Φ_{ts}, Π_t and other time-indexed quantities.

First, to estimate the variance of $\mathbf{e}'\Psi\mathbf{y}_0$, letting $\Psi_{ts}^* = \Psi_{ts} K_m$, we have:

$$\begin{aligned}
\mathbf{e}'\Psi\mathbf{y}_0 &= \sum_{t=1}^T \sum_{s=1}^T e_t' \Psi_{ts} y_0 = \sum_{t=1}^T e_t' \Psi_{t+}^* y_0^* \\
&= \sum_{t=1}^T e_t' \Psi_{t+}^{*d} y_0^* + \sum_{t=1}^T e_t' (\Psi_{t+}^{*l} + \Psi_{t+}^{*u}) y_0^* \\
&= \sum_{t=1}^T e_t' \Psi_{t+}^{*d} y_0^* + \sum_{t=1}^T e_t' \xi_t \\
&= \sum_{i=1}^n \Big(\sum_{t=1}^T e_{it} \Psi_{ii,t+}^* y_{0i}^* + \sum_{t=1}^T e_{it}\xi_{it} \Big),
\end{aligned}$$

where $\{\xi_{it}\} = \xi_t = (\Psi_{t+}^{*l} + \Psi_{t+}^{*u}) y_0^*$ and $\Psi_{ii,t+}^*$ is the ith diagonal element of Ψ_{t+}^*, $i = 1, \ldots, n$.

Noting that $e_t = \varepsilon + B_3^{-1} v_t$ and $y_0^* = \varepsilon + \eta_m^* + V_m^*$, we have, $\mathrm{E}(e_{it}' \Psi_{ii,t+}^* y_{0i}^*) = \sigma_{\varepsilon 0}^2 \Psi_{ii,t+}^* \equiv d_{\Psi,it}$, and $\mathrm{E}(e_{it}'\xi_{it}) = 0$. These lead to $\mathbf{e}'\Psi_0\mathbf{y}_0 - \mathrm{E}(\mathbf{e}'\Psi_0\mathbf{y}_0) = \sum_{i=1}^n g_{\Psi,i}$, where

$$g_{\Psi,i} = \sum_{t=1}^T \big[(e_{it}\Psi_{ii,t+}^* y_{0i}^* - d_{\Psi,it}) + e_{it}\xi_{it} \big], \tag{3.3}$$

i.e., $\mathbf{e}'\Psi_0\mathbf{y}_0 - \mathrm{E}(\mathbf{e}'\Psi_0\mathbf{y}_0)$ is decomposed into a sum of n 'gradients'.

Similarly for the terms quadratic in $\mathbf{e}$, we have

$$\begin{aligned}
\mathbf{e}'\Phi\mathbf{e} &= \sum_{t=1}^T \sum_{s=1}^T e_t' \Phi_{ts} e_s = \sum_{t=1}^T \sum_{s=1}^T e_t' (\Phi_{ts}^d + \Phi_{ts}^u + \Phi_{ts}^l) e_s \\
&= \sum_{t=1}^T \sum_{s=1}^T e_t' \Phi_{ts}^d e_s + \sum_{t=1}^T \sum_{s=1}^T e_t' \Phi_{ts}^l e_s + \sum_{t=1}^T \sum_{s=1}^T e_s' \Phi_{ts}^{u\prime} e_t \\
&= \sum_{t=1}^T \sum_{s=1}^T e_t' \Phi_{ts}^d e_s + \sum_{t=1}^T \sum_{s=1}^T e_t' \Phi_{ts}^l e_s + \sum_{t=1}^T \sum_{s=1}^T e_t' \Phi_{st}^{u\prime} e_s \\
&= \sum_{t=1}^T e_t' \sum_{s=1}^T \Phi_{ts}^d e_s + \sum_{t=1}^T e_t' \sum_{s=1}^T (\Phi_{ts}^l + \Phi_{st}^{u\prime}) e_s \\
&= \sum_{t=1}^T e_t' e_t^* + \sum_{t=1}^T e_t' \varphi_t, \\
&= \sum_{i=1}^n \Big(\sum_{t=1}^T e_{it} e_{it}^* + \sum_{t=1}^T e_{it}\varphi_{it} \Big),
\end{aligned}$$

where $e_t^* = \sum_{s=1}^T \Phi_{ts}^d e_s$ with elements e_{it}^*, and $\varphi_t = \sum_{s=1}^T (\Phi_{ts}^l + \Phi_{st}^{u\prime}) e_s$ with elements φ_{it}.

Letting $a'_{i,ts}$, b'_i and $c'_{i,ts}$ be the ith row of $(\Phi^{l}_{ts} + \Phi^{u\prime}_{st})$, B_3^{-1} and $(\Phi^{l}_{ts} + \Phi^{u\prime}_{st})B_3^{-1}$, respectively, we have $e^*_{it} = \Phi_{ii,t+}\varepsilon_i + \sum_{s=1}^{T} \Phi_{ii,ts} b'_i v_s$ and $\varphi_{it} = a'_{i,t+}\varepsilon + \sum_{s=1}^{T} c'_{i,ts} v_s$. It follows that

$$\mathrm{E}(e_{it}e^*_{it}) = \mathrm{E}\big[(\varepsilon_i + b'_i v_t)(\Phi_{ii,t+}\varepsilon_i + \sum_{s=1}^{T} \Phi_{ii,ts} b'_i v_s)\big] = \sigma^2_{\varepsilon 0}\Phi_{ii,t+} + \sigma^2_{v0}\Phi_{ii,tt}(b'_i b_i) \equiv d_{1\Phi,it},$$

$$\mathrm{E}(e_{it}\varphi_{it}) = \mathrm{E}\big[(\varepsilon_i + b'_i v_t)(a'_{i,t+}\varepsilon + \sum_{s=1}^{T} c'_{i,ts} v_s)\big] = \sigma^2_{v0}(b'_i c_{i,tt}) \equiv d_{2\Phi,it}.$$

These lead to $\mathbf{e}'\Phi_0\mathbf{e} - \mathrm{E}(\mathbf{e}'\Phi_0\mathbf{e}) = \sum_{i=1}^{n} g_{\Phi,i}$, where

$$g_{\Phi,i} = \sum_{t=1}^{T}\big[(e_{it}e^*_{it} - d_{1\Phi,it}) + (e_{it}\varphi_{it} - d_{2\Phi,it})\big]. \tag{3.4}$$

Finally, for the terms linear in $\mathbf{e}$, $\mathrm{E}(\Pi'\mathbf{e}) = 0$, and, letting Π'_{it} be the ith row of Π_t,

$$\Pi'\mathbf{e} = \sum_{i=1}^{n}(\sum_{t=1}^{T} \Pi_{it} e_{it}) \equiv \sum_{i=1}^{n} g_{\Pi,i}. \tag{3.5}$$

The decompositions of the three types of quantities into sums with 'gradients' given by (3.3)–(3.5) lead to a 'possible' way for a consistent estimate of the VC matrix of the AQS function.

For each Ψ_r, $r = 1, 2, 3$, defined in (2.15), define $g_{\Psi_r,i}$ according to (3.3); for each Φ_r, $r = 1, \ldots, 6$, defined in (2.15), define $g_{\Phi_r,i}$ according to (3.4); and each Π_r, $r = 1, 2, 3, 4$, defined in (2.15), define $g_{\Pi_r,i}$ according to (3.5). Define,

$$\mathbf{g}_i = \begin{cases} g_{\Pi_1,i}, \\ g_{\Phi_1,i}, \\ g_{\Phi_2,i}, \\ g_{\Pi_2,i} + g_{\Phi_3,i} + g_{\Psi_1,i}, \\ g_{\Pi_3,i} + g_{\Phi_4,i} + g_{\Psi_2,i}, \\ g_{\Pi_4,i} + g_{\Phi_5,i} + g_{\Psi_3,i}, \\ g_{\Phi_6,i}. \end{cases} \tag{3.6}$$

Then, the AQS vector at the true parameter value is $S^*_{\mathrm{SDPD}}(\psi_0) = \sum_{i=1}^{n} \mathbf{g}_i$ and its variance is given by (3.1), i.e., $\mathrm{Var}\big[S^*_{\mathrm{SDPD}}(\psi_0)\big] = \sum_{i=1}^{n} \mathrm{E}(\mathbf{g}_i\mathbf{g}'_i) + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} \mathrm{E}(\mathbf{g}_i\mathbf{g}'_j)$, where the single sum can be estimated by its sample counter part $\sum_{i=1}^{n} \hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i$ with $\hat{\mathbf{g}}_i$ being obtained by replacing ψ_0 and $\mathbf{e}$ in $\mathbf{g}_i$ by their estimates $\hat{\psi}_{\mathrm{M}}$ and $\hat{\mathbf{e}}$, and the double sum is estimated using its semi-analytical form shown in the following lemma.

To simplify the representation and to facilitate the calculations, let $\boldsymbol{\pi}_r$ and $\boldsymbol{\pi}_\nu$ be the column(s) of $\boldsymbol{\Pi} = (\Pi_1, \Pi_2, \Pi_3, \Pi_4)$, for $r, \nu = 1, 2, \ldots, k_\varpi$, where $k_\varpi = \mathtt{dim}(\boldsymbol{\beta}) + 3$, and $g_{\boldsymbol{\pi}_r}$ and $g_{\boldsymbol{\pi}_\nu}$ be the corresponding gradients vectors defined according to (3.5).

Lemma 3.1. *For the gradient pairs* $(g_{\boldsymbol{\pi}_r,i}, g_{\boldsymbol{\pi}_\nu,j})$, $r, \nu = 1, \ldots, k_\varpi$; $(g_{\Phi_r,i}, g_{\Phi_\nu,j})$, $r, \nu = 1, \ldots, 6$; *and* $(g_{\Psi_r,i}, g_{\Psi_\nu,j})$, $r, \nu = 1, 2, 3$, *corresponding to* $(\boldsymbol{\pi}_r, \boldsymbol{\pi}_\nu)$, (Φ_r, Φ_ν), *and* (Ψ_r, Ψ_ν), *respectively, we have under Assumptions A–B, for* $j \neq i\,(= 1, \ldots, n)$ *and* $m \geq 1$,

$$\mathrm{E}(g_{\boldsymbol{\pi}_r,i}\, g_{\boldsymbol{\pi}_\nu,j}) = \sigma^2_{v0}(b'_i b_j)\sum_{t=1}^{T} \boldsymbol{\pi}_{ri,t}\boldsymbol{\pi}_{\nu j,t}, \tag{3.7}$$

$$\begin{aligned}\mathrm{E}(g_{\Phi_r,i}\, g_{\Phi_\nu,j}) = {} & \sigma^4_{v0}\sum_{t=1}^{T}\sum_{s=1}^{T}\big[(b'_j c^*_{ri,ts})(b'_i c^*_{\nu j,st}) + (b'_i b_j)(c^{*\prime}_{ri,ts} c^*_{\nu j,ts})\big] \\ & + \sigma^2_{v0}\sigma^2_{\varepsilon 0}\sum_{t=1}^{T}\big[a_{\nu ji,t+}(b'_j c^*_{ri,+t}) + a_{rij,t+}(b'_i c^*_{\nu j,+t}) + (b'_i b_j)(a^{*\prime}_{ri,t+} a^*_{\nu j,t+})\big] \\ & + (\mu^{(4)}_{v0} - 3\sigma^4_{v0})\sum_{t=1}^{T}\big[(b_i \odot c^*_{ri,tt})'(b_j \odot c^*_{\nu j,tt})\big],\end{aligned} \tag{3.8}$$

$$\mathrm{E}(g_{\Psi_r,i}\, g_{\Psi_\nu,j}) = \sigma^4_{\varepsilon 0}(w_{rij,+} w_{\nu ji,+}) + \sigma^2_{v0}\sum_{t=1}^{T}(b'_i b_j)\mathrm{E}(\xi^*_{ri,t}\xi^*_{\nu j,t}), \tag{3.9}$$

$$\mathrm{E}(g_{\Phi_r,i}\, g_{\boldsymbol{\pi}_v,j}) = \mu_{v0}^{(3)} \sum_{t=1}^{T} (b_i \odot c^*_{ri,tt})' b_j \boldsymbol{\pi}_{vj,t}, \tag{3.10}$$

$$\mathrm{E}(g_{\Psi_r,i}\, g_{\boldsymbol{\pi}_v,j}) = \sigma_{v0}^2 \sum_{t=1}^{T} \boldsymbol{\pi}'_{vj,t} \mathrm{E}(\xi^*_{ri,t})(b'_i b_j), \tag{3.11}$$

$$\begin{aligned}\mathrm{E}(g_{\Phi_r,i}\, g_{\Psi_v,j}) = {} & \sigma_{\varepsilon 0}^2 \sigma_{v0}^2 \sum_{t=1}^{T} \big[(b'_i b_j)(a'_{ri,t+} w^*_{vj,t}) + w_{vji,t}(b'_j c^*_{ri,+t}) + \Phi_{ii,t+} w_{vji,t}(b'_i b_j)\big] \\ & + \sigma_{\varepsilon 0}^4 (w_{vji,+} a_{rij,++}) + \mu_{v0}^{(3)} \sum_{t=1}^{T} (b_i \odot c^*_{ri,tt})' b_j \mathrm{E}(\xi^*_{vj,t}),\end{aligned} \tag{3.12}$$

where $\xi^*_{ri,t} = w^{*\prime}_{ri,t} y^*_0$; $w^{*\prime}_{ri,t}$, $a^{*\prime}_{ri,ts}$ *and* $c^{*\prime}_{ri,ts}$ *are the ith row of* $\Psi^*_{r,t+}$, $\Phi^*_{r,ts} = (\Phi^l_{r,ts} + \Phi^{u\prime}_{r,st} + \Phi^d_{r,ts})$, *and* $\Phi^*_{r,ts} B_3^{-1}$, *respectively; and* $a_{rij,t+}$ *and* $w_{rij,+}$ *are the* (i,j)*th element of* $(\Phi^l_{r,t+} + \Phi^{u\prime}_{r,+t})$ *and* $(\Psi^{*l}_{r,++} + \Psi^{*u}_{r,++})$*, respectively.*

From (3.6), it is clear that $\mathrm{E}(\mathbf{g}_i \mathbf{g}'_j)$, $i \neq j$, can be obtained from the results of Lemma 3.1. Note that the (Π, Φ) terms of $\mathrm{E}(\mathbf{g}_i \mathbf{g}'_j)$ are analytical functions of ψ_0, $\mu_{v0}^{(3)}$ and $\mu_{v0}^{(4)}$, and hence can be estimated by plugging-in consistent estimators of these parameters. However, the Ψ-related terms are also functions of $\mathrm{E}(y_0)$ and $\mathrm{E}(y_0 y'_0)$ that appear in $\mathrm{E}(\xi^*_{ri,t})$ and $\mathrm{E}(\xi^*_{ri,t}\xi^*_{vj,t})$, besides these parameters. Consistent estimators $\hat{\mu}_v^{(3)}$ and $\hat{\mu}_v^{(4)}$ of $\mu_{v0}^{(3)}$ and $\mu_{v0}^{(4)}$ are readily available as seen below, but the estimation of $\mathrm{E}(y_0)$ and $\mathrm{E}(y_0 y'_0)$ is not trivial. Their expressions involve unobservables and thus cannot be used. In this paper, we propose to estimate the terms involving $\mathrm{E}(\xi^*_{ri,t})$ and $\mathrm{E}(\xi^*_{ri,t}\xi^*_{vj,t})$ by their sample analogues and the other analytical terms by plugging-in method, i.e., removing E in the expressions and then replacing (in all terms) ψ_0, $\mu_{v0}^{(3)}$ and $\mu_{v0}^{(4)}$ by $\hat{\psi}_{\mathtt{M}}$, $\hat{\mu}_v^{(3)}$ and $\hat{\mu}_v^{(4)}$. The resulting estimator $\widehat{\mathrm{E}}(\mathbf{g}_i \mathbf{g}'_j)$ of $\mathrm{E}(\mathbf{g}_i \mathbf{g}'_j)$ is thus mixtures of plug-in method and sample analogue method. The resulting estimator of the variance of the estimating functions, Γ^*_{SDPD}, is given as follows,

$$\widehat{\Gamma}^*_{\mathrm{SDPD}} = \frac{1}{nT} \sum_{i=1}^{n} \hat{\mathbf{g}}_i \hat{\mathbf{g}}'_i + \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \widehat{\mathrm{E}}(\mathbf{g}_i \mathbf{g}'_j). \tag{3.13}$$

Its consistency is proved in the following theorem.

Theorem 3.1. *Under the assumptions of Theorem 2.1, we have, as* $n \to \infty$,

$$\widehat{\Gamma}^*_{\mathrm{SDPD}} - \Gamma^*_{\mathrm{SDPD}}(\psi_0) = \frac{1}{nT} \sum_{i=1}^{n} \big[\hat{\mathbf{g}}_i \hat{\mathbf{g}}'_i - \mathrm{E}(\mathbf{g}_i \mathbf{g}'_i)\big] + \frac{1}{nT} \sum_{1=1}^{n} \sum_{j=1, j\neq i}^{n} \big[\widehat{\mathrm{E}}(\mathbf{g}_i \mathbf{g}'_j) - \mathrm{E}(\mathbf{g}_i \mathbf{g}'_j)\big] \xrightarrow{p} 0,$$

and hence, $\Sigma^{*-1}_{\mathrm{SDPD}}(\hat{\psi}_{\mathtt{M}}) \widehat{\Gamma}^*_{\mathrm{SDPD}} \Sigma^{*-1}_{\mathrm{SDPD}}(\hat{\psi}_{\mathtt{M}}) - \Sigma^{*-1}_{\mathrm{SDPD}}(\psi_0) \Gamma^*_{\mathrm{SDPD}}(\psi_0) \Sigma^{*-1}_{\mathrm{SDPD}}(\psi_0) \xrightarrow{p} 0.$

Finally, we present a pair of simple and consistent estimators of the 3rd and 4th moments of v_{it}, $\mu_{v0}^{(3)}$ and $\mu_{v0}^{(4)}$. Let $\bar{e} = \frac{1}{T}\sum_{t=1}^{T} e_t$ and $\bar{v} = \frac{1}{T}\sum_{t=1}^{T} v_t$. Then, we have $v_t - \bar{v} = B_3(e_t - \bar{e})$. Letting $v^*_t = v_t - \bar{v}$, we have $\mathrm{E}(v^{*3}_{it}) = \frac{T^2 - 3T + 2}{T^2} \mu_{v0}^{(3)}$. An estimator of $\mu_{v0}^{(3)}$ is naturally

$$\hat{\mu}_v^{(3)} = \frac{T^2}{T^2 - 3T + 2} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \hat{v}^{*3}_{it}.$$

To estimate $\mu_{v0}^{(4)}$, we take first difference of e_{it} to get rid of the error component related to the CRE term. After first differencing, we have $\Delta v_t = B_3 \Delta e_t$, $t = 2, \ldots, T$, and

$$\mathrm{E}(\Delta v^4_{it}) = \mathrm{E}[(v_{it} - v_{i,t-1})^4] = \mathrm{E}(v^4_{it}) + \mathrm{E}(v^4_{i,t-1}) + 6\mathrm{E}(v^2_{it} v^2_{i,t-1}) = 2\mu_{v0}^{(4)} + 6\sigma^4_{v0}.$$

Therefore an estimator of $\mu_{v0}^{(4)}$ can be: $\hat{\mu}_{v0}^{(4)} = \frac{1}{2n}\sum_{i=1}^{n} \Delta \hat{v}^4_{it} - 3\hat{\sigma}^4_{v0}$, for any $t = 2, \ldots, T$. Obviously, one should combine these to give a pooled estimator:

$$\hat{\mu}_{v0}^{(4)} = \frac{1}{2n(T-1)} \sum_{t=2}^{T} \sum_{i=1}^{n} \Delta \hat{v}^4_{it} - 3\hat{\sigma}^4_{v0}.$$

A computational note. The calculation of the double summation term in (3.13), i.e., $\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \widehat{\mathrm{E}}(\mathbf{g}_i \mathbf{g}'_j)$, is greatly facilitated by writing (3.7)–(3.12) in matrix forms for all i, j, using the Kronecker product $\otimes$ operator, and the Hadamard product operator $\odot$:

$$\Lambda(\boldsymbol{\pi}_r, \boldsymbol{\pi}_v) = \sigma_{v0}^2 \mathbb{B}_3 \odot \Big(\sum_{t=1}^{T} \boldsymbol{\pi}_{rt} \boldsymbol{\pi}'_{vt}\Big), \tag{3.14}$$

$$\Lambda(\Psi_r,\Psi_\nu)=\sigma_{\varepsilon 0}^4(\Psi^*_{r,++}\odot\Psi^{*\prime}_{\nu,++})+\sigma_{v0}^2\sum_{t=1}^T\mathbb{B}_3\odot \mathrm{E}(\xi^*_{r,t}\xi^{*\prime}_{\nu,t}), \tag{3.15}$$

$$\begin{aligned}\Lambda(\Phi_r,\Phi_\nu)=\ &\sigma_{v0}^4\sum_{t=1}^T\sum_{s=1}^T\big[(\Phi^*_{r,ts}\mathbb{B}_3)\odot(\mathbb{B}_3\Phi^{*\prime}_{\nu,st})+\mathbb{B}_3\odot(\Phi^*_{r,ts}\mathbb{B}_3\Phi^{*\prime}_{\nu,ts})\big]\\ &+\sigma_{v0}^2\sigma_{\varepsilon_0}^2\sum_{t=1}^T\big[\Phi^{\circ\prime}_{\nu,t}\odot(\Phi^*_{r,+t}\mathbb{B}_3)+\Phi^{\circ}_{r,t}\odot(\mathbb{B}_3\Phi^{*\prime}_{\nu,+t})+\mathbb{B}_3\odot(\Phi^*_{r,t+}\Phi^{*\prime}_{\nu,t+})\big]\\ &+(\mu_{v0}^{(4)}-3\sigma_{v0}^4)\sum_{t=1}^T\big[B_3^{-1}\odot(\Phi^*_{r,tt}B_3^{-1})\big]\big[B_3^{-1}\odot(\Phi^*_{\nu,tt}B_3^{-1})\big]',\end{aligned} \tag{3.16}$$

$$\Lambda(\Psi_r,\boldsymbol{\pi}_\nu)=\sigma_{v0}^2\mathbb{B}_3\odot\Big[\sum_{t=1}^T\mathrm{E}(\xi^*_{r,t})\boldsymbol{\pi}'_{\nu t}\Big], \tag{3.17}$$

$$\Lambda(\Phi_r,\boldsymbol{\pi}_\nu)=\mu_{v0}^{(3)}\sum_{t=1}^T\big[B_3^{-1}\odot(\Phi^*_{r,tt}B_3^{-1})\big]B_3'^{-1}\mathtt{diag}(\boldsymbol{\pi}_{\nu t}), \tag{3.18}$$

$$\begin{aligned}\Lambda(\Phi_r,\Psi_\nu)=\ &\sigma_{\varepsilon 0}^4(\Phi^\circ_{r,++}\odot\Psi^{*\prime}_{\nu,++})+\sigma_{v0}^2\sigma_{\varepsilon 0}^2\sum_{t=1}^T\big[\mathbb{B}_3\odot(\Phi^\circ_{r,t}\Psi^{*\prime}_{\nu,t+})+(\Phi^\sharp\mathbb{B}_3)\odot\Psi^{*\prime}_{\nu,t+}\big]\\ &+\mu_{v0}^{(3)}\sum_{t=1}^T\big[B_3^{-1}\odot(\Phi^*_{r,tt}B_3^{-1})\big]B_3'^{-1}\mathtt{diag}\big[\mathrm{E}(\xi^*_{\nu,t})\big],\end{aligned} \tag{3.19}$$

where $\mathbb{B}_3=(B_3'B_3)^{-1}$, $\Phi^\sharp_{r,t}=\Phi^*_{r,+t}+\Phi^d_{r,t+}$, and $\Phi^\circ_{r,t}=\Phi^l_{r,t+}+\Phi^{u\prime}_{r,+t}$.

Then, it is easy to see that $\sum_{i=1}^n\sum_{j=1,j\neq i}^n\mathrm{E}(g_{\omega i}\ g_{wj})$ equals the sum of the off-diagonal elements of $\Lambda(\omega,w)$, for $\omega,w=\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_{k_\varpi}$, Ψ_1, Ψ_2, Ψ_3, and $\Phi_1,\ldots,\Phi_6$, which lead immediately to $\sum_{i=1}^n\sum_{j=1,j\neq i}^n\mathrm{E}(\mathbf{g}_i\mathbf{g}_j')$ and its estimate $\sum_{i=1}^n\sum_{j=1,j\neq i}^n\widehat{\mathrm{E}}(\mathbf{g}_i\mathbf{g}_j')$.

A final discussion is given to the case where $m=0$, i.e., y_0 **is exogenous**. In this case, it is obvious that the conditional QML method is valid for parameter estimation. But for the VC matrix estimation, the traditional plug-in method still cannot be applied under fixed T scenario, due to the coexistence of 3rd and 4th moments of the two error components. In contrast, our new method applies and all we need is to re-derive the Ψ-related results of Lemma 3.1 under exogenous y_0, which take the following simple forms:

$$\mathrm{E}(g_{\Psi_r i}\,g_{\Psi_\nu j})=\sigma_{v0}^2\sum_{t=1}^T(b_i'b_j)\mathrm{E}(\xi^\circ_{ri,t}\xi^\circ_{\nu j,t}),\quad r,\nu=1,2,3, \tag{3.20}$$

$$\mathrm{E}(g_{\Psi_r i}\,g_{\boldsymbol{\pi}_\nu j})=\sigma_{v0}^2\sum_{t=1}^T\boldsymbol{\pi}'_{\nu j,t}\mathrm{E}(\xi^\circ_{ri,t})(b_i'b_j),\quad r=1,2,3,\quad \nu=1,\ldots,k_\varpi, \tag{3.21}$$

$$\mathrm{E}(g_{\Phi_r i}\,g_{\Psi_\nu j})=\mu_{v0}^{(3)}\sum_{t=1}^T(b_i\odot c^*_{ri,tt})'b_j\mathrm{E}(\xi^\circ_{\nu j,t}),\quad r=1,\ldots,6,\quad \nu=1,2,3, \tag{3.22}$$

where $\xi^\circ_{it}=w^{\circ\prime}_{it}y_0$, and $w^{\circ\prime}_{it}$ is the ith row of $(\Psi^l_{t+}+\Psi^u_{t+})$.

4. Monte Carlo study

Extensive Monte Carlo experiments are run to investigate the finite sample performance of the proposed M-estimator of the CRE-SDPD model, and the finite sample performance of the proposed estimate of the VC matrix of the M-estimator. As in the special case of a RE-SDPD model with only spatial errors the full QMLE is available from Su and Yang (2015), a comparison is made between the full QMLE and the proposed M-estimator. We use the following three data generating processes (DGPs):

DGP1: $y_t=\rho y_{t-1}+\lambda_1W_1y_t+\lambda_2W_2y_{t-1}+X_t\beta+Z\gamma+\mu+\alpha_t 1_n+u_t$,

DGP2: $y_t=\rho y_{t-1}+\lambda_1W_1y_t+\lambda_2W_2y_{t-1}+X_t\beta+Z\gamma+\varepsilon+\alpha_t 1_n+u_t$,

DGP3: $y_t=\alpha 1_n+\rho y_{t-1}+X_t\beta+Z\gamma+\varepsilon+u_t$,

where $u_t=\lambda_3W_3u_t+v_t$ for all three DGPs, and μ, ε and v_t represent, respectively, the CRE, RE, and idiosyncratic error. Mundlak's specification, $\mu=\bar{X}\pi+\varepsilon$, is adopted.

The elements of X_t are generated in a similar fashion as in Hsiao et al. (2002),[7] and the elements of Z are random draws from Bernoulli (0.5). The elements of ε are random draws from $N(0, 1)$. The spatial weight matrices are generated according to the following schemes: Rook contiguity, Queen contiguity, or group interaction.[8] The error v_t distribution can be (*i*) normal, (*ii*) normal mixture (10%$N(0, 4)$, 90%$N(0, 1)$), or (*iii*) chi-squared with degree of freedom of 3. In both (*ii*) and (*iii*), the generated errors are standardized to have mean zero and variance σ_v^2. We choose $\beta = \gamma = \pi = \sigma_v^2 = \alpha_T = 1$, and generate α_t, $t = 1, \ldots, T-1$, from $N(1, 1)$. We use a set of values for ρ ranging from -0.9 to 0.9, a set of values for $(\lambda_1, \lambda_2, \lambda_3)$ in the similar range, $T = 3$ or 6, and $N = 50, 100, 200, 400$. Each set of Monte Carlo results, corresponding to a combination of the values of $(n, T, m, \rho, \lambda's)$, is based on 2000 samples.

Monte Carlo (empirical) means and standard deviations (sds) are reported for the CQML estimator (CQMLE), the M-estimator, and the full QMLE (DGP3). Empirical averages of the robust standard errors (rses) based on the VC matrix estimate $\Sigma^{*-1}_{\mathrm{SDPD}}(\hat{\psi}_{\mathrm{M}})\widehat{\Gamma}^*_{\mathrm{SDPD}}\Sigma^{*-1}_{\mathrm{SDPD}}(\hat{\psi}_{\mathrm{M}})$ are also reported for the M-estimator, which should be compared with the corresponding empirical sds. A subset of results are reported in Tables 1–5. Monte Carlo results that are involved in the discussions but unreported due to space constraint can be found in the `Supplement Appendix` to this paper, available from http://www.mysmu.edu/faculty/zlyang/.

Tables 1–3 present the results based on DGP1, the CRE-SDPD model with all three types of spatial effects. The results show an excellent performance of the proposed M-estimators of the model parameters, and the rses. The M-estimator of the dynamic parameter is nearly unbiased, whereas the CQMLE can be quite biased and as n increases it does not show a sign of convergence. The M-estimators of the spatial parameters λ_1 and λ_2 also show an excellent finite sample performance. Both CQMLE and M-estimator of the spatial parameter λ_3 show some bias. This is perhaps due to the intrinsic nature of the QML-type estimation of spatial error effects. Increasing T improves its performance as shown in Table 2. The rses are on average very close to the corresponding Monte Carlo sds in general, showing the robustness and good finite sample performance of the proposed VC matrix estimate. The non-robust ses based on $\widehat{\Gamma}^{*-1}_{\mathrm{SDPD}}$ only or $\Sigma^{*-1}_{\mathrm{SDPD}}(\hat{\psi}_{\mathrm{M}})$ only are also simulated and the results (unreported for conserving space) show that when errors are normal, all three methods give averaged standard errors close to the corresponding Monte Carlo sds; but when the errors are not normal the non-robust ses can be quite different from the corresponding Monte Carlo sds in particular in the standard errors of σ_v^2 and ϕ.

Tables 4–5 present the results based on DGP2, the RE-SDPD with all three types of spatial effects. Similar observations hold: the proposed estimation strategy performs excellently and clearly outperforms the conditional QMLEs. The results also show that the proposed estimate of the standard errors of M-estimator also performs very well.

Table 6 presents the results based on DGP3, the RE-SDPD with only spatial error effect. For this model, the full QMLE (FQMLE) is available from Su and Yang (2015). As the main focus of this set of Monte Carlo experiments is to compare M-estimator with FQMLE, the rses of the M-estimator are not reported. The results show that both M-estimator and FQMLE of the dynamic parameter are nearly unbiased whereas the CQMLE is quite different from the true value and does not show a sign of convergence. Three estimators of spatial parameter λ_3 all show some bias, but the M-estimator has the smallest bias among the three. Comparing the empirical sds, we see that the M-estimator is slightly less efficient than the FQMLE, as expected. Computationally, however, the M-estimator is much more efficient.

Under all three DGPs, the Monte Carlo experiments are also run using a 'wrong' value of m and the results show that the M-estimator is quite robust against the choice of m value; more W specifications are considered, and the results show a quite robust performance of our estimation and inference methods; and more cases for $T = 6$ are considered and the general observations from the results are that with a larger value of T the performance of the estimators of λ_3 significantly improved, and the CQMLE perform significantly better but is still clearly dominated by the M-estimator. All Monte Carlo results, upon which these conclusions are drawn, can be found in the `Supplementary Appendix`.

5. Empirical application: Political competition in China

In this section, we apply the estimation and inference methods for the CRE-SDPD model proposed in this paper to investigate strategic interactions in political competitions across Chinese cities. The tournament competition among Chinese local government leaders has been well documented. The competitions have been found over primary policy issues such as economic growth and fiscal budget, as well as over second-dimensional policy issues such as coal mine safety (Li and Zhou, 2005; Yao and Zhang, 2015; Yu et al., 2016; Shi and Xi, 2018). The provincial superiors can evaluate and promote local leaders based on their performance, and local leaders compete with each other for positions at higher levels.

We analyze the annual total investments (in RMB) of 338 prefecture-level cities (of which 80 are autonomous) in the 27 provincial level administrative regions (in short, provinces) in mainland China, from 2010 to 2013. The list of cities can be found in the `Supplementary Appendix`. The time-lagged dependent variable y_{t-1} measures policy stability. The spatial lag term Wy_t captures the competition among cities, reflecting how investment decisions of the neighboring cities affect

7 The detail is: $X_t = \mu_x + gt1_n + \zeta_t$, $(1-\phi_1 L)\zeta_t = \varepsilon_t + \phi_2\varepsilon_{t-1}$, $\varepsilon_t \sim N(0, \sigma_1^2 I_n)$, $\mu_x = e + \frac{1}{T+m+1}\sum_{t=-m}^{T}\varepsilon_t$, and $e \sim N(0, \sigma_2^2 I_n)$. Let $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2)$.

8 The Rook and Queen schemes are standard. For group interaction, we first generate $k = n^{\alpha}$ groups of sizes $n_g \sim U(.5\bar{n}, 1.5\bar{n})$, $g = 1, \ldots, k$, where $0 < \alpha < 1$ and $\bar{n} = n/k$, and then adjust n_g so that $\sum_{g=1}^{k} n_g = n$. The reported results correspond to $\alpha = 0.5$. See Yang (2015) for details in generating these spatial layouts.

Table 1
Empirical Mean(sd) [se] of CQMLE and M-Estimator, **DGP1** (CRE), $T = 3$, $m = 10$ $W_1 = W_3$: **Queen Contiguity**; W_2: **Group Interaction**.

	Normal Error		Normal Mixture		Chi-Square	
ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
	$n = 50$					
1	.999 (.053)	1.003 (.052) [.050]	.998 (.054)	1.001 (.053) [.050]	.999 (.053)	1.003 (.052) [.050]
1	.915 (.319)	.996 (.339) [.322]	.925 (.311)	1.007 (.331) [.320]	.911 (.313)	.992 (.332) [.320]
1	.856 (.163)	.995 (.153) [.136]	.854 (.166)	.991 (.156) [.137]	.850 (.167)	.987 (.157) [.136]
1	.995 (.163)	.940 (.147) [.135]	1.008 (.261)	.952 (.242) [.205]	1.004 (.235)	.949 (.216) [.188]
1	.761 (.374)	.992 (.432) [.374]	.775 (.424)	1.003 (.485) [.397]	.774 (.422)	1.001 (.486) [.390]
.3	.353 (.055)	.300 (.050) [.043]	.355 (.055)	.302 (.050) [.043]	.354 (.055)	.302 (.051) [.043]
.2	.189 (.052)	.193 (.053) [.049]	.186 (.050)	.190 (.051) [.049]	.188 (.053)	.193 (.054) [.049]
.2	.193 (.039)	.197 (.040) [.037]	.191 (.039)	.196 (.040) [.037]	.191 (.039)	.195 (.039) [.036]
.2	.101 (.182)	.096 (.184) [.169]	.098 (.183)	.093 (.188) [.165]	.106 (.174)	.099 (.177) [.165]
	$n = 100$					
1	1.001 (.036)	1.000 (.035) [.035]	1.000 (.036)	.999 (.036) [.035]	1.001 (.036)	1.000 (.036) [.035]
1	.919 (.224)	.997 (.237) [.230]	.916 (.217)	.993 (.229) [.231]	.921 (.220)	.999 (.232) [.230]
1	.843 (.117)	.994 (.106) [.099]	.849 (.119)	1.000 (.109) [.100]	.847 (.118)	.997 (.107) [.100]
1	1.035 (.117)	.979 (.105) [.100]	1.039 (.188)	.983 (.170) [.155]	1.034 (.171)	.978 (.156) [.143]
1	.755 (.254)	.981 (.295) [.259]	.770 (.284)	.996 (.322) [.284]	.766 (.278)	.991 (.315) [.280]
.3	.355 (.037)	.301 (.033) [.030]	.353 (.037)	.300 (.033) [.030]	.354 (.037)	.300 (.033) [.030]
.2	.185 (.025)	.198 (.026) [.025]	.185 (.026)	.198 (.027) [.025]	.185 (.026)	.197 (.027) [.025]
.2	.191 (.028)	.198 (.029) [.028]	.191 (.027)	.198 (.028) [.028]	.189 (.028)	.196 (.029) [.028]
.2	.165 (.118)	.152 (.121) [.116]	.165 (.118)	.151 (.120) [.114]	.162 (.117)	.150 (.120) [.114]
	$n = 200$					
1	.999 (.025)	1.001 (.025) [.025]	.999 (.026)	1.000 (.026) [.025]	1.000 (.025)	1.001 (.025) [.025]
1	.925 (.160)	.997 (.171) [.167]	.929 (.156)	1.001 (.167) [.167]	.927 (.160)	.998 (.172) [.167]
1	.829 (.086)	.996 (.078) [.072]	.833 (.086)	.998 (.077) [.073]	.833 (.087)	.999 (.078) [.072]
1	1.050 (.085)	.987 (.075) [.072]	1.046 (.129)	.984 (.116) [.113]	1.052 (.122)	.989 (.109) [.105]
1	.738 (.179)	.992 (.208) [.190]	.753 (.196)	1.004 (.227) [.207]	.744 (.184)	.996 (.210) [.203]
.3	.361 (.028)	.301 (.024) [.022]	.360 (.028)	.300 (.024) [.023]	.359 (.028)	.300 (.024) [.023]
.2	.186 (.022)	.198 (.023) [.023]	.187 (.022)	.198 (.023) [.023]	.186 (.023)	.198 (.023) [.023]
.2	.185 (.023)	.197 (.023) [.022]	.185 (.022)	.197 (.023) [.022]	.185 (.022)	.198 (.023) [.022]
.2	.186 (.083)	.176 (.085) [.084]	.185 (.084)	.175 (.086) [.083]	.187 (.085)	.178 (.087) [.083]
	$n = 400$					
1	1.000 (.019)	1.000 (.019) [.019]	1.000 (.019)	1.000 (.019) [.019]	1.001 (.019)	1.001 (.018) [.019]
1	.923 (.110)	1.002 (.118) [.117]	.917 (.112)	.995 (.119) [.117]	.920 (.113)	.999 (.120) [.117]
1	.842 (.061)	1.000 (.054) [.051]	.839 (.061)	.998 (.055) [.052]	.839 (.061)	.998 (.055) [.052]
1	1.053 (.060)	.995 (.054) [.051]	1.054 (.094)	.995 (.085) [.082]	1.054 (.088)	.995 (.080) [.075]
1	.756 (.122)	.995 (.140) [.133]	.761 (.143)	1.001 (.163) [.148]	.759 (.141)	.999 (.160) [.144]
.3	.356 (.019)	.300 (.017) [.015]	.357 (.019)	.300 (.017) [.016]	.357 (.019)	.301 (.017) [.016]
.2	.184 (.014)	.200 (.015) [.014]	.184 (.014)	.200 (.014) [.014]	.183 (.014)	.199 (.014) [.014]
.2	.189 (.021)	.199 (.021) [.021]	.189 (.020)	.198 (.021) [.021]	.189 (.021)	.199 (.021) [.021]
.2	.203 (.059)	.187 (.060) [.059]	.205 (.059)	.189 (.061) [.059]	.203 (.059)	.187 (.061) [.059]

Note: $\psi = (\beta, \gamma, \pi, \sigma_v^2, \phi, \rho, \lambda_1, \lambda_2, \lambda_3)'$. The results corresponding to $\breve{\alpha}$ are suppressed to save space, and are reported in Supplement Appendix; X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$.

the own investment level of a city. The competition may also be dynamic in the sense that the own investment decision of a city can depend on the investment level of its neighbors in the past, reflected by the space–time lag term Wy_{t-1}. The unobserved shocks that affect the investment level are likely to be correlated across neighboring cities, reflected by the spatial error term Wu_t. Time specific effects α_t capture macroeconomic conditions general to all cities in each year. The time-varying regressors X_t contain a set of city level variables: population, GDP, fiscal revenue, fiscal expenditure, and fiscal account balance in the previous year. To capture the effect of provincial economic environment, we also include a set of province level variables: fiscal revenue, fiscal expenditure, and public capital investment that is the government funded investment in fixed assets. The time-invariant regressors contain a constant and a dummy variable that indicates if a city is an autonomous city. As the basic spatial units in this study are cities, the model for CRE is constructed using city level time-varying variables: GDP, fiscal revenue, fiscal expenditure and fiscal account balance, with Population being excluded as it does not vary much over time (2010–2013). We considered two types of spatial weight matrices: W_{prov} that treats cities as neighbors if they are in the same province, and W_{geo} that treats cities as neighbors if they share a common border but may not be in the same province. Both weight matrices are row-normalized with zero on the diagonals. Table 7 below summarizes the main empirical findings.

The point estimate of the spatial lag parameter λ_1 is 0.249 and significant at 1% level when W_{prov} is used, suggesting a positive and strong spatial interaction in total investment among cities in the same province. When W_{geo} is used, λ_1 is estimated to be 0.041 and is insignificant. This result is consistent with the theory of tournament competition between local leaders. Dynamic competition does not seem affect the investment level as λ_2 are small and insignificant in both

Table 2
Empirical Mean (sd) [se] of CQMLE and M-Estimator, **DGP1** (CRE), $T = 6$, $m = 10$ $W_1 = W_3$: **Queen Contiguity**; W_2: **Group Interaction**.

	Normal Error		Normal Mixture		Chi-Square	
ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
	$n = 50$					
1	.998 (.028)	1.001 (.028) [.028]	.998 (.029)	1.001 (.029) [.027]	.997 (.029)	1.000 (.029) [.027]
1	.991 (.306)	1.002 (.309) [.295]	.987 (.300)	.998 (.303) [.293]	.996 (.300)	1.007 (.303) [.292]
1	.975 (.087)	.997 (.086) [.082]	.979 (.087)	1.000 (.087) [.082]	.976 (.089)	.997 (.089) [.082]
1	.965 (.088)	.961 (.088) [.085]	.971 (.160)	.967 (.158) [.144]	.971 (.145)	.967 (.144) [.133]
1	.927 (.270)	.960 (.277) [.256]	.939 (.441)	.973 (.468) [.364]	.927 (.395)	.959 (.407) [.337]
.3	.309 (.025)	.300 (.024) [.023]	.308 (.025)	.299 (.025) [.023]	.309 (.025)	.300 (.025) [.023]
.2	.195 (.040)	.195 (.040) [.036]	.196 (.039)	.196 (.039) [.035]	.194 (.039)	.195 (.039) [.036]
.2	.200 (.031)	.198 (.031) [.028]	.199 (.030)	.197 (.030) [.028]	.200 (.031)	.198 (.031) [.028]
.2	.120 (.115)	.119 (.116) [.109]	.120 (.113)	.119 (.113) [.107]	.124 (.113)	.123 (.113) [.107]
	$n = 100$					
1	.997 (.021)	1.001 (.020) [.020]	.996 (.021)	1.000 (.021) [.020]	.996 (.021)	1.000 (.021) [.020]
1	.985 (.222)	.998 (.225) [.213]	.989 (.219)	1.002 (.222) [.212]	.982 (.217)	.995 (.220) [.213]
1	.980 (.062)	1.000 (.062) [.061]	.981 (.062)	1.001 (.062) [.060]	.982 (.064)	1.001 (.064) [.061]
1	.986 (.064)	.981 (.064) [.062]	.989 (.114)	.984 (.113) [.108]	.990 (.102)	.985 (.102) [.099]
1	.945 (.187)	.977 (.192) [.186]	.949 (.312)	.981 (.321) [.279]	.955 (.286)	.988 (.295) [.263]
.3	.308 (.017)	.299 (.017) [.016]	.308 (.017)	.300 (.016) [.016]	.308 (.018)	.299 (.017) [.017]
.2	.197 (.026)	.198 (.026) [.025]	.197 (.025)	.197 (.025) [.024]	.197 (.026)	.198 (.026) [.024]
.2	.200 (.026)	.199 (.026) [.026]	.200 (.027)	.199 (.027) [.026]	.200 (.028)	.198 (.028) [.026]
.2	.164 (.081)	.163 (.081) [.077]	.164 (.080)	.163 (.080) [.076]	.161 (.080)	.160 (.080) [.076]
	$n = 200$					
1	.997 (.014)	1.000 (.014) [.014]	.997 (.014)	1.000 (.014) [.014]	.996 (.014)	1.000 (.014) [.014]
1	.990 (.154)	1.002 (.155) [.152]	.990 (.153)	1.001 (.155) [.152]	.989 (.152)	1.001 (.154) [.151]
1	.982 (.045)	1.001 (.045) [.043]	.980 (.045)	.999 (.045) [.043]	.981 (.044)	1.000 (.044) [.043]
1	.997 (.045)	.993 (.044) [.045]	.995 (.079)	.991 (.078) [.078]	.999 (.073)	.995 (.072) [.072]
1	.959 (.136)	.990 (.140) [.133]	.961 (.219)	.991 (.225) [.209]	.952 (.193)	.982 (.198) [.189]
.3	.307 (.012)	.300 (.012) [.011]	.308 (.012)	.300 (.012) [.011]	.308 (.012)	.300 (.012) [.011]
.2	.199 (.019)	.199 (.019) [.019]	.199 (.019)	.200 (.019) [.019]	.199 (.019)	.199 (.019) [.019]
.2	.199 (.020)	.199 (.020) [.020]	.200 (.020)	.200 (.020) [.020]	.199 (.020)	.199 (.020) [.020]
.2	.181 (.055)	.180 (.055) [.055]	.180 (.056)	.179 (.056) [.055]	.181 (.055)	.180 (.055) [.055]
	$n = 400$					
1	.996 (.010)	1.000 (.010) [.010]	.997 (.010)	1.001 (.010) [.010]	.996 (.010)	1.000 (.010) [.010]
1	.984 (.106)	.995 (.108) [.108]	.987 (.110)	.997 (.111) [.108]	.991 (.107)	1.001 (.109) [.108]
1	.981 (.029)	1.000 (.029) [.030]	.982 (.030)	1.001 (.029) [.030]	.981 (.030)	1.001 (.030) [.030]
1	1.000 (.031)	.996 (.031) [.032]	1.001 (.057)	.997 (.057) [.056]	1.001 (.053)	.997 (.052) [.051]
1	.964 (.093)	.994 (.095) [.095]	.966 (.153)	.996 (.157) [.152]	.964 (.138)	.994 (.142) [.139]
.3	.308 (.008)	.300 (.008) [.008]	.307 (.009)	.299 (.008) [.008]	.308 (.008)	.300 (.008) [.008]
.2	.199 (.012)	.200 (.012) [.012]	.198 (.012)	.200 (.012) [.012]	.198 (.012)	.200 (.012) [.012]
.2	.199 (.016)	.199 (.016) [.016]	.200 (.016)	.200 (.016) [.016]	.200 (.016)	.199 (.016) [.016]
.2	.192 (.038)	.191 (.038) [.038]	.192 (.037)	.191 (.038) [.038]	.192 (.039)	.192 (.039) [.038]

Note: $\psi = (\beta, \gamma, \pi, \sigma_v^2, \phi, \rho, \lambda_1, \lambda_2, \lambda_3)'$. The results corresponding to $\breve{\alpha}$ are suppressed to save space, and are reported in Supplement Appendix; X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$.

specifications. The spatial error parameter λ_3 is estimated to be 0.249 with standard error 0.115 under W_{prov}, and 0.287 with standard error 0.899 under W_{geo}. These provide strong evidence that unobserved shocks are highly correlated among administrative neighbors while less likely to be correlated among geographic neighbors, suggesting that shocks are mainly political and confined within administrative boundaries.

Based on W_{prov}, the coefficient of the time lag investment is estimated to be 0.211 and is significant at 1% level, indicating a positive dependence of investment on its previous level. As expected, the total investment depends positively on population and GDP. Fiscal expenditure have positive and significant impacts on the investment as it contributes to creating investment opportunities and providing pro-business services. Based on our results, budget constraints of the city-level government do not affect the investment level as both parameters of fiscal revenue and fiscal account balance in the previous year are insignificant. On the provincial level, we find negative effects of provincial fiscal revenue and public capital investment, and a positive effect of fiscal expenditure. Moreover, being an autonomous city has a large negative impact on the total investment level. We find that the 'individual-specific' effects are correlated (negatively) with GDP but not with the other time varying regressors. Robustness checks and alternative analyses can be found in Supplementary Appdix.[9] Issues remaining include a better way to define the spatial weight matrices, a better way

9 To see the possible existence of other social/natural effects at regional levels (in addition to the included province level variables and the Autonomous city dummy), we have done robustness checks by adding various regional dummies and the results remain largely unchanged.

`

Table 3
Empirical Mean (sd) [se] of CQMLE and M-Estimator, **DGP1** (CRE), $T = 3$, $m = 10$ $W_1 = W_3$: **Group Interaction**; W_2: **Queen Contiguity**.

	Normal Error		Normal Mixture		Chi-Square	
ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
	$n = 50$					
1	.994 (.048)	1.000 (.048) [.046]	.998 (.049)	1.003 (.048) [.046]	.995 (.051)	1.001 (.050) [.046]
1	.906 (.330)	.991 (.345) [.327]	.911 (.329)	.994 (.343) [.326]	.899 (.331)	.983 (.346) [.326]
1	.883 (.141)	.996 (.135) [.123]	.884 (.147)	.995 (.141) [.121]	.881 (.143)	.993 (.136) [.121]
1	.990 (.158)	.946 (.146) [.136]	.996 (.256)	.953 (.240) [.205]	.997 (.230)	.953 (.214) [.189]
1	.784 (.351)	.965 (.400) [.357]	.811 (.416)	.990 (.465) [.384]	.799 (.411)	.980 (.462) [.377]
.3	.344 (.046)	.302 (.043) [.038]	.343 (.048)	.301 (.045) [.038]	.344 (.047)	.303 (.044) [.038]
.2	.187 (.026)	.197 (.027) [.024]	.187 (.027)	.197 (.027) [.024]	.187 (.027)	.196 (.028) [.024]
.2	.186 (.037)	.196 (.037) [.035]	.188 (.037)	.198 (.037) [.035]	.187 (.038)	.197 (.038) [.035]
.2	.127 (.152)	.123 (.152) [.132]	.122 (.157)	.117 (.159) [.129]	.126 (.151)	.121 (.152) [.129]
	$n = 100$					
1	.997 (.034)	1.001 (.034) [.033]	.997 (.033)	1.001 (.033) [.033]	.997 (.035)	1.001 (.035) [.033]
1	.908 (.225)	.997 (.237) [.234]	.906 (.232)	.995 (.243) [.234]	.908 (.228)	.997 (.238) [.235]
1	.868 (.106)	.997 (.099) [.092]	.869 (.107)	.998 (.098) [.092]	.869 (.108)	.998 (.099) [.092]
1	1.023 (.114)	.974 (.104) [.099]	1.030 (.184)	.980 (.169) [.154]	1.024 (.169)	.974 (.156) [.143]
1	.786 (.257)	.992 (.298) [.260]	.789 (.282)	.992 (.318) [.280]	.798 (.277)	1.005 (.313) [.280]
.3	.348 (.035)	.300 (.032) [.029]	.348 (.034)	.301 (.031) [.029]	.348 (.036)	.300 (.032) [.029]
.2	.187 (.025)	.197 (.026) [.024]	.186 (.024)	.196 (.025) [.024]	.187 (.024)	.197 (.025) [.024]
.2	.191 (.030)	.197 (.031) [.029]	.191 (.030)	.197 (.031) [.029]	.193 (.031)	.199 (.031) [.029]
.2	.148 (.117)	.145 (.118) [.109]	.154 (.117)	.151 (.118) [.108]	.157 (.116)	.154 (.117) [.107]
	$n = 200$					
1	.997 (.025)	.999 (.025) [.025]	.998 (.026)	1.001 (.026) [.025]	.998 (.025)	1.000 (.025) [.025]
1	.940 (.167)	.996 (.177) [.165]	.949 (.157)	1.006 (.165) [.166]	.941 (.165)	.998 (.174) [.165]
1	.876 (.074)	1.001 (.068) [.066]	.877 (.074)	1.001 (.069) [.067]	.873 (.076)	.998 (.071) [.067]
1	1.034 (.083)	.987 (.075) [.071]	1.028 (.125)	.982 (.116) [.112]	1.034 (.118)	.987 (.108) [.104]
1	.800 (.178)	.996 (.201) [.184]	.815 (.197)	1.011 (.220) [.204]	.801 (.185)	.998 (.207) [.198]
.3	.345 (.023)	.300 (.021) [.020]	.344 (.023)	.299 (.021) [.020]	.345 (.024)	.300 (.021) [.020]
.2	.189 (.028)	.195 (.029) [.028]	.191 (.029)	.196 (.030) [.028]	.190 (.028)	.196 (.030) [.028]
.2	.192 (.020)	.199 (.021) [.020]	.192 (.021)	.199 (.021) [.020]	.192 (.021)	.199 (.021) [.020]
.2	.165 (.095)	.163 (.097) [.094]	.162 (.100)	.160 (.102) [.094]	.165 (.095)	.162 (.097) [.094]
	$n = 400$					
1	.996 (.019)	1.000 (.019) [.018]	.997 (.018)	1.001 (.018) [.018]	.996 (.019)	1.000 (.019) [.018]
1	.916 (.108)	.996 (.116) [.117]	.924 (.110)	1.004 (.116) [.117]	.917 (.109)	.997 (.117) [.117]
1	.838 (.059)	.999 (.053) [.051]	.838 (.059)	.999 (.053) [.051]	.840 (.062)	1.000 (.056) [.051]
1	1.055 (.059)	.995 (.053) [.051]	1.053 (.092)	.992 (.083) [.082]	1.054 (.087)	.993 (.078) [.075]
1	.751 (.121)	.996 (.141) [.134]	.757 (.138)	1.003 (.158) [.148]	.754 (.135)	1.000 (.155) [.145]
.3	.358 (.019)	.300 (.016) [.016]	.358 (.019)	.300 (.016) [.016]	.358 (.020)	.300 (.017) [.016]
.2	.184 (.021)	.197 (.022) [.021]	.184 (.021)	.197 (.022) [.021]	.184 (.021)	.197 (.022) [.021]
.2	.187 (.013)	.199 (.014) [.014]	.187 (.013)	.199 (.014) [.014]	.187 (.013)	.200 (.014) [.014]
.2	.182 (.079)	.174 (.080) [.077]	.181 (.079)	.173 (.081) [.076]	.184 (.080)	.176 (.082) [.076]

Note: $\psi = (\beta, \gamma, \pi, \sigma_v^2, \phi, \rho, \lambda_1, \lambda_2, \lambda_3)'$. The results corresponding to $\breve{\alpha}$ are suppressed to save space, and are reported in Supplement Appendix; X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$.

to capture regional effects, etc. While we strive for a rigorous empirical analysis, the main purpose of this study is to illustrate the proposed set of inference methods for the SDPD-CRE model. A comprehensive study on this topic is beyond the scope of the current research.

6. Conclusion and discussion

This paper introduces an M-estimation method for the spatial dynamic panel data (SDPD) model with correlated random effects (CRE), based on the short panel setup. The estimation strategy is based on the adjusted quasi score functions following the fundamental idea of Yang (2018). For statistical inferences, a hybrid method that combines analytical derivations and the feasible sample analogues is proposed for estimating the robust standard errors of the M-estimators. The asymptotic properties of these estimators are studied in detail and Monte Carlo simulation shows that both the M-estimators and the robust standard errors perform very well in finite samples. Clearly, the proposed estimation and inference methods for the CRE-SDPD model provide a useful complement to Yang (2018) for the FE-SDPD model, for their various advantages as discussed in the introduction, in particular, for allowing estimation of effects of time-invariant regressors and prediction in levels.

We end the paper by offering a discussion on possible extensions of our work. In this paper, we have focused on the Mundlak's (1978) CRE specification to ease exposition. The method is differentiated from the FE-approach in the

Table 4
Empirical Mean (sd) [se] of CQMLE and M-Estimator, **DGP2** (RE), $T = 3$, $m = 10$ $W_1 = W_3$: **Queen Contiguity**; W_2: **Group Interaction**.

	Normal Error		Normal Mixture		Chi-Square	
ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
	$n = 50$					
1	.979 (.047)	1.001 (.045) [.042]	.978 (.047)	1.000 (.045) [.042]	.977 (.047)	.999 (.046) [.042]
1	.936 (.321)	1.000 (.335) [.325]	.947 (.315)	1.011 (.330) [.323]	.930 (.314)	.993 (.329) [.322]
1	.971 (.153)	.937 (.144) [.134]	.983 (.248)	.949 (.236) [.203]	.980 (.223)	.947 (.213) [.187]
1	.861 (.373)	1.024 (.424) [.371]	.875 (.419)	1.037 (.474) [.398]	.873 (.416)	1.032 (.469) [.392]
.3	.333 (.042)	.297 (.040) [.035]	.334 (.041)	.298 (.039) [.036]	.334 (.041)	.298 (.040) [.035]
.2	.203 (.076)	.189 (.075) [.067]	.199 (.072)	.184 (.072) [.067]	.203 (.076)	.188 (.075) [.068]
.2	.204 (.061)	.194 (.059) [.055]	.202 (.061)	.193 (.059) [.055]	.202 (.061)	.192 (.058) [.055]
.2	.082 (.194)	.095 (.192) [.175]	.082 (.191)	.095 (.191) [.171]	.088 (.183)	.100 (.182) [.170]
	$n = 100$					
1	.979 (.032)	.999 (.031) [.030]	.979 (.032)	.999 (.031) [.030]	.979 (.032)	.999 (.031) [.029]
1	.956 (.226)	.998 (.234) [.229]	.967 (.221)	1.009 (.229) [.230]	.958 (.229)	.999 (.237) [.229]
1	1.002 (.107)	.974 (.103) [.098]	1.013 (.175)	.984 (.168) [.155]	1.006 (.161)	.978 (.154) [.143]
1	.869 (.244)	.996 (.276) [.252]	.877 (.279)	1.006 (.310) [.278]	.874 (.267)	1.001 (.298) [.272]
.3	.329 (.026)	.300 (.025) [.023]	.329 (.026)	.299 (.026) [.024]	.329 (.026)	.299 (.025) [.024]
.2	.186 (.044)	.195 (.044) [.042]	.186 (.044)	.195 (.045) [.042]	.186 (.045)	.195 (.046) [.042]
.2	.204 (.046)	.196 (.045) [.044]	.204 (.046)	.195 (.045) [.044]	.200 (.047)	.191 (.046) [.044]
.2	.163 (.126)	.154 (.128) [.120]	.161 (.125)	.151 (.126) [.119]	.160 (.126)	.151 (.128) [.119]
	$n = 200$					
1	.977 (.023)	.999 (.022) [.022]	.978 (.024)	1.000 (.023) [.022]	.978 (.023)	1.000 (.022) [.022]
1	.938 (.162)	1.002 (.169) [.168]	.941 (.166)	1.005 (.172) [.168]	.936 (.162)	1.000 (.169) [.168]
1	1.021 (.076)	.991 (.073) [.071]	1.014 (.121)	.984 (.116) [.113]	1.019 (.113)	.989 (.108) [.105]
1	.861 (.169)	.996 (.191) [.180]	.879 (.195)	1.015 (.217) [.202]	.867 (.185)	1.002 (.206) [.195]
.3	.331 (.019)	.300 (.018) [.017]	.331 (.019)	.300 (.019) [.017]	.331 (.019)	.300 (.019) [.017]
.2	.195 (.036)	.194 (.037) [.035]	.199 (.036)	.198 (.036) [.035]	.198 (.035)	.196 (.035) [.035]
.2	.197 (.037)	.196 (.037) [.035]	.198 (.037)	.197 (.036) [.035]	.198 (.036)	.197 (.036) [.035]
.2	.178 (.089)	.180 (.089) [.088]	.173 (.089)	.174 (.089) [.087]	.177 (.088)	.178 (.088) [.087]
	$n = 400$					
1	.979 (.017)	1.000 (.016) [.016]	.979 (.017)	1.001 (.016) [.016]	.979 (.017)	1.000 (.016) [.016]
1	.953 (.112)	.997 (.116) [.117]	.955 (.113)	.998 (.117) [.117]	.957 (.114)	1.001 (.118) [.117]
1	1.023 (.052)	.992 (.050) [.051]	1.025 (.090)	.995 (.086) [.082]	1.023 (.080)	.993 (.076) [.075]
1	.867 (.116)	1.003 (.132) [.128]	.868 (.136)	1.004 (.152) [.142]	.870 (.131)	1.006 (.146) [.139]
.3	.332 (.013)	.300 (.013) [.012]	.331 (.013)	.299 (.013) [.012]	.332 (.013)	.300 (.013) [.012]
.2	.190 (.023)	.199 (.023) [.023]	.189 (.023)	.198 (.024) [.023]	.189 (.023)	.198 (.023) [.023]
.2	.203 (.033)	.199 (.032) [.031]	.200 (.033)	.196 (.032) [.031]	.201 (.033)	.197 (.032) [.031]
.2	.199 (.062)	.188 (.063) [.061]	.196 (.062)	.185 (.063) [.061]	.198 (.062)	.187 (.063) [.061]

Note: $\psi = (\beta, \gamma, \sigma_v^2, \phi, \rho, \lambda_1, \lambda_2, \lambda_3)'$. The results corresponding to $\breve{\alpha}$ are suppressed to save space, and are reported in Supplement Appendix; X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$.

introduction and Supplementary Appendix. The results can be adapted to cover any CRE form that is linearizable in the sense that it can be written as or approximated by a linear model based on the observed time-varying regressors. The most general CRE form may be $\mu = g(X_0, X_1, \ldots, X_T) + \varepsilon$ with an unknown $g(\cdot)$ and an additive error ε, giving a SDPD model with *nonlinear* CRE. Standard semiparametric methods may be used to handle this unknown function and the model estimation may proceed in a similar way as that in this paper. This is an interesting model specification, but a detailed study is clearly beyond the scope of this paper, which will be carried out in a future research.

We have focused on the case where the idiosyncratic errors $\{v_{it}\}$ are iid. While time dependence is already built in the model as a dynamic lag of the response, it may be important to allow time dependence in $\{v_t\}$ as well in case of excessive time dependence of the process. We show that our results can be extended by allowing v_t to follow an MA(1) process:

$$v_t = \nu_t + \tau \nu_{t-1},$$

where $\{\nu_{it}\}$ are iid$(0, \sigma_\nu^2)$. It is easy to see that $\mathrm{E}(v_t v_t') = (1+\tau^2)\sigma_\nu^2 I_n$, and that $\mathrm{E}(v_t v_{t-1}') = \mathrm{E}(v_{t-1} v_t') = \tau\sigma_\nu^2 I_n$, $t = 2, \ldots, T$, so that $\mathrm{E}(\mathbf{v}\mathbf{v}') = \sigma_\nu^2 \Sigma \otimes I_n$, where $\Sigma = (1+\tau^2)I_T + \tau A$ and A is $T \times T$ with its (i, j)th element being 1 if $i = j \pm 1$, and 0 otherwise. Then, letting $\phi = \sigma_\varepsilon^2/\sigma_\nu^2$, the VC matrix of the composite errors, $\mathbf{e} = \boldsymbol{\varepsilon} + \mathbf{B}_3^{-1}\mathbf{v}$, takes a similar form:

$$\mathrm{Var}(\mathbf{e}) = \sigma_\nu^2[\phi(J_T \otimes I_n) + \Sigma \otimes (B_3'B_3)^{-1}] \equiv \sigma_\nu^2 \Omega.$$

With the new parameter τ, the vector of model parameters becomes $\psi = (\boldsymbol{\beta}', \sigma_\nu^2, \phi, \tau, \rho, \lambda')'$. The results of Lemma 2.1 are extended, with $\mathbf{C}$ and $\mathbf{C}_{-1}$ being kept the same, but $\mathbf{D}$ and $\mathbf{D}_{-1}$ taking new and slightly more complicated expressions. The desired AQS functions are then obtained, leading to the M-estimator of ψ (see Supplementary Appendix for details). Theorems 2.1 and 2.2 can be extended as the AQS functions can be written as linear combinations of terms linear,

Table 5
Empirical Mean (sd) [se] of CQMLE and M-Estimator, **DGP2** (RE), $T = 3$, $m = 10$ $W_1 = W_3$: **Group Interaction**; W_2: **Queen Contiguity**.

	Normal Error		Normal Mixture		Chi-Square	
ψ	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
	$n = 50$					
1	.978 (.046)	1.000 (.045) [.042]	.976 (.047)	.998 (.046) [.041]	.977 (.046)	.999 (.045) [.042]
1	.954 (.328)	1.019 (.341) [.324]	.938 (.338)	1.001 (.351) [.323]	.945 (.326)	1.009 (.339) [.325]
1	.971 (.150)	.939 (.143) [.134]	.982 (.247)	.950 (.237) [.204]	.977 (.223)	.945 (.213) [.187]
1	.869 (.373)	1.022 (.426) [.371]	.877 (.420)	1.029 (.470) [.399]	.888 (.417)	1.043 (.473) [.397]
.3	.332 (.040)	.298 (.039) [.035]	.333 (.041)	.300 (.040) [.035]	.331 (.040)	.297 (.039) [.035]
.2	.186 (.051)	.187 (.051) [.047]	.187 (.053)	.187 (.053) [.047]	.187 (.053)	.188 (.053) [.047]
.2	.209 (.075)	.196 (.070) [.067]	.205 (.076)	.192 (.071) [.067]	.207 (.073)	.194 (.069) [.068]
.2	.130 (.156)	.127 (.158) [.134]	.133 (.154)	.130 (.155) [.130]	.131 (.153)	.130 (.154) [.129]
	$n = 100$					
1	.980 (.032)	.999 (.031) [.030]	.981 (.032)	1.000 (.031) [.030]	.981 (.033)	1.000 (.032) [.030]
1	.988 (.223)	1.004 (.231) [.228]	.996 (.226)	1.012 (.234) [.228]	.982 (.232)	.997 (.240) [.228]
1	1.003 (.108)	.977 (.104) [.099]	1.004 (.172)	.978 (.165) [.154]	1.002 (.156)	.976 (.150) [.142]
1	.875 (.246)	.994 (.275) [.251]	.892 (.266)	1.012 (.293) [.278]	.890 (.267)	1.009 (.293) [.273]
.3	.327 (.025)	.300 (.025) [.023]	.327 (.025)	.300 (.025) [.023]	.327 (.025)	.299 (.024) [.023]
.2	.193 (.041)	.191 (.041) [.039]	.191 (.040)	.189 (.041) [.039]	.193 (.040)	.191 (.041) [.039]
.2	.192 (.047)	.194 (.045) [.044]	.193 (.046)	.194 (.045) [.044]	.196 (.046)	.197 (.045) [.043]
.2	.145 (.120)	.147 (.121) [.111]	.148 (.120)	.150 (.120) [.110]	.153 (.117)	.155 (.117) [.107]
	$n = 200$					
1	.977 (.022)	1.000 (.021) [.021]	.977 (.023)	1.000 (.022) [.022]	.978 (.023)	1.000 (.022) [.021]
1	.933 (.161)	1.001 (.168) [.168]	.930 (.162)	.998 (.169) [.168]	.931 (.166)	.998 (.173) [.167]
1	1.018 (.077)	.988 (.074) [.071]	1.014 (.122)	.984 (.117) [.113]	1.019 (.114)	.989 (.109) [.104]
1	.867 (.171)	1.001 (.194) [.180]	.877 (.193)	1.011 (.216) [.201]	.865 (.187)	.997 (.209) [.194]
.3	.332 (.018)	.300 (.018) [.017]	.331 (.019)	.300 (.018) [.017]	.331 (.019)	.300 (.019) [.017]
.2	.186 (.031)	.195 (.031) [.030]	.187 (.030)	.196 (.030) [.030]	.186 (.031)	.195 (.031) [.029]
.2	.204 (.033)	.198 (.033) [.032]	.203 (.033)	.198 (.033) [.032]	.202 (.033)	.197 (.032) [.032]
.2	.167 (.101)	.161 (.103) [.095]	.171 (.096)	.165 (.097) [.094]	.165 (.097)	.159 (.099) [.094]
	$n = 400$					
1	.977 (.017)	1.000 (.016) [.016]	.977 (.017)	1.000 (.016) [.016]	.977 (.017)	1.000 (.016) [.016]
1	.958 (.113)	1.001 (.117) [.116]	.956 (.109)	.998 (.114) [.116]	.958 (.112)	1.000 (.117) [.116]
1	1.026 (.056)	.995 (.053) [.051]	1.025 (.087)	.994 (.083) [.081]	1.025 (.081)	.994 (.078) [.075]
1	.861 (.118)	.999 (.134) [.127]	.867 (.135)	1.005 (.151) [.142]	.862 (.133)	1.001 (.148) [.139]
.3	.332 (.013)	.300 (.013) [.012]	.332 (.013)	.299 (.013) [.012]	.332 (.013)	.300 (.013) [.012]
.2	.197 (.032)	.196 (.033) [.032]	.196 (.033)	.195 (.034) [.032]	.196 (.033)	.195 (.033) [.032]
.2	.198 (.023)	.199 (.023) [.023]	.198 (.023)	.199 (.023) [.023]	.198 (.024)	.199 (.024) [.023]
.2	.172 (.082)	.174 (.083) [.078]	.176 (.081)	.177 (.081) [.078]	.170 (.081)	.172 (.082) [.078]

Note: $\psi = (\beta, \gamma, \sigma_v^2, \phi, \rho, \lambda_1, \lambda_2, \lambda_3)'$. The results corresponding to $\breve{\alpha}$ are suppressed to save space, and are reported in Supplement Appendix; X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$.

quadratic, and bilinear in $\boldsymbol{\nu}$, $\boldsymbol{\nu}_{-1}$, ε, and V_m. Lemma 3.1 and Theorem 3.1 can be extended as well by re-defining the $\mathbf{g}_i$ functions and re-deriving the results in Lemma 3.1. While fundamental ideas are the same, these extensions require additional complicated algebra and proofs, and need to be handled by a separate research.

So far, the time heterogeneity appears in the model in the form of time-specific effects $\{\alpha_t\}$. It may be of interest to allow more extensive forms of time-heterogeneity such as time-varying regression coefficients, time-varying spatial coefficients, time-varying spatial weight matrices, etc. From the theoretical developments, we see that our results may be extended to allow for time-varying regression coefficients, but may not be for the other types of time-heterogeneity. Finally, the cross-sectional heteroskedasticity (space-varying error variances) in the CRE-SDPD model is another interesting extension to consider. It requires an entirely different way to adjust the conditional quasi scores so that the AQS functions obtained are not only (asymptotically) unbiased but also robust against unknown cross-sectional heteroskedasticity. These models and methods would be much more challenging than the already quite challenging works presented in this paper, and will be the topics of our future research.

Finally, it should be pointed out that moving the $Z\gamma$ term in (1.1) to (1.2), i.e., letting $\mu = \bar{X}\pi + Z\gamma + \varepsilon$, gives an equivalent model specification, and all results carry over, including the equivalence between FE and CRE estimators of the coefficients of time-varying regressors under a simple static panel data model as discussed in the introduction (see also Krishnakumar, 2006). In this case, μ is explained as having two components: observable and unobservable (see Hausman and Taylor, 1981, p.1378). However, common perception on μ is that it represents unobservable individual-specific effects such as ability or managerial skill, and hence the original specification in (1.1) and (1.2) would be more sensible. Furthermore, the time-invariant variables Z are assumed to be strictly exogenous. This is reasonable but not entirely necessary, because some variables in Z may be linearly correlated with μ through X_t and this type of endogeneity

Table 6
Empirical Mean (sd) of CQMLE, M-Estimator, and FQMLE, **DGP3** (RE), $T = 3$, $m = 10$; W_3: **Queen Contiguity**.

		Normal Error			Normal Mixture			Chi-Square		
n	ψ	CQMLE	M-Est	FQMLE	CQMLE	M-Est	FQMLE	CQMLE	M-Est	FQMLE
50	1	.971 (.045)	.998 (.044)	.999 (.044)	.974 (.046)	1.001 (.045)	1.002 (.044)	.973 (.045)	1.000 (.043)	1.001 (.043)
	1	.888 (.317)	1.000 (.338)	.991 (.337)	.889 (.335)	1.001 (.359)	.993 (.357)	.879 (.316)	.990 (.338)	.982 (.338)
	1	1.006 (.076)	.984 (.073)	.983 (.071)	1.003 (.124)	.982 (.120)	.980 (.117)	1.006 (.117)	.985 (.113)	.983 (.111)
	1	.785 (.325)	.992 (.400)	.994 (.366)	.839 (.400)	1.052 (.474)	1.058 (.446)	.815 (.391)	1.020 (.458)	1.033 (.435)
	.5	.539 (.032)	.500 (.032)	.499 (.031)	.538 (.032)	.499 (.032)	.498 (.030)	.540 (.032)	.501 (.032)	.499 (.031)
	.3	.265 (.142)	.269 (.142)	.235 (.139)	.266 (.151)	.271 (.151)	.236 (.146)	.262 (.146)	.267 (.147)	.233 (.141)
100	1	.975 (.032)	.999 (.031)	1.000 (.031)	.974 (.032)	.999 (.032)	.999 (.031)	.975 (.032)	.999 (.032)	1.000 (.031)
	1	.918 (.224)	.996 (.237)	1.000 (.236)	.924 (.227)	1.002 (.240)	1.005 (.239)	.922 (.228)	1.000 (.241)	1.005 (.240)
	1	1.010 (.052)	.993 (.051)	.993 (.050)	1.009 (.086)	.992 (.084)	.993 (.083)	1.011 (.080)	.994 (.078)	.994 (.075)
	1	.830 (.229)	.992 (.274)	.996 (.247)	.849 (.259)	1.012 (.302)	1.014 (.277)	.841 (.252)	1.004 (.295)	1.010 (.273)
	.5	.531 (.020)	.500 (.021)	.499 (.020)	.531 (.020)	.500 (.020)	.499 (.019)	.531 (.020)	.500 (.020)	.499 (.020)
	.3	.283 (.106)	.285 (.107)	.289 (.098)	.283 (.104)	.285 (.104)	.290 (.096)	.281 (.106)	.284 (.106)	.288 (.096)
200	1	.968 (.023)	.999 (.023)	1.000 (.022)	.969 (.023)	1.000 (.023)	1.001 (.022)	.969 (.023)	1.000 (.023)	1.001 (.023)
	1	.954 (.157)	.998 (.166)	.996 (.166)	.950 (.157)	.993 (.166)	.991 (.166)	.951 (.158)	.994 (.167)	.992 (.167)
	1	1.014 (.037)	.996 (.036)	.995 (.035)	1.012 (.061)	.994 (.060)	.993 (.059)	1.012 (.056)	.994 (.054)	.993 (.053)
	1	.821 (.156)	.990 (.189)	.993 (.176)	.837 (.181)	1.007 (.213)	1.009 (.204)	.833 (.175)	1.004 (.208)	1.007 (.193)
	.5	.533 (.014)	.500 (.014)	.500 (.014)	.533 (.014)	.500 (.014)	.499 (.014)	.533 (.014)	.500 (.014)	.500 (.014)
	.3	.286 (.073)	.289 (.074)	.266 (.072)	.292 (.075)	.295 (.076)	.271 (.074)	.288 (.073)	.291 (.074)	.268 (.071)
400	1	.972 (.015)	1.000 (.015)	1.002 (.016)	.973 (.016)	1.000 (.015)	1.002 (.016)	.973 (.015)	1.000 (.015)	1.002 (.016)
	1	.933 (.112)	1.000 (.120)	1.010 (.120)	.935 (.110)	1.002 (.117)	1.013 (.118)	.932 (.114)	.998 (.121)	1.008 (.122)
	1	1.018 (.027)	.998 (.026)	1.029 (.027)	1.016 (.043)	.997 (.042)	1.027 (.036)	1.017 (.041)	.997 (.040)	1.028 (.035)
	1	.813 (.113)	.996 (.137)	.955 (.211)	.823 (.127)	1.007 (.151)	.971 (.212)	.819 (.126)	1.001 (.150)	.961 (.213)
	.5	.535 (.010)	.500 (.010)	.498 (.012)	.535 (.011)	.500 (.011)	.497 (.012)	.535 (.011)	.500 (.011)	.498 (.012)
	.3	.293 (.051)	.297 (.051)	.285 (.053)	.291 (.053)	.295 (.053)	.283 (.053)	.293 (.052)	.296 (.053)	.284 (.053)

Note: $\psi = (\beta, \gamma, \sigma_v^2, \phi, \rho, \lambda_3)'$. The results corresponding to $\breve{\alpha}$ are suppressed to save space, and are reported in Supplement Appendix; X_t values are generated with $\theta_x = (g, \phi_1, \phi_2, \sigma_1, \sigma_2) = (.01, .5, .5, 2, 1)$.

Table 7
Spatial and dynamic interaction/competition in investments among Chinese cities.

	W_{prov}		W_{geo}	
Population	0.263***	[0.119]	0.424***	[0.206]
GDP	0.489***	[0.058]	0.412***	[0.174]
Fiscal revenue	0.575	[0.487]	3.454*	[1.370]
Fiscal expenditure	0.635***	[0.247]	0.906	[1.241]
Fiscal account balance	−0.337	[0.293]	0.297	[1.177]
Provincial fiscal revenue	−0.162***	[0.041]	−0.239***	[0.121]
Provincial fiscal expenditure	0.074***	[0.025]	0.194***	[0.093]
Public capital investment	−0.058*	[0.036]	−0.140	[0.119]
Autonomous city	−97.861***	[37.057]	−187.332**	[103.53]
2011	−25.562	[28.248]	−344.618	[638.06]
2012	−9.296	[15.918]	−149.003	[288.44]
Dynamic parameter				
Time lag (ρ)	0.211***	[0.061]	−0.803	[2.556]
Spatial parameters				
SL (λ_1)	0.249***	[0.068]	0.041	[0.712]
STL (λ_2)	−0.042	[0.070]	0.043	[0.218]
SE (λ_3)	0.293***	[0.115]	0.287	[0.899]
Correlated random effects				
Average GDP	−0.165*	[0.108]	0.325	[1.164]
Average fiscal revenue	−0.206	[0.777]	−1.785	[5.237]
Average fiscal expenditure	−0.583	[1.092]	−1.750	[2.430]
Average fiscal account balance	−0.059	[0.424]	−0.346	[1.783]

Note: Population is measured in 10^4, and other variables excluding dummies are measured in 10^8.
*Significance at 10% level.
**Significance at 5% level.
***Significance at 1% level.

may be captured by Mundlak's or Chamberlain's specification. If not, a more general CRE specification in line with the above discussion may help. See Hausman and Taylor (1981) for a general discussion on the endogeneity in Z. A full treatment of the issue of endogeneity in the components of Z under the current SDPD-CRE setting would be an interesting topic of future research.

Appendix A. Some basic Lemmas

Lemma A.1 (Kelejian and Prucha, 1999; Lee, 2002). *Let $\{A_n\}$ and $\{B_n\}$ be two sequences of $n \times n$ matrices that are uniformly bounded in both row and column sums. Let C_n be a sequence of conformable matrices whose elements are uniformly $O(h_n^{-1})$. Then*

(i) the sequence $\{A_nB_n\}$ are uniformly bounded in both row and column sums,
(ii) the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$, and
(iii) the elements of A_nC_n and C_nA_n are uniformly $O(h_n^{-1})$.

Lemma A.2 (Lee, 2004, p.1918). *For W_1 and B_1 defined in Model (1.1), if $\|W_1\|$ and $\|B_{10}^{-1}\|$ are uniformly bounded, where $\|\cdot\|$ is a matrix norm, then $\|B_1^{-1}\|$ is uniformly bounded in a neighborhood of λ_{10}.*

Lemma A.3 (Lee, 2004, p.1918). *Let X_n be an $n \times p$ matrix. If the elements X_n are uniformly bounded and $\lim_{n\to\infty} \frac{1}{n}X_n'X_n$ exists and is nonsingular, then $P_n = X_n(X_n'X_n)^{-1}X_n'$ and $M_n = I_n - P_n$ are uniformly bounded in both row and column sums.*

Lemma A.4 (Lemma A.4, Yang, 2018). *Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in either row or column sums. Suppose that the elements $a_{n,ij}$ of A_n are $O(h_n^{-1})$ uniformly in all i and j. Let v_n be a random n-vector of iid elements with mean zero, variance σ^2 and finite 4th moment, and b_n a constant n-vector of elements of uniform order $O(h_n^{-1/2})$. Then*

(i) $\mathrm{E}(v_n'A_nv_n) = O(\frac{n}{h_n})$, *(ii) $v\mathrm{Var}(v_n'A_nv_n) = O(\frac{n}{h_n})$,*
(iii) $\mathrm{Var}(v_n'A_nv_n + b_n'v_n) = O(\frac{n}{h_n})$, *(iv) $v_n'A_nv_n = O_p(\frac{n}{h_n})$,*
(v) $v_n'A_nv_n - \mathrm{E}(v_n'A_nv_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, *(vi) $v_n'A_nb_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$,*

and (vii), the results (iii) and (vi) remain valid if b_n is a random n-vector independent of v_n such that $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$.

Lemma A.5 (Lemma A.5, Yang, 2018). *Let $\{\Phi_n\}$ be a sequence of $n \times n$ matrices with row and column sums uniformly bounded, and elements of uniform order $O(h_n^{-1})$. Let $v_n = (v_1, \ldots, v_n)'$ be a random vector of iid elements with mean zero, variance σ_v^2, and finite $(4+2\epsilon_0)$th moment for some $\epsilon_0 > 0$. Let $b_n = \{b_{ni}\}$ be an $n \times 1$ random vector, independent of v_n, such that (i) $\{\mathrm{E}(b_{ni}^2)\}$ are of uniform order $O(h_n^{-1})$, (ii) $sup_iE|b_{ni}|^{2+\epsilon_0} < \infty$, (iii)$\frac{h_n}{n}\sum_{i=1}^n[\phi_{n,ii}(b_{ni} - \mathrm{E}b_{ni})] = o_p(1)$ where $\{\phi_{n,ii}\}$ are the diagonal elements of Φ_n, and (iv)$\frac{h_n}{n}\sum_{i=1}^n[b_{ni}^2 - \mathrm{E}(b_{ni}^2)] = o_p(1)$. Define the bilinear–quadratic form:*

$$Q_n = b_n'v_n + v_n'\Phi_nv_n - \sigma_v^2 tr(\Phi_n),$$

and let $\sigma_{Q_n}^2$ be the variance of Q_n. If $\lim_{n\to\infty}h_n^{1+2/\epsilon_0}/n = 0$ and $\{\frac{h_n}{n}\sigma_{Q_n}^2\}$ are bounded away from zero, then $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1)$.

Lemma A.6. *Under Assumption F, for an $n \times n$ matrix Φ uniformly bounded in either row or column sums, with elements of uniform order h_n^{-1}, and an $n \times 1$ vector ϕ with elements of uniform order $h_n^{-1/2}$, we have:*
(i) $\frac{h_n}{n}y_0'\Phi y_0 = O_p(1)$; (ii) $\frac{h_n}{n}[y_0 - \mathrm{E}(y_0)]'\phi = o_p(1)$; (iii) $\frac{h_n}{n}[y_0'\Phi y_0 - \mathrm{E}(y_0'\Phi y_0)] = o_p(1)$.

Appendix B. Proofs for Section 2

Proof of Lemma 2.1. By (2.1), backward substitution leads to, for $t = -m+1, \ldots, T$,

$$\begin{aligned}\mathrm{E}(y_t\varepsilon') &= B_1^{-1}B_2E(y_{t-1}\varepsilon') + B_1^{-1}E(\varepsilon\varepsilon') + B_1^{-1}B_3^{-1}\mathrm{E}(v_t\varepsilon') \\ &= \mathcal{B}^t\mathrm{E}(y_0\varepsilon') + (\sum_{i=0}^{t-1}\mathcal{B}^i)B_1^{-1}\mathrm{E}(\varepsilon\varepsilon') = (\sum_{i=0}^{t+m-1}\mathcal{B}^i)B_1^{-1}\sigma_{\varepsilon0}^2.\end{aligned}$$

Therefore, $\mathrm{E}(Y_{-1}\boldsymbol{\varepsilon}') = \sigma_{\varepsilon0}^2\mathbf{C}_{-1}$ and $\mathrm{E}(Y\boldsymbol{\varepsilon}') = \sigma_{\varepsilon0}^2\mathbf{C}$.

For $t, s = 1, \ldots, T$, we have $\mathrm{E}(y_tv_t') = B_1^{-1}B_2E(y_{t-1}v_t') + B_1^{-1}B_3^{-1}\mathrm{E}(v_tv_t') = \sigma_{v0}^2B_1^{-1}B_3^{-1}$; $\mathrm{E}(y_tv_s') = 0$ when $t < s$; and

$$\begin{aligned}\mathrm{E}(y_tv_s') &= B_1^{-1}B_2E(y_{t-1}v_s') + B_1^{-1}B_3^{-1}E(v_tv_s') = \mathcal{B}^2E(y_{t-2}v_s') = \cdots \\ &= \mathcal{B}^{t-s}E(y_sv_s') = \mathcal{B}^{t-s}E(B_1^{-1}B_3^{-1}v_sv_s') = \mathcal{B}^{t-s}B_1^{-1}B_3^{-1}\sigma_{v0}^2,\end{aligned}$$

when $t > s$. Therefore, $\mathrm{E}(Y_{-1}\mathbf{v}')(\mathbf{B}_3^{-1})' = \sigma_{v0}^2\mathbf{D}_{-1}$ and $\mathrm{E}(Y\mathbf{v}')(\mathbf{B}_3^{-1})' = \sigma_{v0}^2\mathbf{D}$. Combining these results, we obtain the results of Lemma 2.1. ■

Proof of Theorem 2.1. The proof of this theorem uses similar ideas as in the proof of Theorem 3.1 of Yang (2018). Rather than working with differences series, levels are used and account need to be taken of additional randomness from ε. Under

Assumption G, by Theorem 5.9 of van der Vaart (1998) the consistency of $\hat{\delta}_{\mathrm{M}}$ follows if $\sup_{\delta\in\Delta}\frac{1}{nT}\|S^{*c}_{\mathrm{SDPD}}(\delta)-\bar{S}^{*c}_{\mathrm{SDPD}}(\delta)\| \xrightarrow{p} 0$ as $n\to\infty$, where $S^{*c}_{\mathrm{SDPD}}(\delta)$ is the concentrated AQS function of δ defined below (2.13), and $\bar{S}^{*c}_{\mathrm{SDPD}}(\delta)$ is its population counterpart defined above Theorem 2.1, given below

$$S^{*c}_{\mathrm{SDPD}}(\delta)=\begin{cases}\frac{1}{2\hat{\sigma}_v^2(\delta)}\hat{e}'(\delta)\Omega^{-1}(J_T\otimes I_n)\Omega^{-1}\hat{e}(\delta)-\frac{1}{2}\mathrm{tr}[\Omega^{-1}(J_T\otimes I_n)],\\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{e}'(\delta)\Omega^{-1}Y_{-1}-\mathrm{tr}[(\phi\mathbf{C}_{-1}+\mathbf{D}_{-1})\Omega^{-1}],\\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{e}'(\delta)\Omega^{-1}\mathbf{W}_1Y-\mathrm{tr}[(\phi\mathbf{C}+\mathbf{D})\Omega^{-1}\mathbf{W}_1],\\ \frac{1}{\hat{\sigma}_v^2(\delta)}\hat{e}'(\delta)\Omega^{-1}\mathbf{W}_2Y_{-1}-\mathrm{tr}[(\phi\mathbf{C}_{-1}+\mathbf{D}_{-1})\Omega^{-1}\mathbf{W}_2],\\ \frac{1}{2\hat{\sigma}_v^2(\delta)}\hat{e}'(\delta)\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}\hat{e}(\delta)-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\dot{\Omega}_{\lambda_3}),\end{cases}\tag{B.1}$$

$$\bar{S}^{*c}_{\mathrm{SDPD}}(\delta)=\begin{cases}\frac{1}{2\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{\mathbf{e}}'(\delta)\Omega^{-1}(J_T\otimes I_n)\Omega^{-1}\bar{\mathbf{e}}(\delta)]-\frac{1}{2}\mathrm{tr}[\Omega^{-1}(J_T\otimes I_n)],\\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{\mathbf{e}}'(\delta)\Omega^{-1}Y_{-1}]-\mathrm{tr}[(\phi\mathbf{C}_{-1}+\mathbf{D}_{-1})\Omega^{-1}],\\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{\mathbf{e}}'(\delta)\Omega^{-1}\mathbf{W}_1Y]-\mathrm{tr}[(\phi\mathbf{C}+\mathbf{D})\Omega^{-1}\mathbf{W}_1],\\ \frac{1}{\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{\mathbf{e}}'(\delta)\Omega^{-1}\mathbf{W}_2Y_{-1}]-\mathrm{tr}[(\phi\mathbf{C}_{-1}+\mathbf{D}_{-1})\Omega^{-1}\mathbf{W}_2],\\ \frac{1}{2\bar{\sigma}_v^2(\delta)}\mathrm{E}[\bar{\mathbf{e}}'(\delta)\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}\bar{\mathbf{e}}(\delta)]-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\dot{\Omega}_{\lambda_3}),\end{cases}\tag{B.2}$$

where $\hat{\sigma}_v^2(\delta)$ is defined in (2.13), and $\bar{\sigma}_v^2(\delta)$ is defined above Theorem 2.1. With (B.1) and (B.2), the proof of consistency of $\hat{\delta}_{\mathrm{M}}$ boils down to the proofs of the following:

(a) $\inf_{\delta\in\boldsymbol{\Delta}}\bar{\sigma}_v^2(\delta)$ is bounded away from zero,

(b) $\sup_{\delta\in\boldsymbol{\Delta}}\left|\hat{\sigma}_v^2(\delta)-\bar{\sigma}_v^2(\delta)\right|=o_p(1)$,

(c) $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{nT}\left|\hat{e}'(\delta)\Omega^{-1}(J_T\otimes I_n)\Omega^{-1}\hat{e}(\delta)-\mathrm{E}[\bar{e}'(\delta)\Omega^{-1}(J_T\otimes I_n)\Omega^{-1}\bar{e}(\delta)]\right|=o_p(1)$,

(d) $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{nT}\left|\hat{e}'(\delta)\Omega^{-1}Y_{-1}-\mathrm{E}[\bar{e}'(\delta)\Omega^{-1}Y_{-1}]\right|=o_p(1)$,

(e) $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{nT}\left|\hat{e}'(\delta)\Omega^{-1}\mathbf{W}_1Y-\mathrm{E}[\bar{e}'(\delta)\Omega^{-1}\mathbf{W}_1Y]\right|=o_p(1)$,

(f) $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{nT}\left|\hat{e}'(\delta)\Omega^{-1}\mathbf{W}_2Y_{-1}-\mathrm{E}[\bar{e}'(\delta)\Omega^{-1}\mathbf{W}_2Y_{-1}]\right|=o_p(1)$,

(g) $\sup_{\delta\in\boldsymbol{\Delta}}\frac{1}{nT}\left|\hat{e}'(\delta)\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}\hat{e}(\delta)-\mathrm{E}[\bar{e}'(\delta)\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}\hat{e}(\delta)]\right|=o_p(1)$.

Let $\Omega^{\frac{1}{2}}$ be a square-root matrix of Ω. Define $\bar{\mathbf{e}}^*(\delta)=\Omega^{-\frac{1}{2}}\bar{\mathbf{e}}(\delta)$, $\hat{\mathbf{e}}^*(\delta)=\Omega^{-\frac{1}{2}}\hat{\mathbf{e}}(\delta)$, and $\mathbf{B}_r^*=\Omega^{-\frac{1}{2}}\mathbf{B}_r$, $r=1,2$. Let $Y^\circ=Y-\mathrm{E}(Y)$ and $Y_{-1}^\circ=Y_{-1}-\mathrm{E}(Y_{-1})$. Define the projection matrices: $\mathbf{M}=I_{nT}-\Omega^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}$ and $\mathbf{P}=I_{nT}-\mathbf{M}$. We have:

$$\bar{\mathbf{e}}^*(\delta)=\mathbf{M}(\mathbf{B}_1^*Y-\mathbf{B}_2^*Y_{-1})+\mathbf{P}(\mathbf{B}_1^*Y^\circ-\mathbf{B}_2^*Y_{-1}^\circ),\tag{B.3}$$

$$\hat{\mathbf{e}}^*(\delta)=\mathbf{M}(\mathbf{B}_1^*Y-\mathbf{B}_2^*Y_{-1}).\tag{B.4}$$

Proof of (a). Recall that $\bar{\sigma}_v^2(\delta)=\frac{1}{nT}\mathrm{E}[\bar{\mathbf{e}}^{*\prime}(\delta)\bar{\mathbf{e}}^*(\delta)]$, by (B.3) and the orthogonality between $\mathbf{M}$ and $\mathbf{P}$, we can write

$$\bar{\sigma}_v^2(\delta)=\frac{1}{nT}\mathrm{tr}[\mathrm{Var}(\mathbf{B}_1^*Y-\mathbf{B}_2^*Y_{-1})]+\frac{1}{nT}(\mathbf{B}_1^*\mathrm{E}Y-\mathbf{B}_2^*\mathrm{E}Y_{-1})'\mathbf{M}(\mathbf{B}_1^*\mathrm{E}Y-\mathbf{B}_2^*\mathrm{E}Y_{-1}).$$

As $\mathbf{M}$ is positive semi-definite (p.s.d), the second term is non-negative uniformly in $\delta\in\boldsymbol{\Delta}$. By Assumptions C, and E (iv), $\inf_{\delta\in\boldsymbol{\Delta}}\gamma_{\max}(\Omega)\le\sup_{\delta\in\boldsymbol{\Delta}}\gamma_{\max}(\Omega)\le\phi+\frac{1}{c_3}$. Therefore the first term is $\frac{1}{nT}\mathrm{tr}[\Omega^{-1}\mathrm{Var}(\mathbf{B}_1Y-\mathbf{B}_2Y_{-1})]\ge\frac{1}{nT}\gamma_{\max}^{-1}(\Omega)\mathrm{tr}[\mathrm{Var}(\mathbf{B}_1Y-\mathbf{B}_2Y_{-1})]\ge c>0$, uniformly in $\delta\in\boldsymbol{\Delta}$. It follows that $\inf_{\delta\in\boldsymbol{\Delta}}\bar{\sigma}_v^2(\delta)>c>0$.[10]

Proof of (b). By (B.3) and (B.4), we can decompose $\hat{\sigma}_v^2(\delta)-\bar{\sigma}_v^2(\delta)$ into four terms

$$\hat{\sigma}_v^2(\delta)-\bar{\sigma}_v^2(\delta)=(Q_1-\mathrm{E}Q_1)+(Q_2-\mathrm{E}Q_2)+2(Q_3-\mathrm{E}Q_3)+\mathrm{E}Q_4.\tag{B.5}$$

where $Q_1=\frac{1}{nT}Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_1^*Y$, $Q_2=\frac{1}{nT}Y_{-1}'\mathbf{B}_2^{*\prime}\mathbf{M}\mathbf{B}_2^*Y_{-1}$, $Q_3=-\frac{2}{nT}Y'\mathbf{B}_1^{*\prime}\mathbf{M}\mathbf{B}_2^*Y_{-1}$, and $Q_4=-\frac{1}{nT}(\mathbf{B}_1^*Y^\circ-\mathbf{B}_2^*Y_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*Y^\circ-\mathbf{B}_2^*Y_{-1}^\circ)$. The results follows if $Q_j-\mathrm{E}Q_j\xrightarrow{p}0$, $j=1,2,3$, and $\mathrm{E}Q_4\to 0$, uniformly in $\delta\in\boldsymbol{\Delta}$. By (2.14) and letting $\mathbf{M}^*=\Omega^{-\frac{1}{2}}\mathbf{M}\Omega^{-\frac{1}{2}}$, we have

$$Q_1=\sum_{\ell=1}^{9}Q_{1,\ell}+\frac{1}{nT}\boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_1\boldsymbol{\eta},$$

[10] **Note**: (i) eigenvalues of a projection matrix are either 0 or 1; (ii) eigenvalues of a positive definite matrix are strictly positive; (iii) for symmetric matrix A and positive semidefinite (p.s.d.) matrix B, $\gamma_{\min}(A)\mathrm{tr}(B)\le\mathrm{tr}(AB)\le\gamma_{\max}(A)\mathrm{tr}(B)$; ($iv$) for symmetric matrices A and B, $\gamma_{\max}(A+B)\le\gamma_{\max}(A)+\gamma_{\max}(B)$; and ($v$) for p.s.d. matrices A and B, $\gamma_{\max}(AB)\le\gamma_{\max}(A)\gamma_{\max}(B)$. See, e.g., Bernstein (2009).

$$Q_2 = \sum_{\ell=1}^{9} Q_{2,\ell} + \frac{1}{nT}\boldsymbol{\eta}_{-1}'\mathbf{B}_2'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1},$$

$$Q_3 = \sum_{\ell=1}^{14} Q_{3,\ell} + \frac{1}{nT}\boldsymbol{\eta}'\mathbf{B}_1'\mathbf{M}^*\mathbf{B}_2\boldsymbol{\eta}_{-1},$$

where $Q_{k\ell}$ takes one of the forms: $\frac{1}{nT}\mathbf{y}_0'\Phi_1(\delta)\mathbf{y}_0$, $\frac{1}{nT}\mathbf{v}'\Phi_2(\delta)\mathbf{v}$, $\frac{1}{nT}\boldsymbol{\varepsilon}'\Phi_3(\delta)\boldsymbol{\varepsilon}$, $\frac{1}{nT}\mathbf{y}_0'\Psi_1(\delta)\mathbf{v}$, $\frac{1}{nT}\mathbf{y}_0'\Psi_2(\delta)\boldsymbol{\varepsilon}$, $\frac{1}{nT}\boldsymbol{\varepsilon}'\Psi_3(\delta)\mathbf{v}$, $\frac{1}{nT}\mathbf{y}_0'\Pi_1(\delta)$, $\frac{1}{nT}\mathbf{v}'\Pi_2(\delta)$, and $\frac{1}{nT}\boldsymbol{\varepsilon}'\Pi_3(\delta)$. The matrices $\Phi_r(\delta)$ and $\Psi_r(\delta)$, and vectors $\Pi_r(\delta)$, $r = 1, 2, 3$, depend on δ through $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{M}^*$, and involve $\mathbb{Q}$, $\mathbb{Q}_{-1}$, $\mathbb{S}$, $\mathbb{S}_{-1}$, $\mathbb{B}$, $\mathbb{B}_{-1}$, $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_{-1}$, which are all matrix or vector functions of true parameters.

By Lemma A.1, Assumption E and the expressions in (2.15) and (2.16), the $nT \times nT$ matrices $\mathbb{R}$, $\mathbb{R}_{-1}$, $\mathbb{S}$, $\mathbb{S}_{-1}$, $\mathbb{B}$ and $\mathbb{B}_{-1}$ are uniformly bounded in both row and column sums, and the elements of the $nT \times 1$ vectors $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_{-1}$ are uniformly bounded. By Assumption E(iii) and (v), Assumption D, Lemma A.1 and Lemma A.3, $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{M}^*$ are uniformly bounded in either row or column sums. By Lemma A.6, it can be easily shown that $\frac{1}{nT}[\mathbf{y}_0'\Phi_1(\delta)\mathbf{y}_0 - \mathrm{E}(\mathbf{y}_0'\Phi_1(\delta)\mathbf{y}_0)] = o_p(1)$, and $\frac{1}{nT}[\mathbf{y}_0'\Pi_1(\delta) - \mathrm{E}(\mathbf{y}_0')\Pi_1(\delta)] = o_p(1)$. The point wise convergence of the quadratic terms $\frac{1}{nT}\mathbf{v}'\Phi_2(\delta)\mathbf{v}$ and $\frac{1}{nT}\boldsymbol{\varepsilon}'\Phi_3(\delta)\boldsymbol{\varepsilon}$, and the bilinear term $\frac{1}{nT}\mathbf{y}_0'\Psi_1(\delta)\mathbf{v}$, can be established by Assumption B and Lemma A.4. The result $\frac{1}{nT}\{\mathbf{y}_0'\Psi_2(\delta)\boldsymbol{\varepsilon} - \mathrm{E}[\mathbf{y}_0'\Psi_2(\delta)\boldsymbol{\varepsilon}]\} = o_p(1)$ is proved by decomposing y_0 into three terms using (2.17) and then applying Lemma A.4 under Assumptions B and F. The point wise convergence of the linear terms $\frac{1}{nT}\mathbf{v}'\Pi_2(\delta)$ and $\frac{1}{nT}\boldsymbol{\varepsilon}'\Pi_3(\delta)$ are proved by Chebyshev's inequality. Therefore, for $k = 1, 2, 3$, and all ℓ,

$$Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0, \quad \text{for each } \delta \in \Delta.$$

Now, all the $Q_{k,\ell}(\delta)$ terms are linear or quadratic in ρ, λ_1 and λ_2, and it is easy to show that $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\omega}Q_{k,\ell}(\delta)| = O_p(1)$, for $\omega = \rho, \lambda_1, \lambda_2$. Note that only matrix $\mathbf{M}^*$ involves λ_3 and ϕ. For $\omega = \phi, \lambda_3$, some algebra leads to the following simple expression $\frac{d}{d\omega}\mathbf{M}^* = -\mathbf{M}^*\dot{\Omega}_\omega\mathbf{M}^*$, where $\dot{\Omega}_{\lambda_3} = \frac{\partial}{\partial\lambda_3}\Omega = I_T \otimes (B_3'B_3)^{-1}(B_3'W_3 + W_3'B_3)(B_3'B_3)^{-1}$ and $\dot{\Omega}_\phi = \frac{\partial}{\partial\phi}\Omega = J_T \otimes I_n$. Thus, by applying Lemmas A.1, A.4 and A.6, repeatedly, it is easy to show that, for $k = 1, 2, 3$, and all ℓ, $\sup_{\delta\in\boldsymbol{\Delta}} |\frac{\partial}{\partial\lambda_3}Q_{k,\ell}(\delta)| = O_p(1)$. It follows that $Q_{k,\ell}(\delta)$ are stochastically equicontinuous. The pointwise convergence and stochastic equicontinuity therefore lead to,

$$Q_{k,\ell}(\delta) - \mathrm{E}Q_{k,\ell}(\delta) \xrightarrow{p} 0, \quad \text{uniformly in } \delta \in \boldsymbol{\Delta},$$

by Theorem 2.1 of Newey (1991).

It left to show $\mathrm{E}Q_4(\delta) = \frac{1}{nT}\mathrm{E}[(\mathbf{B}_1^*Y^\circ - \mathbf{B}_2^*Y_{-1}^\circ)'\mathbf{P}(\mathbf{B}_1^*Y^\circ - \mathbf{B}_2^*Y_{-1}^\circ)] \to 0$, uniformly in $\delta \in \boldsymbol{\Delta}$. By Assumption D, $\gamma_{\min}(\frac{\mathbf{X}'\mathbf{X}}{nT}) > \underline{c}_x$. By Assumption E,

$$\sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\min}(\Omega) \geq \inf_{\delta\in\boldsymbol{\Delta}} \gamma_{\min}(\Omega) \geq \inf_{\lambda_3\in\Lambda_3} \gamma_{\max}^{-1}(B_3'B_3) \geq \sup_{\lambda_3\in\Lambda_3} \gamma_{\max}^{-1}(B_3'B_3) \geq \frac{1}{\bar{c}_3}.$$

Hence, $\sup_{\delta\in\boldsymbol{\Delta}} \gamma_{\min}(\frac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{nT}) \geq \sup_{\phi\in\Phi} \frac{\underline{c}_3}{\phi\bar{c}_3+1}\underline{c}_x \geq c \geq 0$. Therefore, we have by the assumptions in Theorem 2.1 and Assumption D,

$$\begin{aligned}\mathrm{E}Q_4 &= \tfrac{1}{nT}\mathrm{tr}[\Omega^{-1}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathrm{Var}(\mathbf{B}_1Y - \mathbf{B}_2Y_{-1})] \\ &\leq \tfrac{1}{nT}\gamma_{\min}^{-2}(\Omega)\gamma_{\min}^{-1}\big(\tfrac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{nT}\big)\bar{c}_y\ \tfrac{1}{nT}\mathrm{tr}[\mathbf{X}'\mathbf{X}] = O(n^{-1}),\end{aligned}$$

Hence, $\hat{\sigma}_v^2(\delta) - \bar{\sigma}_v^2(\delta) \xrightarrow{p} 0$, uniformly in $\delta \in \boldsymbol{\Delta}$, completing the proof of **(b)**.

Proofs of (c)–(g). Using the expressions (B.3) and (B.4) and the representation (2.16), all the quantities inside $|\cdot|$ in **(c)–(g)** can all be expressed in the forms similar to (B.5). Thus, the proofs of **(c)–(g)** follow the proof of **(b)**. See the `Supplementary Appendix` for more details on the proof of Theorem 2.1. ■

Proof of Theorem 2.2. We have by the mean value theorem (henceforth MVT),

$$0 = \frac{1}{\sqrt{nT}}S_{\text{SDPD}}^*(\hat{\psi}_{\text{SDPD}}) = \frac{1}{\sqrt{nT}}S_{\text{SDPD}}^*(\psi_0) + \Big[\frac{1}{nT}\frac{\partial}{\partial\psi'}S_{\text{SDPD}}^*(\bar{\psi})\Big]\sqrt{nT}(\hat{\psi}_{\text{M}} - \psi_0),$$

where $\bar{\psi}$ lies elementwise between $\hat{\psi}_{\text{M}}$ and ψ_0. The result of the theorem follows if

(a) $\frac{1}{\sqrt{nT}}S_{\text{SDPD}}^*(\psi_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty}\Gamma_{\text{SDPD}}^*(\psi_0)\big]$,

(b) $\frac{1}{nT}\big[\frac{\partial}{\partial\psi'}S_{\text{SDPD}}^*(\bar{\psi}) - \frac{\partial}{\partial\psi'}S_{\text{SDPD}}^*(\psi_0)\big] \xrightarrow{p} 0$, and

(c) $\frac{1}{nT}\big[\frac{\partial}{\partial\psi'}S_{\text{SDPD}}^*(\psi_0) - \mathrm{E}\big(\frac{\partial}{\partial\psi'}S_{\text{SDPD}}^*(\psi_0)\big)\big] \xrightarrow{p} 0$.

Proof of (a). By $\mathbf{e} = \boldsymbol{\varepsilon} + \mathbf{B}_{30}^{-1}\mathbf{v}$ and letting $\Pi_{r0}^{\circ} = \mathbf{B}_{30}^{\prime-1}\Pi_{r0}, r = 1, \ldots, 4$, $\Psi_{r0}^{\circ} = \mathbf{B}_{30}^{\prime-1}\Psi_{r0}, r = 1, 2, 3$, and $\Phi_{r0}^{\circ} = \mathbf{B}_{30}^{\prime-1}\Phi_{r0}\mathbf{B}_{30}^{-1}, r = 1, \ldots, 6$, and $\Phi_{r0}^{\circ} = \mathbf{B}_{30}^{\prime-1}\Phi_{r0}$ and $\Phi_{r0}^{\diamond} = \mathbf{B}_{30}^{\prime-1}\Phi_{r0}\mathbf{B}_{30}^{-1}, r = 1, \ldots, 6$, dropping subscript "0" to simplify notations, the AQS functions given by (2.15) can be further expressed as follows,

$$S_{\mathrm{SDPD}}^{*}(\psi_0) = \begin{cases} \boldsymbol{\Pi}_1'\varepsilon + \Pi_1^{\circ\prime}\mathbf{v}, \\ \boldsymbol{\varepsilon}'\Phi_1\boldsymbol{\varepsilon} + \mathbf{v}'\Phi_1^{\diamond}\mathbf{v} + 2\mathbf{v}'\Phi_1^{\circ}\boldsymbol{\varepsilon} - \mu_{\sigma_v^2}, \\ \boldsymbol{\varepsilon}'\Phi_2\boldsymbol{\varepsilon} + \mathbf{v}'\Phi_2^{\diamond}\mathbf{v} + 2\mathbf{v}'\Phi_2^{\circ}\boldsymbol{\varepsilon} - \mu_{\phi}, \\ \boldsymbol{\varepsilon}'\Psi_1\mathbf{y}_0 + \mathbf{v}'\Psi_1^{\circ}\mathbf{y}_0 + \Pi_2'\boldsymbol{\varepsilon} + \Pi_2^{\circ\prime}\mathbf{v} + \boldsymbol{\varepsilon}'\Phi_3\boldsymbol{\varepsilon} + \mathbf{v}'\Phi_3^{\diamond}\mathbf{v} + 2\mathbf{v}'\Phi_3^{\circ}\boldsymbol{\varepsilon} - \mu_{\rho}, \\ \boldsymbol{\varepsilon}'\Psi_2\mathbf{y}_0 + \mathbf{v}'\Psi_2^{\circ}\mathbf{y}_0 + \Pi_3'\boldsymbol{\varepsilon} + \Pi_3^{\circ\prime}\mathbf{v} + \boldsymbol{\varepsilon}'\Phi_4\boldsymbol{\varepsilon} + \mathbf{v}'\Phi_4^{\diamond}\mathbf{v} + 2\mathbf{v}'\Phi_4^{\circ}\boldsymbol{\varepsilon} - \mu_{\lambda_1}, \\ \boldsymbol{\varepsilon}'\Psi_3\mathbf{y}_0 + \mathbf{v}'\Psi_3^{\circ}\mathbf{y}_0 + \Pi_4'\boldsymbol{\varepsilon} + \Pi_4^{\circ\prime}\mathbf{v} + \boldsymbol{\varepsilon}'\Phi_5\boldsymbol{\varepsilon} + \mathbf{v}'\Phi_5^{\diamond}\mathbf{v} + 2\mathbf{v}'\Phi_5^{\circ}\boldsymbol{\varepsilon} - \mu_{\lambda_{12}}, \\ \boldsymbol{\varepsilon}'\Phi_6\boldsymbol{\varepsilon} + \mathbf{v}'\Phi_6^{\diamond}\mathbf{v} + 2\mathbf{v}'\Phi_6^{\circ}\boldsymbol{\varepsilon} - \mu_{\lambda_3}, \end{cases} \tag{B.6}$$

where $\mu_{\sigma_v^2} = \frac{nT}{2\sigma_{v0}^2}$, $\mu_{\phi} = \frac{1}{2}\mathrm{tr}[\Omega_0^{-1}(J_T \otimes I_n)]$, $\mu_{\rho} = \mathrm{tr}[(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10})\Omega_0^{-1}]$, $\mu_{\lambda_1} = \mathrm{tr}[(\phi_0\mathbf{C}_0 + \mathbf{D}_0)\Omega_0^{-1}\mathbf{W}_1]$, $\mu_{\lambda_2} = \mathrm{tr}[(\phi_0\mathbf{C}_{-10} + \mathbf{D}_{-10})\Omega_0^{-1}\mathbf{W}_2]$, and $\mu_{\lambda_3} = \mathrm{tr}(\Omega_0^{-1}\dot{\Omega}_{\lambda_3 0})$.

Partition the vectors or matrices Π_r and Π_r° according to $t = 1, \ldots, T$, and denote the partitioned vectors or matrices, respectively, by $\{\Pi_{rt}\}$ and $\{\Pi_{rt}^{\circ}\}$; partition the matrices Φ_r, Φ_r°, $\Phi_r^{\diamond}$, Ψ_r, and Ψ_r° according to $t, s = 1, \ldots, T$, and denote the partitioned matrices, respectively, by $\{\Phi_{rts}\}$, $\{\Phi_{rts}^{\circ}\}$, $\{\Phi_{rts}^{\diamond}\}$, $\{\Psi_{rts}\}$, and $\{\Phi_{rts}^{\circ}\}$. As $\boldsymbol{\varepsilon} = 1_T \otimes \varepsilon$ and $\mathbf{y}_0 = 1_T \otimes y_0$, denoting $\Pi_{r+} = \sum_{t=1}^{T}\Pi_{rt}$, $\Phi_{rt+}^{\circ} = \sum_{s=1}^{T}\Phi_{rts}^{\circ}$, $\Phi_{r++} = \sum_{s=1}^{T}\sum_{s=1}^{T}\Phi_{rts}$, we have

$$\Pi_r'\boldsymbol{\varepsilon} = \Pi_{r+}\varepsilon, \quad \boldsymbol{\varepsilon}'\Phi\boldsymbol{\varepsilon} = \varepsilon'\Phi_{r++}\varepsilon, \quad \boldsymbol{\varepsilon}'\Psi\mathbf{y}_0 = \varepsilon'\Psi_{r++}y_0, \quad \mathbf{v}'\Psi_r^{\circ}\mathbf{y}_0 = \mathbf{v}'\Psi_{r+}^{\circ}y_0, \quad \mathbf{v}'\Phi_r^{\circ}\boldsymbol{\varepsilon} = \mathbf{v}'\Phi_{r+}^{\circ}\varepsilon.$$

where $\Psi_{r+}^{\circ} = \Psi_r^{\circ}(1_T \otimes I_n)$ and $\Phi_{r+}^{\circ} = \Phi_r^{\circ}(1_T \otimes I_n)$. Now, by (3.2), the terms bilinear in ε and y_0, and the terms bilinear in $\mathbf{v}$ and y_0 can be expressed as

$$\varepsilon'\Psi_{r++}y_0 = \varepsilon'\Psi_{r++}K_m\varepsilon + \varepsilon'\Psi_{r++}K_m(\eta_m^* + V_m^*), \quad \text{and} \quad \mathbf{v}'\Psi_{r+}^{\circ}y_0 = \mathbf{v}'\Psi_{r+}^{\circ}K_m\varepsilon + \mathbf{v}'\Psi_{r+}^{\circ}K_m(\eta_m^* + V_m^*).$$

Therefore, the AQS vector at the true parameters consists of terms linear-quadratic in $\mathbf{v}$, linear-quadratic in ε, and bilinear in ε and $\mathbf{v}$. Thus, for every non-zero $\dim(\psi) \times 1$ vector of constants c, $c'S_{\mathrm{SDPD}}^{*}(\psi_0)$ can be expressed as

$$c'S_{\mathrm{SDPD}}^{*}(\psi_0) = \mathbf{v}'A\mathbf{v} + \mathbf{v}'\zeta + \varepsilon'B\varepsilon + \varepsilon'\varphi + \mathbf{v}D\varepsilon - c'\mu_{\psi},$$

for suitably defined non-stochastic matrices A, B and D, and (random) vectors ζ and φ, where $\mu_{\psi} = \{0_{\dim(\boldsymbol{\beta})}', \mu_{\sigma_v^2}, \mu_{\phi}, \mu_{\rho}, \mu_{\lambda_1}, \mu_{\lambda_2}, \mu_{\lambda_3}\}'$. Both ζ and φ are measurable functions of V_m, and hence are independent of ε and $\mathbf{v}$. Putting $c'S_{\mathrm{SDPD}}^{*}(\psi_0)$ in a more compact form: $\mathbb{V}'\mathbb{A}\mathbb{V} + \mathbb{V}'\varpi - c'\mu_{\psi}$, where $\mathbb{V} = (\mathbf{v}', \varepsilon')'$, $\mathbb{A} = \{A, D; \mathbf{0}, B\}$, $\varpi = (\zeta', \varphi)'$, and $\mathbf{0}$ denotes a matrix of zeros, the asymptotic normality of $\frac{1}{\sqrt{nT}}c'S_{\mathrm{SDPD}}^{*}(\psi_0)$ follows from Lemma A.5. Finally, the Cramér–Wold devise leads to the joint asymptotic normality of $\frac{1}{\sqrt{nT}}S_{\mathrm{SDPD}}^{*}(\psi_0)$.

Proof of (b). The Hessian matrix, $H_{\mathrm{SDPD}}^{*}(\psi) = \frac{\partial}{\partial\psi'}S_{\mathrm{SDPD}}^{*}(\psi)$, has the distinct elements:

$H_{\beta\beta}^{*} = -\frac{1}{\sigma_v^2}X'\Omega^{-1}X$, $\quad H_{\beta\sigma_v^2}^{*} = -\frac{1}{\sigma_v^4}X'\Omega^{-1}\mathbf{e}(\theta)$, $\quad H_{\beta\phi}^{*} = \frac{1}{\sigma_v^2}X'\dot{\Omega}_{\phi}^{-}\mathbf{e}(\theta)$, $\quad H_{\beta\rho}^{*} = -\frac{1}{\sigma_v^2}X'\Omega^{-1}Y_{-1}$,

$H_{\beta\lambda_1}^{*} = -\frac{1}{\sigma_v^2}X'\Omega^{-1}W_1Y$, $\quad H_{\beta\lambda_2}^{*} = -\frac{1}{\sigma_v^2}X'\Omega^{-1}W_2Y_{-1}$ $\quad H_{\sigma_v^2\sigma_v^2}^{*} = -\frac{1}{\sigma_v^6}\mathbf{e}'(\theta)\Omega^{-1}\mathbf{e}(\theta) + \frac{nT}{2\sigma_v^4}$, $\quad H_{\beta\lambda_3}^{*} = \frac{1}{\sigma_v^2}X'\dot{\Omega}_{\lambda_3}^{-}\mathbf{e}(\theta)$,

$H_{\sigma_v^2\phi}^{*} = \frac{1}{2\sigma_v^4}\mathbf{e}'(\theta)\dot{\Omega}_{\phi}^{-}\mathbf{e}(\theta)$, $\quad H_{\sigma_v^2\rho}^{*} = -\frac{1}{\sigma_v^4}\mathbf{e}'(\theta)\Omega^{-1}Y_{-1}$, $\quad H_{\sigma_v^2\lambda_2}^{*} = -\frac{1}{\sigma_v^4}\mathbf{e}'(\theta)\Omega^{-1}W_2Y_{-1}$, $\quad H_{\sigma_v^2\lambda_1}^{*} = -\frac{1}{\sigma_v^4}\mathbf{e}'(\theta)\Omega^{-1}W_1Y$,

$H_{\sigma_v^2\lambda_3}^{*} = \frac{1}{2\sigma_v^4}\mathbf{e}'(\theta)\dot{\Omega}_{\lambda_3}^{-}\mathbf{e}(\theta)$, $\quad H_{\phi\rho}^{*} = \frac{1}{\sigma_v^2}\mathbf{e}'(\theta)\dot{\Omega}_{\phi}^{-}Y_{-1}$, $\quad H_{\phi\lambda_1}^{*} = \frac{1}{\sigma_v^2}\mathbf{e}'(\theta)\dot{\Omega}_{\phi}^{-}W_1Y$, $\quad H_{\phi\lambda_2}^{*} = \frac{1}{\sigma_v^2}\mathbf{e}'(\theta)\dot{\Omega}_{\phi}^{-}W_2Y_{-1}$,

$H_{\phi\phi}^{*} = -\frac{1}{2\sigma_v^2}\mathbf{e}'(\theta)\ddot{\Omega}_{\phi}^{-}\mathbf{e}(\theta) - \frac{1}{2}\mathrm{tr}[\dot{\Omega}_{\phi}^{-}(J_T \otimes I_n)]$, $\quad H_{\phi\lambda_3}^{*} = -\frac{1}{2\sigma_v^2}\mathbf{e}'(\theta)\ddot{\Omega}_{\phi,\lambda_3}^{-}\mathbf{e}(\theta) - \frac{1}{2}\mathrm{tr}[\dot{\Omega}_{\lambda_3}^{-}(J_T \otimes I_n)]$,

$H_{\rho\rho}^{*} = -\frac{1}{\sigma_v^2}Y_{-1}'\Omega^{-1}Y_{-1} - \mathrm{tr}[(\phi\dot{\mathbf{C}}_{-1,\rho} + \dot{\mathbf{D}}_{-1,\rho})\Omega^{-1}]$, $\quad H_{\rho\lambda_1}^{*} = -\frac{1}{\sigma_v^2}Y'W_1'\Omega^{-1}Y_{-1} - \mathrm{tr}[(\phi\dot{\mathbf{C}}_{-1,\lambda_1} + \dot{\mathbf{D}}_{-1,\lambda_1})\Omega^{-1}]$,

$H_{\rho\lambda_2}^{*} = -\frac{1}{\sigma_v^2}Y_{-1}'W_2'\Omega^{-1}Y_{-1} - \mathrm{tr}[(\phi\dot{\mathbf{C}}_{-1,\lambda_2} + \dot{\mathbf{D}}_{-1,\lambda_2})\Omega^{-1}]$, $\quad H_{\lambda_1\lambda_2}^{*} = -\frac{1}{\sigma_v^2}Y_{-1}'W_2'\Omega^{-1}W_1Y - \mathrm{tr}[(\phi\dot{\mathbf{C}}_{\lambda_2} + \dot{\mathbf{D}}_{\lambda_2})\Omega^{-1}W_2]$,

$H_{\rho\lambda_3}^{*} = \frac{1}{\sigma_v^2}\mathbf{e}'(\theta)\dot{\Omega}_{\lambda_3}^{-}Y_{-1} - \mathrm{tr}[(\dot{D}_{-1,\lambda_3}\Omega^{-1}) + (\phi\mathbf{C}_1 + \mathbf{D}_1)\dot{\Omega}_{\lambda_3}^{-}]$, $\quad H_{\lambda_1\rho}^{*} = -\frac{1}{\sigma_v^2}Y'W_1'\Omega^{-1}Y_{-1} - \mathrm{tr}[(\phi\dot{\mathbf{C}}_{\rho} + \dot{\mathbf{D}}_{\rho})\Omega^{-1}W_1]$,

$H_{\lambda_1\lambda_1}^{*} = -\frac{1}{\sigma_v^2}Y'W_1'\Omega^{-1}W_1Y - \mathrm{tr}[(\phi\dot{\mathbf{C}}_{\lambda_1} + \dot{\mathbf{D}}_{\lambda_1})\Omega^{-1}W_1]$, $\quad H_{\lambda_1\lambda_3}^{*} = \frac{1}{\sigma_v^2}\mathbf{e}'(\theta)\dot{\Omega}_{\lambda_3}^{-}W_1Y - \mathrm{tr}\{[\dot{D}_{\lambda_3}\Omega^{-1} + (\phi\mathbf{C} + \mathbf{D})\dot{\Omega}_{\lambda_3}^{-}]W_1\}$,

$$
\begin{aligned}
H^*_{\lambda_2\rho} &= -\tfrac{1}{\sigma_v^2} Y'_{-1} W'_2 \Omega^{-1} Y_{-1} - \mathrm{tr}[(\phi \dot{\mathbf{C}}_{-1,\rho} + \dot{\mathbf{D}}_{-1,\rho})\Omega^{-1} W_2],\\
H^*_{\lambda_2\lambda_1} &= -\tfrac{1}{\sigma_v^2} Y'_{-1} W'_2 \Omega^{-1} W_1 Y - \mathrm{tr}[(\phi \dot{\mathbf{C}}_{-1,\lambda_1} + \dot{\mathbf{D}}_{-1,\lambda_1})\Omega^{-1} W_2],\\
H^*_{\lambda_2\lambda_2} &= -\tfrac{1}{\sigma_v^2} Y'_{-1} W'_2 \Omega^{-1} W_2 Y_1 - \mathrm{tr}[(\phi \dot{\mathbf{C}}_{-1,\lambda_2} + \dot{\mathbf{D}}_{-1,\lambda_2})\Omega^{-1} W_2],\\
H^*_{\lambda_2\lambda_3} &= \tfrac{1}{\sigma_v^2} \mathbf{e}'(\theta) \dot{\Omega}^-_{\lambda_3} W_2 Y_{-1} - \mathrm{tr}\{[\dot{D}_{-1\lambda_3}\Omega^{-1} + (\phi \mathbf{C}_1 + \mathbf{D}_1)\dot{\Omega}^-_{\lambda_3}] W_2\},\\
H^*_{\lambda_3\lambda_3} &= -\tfrac{1}{2\sigma_v^2} \mathbf{e}'(\theta) \ddot{\Omega}^-_{\lambda_3} \mathbf{e}(\theta) - \tfrac{1}{2}\mathrm{tr}(\dot{\Omega}^-_{\lambda_3}\dot{\Omega}_{\lambda_3} + \Omega^{-1}\ddot{\Omega}_{\lambda_3}),\\
H^*_{\lambda_3\rho} &= \tfrac{1}{\sigma_v^2} \mathbf{e}'(\theta)\dot{\Omega}^-_{\lambda_3} Y_{-1}, \quad H^*_{\lambda_3\lambda_1} = \tfrac{1}{\sigma_v^2} \mathbf{e}'(\theta)\dot{\Omega}^-_{\lambda_3} W_1 Y, \quad H^*_{\lambda_3\lambda_2} = \tfrac{1}{\sigma_v^2} \mathbf{e}'(\theta)\dot{\Omega}^-_{\lambda_3} W_2 Y_{-1},
\end{aligned}
$$

where $\dot{\mathbf{C}}_\omega = \frac{\partial \mathbf{C}}{\partial \omega}$, $\dot{\mathbf{D}}_\omega = \frac{\partial \mathbf{D}}{\partial \omega}$, $\dot{\mathbf{C}}_{-1,\omega} = \frac{\partial \mathbf{C}_{-1}}{\partial \omega}$, $\dot{\mathbf{D}}_{-1,\omega} = \frac{\partial \mathbf{D}_{-1}}{\partial \omega}$, for $\omega = \rho, \lambda_1, \lambda_2, \lambda_3$, and these expressions can easily be obtained from the expressions of $\mathbf{C}$, $\mathbf{C}_{-1}$, $\mathbf{D}$, and $\mathbf{D}_{-1}$ given in Lemma 2.1; and further,

$$
\begin{aligned}
\dot{\Omega}_{\lambda_3} &= \tfrac{\partial \Omega_{\lambda_3}}{\partial \lambda_3} = (\mathbf{B}'_3\mathbf{B}_3)^{-1}(\mathbf{B}'_3\mathbf{W}_3 + \mathbf{W}'_3\mathbf{B}_3)(\mathbf{B}'_3\mathbf{B}_3)^{-1},\\
\ddot{\Omega}_{\lambda_3} &= \tfrac{\partial \dot{\Omega}_{\lambda_3}}{\partial \lambda_3} = 2[\dot{\Omega}_{\lambda_3}(\mathbf{B}'_3\mathbf{W}_3 + \mathbf{W}'_3\mathbf{B}_3)(\mathbf{B}'_3\mathbf{B}_3)^{-1} - (\mathbf{B}'_3\mathbf{B}_3)^{-1}(\mathbf{W}'_3\mathbf{W}_3)(\mathbf{B}'_3\mathbf{B}_3)^{-1}],\\
\dot{\Omega}^-_{\lambda_3} &= \tfrac{\partial \Omega^{-1}}{\partial \lambda_3} = -\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}, \qquad \ddot{\Omega}^-_{\lambda_3} = \tfrac{\partial \dot{\Omega}^-_{\lambda_3}}{\partial \lambda_3} = -2\Omega^{-1}\dot{\Omega}_{\lambda_3}\dot{\Omega}^-_{\lambda_3} - \Omega^{-1}\ddot{\Omega}_{\lambda_3}\Omega^{-1},\\
\dot{\Omega}^-_{\phi} &= \tfrac{\partial \Omega^{-1}}{\partial \phi} = \Omega^{-1}(J_T \otimes I_n)\Omega^{-1}, \qquad \ddot{\Omega}^-_{\phi} = \tfrac{\partial \dot{\Omega}^{-1}_{\phi}}{\partial \phi} = 2\Omega^{-1}(J_T \otimes I_n)\Omega^{-1}(J_T \otimes I_n)\Omega^{-1},\\
\ddot{\Omega}^-_{\phi,\lambda_3} &= \tfrac{\partial \dot{\Omega}^-_{\phi}}{\partial \lambda_3} = 2\Omega^{-1}\dot{\Omega}_{\lambda_3}\Omega^{-1}(J_T \otimes I_n)\Omega^{-1}.
\end{aligned}
$$

It is easy to show that $\frac{1}{nT}H^*_{\mathrm{SDPD}}(\psi_0) = O_p(1)$ by Lemma A.1 and the model assumptions. Thus, $\frac{1}{nT}H^*_{\mathrm{SDPD}}(\bar{\psi}) = O_p(1)$ because $\bar{\psi} - \psi_0 = o_p(1)$, which is implied by $\hat{\psi}_{\mathrm{M}} \xrightarrow{p} \psi_0$. As $\bar{\sigma}_v^2 \xrightarrow{p} \sigma_{v0}^2$, $\bar{\sigma}_v^{-r} = \sigma_{v0}^{-r} + o_p(1)$, $r = 2, 4, 6$. As σ_v^r appears in $H^*_{\mathrm{SDPD}}(\psi)$ multiplicatively,

$$
\tfrac{1}{n(T-1)} H^*_{\mathrm{SDPD}}(\bar{\psi}) = \tfrac{1}{n(T-1)} H^*_{\mathrm{SDPD}}(\bar{\boldsymbol{\beta}}, \sigma_{v0}^2, \bar{\phi}, \bar{\rho}, \bar{\lambda}) + o_p(1).
$$

The proof of (**b**) is thus equivalent to the proof of

$$
\tfrac{1}{n(T-1)}\left[H^*_{\mathrm{SDPD}}(\bar{\boldsymbol{\beta}}, \sigma_{v0}^2, \bar{\phi}, \bar{\rho}, \bar{\lambda}) - H^*_{\mathrm{SDPD}}(\psi_0)\right] \xrightarrow{p} 0.
$$

Writing $\mathbf{e}(\theta) = \mathbf{e} - (\lambda_1 - \lambda_{10})\mathbf{W}_1 Y - (\rho - \rho_0)Y_{-1} - (\lambda_2 - \lambda_{20})\mathbf{W}_2 Y_{-1} - \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, and by the representations for Y and Y_{-1} given in (2.14), we see that all the random elements of $H^*_{\mathrm{SDPD}}(\psi)$ can be written as linear combinations of terms:

$$
\begin{aligned}
&\textit{quadratic in } \mathbf{e}: && (\varpi - \varpi_0)^j(\omega - \omega_0)^k \mathbf{e}'\mathbb{A}\mathbf{G}(\phi, \lambda_3)\mathbb{B}\mathbf{e},\\
&\textit{quadratic in } \mathbf{y}_0: && (\varpi - \varpi_0)^j(\omega - \omega_0)^k \mathbf{y}'_0\mathbb{A}\mathbf{G}(\phi, \lambda_3)\mathbb{B}\mathbf{y}_0,\\
&\textit{linear in } \mathbf{e}: && (\varpi - \varpi_0)^j \mathbf{e}'\mathbb{A}\mathbf{G}(\phi, \lambda_3)\mathbb{B}\mathbb{Z},\\
&\textit{linear in } \mathbf{y}_0: && (\varpi - \varpi_0)^j \mathbf{y}'_0\mathbb{A}\mathbf{G}(\phi, \lambda_3)\mathbb{B}\mathbb{Z},\\
&\textit{bylinear in } \mathbf{e} \textit{ and } \mathbf{y}_0: && (\varpi - \varpi_0)^j(\omega - \omega_0)^k \mathbf{e}'\mathbb{A}\mathbf{G}(\phi, \lambda_3)\mathbb{B}\mathbf{y}_0,
\end{aligned}
$$

for $j, k = 0, 1$, $\varpi, \omega = \rho, \lambda_1, \lambda_2$, where $\mathbb{A}$ and $\mathbb{B}$ denote generically $nT \times nT$ non-stochastic matrices, and $\mathbb{Z}$ generically $nT \times d$ non-stochastic vector or matrices, all from (2.14) and free from parameters; and $\mathbf{G}(\phi, \lambda_3)$ can be Ω^{-1}, $\dot{\Omega}^-_{\lambda_3}$, $\ddot{\Omega}^-_{\lambda_3}$, $\dot{\Omega}^-_{\phi}$, $\ddot{\Omega}^-_{\phi}$, and $\ddot{\Omega}^-_{\phi,\lambda_3}$.

Take a typical quadratic term of $\mathbf{e}$, $\mathbf{e}'\mathbb{A}\mathbf{G}(\phi, \lambda_3)\mathbb{B}\mathbf{e}$, for example. Letting (ϕ^*, λ_3^*) be between $(\bar{\phi}, \bar{\lambda}_3)$ and (ϕ_0, λ_{30}), we have by MVT,

$$
\frac{1}{nT}[\mathbf{e}'\mathbb{A}\mathbf{G}(\bar{\lambda}_3, \bar{\phi})\mathbb{B}\mathbf{e} - \mathbf{e}'\mathbb{A}\mathbf{G}(\lambda_{30}, \phi_0)\mathbb{B}\mathbf{e}] = \frac{\bar{\phi} - \phi_0}{nT}\mathbf{e}'\mathbb{A}\dot{\mathbf{G}}_{\phi^*}\mathbb{B}\mathbf{e} + \frac{\bar{\lambda}_3 - \lambda_0}{nT}\mathbf{e}'\mathbb{A}\dot{\mathbf{G}}_{\lambda_3^*}\mathbb{B}\mathbf{e},
$$

where $\dot{\mathbf{G}}_{\phi}$ and $\dot{\mathbf{G}}_{\lambda_3}$ are the partial derivatives of $\mathbf{G}(\phi, \lambda_3)$ evaluated at (ϕ^*, λ_3^*). Noting that $\mathbf{G}$ is a linear combination of the matrices Ω^{-1}, $\mathbf{B}_3^{-1}$ and $\mathbf{W}_3$, and their products, its partial derivatives evaluated at (ϕ, λ_3) are linear combinations of Ω^{-1}, $\mathbf{B}_3^{-1}$ and $\mathbf{W}_3$, and their products as well, and hence are uniformly bounded in both row and column sums for (ϕ, λ_3) in a neighborhood of (ϕ_0, λ_{30}). By Lemma A.4(i) and the consistency of $\hat{\psi}_{\mathrm{M}}$, $\frac{1}{nT}[\mathbf{e}'\mathbb{A}\mathbf{G}(\bar{\phi}, \bar{\lambda}_3)\mathbb{B}\mathbf{e} - \mathbf{e}'\mathbb{A}\mathbf{G}(\phi_0, \lambda_{30})\mathbb{B}\mathbf{e}] \xrightarrow{p} 0$. The convergence of all other terms can be shown similarly by using Lemma A.4, Assumption F, and the consistency of $\hat{\psi}_{\mathrm{M}}$.

It is left to show that all the 'trace' terms in $\frac{1}{nT}\left[H^*_{\mathrm{SDPD}}(\bar{\boldsymbol{\beta}}, \sigma_{v0}^2, \bar{\phi}, \bar{\rho}, \bar{\lambda}) - H^*_{\mathrm{SDPD}}(\psi_0)\right]$ are $o_p(1)$. Consider, for example, its $\rho\rho$-element. Denote $\mathbf{E}_{-1} \equiv \mathbf{E}_{-1}(\phi, \rho, \lambda) = \phi\mathbf{C}_{-1} + \mathbf{D}_{-1}$ and let $\dot{\mathbf{E}}_{-1,\rho}(\phi, \rho, \lambda)$ be its partial derivative w.r.t. ρ. For

$(\phi^*, \rho^*, \lambda^*)$ be between $(\bar\phi, \bar\rho, \bar\lambda)$ and $(\phi_0, \rho_0, \lambda_0)$, we have by MVT,

$$
\begin{aligned}
&\tfrac{1}{nT}\{\mathrm{tr}[\dot{\mathbf{E}}_{-1,\rho}(\bar\phi,\bar\rho,\bar\lambda)\Omega^{-1}(\bar\phi,\bar\lambda_3)] - \mathrm{tr}[\dot{\mathbf{E}}_{-1,\rho}(\phi_0,\rho_0,\lambda_0)\Omega^{-1}(\phi_0,\lambda_{30})]\}\\
&= \tfrac{\bar\phi-\phi_0}{nT}\mathrm{tr}[\phi^*\Omega^{-1}(\phi^*,\lambda_3^*) + \dot{\mathbf{E}}_{-1,\rho^*}\dot\Omega^{-1}_{\phi^*}] + \tfrac{\bar\rho-\rho_0}{nT}\mathrm{tr}[\ddot{\mathbf{E}}^{\rho^*}_{-1,\rho}\Omega^{-1}(\phi^*,\lambda_3^*)]\\
&\quad + \tfrac{\bar\lambda_1-\lambda_{10}}{nT}\mathrm{tr}[\ddot{\mathbf{E}}^{\lambda_1^*}_{-1,\rho}\Omega^{-1}(\phi^*,\lambda_3^*)] + \tfrac{\bar\lambda_2-\lambda_{20}}{nT}\mathrm{tr}[\ddot{\mathbf{E}}^{\lambda_2^*}_{-1,\rho}\Omega^{-1}(\phi^*,\lambda_3^*)]\\
&\quad + \tfrac{\bar\lambda_3-\lambda_{30}}{nT}\mathrm{tr}[\ddot{\mathbf{E}}^{\lambda_3^*}_{-1,\rho}\Omega^{-1}(\phi^*,\lambda_3^*) + \dot{\mathbf{E}}_{-1,\rho}\dot\Omega^{-1}(\lambda_3^*)],
\end{aligned}
$$

where $\ddot{\mathbf{E}}^{r^*}_{-1,\rho}$, $r = \phi, \rho, \lambda_1, \lambda_2, \lambda_3$, are the partial derivatives of $\dot{\mathbf{E}}_{-1,\rho}$ evaluated at $(\phi^*, \rho^*, \lambda^*)$. Consider W.L.O.G. $T = 2$. Recall the definitions of C and D, we have,

$$
\mathbf{D}(\rho,\lambda_1,\lambda_2,\lambda_3) = \begin{pmatrix} B_1^{-1}(B_3'B_3)^{-1}, & B_1^{-1}(B_3'B_3)^{-1}\\ \mathcal{B}B_1^{-1}(B_3'B_3)^{-1}, & \mathcal{B}B_1^{-1}(B_3'B_3)^{-1}\end{pmatrix},
$$

$$
\mathbf{C}(\rho,\lambda_1,\lambda_2) = \begin{pmatrix} (\sum_{i=0}^{m}\mathcal{B}^i)B_1^{-1}, & (\sum_{i=0}^{m}\mathcal{B}^i)B_1^{-1}\\ (\sum_{i=0}^{m+1}\mathcal{B}^i)B_1^{-1}, & (\sum_{i=0}^{m}\mathcal{B}^i)B_1^{-1}\end{pmatrix}.
$$

This shows that the elements of $\mathbf{E}_{-1}$ and $\mathbf{E}_{-1,\rho}$ are linear combinations of the matrices W_1, B_1^{-1}, B_2 and B_3^{-1}, and their products. Therefore, $\ddot{\mathbf{E}}^{r}_{-1,\rho}$ has elements being linear combinations of W_1, W_2, W_3, B_1^{-1}, B_2, and B_3^{-1}, and their products, and hence are uniformly bounded in both row and column sums for (ρ, λ) in a neighborhood of (ρ_0, λ_0) by Lemmas A.1 and A.2. Therefore, each trace term in the equation above divided by nT, such as $\frac{1}{nT}\mathrm{tr}[\phi^*\Omega^{-1}(\phi^*,\lambda_3^*) + \dot{\mathbf{E}}_{-1,\rho^*}\dot\Omega^{-1}_{\phi^*}]$, is $O_p(1)$. This completes the proof **(b)**.

Proof of (c). By the representations given in (2.16), the elements of Hessian matrix can be written as linear combinations of quadratic and linear terms of $\mathbf{v}$ and $\boldsymbol{\varepsilon}$, quadratic and linear terms of $\mathbf{y}_0$, bilinear terms of $\mathbf{v}$ and $\mathbf{y}_0$, $\boldsymbol{\varepsilon}$ and $\mathbf{y}_0$, $\mathbf{v}$ and $\boldsymbol{\varepsilon}$. Thus, the results follow by repeatedly applying Lemmas A.1, A.4, and A.6. ■

Appendix C. Proofs for Section 3

Proof of Lemma 3.1. The result (3.7) is obvious. **To show** (3.8), we need the result:

$$
\mathrm{E}[(a'v_t)(b'v_t)(c'v_t)(d'v_t)] = (\mu^{(4)}_{v0} - 3\sigma^4_{v0})(a\odot b)'(c\odot d) + \sigma^4_{v0}[(a'b)(c'd) + (a'c)(b'd) + (a'd)(b'c)],
$$

where $\odot$ denotes the Hadamard product, and a, b, c, and d are $n\times 1$ vectors. Write

$$
g_{\Phi_r i} = \sum_{t=1}^{T}(e_{it}e^*_{ri,t} - d_{1\Phi_r it}) + \sum_{t=1}^{T}(e_{it}\varphi_{ri,t} - d_{2\Phi_r it}) = Q_{r1,i} + Q_{r2,i},\ \ r = 1, 2, \ldots, 6.
$$

As $e^*_{ri,t} = \Phi_{rii,t+}\varepsilon_i + \sum_{s=1}^{T}\Phi_{rii,ts}b_i'v_s$ and $\varphi_{ri,t} = a'_{ri,t+}\varepsilon + \sum_{s=1}^{T}c'_{ri,ts}v_s$, we have

$$
\begin{aligned}
\mathrm{E}(Q_{r1,i}Q_{v1,j}) &= \sum_{t=1}^{T}\mathrm{Cov}(e_{it}e^*_{rit}, e_{jt}e^*_{vjt}) + \sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\mathrm{Cov}(e_{it}e^*_{rit}, e_{js}e^*_{vjs})\\
&= \sigma^4_{v0}\sum_{t=1}^{T}\sum_{s=1}^{T}\big[\Phi_{rii,ts}\Phi_{vjj,st}(b_i'b_j)^2 + \Phi_{rii,ts}\Phi_{vjj,ts}(b_i'b_j)^2\big] + (\mu^{(4)}_{v0} - 3\sigma^4_{v0})\sum_{t=1}^{T}\Phi_{rii,tt}\Phi_{vjj,tt}(b_i\odot b_i)'(b_j\odot b_j),\\
\mathrm{E}(Q_{r1i}Q_{v2j}) &= \sigma^4_{v0}\sum_{t=1}^{T}\sum_{s=1}^{T}[\Phi_{rii,ts}(b_i'c_{vj,st})(b_i'b_j) + \Phi_{rii,ts}(b_i'c_{vj,ts})(b_i'b_j)] + \sigma^2_{v0}\sigma^2_{\varepsilon 0}\sum_{t=1}^{T}(\Phi_{rii,t+} + \Phi_{rii,+t})(1_i'a_{vj,t+})(b_i'b_j)\\
&\quad + (\mu^{(4)}_{v0} - 3\sigma^4_{v0})\sum_{t=1}^{T}\Phi_{rii,tt}(b_i\odot b_i)'(b_j\odot c_{vj,tt}),\\
\mathrm{E}(Q_{r2i}Q_{v1j}) &= \sigma^4_{v0}\sum_{t=1}^{T}\sum_{s=1}^{T}[\Phi_{vjj,ts}(b_j'c_{ri,st})(b_i'b_j) + \Phi_{vjj,ts}(b_j'c_{ri,ts})(b_i'b_j)] + \sigma^2_{v0}\sigma^2_{\varepsilon 0}\sum_{t=1}^{T}(\Phi_{vjj,t+} + \Phi_{vjj,+t})(1_j'a_{ri,t+})(b_i'b_j)\\
&\quad + (\mu^{(4)}_{v0} - 3\sigma^4_{v0})\sum_{t=1}^{T}\Phi_{vjj,tt}(b_j\odot b_j)'(b_i\odot c_{ri,tt}),\\
\mathrm{E}(Q_{r2i}Q_{v2j}) &= \sigma^4_{v0}\sum_{t=1}^{T}\sum_{s=1}^{T}[(b_i'c_{vj,st})(b_j'c_{ri,ts}) + (b_i'b_j)(c'_{ri,ts}c_{vj,ts})]\\
&\quad + \sigma^2_{v0}\sigma^2_{\varepsilon 0}\sum_{t=1}^{T}[(1_j'a_{ri,t+})(b_i'c_{vj,+t}) + (1_i'a_{vj,t+})(b_j'c_{ri,+t}) + (a'_{ri,t+}a_{vj,t+})(b_i'b_j)]\\
&\quad + (\mu^{(4)}_{v0} - 3\sigma^4_{v0})\textstyle\sum_{t=1}^{T}(b_i\odot c_{ri,tt})'(b_j\odot c_{vj,tt}).
\end{aligned}
$$

Summarizing and simplifying by letting $c^*_{ri,ts}$ be the ith row of $(\Phi^l_{r,ts} + \Phi^{u\prime}_{r,st} + \Phi^d_{r,ts})B_3^{-1}$, we obtain the result for $\mathrm{E}(g_{\Phi_r i} g_{\Phi_v j})$, i.e., (3.8) in Lemma 3.1.

To show (3.9), write $g_{\Psi_r i} = \sum_{t=1}^T (e_{it}\Psi^*_{rii,t+} y^*_{0i} - d_{\Psi_r it}) + \sum_{t=1}^T e'_{it}\xi_{ri,t} = Q_{r1,i} + Q_{r2,i}$. Using $e_{it} = \varepsilon_i + b'_i v_t$, and $y^*_0 = \varepsilon + \eta^*_m + V^*_m$, we obtain

$$
\begin{aligned}
\mathrm{E}(Q_{r1,i}Q_{v1,j}) &= \sum_{t=1}^T \mathrm{Cov}(e'_{it}\Psi^*_{rii,t+}y^*_{0i}, e'_{jt}\Psi^*_{vjj,t+}y^*_{0j}) + \sum_{t=1}^T \sum_{s(\neq t)} \mathrm{Cov}(e'_{it}\Psi^*_{rii,t+}y^*_{0i}, e'_{js}\Psi^*_{vjj,s+}y^*_{0j}) \\
&= \sum_{t=1}^T \mathrm{E}\left[(\varepsilon_i + b'_i v_t)\Psi^*_{rii,t+}y^*_{0i}(\varepsilon_j + b'_j v_t)\Psi^*_{vjj,t+}y^*_{0j}\right] - d_{\Psi_r it} d_{\Psi_v jt} \\
&\quad + \sum_{t=1}^T \sum_{s(\neq t)} \mathrm{E}\left[(\varepsilon_i + b'_i v_t)\Psi^*_{rii,t+}y^*_{0i}(\varepsilon_j + b'_j v_s)\Psi^*_{vjj,s+}y^*_{0j}\right] - d_{\Psi_r it} d_{\Psi_v js} \\
&= \sigma^2_{v0}(b'_i b_j) \sum_{t=1}^T (\Psi^*_{rii,t+}\Psi^*_{vjj,t+})\mathrm{E}(y^*_{0i}y^*_{0j}),
\end{aligned}
$$

where the double summation part vanishes, because for $i \neq j$ and $t \neq s$, $e'_{it}\Psi^*_{rii,t+}y^*_{0i}$ and $e'_{js}\Psi^*_{vjj,s+}y^*_{0j}$ are conditionally independent given V_m as they are, respectively, measurable-$(\varepsilon_i, v_t, V_m)$ and measurable-$(\varepsilon_j, v_s, V_m)$. Similarly, using $\xi_{ri,t} = w'_{ri,t}y^*_0$, we show that

$$
\begin{aligned}
\mathrm{E}(Q_{r1,i}Q_{v2,j}) &= \sigma^2_{v0}(b'_i b_j) \sum_{t=1}^T \Psi^*_{rii,t+}\mathrm{E}(y^*_{0i}\xi_{vj,t}), \\
\mathrm{E}(Q_{r2,i}Q_{v1,j}) &= \sigma^2_{v0}(b'_i b_j) \sum_{t=1}^T \Psi^*_{vjj,t+}\mathrm{E}(y^*_{0j}\xi_{ri,t}), \\
\mathrm{E}(Q_{r2,i}Q_{v2,j}) &= \sigma^2_{v0}(b'_i b_j) \sum_{t=1}^T \mathrm{E}(\xi_{ri,t}\xi_{vj,t}) + \sigma^4_{\varepsilon 0}(1'_j w_{ri,+})(1'_i w_{vj,+}),
\end{aligned}
$$

where 1_i denotes an $n \times 1$ vector of element 1 at the ith position and zero elsewhere. Summarizing and simplifying, we have the result for $\mathrm{E}(g_{\Psi_r i}\, g_{\Psi_v j})$ given in (3.9).

To show (3.11), write $g_{\Pi_v i} = \sum_{t=1}^T \Pi'_{vj,t}e_{jt} = P_{vi}$, $v = 1, 2, \ldots, k_\varpi$. Using this and $g_{\Psi_r i} = Q_{r1i} + Q_{r2i}$, $r = 1, 2, 3$, given above, we obtain

$$
\mathrm{E}(Q_{r1,i}P_{vj}) = \sigma^2_{v0} \sum_{t=1}^T (b'_i b_j)\Pi'_{vj,t}\Psi^*_{rii,t+}\mathrm{E}(y^*_{0i}) \text{ and } \mathrm{E}(Q_{r2,i}P_{vj}) = \sigma^2_{v0} \sum_{t=1}^T (b'_i b_j)\Pi'_{vj,t}\mathrm{E}(\xi_{ri,t}),
$$

leading to $\mathrm{E}(g_{\Psi_r i} g'_{\Pi_v j}) = \sigma^2_{v0} \sum_{t=1}^T \Pi'_{vj,t}\mathrm{E}(\xi^*_{ri,t})(b'_i b_j)$.

Result (3.10) and (3.12) are derived in a similar way in which we separate each $g_{\Phi_r i}$ and $g_{\Psi_r i}$ into two terms, and calculate the covariances of each pair of the terms and then sum them up. Details on these can be found in the `Supplementary Appendix`. ■

Proof of Theorem 3.1. First, the result $\Sigma^*_{\mathrm{SDPD}}(\hat\psi_{\mathrm{M}}) - \Sigma^*_{\mathrm{SDPD}}(\psi_0) \xrightarrow{p} 0$ is implied by the result **(b)** in the proof of Theorem 2.2. The result $\widehat{\Gamma}^*_{\mathrm{SDPD}} - \Gamma^*_{\mathrm{SDPD}}(\psi_0) \xrightarrow{p} 0$ follows from

(a) $\frac{1}{nT}\sum_{i=1}^n \left[\hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i - \mathrm{E}(\mathbf{g}_i\mathbf{g}'_i)\right] \xrightarrow{p} 0$,

(b) $\frac{1}{nT}\sum_{1=1}^n \sum_{j=1, j\neq i}^n \left[\widehat{\mathrm{E}}(\mathbf{g}_i\mathbf{g}'_j) - \mathrm{E}(\mathbf{g}_i\mathbf{g}'_j)\right] \xrightarrow{p} 0$.

To show (a): the result follows if $(i)\frac{1}{nT}\sum_{i=1}^n (\hat{\mathbf{g}}_i\hat{\mathbf{g}}'_i - \mathbf{g}_i\mathbf{g}'_i) \xrightarrow{p} 0$, and $(ii)\frac{1}{nT}\sum_{i=1}^n [\mathbf{g}_i\mathbf{g}'_i - \mathrm{E}(\mathbf{g}_i\mathbf{g}'_i)] \xrightarrow{p} 0$. The proof of (i) is straightforward by MVT. We focus on the proof of (ii).

The components of $S^*_{\mathrm{SDPD}}(\psi_0)$ are mixtures of terms of the forms: $\Pi'\mathbf{e} = \sum_{i=1}^n g_{\Pi i}$, $\mathbf{e}'\Psi\mathbf{y}_0 - \mathrm{E}(\mathbf{e}'\Psi\mathbf{y}_0) = \sum_{i=1}^n g_{\Psi i}$, and $\mathbf{e}'\Phi\mathbf{e} - \mathrm{E}(\mathbf{e}'\Phi\mathbf{e}) = \sum_{i=1}^n g_{\Phi i}$. It suffices to show that

$$
\frac{1}{nT}\sum_{i=1}^n [g_{ki}g'_{ri} - \mathrm{E}(g_{ki}g'_{ri})] = o_p(1), \quad \text{for } g_{ki}, g_{ri} = g_{\Pi i}, g_{\Psi i}, g_{\Phi i}. \tag{C.1}
$$

First, we show $\frac{1}{nT}\sum_{i=1}^{n}[g_{\Pi i}^2 - \mathrm{E}(g_{\Pi i}^2)] \xrightarrow{p} 0$. Assuming, W.L.O.G, Π_{it} are scalars, write

$$g_{\Pi i} = \sum_{t=1}^{T} \Pi_{it}\mathbf{e}_{it} = \sum_{t=1}^{T} \Pi_{it}(\varepsilon_i + b_i' v_t) = \Pi_{i+}\varepsilon_i + b_i'\mathbf{v}_i, \tag{C.2}$$

where $\Pi_{i+} = \sum_{t=1}^{T}\Pi_{it}$ and $\mathbf{v}_i = \sum_{t=1}^{T}\Pi_{it}v_t$. We have $\frac{1}{nT}\sum_{i=1}^{n}\big[g_{\Pi i}^2 - \mathrm{E}(g_{\Pi i}^2)\big] \equiv U_1 + U_2 + U_3$, where $U_1 = \frac{1}{nT}\sum_{i=1}^{n}\Pi_{i+}^2(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)$, $U_2 = \frac{2}{nT}\sum_{i=1}^{n}(\Pi_{i+}\varepsilon_i)(b_i'\mathbf{v}_i)$ and $U_3 = \frac{1}{nT}\sum_{i=1}^{n}\big[(b_i'\mathbf{v}_i)^2 - \sigma_{v0}^2(\sum_{t=1}^{T}\Pi_{it}^2)(b_i'b_i)\big]$. Now, it is straightforward to show that $U_r = o_p(1)$, for $r = 1, 2, 3$, by applying Lemmas A.1 and A.4, and Chebyshev's inequality.

Second, we show $\frac{1}{nT}\sum_{i=1}^{n}[g_{\Phi i}^2 - \mathrm{E}(g_{\Phi i}^2)] \xrightarrow{p} 0$. Using (3.4), we can write

$$g_{\Phi i} = k_i(\varepsilon_i^2 - \sigma_\varepsilon^2) + \varepsilon_i z_{1i} + \varepsilon_i(r_i'\varepsilon) + (u_i - \mu_{u_i}) + \sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t), \tag{C.3}$$

where k_i are scalar constants that are uniformly bounded; $z_{1i} = \sum_{t=1}^{T}p_{it}'v_t$ with p_{it}' being the ith row of some non-stochastic matrix uniformly bounded in row and column sums; $u_i = \sum_{t=1}^{T}\sum_{s=1}^{T}v_t'A_{i,ts}v_s$ with mean $\mu_{u_i} = \sigma_v^2\sum_{t=1}^{T}\mathtt{tr}(A_{i,tt})$, where $A_{i,ts} = \Phi_{ii,ts}(b_ib_i') + (b_ic_{i,ts}')$; b_i are defined as before, and r_i' and q_{it}' represent ith row of some non-stochastic strictly lower triangular matrices which are uniformly bounded in both row and column sums. Noticing that the five terms in (C.3) are uncorrelated, it follows that

$$\frac{1}{nT}\sum_{i=1}^{n}\big[g_{\Phi i}^2 - \mathrm{E}(g_{\Phi i}^2)\big] = \sum_{r=1}^{15}U_r, \tag{C.4}$$

where $U_1 = \frac{1}{nT}\sum_{i=1}^{n}k_i^2\{(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)^2 - \mathrm{E}[(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)^2]\}$, $\quad U_2 = \frac{2}{nT}\sum_{i=1}^{n}k_i(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)\varepsilon_i(r_i'\varepsilon)$,

$$U_3 = \tfrac{1}{nT}\sum_{i=1}^{n}[\varepsilon_i^2(r_i'\varepsilon)^2 - \sigma_{\varepsilon 0}^4\sum_{j=1}^{n}r_{ij}^2], \qquad U_4 = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i(r_i'\varepsilon)\sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t),$$

$$U_5 = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i^2(r_i'\varepsilon)z_{1i}, \qquad U_6 = \tfrac{2}{nT}\sum_{i=1}^{n}k_i(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)\varepsilon_i z_{1i},$$

$$U_7 = \tfrac{2}{nT}\sum_{i=1}^{n}k_i(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)\sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t), \quad U_8 = \tfrac{2}{nT}\sum_{i=1}^{n}[\varepsilon_i z_{1i}\sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t)],$$

$$U_9 = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i(r_i'\varepsilon)(u_i - \mu_{u_i}), \qquad U_{10} = \tfrac{2}{nT}\sum_{i=1}^{n}k_i(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)(u_i - \mu_{u_i}),$$

$$U_{11} = \tfrac{1}{nT}\sum_{i=1}^{n}[\varepsilon_i^2 z_{1i}^2 - (\sum_{t=1}^{T}p_{it}'p_{it})\sigma_{v0}^2\sigma_{\varepsilon 0}^2], \quad U_{13} = \tfrac{1}{nT}\sum_{i=1}^{n}\varepsilon_i z_{1i}(u_i - \mu_{u_i}),$$

$$U_{12} = \frac{1}{nT}\sum_{i=1}^{n}\{[\sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t)]^2 - \sigma_{v0}^2\sigma_{\varepsilon 0}^2(\sum_{t=1}^{T}q_{it}'q_{it})(b_i'b_i)\},$$

$$U_{14} = \tfrac{2}{nT}\sum_{i=1}^{n}(u_i - \mu_{u_i})\sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t), \qquad U_{15} = \tfrac{1}{nT}\sum_{i=1}^{n}\{(u_i - \mu_{u_i})^2 - \mathrm{E}[(u_i - \mu_{u_i})^2]\}.$$

To show each of the fifteen terms above is $o_p(1)$, we write it as the sum of a martingale differences (MD) array and thus the weak law of large numbers (WLLN) for an MD array, e.g., Theorem 19.7 of Davidson (1994, p.299), can be applied to prove its convergence in probability to zero. As the full proof is tedious, we present details for a few typical terms: U_2, U_4, U_{10} and U_{15}. More details are put in the `Supplementary Appendix`.

Write $U_2 = \frac{2}{nT}\sum_{i=1}^{n}k_i(\varepsilon_i^3 - \mu_{\varepsilon 0}^{(3)})(r_i'\varepsilon) + \frac{2}{nT}\sum_{i=1}^{n}k_i\mu_{\varepsilon 0}^{(3)}(r_i'\varepsilon) - \frac{2}{nT}\sigma_{\varepsilon 0}^2\sum_{i=1}^{n}k_i\varepsilon_i(r_i'\varepsilon) \equiv \frac{2}{nT}\sum_{r=1}^{3}\sum_{i=1}^{n}V_{rn,i}$. Let $\mathcal{F}_{ni}^{\varepsilon}$ be the increasing σ-field generated by $(\varepsilon_1, \ldots, \varepsilon_i)$. As $(r_i'\varepsilon)$ is $\mathcal{F}_{n,i-1}^{\varepsilon}$ measurable, we have for $r = 1, 3$, $\mathrm{E}(V_{rn,i}|\mathcal{F}_{n,i-1}^{\varepsilon}) = 0$, and thus $\{V_{rn,i}, \mathcal{F}_{n,i}^{\varepsilon}\}$ forms an MD array. As k_i are uniformly bounded, it is easy to see $\{V_{1n,i}\}$ and $\{V_{3n,i}\}$ are uniformly integrable. With constant coefficients $\frac{1}{nT}$, the other two conditions of WLLN for MD array of Davidson are satisfied. So $\frac{1}{nT}\sum_{i=1}^{n}V_{rn,i} = o_p(1)$, $r = 1, 3$, by Davidson's WLLN for MD arrays. Finally, $\frac{1}{nT}\sum_{i=1}^{n}V_{2n,i} = \sum_{i=1}^{n}a_{ni}\varepsilon_i$, for some constants a_{ni}. Therefore, $U_2 \xrightarrow{p} 0$.

Write $U_4 = \frac{2}{nT}\sum_{i=1}^{n}\varepsilon_i(r_i'\varepsilon)\sum_{t=1}^{T}(q_{it}'\varepsilon)(b_i'v_t) \equiv \sum_{i=1}^{n}V_{ni}$. Let $\mathcal{G}_{ni}$ be the increasing σ-field generated by $(\mathbf{v}, \varepsilon_1, \ldots, \varepsilon_i)$. We have $\mathrm{E}(V_{ni}|\mathcal{G}_{n,i-1}) = 0$, and thus $\{V_{ni}, \mathcal{G}_{ni}\}$ form an MD array. By Assumption B and Lemma A.1 we have $\mathrm{E}(V_{ni}^2) = \sum_{t=1}^{T}(b_i'b_i)\sigma_{\varepsilon 0}^2\sigma_{v0}^2\{(\mu_{\varepsilon 0}^{(4)} - 3\sigma_{\varepsilon 0}^4)(r_i \odot r_i)'(q_{it} \odot q_{it}) + \sigma_{\varepsilon 0}^4[(r_i'r_i)(q_{it}'q_{it}) + 2(r_i'q_{it})^2]\} \le K \le \infty$. Therefore, $\{V_{ni}\}$ is uniformly integrable. The other two conditions of the WLLN for MD arrays of Davidson are satisfied with constant coefficients $\frac{1}{nT}$. So we have $U_4 \xrightarrow{p} 0$.

Write $U_{10} = \frac{2}{nT}\sum_{i=1}^n k_i(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)u_i - \frac{2}{nT}\sum_{i=1}^n k_i(\varepsilon_i^2 - \sigma_{\varepsilon 0}^2)\mu_{u_i} \equiv \frac{2}{nT}\sum_{r=1}^2\sum_{i=1}^n V_{rn,i}$. As k_i and μ_{u_i} are uniformly bounded, we immediately have $\frac{2}{nT}\sum_{i=1}^n V_{2n,i} \xrightarrow{p} 0$ by Kolmogorov's law of large numbers (LLN). For $V_{1n,i}$, first we notice that u_i depends only on $\mathbf{v}$, and thus is independent of ε_i for all i. So, $\{V_{1n,i}, \mathcal{G}_{n,i}\}$ form an MD array. We have

$$\mathrm{E}(u_i^2) = \sigma_{v0}^4 \sum_{t=1}^{T}\sum_{s\neq t}\left[\mathrm{tr}(A_{i,tt})\mathrm{tr}(A_{i,ss}) + \mathrm{tr}(A_{i,ts}A'_{i,ts}) + \mathrm{tr}(A_{i,ts}A_{i,st})\right] + (\mu_{v_0}^{(4)} - 3\sigma_{v0}^4)\sum_{t=1}^{T}\sum_{j=1}^{n} a_{itt,jj}^2.$$

where $a_{itt,jj}$ denotes the (j,j) element of $A_{i,tt}$. As $A_{i,ts} = \Phi_{ii,ts}(b_i b_i') + (b_i c'_{i,ts})$, we have $\mathrm{tr}(A_{i,ts}) = c*\prime_{i,ts} b_i$, which is the (i,i) element of $\Phi_{ts}^*(B_3'B_3)^{-1}$. So $\mathrm{tr}(A_{i,ts}) = O(h_n^{-1})$ by Lemma A.1 and Assumption D. Similarly we have $\sum_{j=1}^n a_{itt,jj}^2 \le \mathrm{tr}(A_{i,ts}A'_{i,ts}) = O(h_n^{-1})$ and $\mathrm{tr}(A_{i,ts}A_{i,st}) = O(h_n^{-1})$. Therefore, the condition, $\mathrm{E}(|V_{1n,i}|^{1+\epsilon}) < K_v < \infty$ for some $\epsilon > 0$, is satisfied. With constant coefficients $\frac{1}{nT}$, the other two conditions of WLLN for MD array of Davidson are satisfied. So we have $\frac{2}{nT}\sum_{i=1}^n V_{1n,i} \xrightarrow{p} 0$ and thus, $U_{10} \xrightarrow{p} 0$.

Write $U_{15} = \frac{1}{nT}\sum_{i=1}^n[u_i^2 - \mathrm{E}(u_i^2)] - \frac{1}{nT}\sum_{i=1}^n \mu_{u_i}(u_i - \mu_{u_i})$. The convergence of the second term follows from Lemma A.4. Now, write $u_i^2 = (\sum_{t=1}^T\sum_{s=1}^T v_t'A_{i,ts}v_s)^2 = \sum_{r=1}^4 H_{r,ni}$, where

$$H_{1,ni} = \sum_t\sum_s\sum_k\sum_{\ell\neq t\neq s\neq k} v_t'A_{i,ts}v_s v_k'A_{i,k\ell}v_\ell; \quad H_{3,ni} = \sum_t\sum_{s\neq t} v_t'A_{i,ts}v_s v_t'A_{i,ts}v_s;$$
$$H_{2,ni} = \sum_t\sum_{s\neq t} v_t'A_{i,tt}v_t v_s'A_{i,ss}v_s; \quad H_{4,ni} = \sum_t v_t'A_{i,tt}v_t v_t'A_{i,tt}v_t.$$

Write $H_{1,ni} = \sum_\ell v_\ell'\varphi_{i\ell}$, where $\varphi_{i\ell} = \sum_{t\neq\ell}\sum_{s\neq\ell}\sum_{k\neq\ell} A'_{i,k\ell}v_k v_t'A_{i,ts}v_s$. We have $\mathrm{E}(v_\ell'\varphi_{i\ell}) = 0$ as v_ℓ and $\varphi_{i\ell}$ are independent. For each ℓ, we can write $\frac{1}{n}\sum_{i=1}^n v_\ell'\varphi_{i\ell} = \frac{1}{n}v_\ell'\sum_{i=1}^n \varphi_{i\ell}$, which is a bilinear form. Therefore, by Assumptions B and D, it is easy to verify the conditions of Lemma A.5. As T is fixed, we have $\frac{1}{nT}\sum_{i=1}^n H_{1,ni} = o_p(1)$.

Rewrite $H_{2,ni} = \sum_t\sum_{s\neq t} u_{it}u_{is}$. For each t and s, write $\frac{1}{nT}\sum_{i=1}^n[u_{it}u_{is} - \mathrm{E}(u_{it})\mathrm{E}(u_{is})] = \frac{1}{nT}\sum_{i=1}^n[u_{it} - \mathrm{E}(u_{it})]\mathrm{E}(u_{is}) + \frac{1}{nT}\sum_{i=1}^n[u_{is} - \mathrm{E}(u_{is})]u_{it} \equiv \frac{1}{nT}\sum_{i=1}^n V_{1n,i} + \frac{1}{nT}\sum_{i=1}^n V_{2n,i}$. Let $v_t^* = [\sum_{i=1}^n A_{i,tt}^d \mathrm{tr}(A_{i,ss})]v_t$ and $\xi_t = [\sum_{i=1}^n(A_{i,tt}^l + A_{i,tt}^{u\prime})]v_t$. The $V_{1n,i}$ is decomposed into:

$$\begin{aligned}
\frac{1}{nT}\sum_{i=1}^n V_{1n,i} &= \frac{1}{nT}\sum_{i=1}^n\left[\left(v_t'A_{i,tt}^d v_t - \sigma_{v0}^2\mathrm{tr}(A_{i,tt})\right) + v_t'(A_{i,tt}^l + A_{i,tt}^{u\prime})v_t\right]\mathrm{E}(u_{is}) \\
&= \frac{\sigma_{v0}^2}{nT}v_t'\Big[\sum_{i=1}^n A_{i,tt}^d\mathrm{tr}(A_{i,ss})\Big]v_t - \frac{\sigma_{v0}^4}{nT}\sum_{i=1}^n\mathrm{tr}(A_{i,tt})\mathrm{tr}(A_{i,ss}) + \frac{1}{nT}v_t'\Big[\sum_{i=1}^n(A_{i,tt}^l + A_{i,tt}^{u\prime})\Big]v_t \\
&= \frac{\sigma_{v0}^2}{nT}[v_t'v_t^* - \mathrm{E}(v_t'v_t^*)] + \frac{1}{nT}v_t'\xi_t = \frac{\sigma_{v0}^2}{nT}\sum_{j=1}^n\left(v_{jt}v_{jt}^* - \mathrm{E}(v_{jt}v_{jt}^*)\right) + \frac{1}{nT}\sum_{j=1}^n v_{jt}\xi_{jt}.
\end{aligned}$$

Clearly, the first term is the average of n independent terms. The second term can be seen to be the average of the MD array $\{v_{jt}\xi_{jt}\}$ with respect to the increasing σ-field, $\mathcal{F}_{nj}^v$, generated by $\{v_{1t}, \ldots, v_{jt}, t = 1, \ldots, T\}$. The ξ_{jt} is $\mathcal{F}_{n,j-1}^v$-measurable and the conditions of WLLN of Davidson are easily verified. Hence, $\frac{1}{nT}\sum_{i=1}^n V_{1n,i} = o_p(1)$. Similarly but more tediously, we show that $\frac{1}{nT}\sum_{i=1}^n V_{2n,i} = o_p(1)$. Therefore, $\frac{1}{nT}\sum_{i=1}^n(H_{2,ni} - \mathrm{E}H_{2,ni}) = o_p(1)$.

The result $\frac{1}{nT}\sum_{i=1}^n[H_{3,ni} - \mathrm{E}(H_{3,ni})] \xrightarrow{p} 0$ can be shown in a similar way.

As $H_{4,ni} = \sum_t(v_t'v_{it}^* + v_t'\xi_{it})^2$, where $v_{it}^* = A_{i,tt}^d v_t$ and $\xi_{it} = (A_{i,tt}^l + A_{i,tt}^{u\prime})v_t$, we have, for each t, $\frac{1}{nT}\sum_{i=1}^n(v_t'v_{it}^* + v_t'\xi_{it})^2 = \frac{1}{nT}\sum_{i=1}^n(v_t'v_{it}^*)^2 + \frac{1}{nT}\sum_{i=1}^n(v_t'\xi_{it})^2 + \frac{2}{nT}\sum_{i=1}^n v_t'v_{it}^*v_t'\xi_{it} \equiv \frac{1}{nT}\sum_{r=1}^3\sum_{i=1}^n V_{rn,it}$. By Assumptions B and E, and Lemma A.1, it is easy to show that $\frac{1}{nT}\sum_{i=1}^n[V_{1n,it} - \mathrm{E}(V_{1n,it})] = \frac{1}{nT}\sum_{j=1}^n(v_{jt}^4 - \mu_v^{(4)})a_{jj} + \frac{1}{nT}\sum_{j=1}^n\sum_{k\neq j}(v_{jt}^2v_{kt}^2 - \sigma_{v0}^4)a_{kj} = o_p(1)$. Similarly, $\frac{1}{nT}\sum_{i=1}^n[V_{3n,it} - \mathrm{E}(V_{3n,it})] = o_p(1)$. Decompose $\frac{1}{nT}\sum_{i=1}^n[V_{2n,it} - \mathrm{E}(V_{2n,it})]$ as:

$$\frac{1}{nT}\sum_{j=1}^n(v_{jt}^2 - \sigma_{v0}^2)\Big(\sum_{i=1}^n\xi_{it,j}^2\Big) + \frac{1}{nT}\sum_{j=1}^n\Big[\Big(\sum_{i=1}^n\xi_{it,j}^2\Big) - \mathrm{E}\Big(\sum_{i=1}^n\xi_{it,j}^2\Big)\Big] + \frac{\sigma_{v0}^2}{nT}\sum_{j=1}^n v_{jt}\Big[\sum_{k\neq j}v_{kt}\Big(\sum_{i=1}^n\xi_{it,j}\xi_{it,k}\Big)\Big].$$

The first and third terms can be shown to be $o_p(1)$ by WLLN for MD arrays as $\xi_{it,j}$ is $\mathcal{F}_{n,j-1}^v$-measurable. Let $a'_{i,j}$ be the j th row of $A_{i,tt}^l + A_{i,tt}^{u\prime}$, then we have $\xi_{it,j} = a'_{i,j}v_t$. The second term becomes $\frac{1}{nT}\sum_{j=1}^n\left[v_t'(\sum_{i=1}^n a_{i,j}a'_{i,j})v_t - \sigma_{v0}^2\mathrm{tr}(\sum_{i=1}^n a_{i,j}a'_{i,j})\right] = o_p(1)$ by Lemmas A.1, A.2 and A.4. Therefore, $\frac{1}{nT}\sum_{i=1}^n H_{4,ni} = o_p(1)$. Combining these results, we have $U_{15} = o_p(1)$. Other terms can be proved similarly, and therefore, $\frac{1}{nT}\sum_{i=1}^n[g_{\Phi i}^2 - \mathrm{E}(g_{\Phi i}^2)] = o_p(1)$.

Third, we show $\frac{1}{nT}\sum_{i=1}^{n}[g_{\Psi i}^2 - \mathrm{E}(g_{\Psi i}^2)] \xrightarrow{p} 0$. Write, using $y_0^* = \eta_m^* + \varepsilon + V_m^*$ and (3.3),

$$g_{\Psi i} = \varepsilon_i h_i + \Psi_{ii+}^*(\varepsilon_i^2 - \sigma_\varepsilon^2) + \varepsilon_i(w_{i+}'\varepsilon) + z_{2i} + \sum_{t=1}^{T}(b_i'v_t)(w_{it}'\varepsilon), \tag{C.5}$$

where $h_i = a_i'V_m^* + \sum_{t=1}^{T} c_{it}'v_t$, $z_{2i} = \sum_{t=1}^{T} s_{it}'v_t$, and a_i', s_i', and c_{it}' are ith row of some non-stochastic matrices that are uniformly bounded in both row and column sums. Recall w_{it}' is the ith row of $(\Psi_{t+}^{*l} + \Psi_{t+}^{*u})$ and note that the ith element of w_{it}' is 0. We have,

$$\frac{1}{nT}\sum_{i=1}^{n}[g_{\Psi i}^2 - \mathrm{E}(g_{\Psi i}^2)] = \sum_{r=1}^{15} U_r, \quad \text{where} \tag{C.6}$$

$$U_1 = \tfrac{1}{nT}\sum_{i=1}^{n}[\varepsilon_i^2 h_i^2 - \sigma_{\varepsilon 0}^2\mathrm{E}(h_i^2)], \qquad U_2 = \tfrac{1}{nT}\sum_{i=1}^{n}\Psi_{ii+}^{*2}\{(\varepsilon_i^2-\sigma_\varepsilon^2)^2 - \mathrm{E}[(\varepsilon_i^2-\sigma_\varepsilon^2)^2]\},$$

$$U_3 = \tfrac{1}{nT}\sum_{i=1}^{n}[\varepsilon_i^2(w_{i+}'\varepsilon)^2 - \sigma_{\varepsilon 0}^4\sum_{j=1}^{n} w_{ij}^2], \quad U_4 = \tfrac{1}{nT}\sum_{i=1}^{n}[z_{2i}^2 - \mathrm{E}(z_{2i}^2)],$$

$$U_6 = \tfrac{2}{nT}\sum_{i=1}^{n}\Psi_{ii+}^*(\varepsilon_i^2-\sigma_\varepsilon^2)\varepsilon_i h_i, \qquad U_7 = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i^2(w_{i+}'\varepsilon)h_i,$$

$$U_8 = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i h_i z_{2i}, \qquad U_9 = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i h_i\sum_{t=1}^{T}(b_i'v_t)(w_{it}'\varepsilon),$$

$$U_{10} = \tfrac{2}{nT}\sum_{i=1}^{n}\Psi_{ii+}^*(\varepsilon_i^2-\sigma_\varepsilon^2)\varepsilon_i(w_{i+}'\varepsilon), \quad U_{11} = \tfrac{2}{nT}\sum_{i=1}^{n}\Psi_{ii+}^*(\varepsilon_i^2-\sigma_\varepsilon^2)\sum_{t=1}^{T}(b_i'v_t)(w_{it}'\varepsilon),$$

$$U_{12} = \tfrac{2}{nT}\sum_{i=1}^{n}\Psi_{ii+}^*(\varepsilon_i^2-\sigma_\varepsilon^2)z_{2i}, \qquad U_{13} = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i(w_{i+}'\varepsilon)\sum_{t=1}^{T}(b_i'v_t)(w_{it}'\varepsilon),$$

$$U_{14} = \tfrac{2}{nT}\sum_{i=1}^{n}\varepsilon_i(w_{i+}'\varepsilon)z_{2i}, \qquad U_{15} = \tfrac{2}{nT}\sum_{i=1}^{n}z_{2i}\sum_{t=1}^{T}(b_i'v_t)(w_{it}'\varepsilon).$$

$$U_5 = \frac{1}{nT}\sum_{i=1}^{n}[(\sum_{t=1}^{T}(b_i'v_t)(w_{it}'\varepsilon))^2 - \sigma_{v0}^2\sigma_{\varepsilon 0}^2(b_i'b_i)\sum_{t=1}^{T}(w_{it}'w_{it})].$$

By Assumptions A and B, V_m^* is independent of ε and v_t, and η_m^* is exogenous. With Assumption F, the terms in (C.6) are similar to those in (C.4), and therefore their convergence is proved similarly. Using (C.2), (C.3) and (C.5), the convergence of the cross product terms $\frac{1}{nT}\sum_{i=1}^{n}\big[g_{\Pi i}g_{\Phi i} - \mathrm{E}(g_{\Pi i}g_{\Phi i})\big] = o_p(1)$, $\frac{1}{nT}\sum_{i=1}^{n}\big[g_{\Phi i}g_{\Psi i} - \mathrm{E}(g_{\Psi i}g_{\Psi i})\big] = o_p(1)$, and $\frac{1}{nT}\sum_{i=1}^{n}\big[g_{\Pi i}g_{\Psi i} - \mathrm{E}(g_{\Psi i}g_{\Psi i})\big] = o_p(1)$ can also be shown similarly. Detailed expressions for the cross product terms can be found in the `Supplementary Appendix`. These complete the prove of convergence in the single summation part of Theorem 3.1.

To show (b): the result $\frac{1}{nT}\sum_{1=1}^{n}\sum_{j=1,j\neq i}^{n}\big[\widehat{\mathrm{E}}(\mathbf{g}_i\mathbf{g}_j') - \mathrm{E}(\mathbf{g}_i\mathbf{g}_j')\big] \xrightarrow{p} 0$ follows if

$$(\mathbf{i})\ \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(\widehat{\mathrm{E}}(\mathbf{g}_i\mathbf{g}_j') - \Upsilon_{ij}) \xrightarrow{p} 0, \text{ and } (\mathbf{ii})\ \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}[\Upsilon_{ij} - \mathrm{E}(\mathbf{g}_i\mathbf{g}_j')] \xrightarrow{p} 0,$$

where Υ_{ij} is from $\mathrm{E}(\mathbf{g}_i\mathbf{g}_j')$ by removing $\mathrm{E}(\cdot)$ in $\mathrm{E}(\xi_{ri,t}^*)$ and $\mathrm{E}(\xi_{ri,t}^*\xi_{vj,t}^*)$. As each element of Υ_{ij} is a linear combination of the terms in (3.7)–(3.12), only the consistency of them matters.

Proof of (i): (1). By Lemma 3.1 we have, $\mathrm{E}(g_{\pi_r}g_{\pi_v}') = \sigma_{v0}^2\sum_{t=1}^{T}(b_i'b_j)\boldsymbol{\pi}_{r,it}'\boldsymbol{\pi}_{v,jt}$. As T is fixed, and as σ_{v0}^2 enters linearly and $\hat\sigma_v^2$ is consistent, it suffices to prove

$$Q_0^t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat b_i'\hat b_j)\hat{\boldsymbol{\pi}}_{it}\hat{\boldsymbol{\pi}}_{jt} - (b_i'b_j)\boldsymbol{\pi}_{it}\boldsymbol{\pi}_{jt}] \xrightarrow{p} 0, \quad \text{for each } t = 1, \ldots, T,$$

which is done by applying Holder's inequality, Lemmas A.1 and A.2, and Assumption E.

(2). By the expression of $\mathrm{E}(g_{\Phi_r i} g_{\Phi_\nu j})$ given in (3.8) in Lemma 3.1, dropping r and ν,

$$
\begin{aligned}
Q_1^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_j'\hat{c}_{i,ts}^*)(\hat{b}_i'\hat{c}_{j,st}^*) - (b_j'c_{i,ts}^*)(b_i'c_{j,st}^*)] \xrightarrow{p} 0,\\
Q_2^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_i'\hat{b}_j)(\hat{c}_{i,ts}^{*\prime}\hat{c}_{j,ts}^*) - (b_i'b_j)(c_{i,ts}^{*\prime}c_{j,ts}^*)] \xrightarrow{p} 0,\\
Q_3^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[\hat{a}_{ji,t+}(\hat{b}_j'\hat{c}_{i,+t}^*) - a_{ji,t+}(b_j'c_{i,+t}^*)] \xrightarrow{p} 0,\\
Q_4^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[\hat{a}_{ij,t+}(b_i'\hat{c}_{j,+t}^*) - a_{ij,t+}(b_i'c_{j,+t}^*)] \xrightarrow{p} 0,\\
Q_5^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{a}_{i,t+}^{*\prime}\hat{a}_{j,t+}^*)(\hat{b}_i'\hat{b}_j) - (a_{i,t+}^{*\prime}a_{j,t+}^*)(b_i'b_j)] \xrightarrow{p} 0,\\
Q_6^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_i\odot\hat{c}_{i,tt}^*)'(\hat{b}_j\odot\hat{c}_{j,tt}^*) - (b_i\odot c_{i,tt}^*)'(b_j\odot c_{j,tt}^*)] \xrightarrow{p} 0.
\end{aligned}
$$

(3). $\mathrm{E}(g_{\Psi_r,i}g_{\Psi_\nu,j}) = \sigma_{\varepsilon 0}^4(w_{rij,+}w_{\nu ji,+}) + \sigma_{v0}^2\sum_{t=1}^{T}(b_i'b_j)\mathrm{E}(\xi_{ri,t}^*\xi_{\nu j,t}^*)$. We need to show:

$$
\begin{aligned}
Q_7^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}(\hat{w}_{ij,+}\hat{w}_{ji,+} - w_{ij,+}w_{ji,+}) \xrightarrow{p} 0, \quad \text{and}\\
Q_8^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_i'\hat{b}_j)(\hat{\xi}_{i,t}^*\hat{\xi}_{j,t}^*) - (b_i'b_j)(\xi_{i,t}^*\xi_{j,t}^*)] \xrightarrow{p} 0.
\end{aligned}
$$

(4). $\mathrm{E}(g_{\Phi i}g_{\Pi' j}) = \mu_{v0}^{(3)}\sum_{t=1}^{T}(b_i\odot c_{i,tt}^*)'b_j\boldsymbol{\pi}_{j,t}$ from Lemma 3.1, and thus we need to show:

$$
Q_9^t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_i\odot\hat{c}_{i,tt}^*)'\hat{b}_j\hat{\boldsymbol{\pi}}_{j,t} - (b_i\odot c_{i,tt}^*)'b_j\boldsymbol{\pi}_{j,t}] \xrightarrow{p} 0.
$$

(5). $\mathrm{E}(g_{\Psi i}g_{\Pi' j}) = \sigma_{v0}^2\sum_{t=1}^{T}\boldsymbol{\pi}_{j,t}\mathrm{E}(\xi_{ri,t}^*)(b_i'b_j)$ from Lemma 3.1, and thus we need to show:

$$
Q_{10}^t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{\boldsymbol{\pi}}_{j,t}\hat{\xi}_{i,t}^*)(\hat{b}_i'\hat{b}_j) - (\boldsymbol{\pi}_{j,t}\xi_{ri,t}^*)(b_i'b_j)] \xrightarrow{p} 0.
$$

(6). Finally by Lemma 3.1, $\mathrm{E}(g_{\Phi i}g_{\Psi j}) = \sigma_{\varepsilon 0}^2\sigma_{v0}^2\sum_{t=1}^{T}[(b_i'b_j)(a_{ri,t+}'w_{\nu j,t}^*) + w_{\nu ji,+}(b_j'c_{ri,+t}^\circ)] + \sigma_{\varepsilon 0}^4(w_{ji,+}a_{ij,++}) + \mu_{v0}^{(3)}\sum_{t=1}^{T}(b_i\odot c_{ri,tt}^*)'b_j\mathrm{E}(\xi_{\nu j,t}^*)$, and thus we need to show:

$$
\begin{aligned}
Q_{11}^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_i'\hat{b}_j)(\hat{a}_{i,t+}'\hat{w}_{j,t}^*) - (b_i'b_j)(a_{i,t+}'w_{j,t}^*)] \xrightarrow{p} 0,\\
Q_{12}^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[\hat{w}_{ji,+}(\hat{b}_j'\hat{c}_{i,+t}^\circ) - w_{ji,+}(b_j'c_{i,+t}^\circ)] \xrightarrow{p} 0,\\
Q_{13}^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{w}_{ji,+}\hat{a}_{ij,++}) - (w_{ji,+}a_{ij,++})] \xrightarrow{p} 0,\\
Q_{14}^t &= \tfrac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(\hat{b}_i\odot\hat{c}_{i,tt}^*)'\hat{b}_j(\hat{\xi}_{j,t}^*) - (b_i\odot c_{i,tt}^*)'b_j(\xi_{j,t}^*)] \xrightarrow{p} 0.
\end{aligned}
$$

Following (3.2) and Assumption F, we can see that all the terms in (3)-(6) are similar to the terms in (i) and (ii), and therefore their convergence in probability to zero is proved similarly to that of the terms in (i) and (ii).

Proof of (ii): First we note that (ii) is not needed for the terms not involving y_0^*. For the terms which involve y_0^*, we need to prove:

$$R_1^t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}(b_i'b_j)[\xi_{i,t}^*\xi_{j,t}^* - \mathrm{E}(\xi_{i,t}^*\xi_{j,t}^*)] \xrightarrow{p} 0,$$

$$R_2^t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}\boldsymbol{\pi}_{j,t}[\xi_{i,t}^* - \mathrm{E}(\xi_{i,t}^*)] \xrightarrow{p} 0,$$

$$R_3^t = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}[(b_i \odot c_{i,tt}^*)'b_j][\xi_{j,t}^* - \mathrm{E}(\xi_{j,t}^*)] \xrightarrow{p} 0.$$

Recall that $\xi_{i,t}^* = w_{it}^{*\prime}y_0^*$ where $w_{it}^{*\prime}$ is the ith row of Ψ_t^*. The convergence of R_2^t and R_3^t immediately follow by Lemmas A.1 and A.6. To show $R_1^t \xrightarrow{p} 0$, note that $\Psi_t^* = \Psi_t K_m$, and $y_0^* = K_m^{-1}y_0$, so we can write, $\xi_{i,t}^* = a_{it}'y_0$, where a_{it}' is the ith row of Ψ_t. Then we have, $\sum_{i=1}^n\sum_{j\neq i}(b_i'b_j)\xi_{i,t}^*\xi_{j,t}^* = y_0'[\sum_{i=1}^n(a_{it}b_i')\sum_{j\neq i}(b_ja_{jt}')]y_0 = y_0'A_ty_0$, where $A_t = \Psi_t'\mathbb{B}\Psi_t - \Psi_t'\mathtt{diag}(\mathbb{B})\Psi_t$, and $\mathbb{B} = (B_3'B_3)^{-1}$. Clearly, A_t is bounded in both row and column sums by Assumption E(iii) and Lemma A.1(i). Therefore, $R_1^t = \frac{1}{n}[y_0'A_ty_0 - \mathrm{E}(y_0'A_ty_0)] = o_p(1)$, by Lemma A.6. These complete the proof of the convergence of the double summation part in Theorem 3.1, and therefore complete the proof of Theorem 3.1. ■

Additional details on the proof of Theorem 3.1, in particular, the proof of **(b)**, can be found in the `Supplementary Appendix` available at http://www.mysmu.edu/faculty/zlyang/.

References

Anselin, L., Le Gallo, J., Jayet, J., 2008. Spatial panel econometrics. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. Springer-Verlag, Berlin Heidelberg, pp. 625–666.
Bai, J., Li, K., 2015. Dynamic spatial panel data models with common shocks. Working paper. Columbia University. New York.
Baltagi, B.H., 2013. Econometric Analysis of Panel Data, fifth ed. John Wiley & Sons Ltd..
Bernstein, D.S., 2009. Matrix Mathematics: Theory, Facts, and Formulas, second ed. Princeton University Press, Princeton.
Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, Cambridge.
Chamberlain, G., 1982. Multivariate regression models for panel data. J. Econometrics 18, 5–46.
Chamberlain, G., 1984. Panel data. In: Griliches, Z., Intrilligator, M. (Eds.), Handbook of Econometrics, Vol. 2. North-Holland, Amsterdam, pp. 1247–1318.
Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.
Debarsy, N., 2012. The mundlak approach in the spatial Durbin panel data model. Spatial Economic Analysis 7, 109–131.
Elhorst, J.P., 2010. Dynamic panels with endogenous interaction effects when T is small. Reg. Sci. Urban Econ. 40, 272–282.
Elhorst, J.P., 2012. Dynamic spatial panels: models, methods, and applications. J. Geogr. Syst. 14, 5–28.
Hausman, J.A., Taylor, W., 1981. Panel data and unobservable individual effects. Econometrica 46, 1377–1398.
Hsiao, C., 2014. Analysis of Panel Data, third ed. Cambridge University Press.
Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J. Econometrics 109, 107–150.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Internat. Econom. Rev. 40, 509–533.
Krishnakumar, J., 2006. Time invariant variables and panel data models: a generalised Frisch–Waugh Theorem and its implications. In: Baltagi, B. (Ed.), Panel Data Econometrics: Theoretical Contributions and Empirical Applications. Elsevier, Amsterdam, pp. 119–132.
Kuersteiner, G.M., Prucha, I.R., 2018. Dynamic panel data models: networks, common shocks, and sequential exogeneity. Working Paper, University of Maryland, College Park.
Lee, M.-J., 2002. Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables. Academic Press, San Diego.
Lee, L.-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. Econom. Theory 18, 252–277.
Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.
Lee, L.-F., Yu, J., 2010a. A spatial dynamic panel data model with both time and individual fixed effects. Econom. Theory 26, 564–597.
Lee, L.-F., Yu, J., 2010b. Some recent developments in spatial panel data models. Reg. Sci. Urban Econ. 40, 255–271.
Lee, L.-F., Yu, J., 2012. Spatial panels: random components vs. fixed effects. Int. Econ. Rev. 53, 1369–1412.
Lee, L.-F., Yu, J., 2014. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. J. Econometrics 180, 174–197.
Lee, L.-F., Yu, J., 2015. Spatial panel data models. In: Baltagi, B.H. (Ed.), The Oxford Handbook of Panel Data. Oxford University Press, Oxford, pp. 363–401.
Lee, L.-F., Yu, J., 2016. Identification of spatial Durbin panel models. J. Appl. Econometrics 31, 133–162.
Li, H., Zhou, L.A., 2005. Political turnover and economic performance: the incentive role of personnel control in China. J. Publ. Econ. 89, 1743–1762.
Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.
Mutl, J., 2006. Dynamic Panel Data Models with Spatially Correlated Disturbances (Ph.D. thesis). University of Maryland, College Park.
Newey, W.K., 1991. Uniform convergence in probability and stochastic equicontinuity. Econometrica 59, 1161–1167.
Qu, X., Wang, X., Lee, L.F., 2016. Instrumental variable estimation of a spatial dynamic panel model with endogenous spatial weights when T is small. Econom. J. 19, 261–290.
Shi, W., Lee, L.F., 2017. Spatial dynamic panel data models with interactive fixed effects. J. Econometrics 197, 323–347.
Shi, X., Xi, T., 2018. Race to safety: political competition, neighborhood effects, and coal mine deaths in china. J. Dev. Econ. 131, 79–95.
Su, L., Yang, Z.L., 2015. QML estimation of dynamic panel data models with spatial errors. J. Econometrics 185, 230–258.
van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Massachusetts.
Yang, Z.L., 2015. A general method for third-order bias and variance correction on a nonlinear estimator. J. Econometrics 186, 178–200.
Yang, Z.L., 2018. Unified M-estimation of fixed-effects spatial dynamic models with short panels. J. Econometrics 205, 423–447.
Yang, Z.L., Li, C., Tse, Y.K., 2006. Functional form and spatial dependence in dynamic panels. Econom. Lett. 91, 138–145.

Yao, Y., Zhang, M., 2015. Subnational leaders and economic growth: evidence from Chinese cities. J. Econ. Growth 20, 405–436.
Yu, J., de Jong, R., Lee, L.-F., 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both *n* and *T* are large. J. Econometrics 146, 118–134.
Yu, J., Lee, L.-F., 2010. Estimation of unit root spatial dynamic panel data models. Econom. Theory 26, 1332–1362.
Yu, J., Zhou, L.A., Zhu, G., 2016. Strategic interaction in political competition: Evidence from spatial effects across Chinese cities. Reg. Sci. Urban Econ. 57, 23–37.