Unified M-estimation of Matrix Exponential Dynamic Panel Specification *

Ye Yang†

Abstract

In this paper unified M-estimation method is proposed for the matrix exponential spatial dynamic panel specification (MESDPS) with fixed effect in short panels. The quasi-maximum likelihood (QML) estimation for dynamic panel data (DPD) model has long been known to have the initial condition specification difficulty, which leads to bias and inconsistency. The MESDPS also suffers from this problem. The initial-condition free M-estimator in this paper solves this problem and is proved to be consistent and asymptotic normal. An outer product of martingale difference (OPMD) estimator for the variance-covariance (VC) matrix of the M-estimator is also derived and proved to be consistent. MESDPS with a matrix exponential spatial specification (MESS) in the dependent variable, the lagged dependent variable and the disturbances are represented by MESDPS(1,1,1). Monte Carlo experiments results for finite sample properties of the M-estimator and the OPMD estimator of MESDPS(1,1,1) and various submodels are reported. The method is applied to US outward FDI data to show its validity.

Key Words: MESDPS, M-estimation, Martingale difference, OPMD, Initial-condition free

JEL Codes: C10, C13, C15, C21, C23.

*This research was supported, in part, by a grant of computer time from the City University of New York High Performance Computing Center under NSF Grants CNS-0958379, CNS-0855217 and ACI 1126113.

†Department of Economics, The Graduate School and University Center, City University of New York, 365 Fifth Avenue, New York, NY, 10016. email:yyang1@gradcenter.cuny.edu

1 Introduction

Panel data (PD) models and dynamic panel data (DPD) models are important elements in Economics literature. Adding spatial dependence into PD and DPD models, the resulting spatial panel data (SPD) models (Baltagi et al. (2003), Anselin (2013)) and spatial dynamic panel data (SDPD) models (Anselin (2001), Yu et al. (2008), Lee & Yu (2010*c*), Yu & Lee (2010), Elhorst (2010), Su & Yang (2015)) have gained much attention. Some papers (Lee & Yu (2010*b*), Lee & Yu (2013), Xu & Lee (2019)) provide excellent survey on these models.

For both SPD and SDPD models, two estimation methods—general method of moment (GMM) (Mutl (2006)) and quasi maximum likelihood (QML) (Elhorst (2010), Lee & Yu (2010*a*), Su & Yang (2015), Yang (2018))—are available. The QML estimators (QMLE) are more efficient than GMM estimators (GMME). However there is one difficulty with QML estimation for SDPD models with short panels: the "initial condition" problem. The first observation Δy_1 for the first differenced data is endogenous in models with fixed effects, no matter whether y_0 is endogenous or exogenous. To solve this problem, the traditional way is to use the predicted value obtained from the values of regressors (Elhorst (2010), Su & Yang (2015)). But this method has its disadvantages. First process starting time is unknown and the time-varying regressors need to be trend or first-difference stationary. Second, the method does not apply to SDPD models with only spatial lags (SL) because the initial difference contains spatial effect in the exogenous part when expanded using backward substitutions. To deal with these problems, Yang (2018) proposes an initial-condition free M-estimator. The estimator is derived from a set of estimating equations based on the unbiased adjusted quasi score (AQS) functions and is consistent and asymptotically normal. It also proposes an outer product of martingale difference (OPMD) estimator for the variance-covariance (VC) matrix of the M-estimator and proves that it is consistent.

On the other hand, matrix exponential spatial specification (MESS) is first proposed by LeSage & Pace (2007). It has advantages over traditional spatial autoregressive (SAR) models: a simpler log-likelihood function without the Jacobian matrix and an unrestricted parameter space for its spatial coefficients. Debarsy et al. (2015) derive QMLE and GMME of MESS(1,1) models and explore their large sample properties, with 1's representing MESS in the dependent variable and disturbances. Similar to SPD models, MESS can be extended to the panel models (Figueiredo & Da Silva (2015), LeSage & Chih (2018), Zhang et al. (2019)).

In this paper, the M-estimator in Yang (2018) is extended to the matrix exponential spatial dynamic panel specification (MESDPS) with fixed effect in short panels. Similar to SDPD model, the MESDPS also suffers from the "initial condition" problem. As discussed above, the traditional way of solving this problem, which is to use the predicted value derived from the values of the regressors, does not provide a satisfactory solution. A consistent way to estimate the coefficients and its VC matrix is needed. We first derive a set of conditional quasi score (CQS) functions treating the initial differences as exogenous, even if they are not. Then we modify these score functions to get the adjusted quasi score (AQS) functions which are unbiased. M-estimators thus are derived by setting AQS functions equaling to zero. To get a consistent estimate for the VC matrix of the M-estimator, a martingale difference (M.D.) of the AQS at the true value is established. The average of the outer product of M.D. (OPMD), referred to as the OPMD estimator, is shown to generate a consistent estimate of the VC matrix when substituted into the "sandwich" estimate of it. In Monte Carlo simulations six types of submodels, MESDPS(1,0,0), MESDPS(0,1,0), MESDPS(1,1,0), MESDPS(1,0,1), MESDPS(0,1,1) and MESDPS(1,1,1) are estimated, where 1's denote

MESS in the dependent variable, the lagged dependent variable and the disturbances respectively. The results show that the M-estimator has good finite sample properties and is robust to the way the initial observation is generated, which implies that it solves the "initial condition" difficulty. The OPMD estimator of the VC matrix generates asymptotic standard errors that's much closer to the true standard deviation than other candidates, especially when the disturbance is non-normal, emphasizing its importance in research when the normality of disturbances is in doubt. Different types of MESDPS are also applied on US outward FDI data to examine the validity of the model. Blonigen et al. (2007) propose four types of FDI based on different motivations of multinational enterprises and distinguish them by the sign of spatial lag term and surrounding market-potential. A modified gravity model that incorporates MESDPS is established to show the validity of the estimation. STLE from Yang (2018) is also reported to emphasize the relation for the spatial coefficients of these two models.

The contribution of this paper is two-fold. First the unified M-estimator is extended to MESDPS. MESS is not simply considered as substitute for SAR models. Although the relation for the spatial coefficients exists under row-normalized spatial weight matrix,[1] the difference in the parameter spaces does not grant the equivalency of these two models. The spatial coefficients in SARAR models are restricted to the range $(-1, 1)$, but the spatial coefficients in MESS has range $(-\infty, \infty)$. So it remains to be explored whether the M-estimation designed for SDPD model in Yang (2018) can be extended to MESDPS. Second, to our best knowledge, this is the first paper to consider MESS in a dynamic panel setting. Previous literature (Figueiredo & Da Silva (2015), LeSage & Chih (2018), Zhang et al. (2019)) study MESS in a panel data model. As mentioned previously, the "initial condition" problem remains when the spatial effects in the dynamic panel data model are in forms of MESS, so a consistent estimator for the coefficients and corresponding standard errors need to be designed, which is accomplished in this paper.

The rest of the paper is organized as follows. Section 2 introduces the M-estimation method. Section 3 presents the asymptotic distribution of the M-estimator and introduces the OPMD estimator of its VC matrix. Section 4 presents Monte Carlo simulation results. Section 5 applies the model to US outward FDI. Section 6 concludes.

2 M-estimation of Matrix Exponential Spatial Dynamic Panel Specification

In this section we first discuss the literature that incorporate MESS in panel data setting. Although these literature are in panel data instead of dynamic panel data setting, we include them in the review to underline the importance of our study, i.e., MESDPS has not been explored in the literature. The M-estimation and the OPMD estimator thus provide researchers who want to work with MESDPS a reliable method to estimate the parameters and conduct inference. In the second subsection we present the M-estimation in MESDPS(1,1,1) in short panel. Short panel assumes large n and small T, which is typical for most real world datasets. M-estimation first formulates a set of conditional quasi score (CQS) functions assuming that the initial difference is exogenous, and then modify it to get a set of adjusted quasi score (AQS) functions which result in consistent parameter estimates. If T becomes large but grows at the same rate with n, the approach is shown

[1] In cross-section context, the spatial coefficients in SARAR(1,1) model $y_n = \lambda W_n y_n + X_n\beta + u_n$, $u_n = \tau M_n u_n + \epsilon_n$ and MESS(1,1) model $e^{\alpha W_n} y_n = X_n\beta + u_n$, $e^{\tau M_n} u_n = \epsilon_n$ have the relation $\lambda = 1 - e^{\alpha}$. See Debarsy et al. (2015), p3.

to work perfectly as well.

2.1 Matrix Exponential Spatial Dynamic Panel Specification

The matrix exponential spatial dynamic panel specification with fixed effect is given by

$$e^{\alpha_1 W_1} y_t = \tau y_{t-1} + e^{\alpha_2 W_2} y_{t-1} + X_t \beta + Z\gamma + \mu + \lambda_t l_n + u_t, \; e^{\alpha_3 W_3} u_t = \epsilon_t, \; t = 1, 2, ..., T, \quad (2.1)$$

where y_t is an $n \times 1$ vector of observations on the dependent variable; W_r for $r = 1, 2, 3$ are three $n \times n$ spatial weight matrices, with corresponding spatial coefficients α_r capturing MESS in the dependent variable, lagged dependent variable and disturbances; y_{t-1} is the lagged vector of y_t with coefficient τ capturing the dynamic effect; X_t is an $n \times k$ matrix of time-varying exogenous variables with corresponding coefficient vector β; Z is an $n \times p$ matrix of time-invariant exogenous variables, which might include the intercept, with corresponding coefficient vector γ; μ is an $n \times 1$ vector of unobserved fixed effects; λ_t is the time-specific effects; l_n is an $n \times 1$ vector of 1; and ϵ_t is a vector of disturbances independent and identically distributed across i and t with mean zero and variance σ_ϵ^2. The matrix exponential $e^{\alpha_r W_r}$ is defined as $\Sigma_{j=0}^{\infty} \frac{\alpha_r^j W_r^j}{j!}$ for $r = 1, 2$ and 3 and is always invertible with inverse $e^{-\alpha_r W_r}$ (Chiu et al. (1996)). The specification in (2.1) is comprehensive. It incorporates different submodels by setting the spatial coefficients $\alpha_r = 0$ for $r = 1, 2$ or 3.

By setting $\alpha_2 = 0$, we have MESDPS(1,0,1) with MESS in the dependent variable and disturbances:

$$e^{\alpha_1 W_1} y_t = (\tau + 1) y_{t-1} + X_t \beta + Z\gamma + \mu + \lambda_t l_n + u_t, \; e^{\alpha_3 W_3} u_t = \epsilon_t, \qquad t = 1, 2, ..., T. \qquad (2.2)$$

Without $(\tau+1)y_{t-1}$, $Z\gamma$ and merging $\lambda_t l_n$ into $X_t\beta$, Zhang et al. (2019) study the QML estimation of (2.2) in panel data setting under heteroskedasticity. They allow large n and small or large T and establish the consistency and asymptotic normality under unknown heteroskedasticity when the spatial weight matrices in y_t and u_t are commutable.

By setting $\alpha_2 = 0$ and $\alpha_3 = 0$, we get MESDPS(1,0,0):

$$e^{\alpha_1 W_1} y_t = (\tau + 1) y_{t-1} + X_t \beta + Z\gamma + \mu + \lambda_t l_n + \epsilon_t, \qquad t = 1, 2, ..., T. \qquad (2.3)$$

Figueiredo & Da Silva (2015) discusses (2.3) without $(\tau + 1)y_{t-1}$ and $Z\gamma$. It uses the deviation from mean operator to get rid of the individual and time effect and present the ML estimation of the transformed model. This approach, however, results in linear dependent disturbance after transformation. Instead, we can pre- and post-multiply the model by the orthonormal eigenvector matrix of the individual and time mean deviation operators respectively (Lee & Yu (2010*a*)).

The literature above incorporate MESS into a panel data model. To the best of our knowledge, MESS in a dynamic panel setting has not been studied in previous literature. The M-estimation proposed in this paper provides consistent and asymptotically normal estimates. The OPMD estimate for the VC matrix also provides excellent finite sample properties. The method is useful for those who want to utilize MESDPS in empirical research.

2.2 M-estimation of MESDPS with Fixed Effect

Different from the geometrical decay in SDPD model, (2.1) has an exponential decay. It also has a simpler quasi log-likelihood function without the Jacobian of the transformation and has an unrestricted parameter space for α_r for $r = 1, 2, 3$. The MESS can be extended to contain multiple spatial weight matrices, i.e., $e^{\Sigma_{s=1}^{q} \alpha_{rs} W_{rs}}$ for $r = 1, 2$ and 3[2]. However, they suffer from the "initial condition" problem discussed below.

[2]See Debarsy et al. (2015), Appendix B.1.

Denote the true value of the parameter vector by $\theta_0 = (\beta_0', \sigma_{\epsilon0}^2, \tau_0, \alpha_0')'$, where $\alpha_0 = (\alpha_{10}, \alpha_{20}, \alpha_{30})'$. Let $A_{20} = \tau_0 I_n + e^{\alpha_{20}W_2}$. Taking first difference for (2.1), we get:

$$e^{\alpha_{10}W_1}\Delta y_t = A_{20}\Delta y_{t-1} + \Delta X_t\beta_0 + \Delta u_t,\ e^{\alpha_{30}W_3}\Delta u_t = \Delta\epsilon_t,\ t = 2, 3, ..., T. \tag{2.4}$$

where $\Delta\lambda_{t0}l_n$ is merged into $\Delta X_t\beta_0$. Note (2.4) is not defined for $t = 1$ because Δy_1 depends on Δy_0 and the latter is not observed. So even if y_0 and Δy_0 is exogenous, the likelihood function which conditions on Δy_0 cannot be formulated. Also y_1 and thus Δy_1 are not exogenous. This "initial condition" problem prevents us from deriving consistent estimates of MESDPS. The traditional way is to use the predicted values based on the observed values of regressors. However, it requires that the time-varying regressors be trend or first-difference stationary. Besides, for MESDPS with MESS in the dependent variable, for example MESDPS(1,0,0), the first differenced equation is given by $e^{\alpha_{10}W_1}\Delta y_t = (\tau_0+1)\Delta y_{t-1} + \Delta X_t\beta_0 + \Delta\epsilon_t$. By backward substitution, we get $\Delta y_1 = (e^{-\alpha_{10}W_1})^m\Delta y_{-m+1} + \Sigma_{i=0}^{m-1}(\tau_0+1)^i(e^{-\alpha_{10}W_1})^{i+1}\Delta X_{-i+1}\beta_0 + \Sigma_{i=0}^{m-1}(\tau_0+1)^i(e^{-\alpha_{10}W_1})^{i+1}\Delta\epsilon_{-i+1}$, where $-m$ is the process starting time. Note the exogenous part contains the spatial effect $e^{-\alpha_{10}W_1}$. The linear structure no longer exists due to the existence of the spatial effect and the linear projection method fails. Thus we need a unified way to estimate the model.

To express the model in vector form, we define the following matrices: $\Delta Y = (\Delta y_2', ..., \Delta y_T')'$, $\Delta Y_{-1} = (\Delta y_1', ..., \Delta y_{T-1}')'$, $\Delta X = (\Delta X_2', ..., \Delta X_T')'$, $\Delta u = (\Delta u_2', ..., \Delta u_T')'$, $\Delta\epsilon = (\Delta\epsilon_2', ..., \Delta\epsilon_T')$, $\boldsymbol{A}_{20} = I_{T-1}\otimes A_{20}$ and $\boldsymbol{e^{\alpha_{r0}W_r}} = I_{T-1}\otimes e^{\alpha_{r0}W_r}$ for $r = 1, 2$ and 3. Stacking the observations vertically, the model can be expressed as:

$$\boldsymbol{e^{\alpha_{10}W_1}}\Delta Y = \boldsymbol{A}_{20}\Delta Y_{-1} + \Delta X\beta_0 + \Delta u,\ \boldsymbol{e^{\alpha_{30}W_3}}\Delta u = \Delta\epsilon. \tag{2.5}$$

So $var(\Delta u) = var(\boldsymbol{e^{-\alpha_{30}W_3}}\Delta\epsilon) = \sigma_{\epsilon0}^2(B\otimes e^{-\alpha_{30}W_3}e^{-\alpha_{30}W_3'}) = \sigma_{\epsilon0}^2\Sigma(\alpha_{30})$, where

$$B = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

Under normally distributed ϵ_t, the joint distribution of Δu_t can be used to derive the log-likelihood function of parameters θ:

$$\begin{aligned}\ell(\theta) = &-\tfrac{n(T-1)}{2}log(2\pi) - \tfrac{n(T-1)}{2}log(\sigma_\epsilon^2) - \tfrac{1}{2}log|\Sigma(\alpha_3)| + log|\boldsymbol{e^{\alpha_1W_1}}| \\ &-\tfrac{1}{2\sigma_\epsilon^2}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi),\end{aligned} \tag{2.6}$$

with $\theta = (\beta', \sigma_\epsilon^2, \tau, \alpha')'$ and $\phi = (\beta', \tau, \alpha_1, \alpha_2)'$ where ϕ are the parameters in $\Delta u(\phi) = \boldsymbol{e^{\alpha_1W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1} - \Delta X\beta$. Note $log|\Sigma(\alpha_3)| = nlog|B| + 2(T-1)log(e^{-\alpha_3 tr(W_3)}) = nlog|B|$ which is a constant and $log(|\boldsymbol{e^{\alpha_1W_1}}|) = (T-1)log(|e^{\alpha_1 tr(W_1)}|) = 0$ because the spatial weight matrices have zero diagonals. So we can ignore the constants and simplify the log-likelihood function to:

$$\ell(\theta) = -\tfrac{n(T-1)}{2}log(\sigma_\epsilon^2) - \tfrac{1}{2\sigma_\epsilon^2}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi). \tag{2.7}$$

Given $\zeta = (\tau, \alpha')'$ with $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$, we can derive the constrained estimator of β and σ_ϵ^2 as following:

$$\tilde\beta(\zeta) = (\Delta X'\Sigma(\alpha_3)^{-1}\Delta X)^{-1}\Delta X'\Sigma(\alpha_3)^{-1}(\boldsymbol{e^{\alpha_1W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1}), \tag{2.8}$$

$$\tilde\sigma_\epsilon^2(\zeta) = \tfrac{1}{n(T-1)}\Delta\tilde u(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\tilde u(\zeta), \tag{2.9}$$

where $\Delta\tilde u(\zeta) = \boldsymbol{e^{\alpha_1W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1} - \Delta X\tilde\beta(\zeta)$. Substituting them back into (2.7), ignoring constant, we get the concentrated log-likelihood function:

$$l^c(\zeta) = -\frac{n(T-1)}{2} log[\Delta\tilde{u}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\tilde{u}(\zeta)] \tag{2.10}$$

The unconstrained conditional QML (CQML) estimators $\tilde{\zeta} = (\tilde{\tau}, \tilde{\alpha}')'$ are then derived by maximizing (2.10). The unconstrained CQML estimators $\tilde{\beta} = \tilde{\beta}(\tilde{\zeta})$ and $\tilde{\sigma}_\epsilon^2 = \tilde{\sigma}_\epsilon^2(\tilde{\zeta})$ are subsequently derived by substituting $\tilde{\zeta}$ into (2.8) and (2.9).

Consider the STLE model in Yang (2018), i.e., $y_t = \rho y_{t-1} + \lambda_1 W_1 y_t + \lambda_2 W_2 y_{t-1} + X_t\beta + Z\gamma + \mu + \alpha_t l_n + u_t$, $u_t = \lambda_3 W_3 u_t + \epsilon_t$, the log-likelihood function (2.7) and the concentrated log-likelihood function (2.10) are simpler without the Jacobian $log|\boldsymbol{B}_1(\lambda_1)|$ where $\boldsymbol{B}_1(\lambda_1) = I_{T-1} \otimes B_1(\lambda_1)$ and $B_1(\lambda_1) = I_n - \lambda_1 W_1$. It makes the MESDPS computationally easier, especially for large sample sizes. A correspondence of relation for the parameters also exist for MESDPS and STLE model. Consider (2.1), assuming the spatial weight matrix is row-normalized, the total impact of a shock Δx_t on the kth independent variable X_{tk} for MESDPS is given by $\Delta y_t = e^{-\alpha_1 W_1} l_n \Delta x_t \beta_k$, so the average total impact is $\frac{1}{n} l_n' \Delta y_t = e^{-\alpha_1}\Delta x_t \beta_k$, where $W_1^k l_n = W_1 l_n = l_n$ is used. Similarly for STLE, the average total impact is given by $\frac{1}{1-\lambda_1}\Delta x_t \beta_k$. Equating them gives us the relation $\lambda_1 = 1 - e^{\alpha_1}$. For y_{t-1}, a shock $\Delta\nu_{t-1}$ leads to the average total impact $e^{-\alpha_1}(\tau + e^{\alpha_2})\Delta\nu_{t-1}$ for MESDPS and $\frac{\rho+\lambda_2}{1-\lambda_1}\Delta\nu_{t-1}$ for STLE. So $\tau + e^{\alpha_2} = \rho + \lambda_2$. Setting $\alpha_2 = 0$ and $\lambda_2 = 0$ gives us $\rho = \tau + 1$, which implies $\lambda_2 = e^{\alpha_2} - 1$. On contrary to the negative relation between α_1 and λ_1, the relation between α_2 and λ_2 is positive. When $-1 < \lambda_2 < 0$, α_2 also takes negative values and vice versa. However, the two models cannot be considered as substitutes of each other. When α_1 or α_2 is bigger than $ln(2)$, the corresponding values for λ_1 is less than -1 and for λ_2 is bigger than 1. So we need to explore whether the M-estimation and the OPMD estimator of standard error can be applied to MESDPS.

The CQML estimator $\tilde{\theta} = (\tilde{\beta}', \tilde{\sigma}_\epsilon^2, \tilde{\tau}, \tilde{\alpha}')'$ derived above encounters a bias when T is small and an asymptotic bias even when T is large but grows at the same rate with n as shown below. We simplify the notation by denoting $\Sigma = \Sigma(\alpha_3)$ and $\Sigma_0 = \Sigma(\alpha_{30})$. The conditional quasi score (CQS) function $S(\theta) = \frac{\partial \ell(\theta)}{\partial\theta}$ is given by

$$S(\theta) = \begin{cases} \frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma^{-1}\Delta u(\phi), \\ -\frac{n(T-1)}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}\Delta u(\phi)'\Sigma^{-1}\Delta u(\phi), \\ \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\Delta Y_{-1}, \\ -\frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_1 \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y, \\ \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_2 \boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}, \\ -\frac{1}{2\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi), \end{cases} \tag{2.11}$$

where $E_3 = e^{\alpha_3 W_3'}(W_3 + W_3')e^{\alpha_3 W_3}$. We will show that τ, α_1 and α_2 elements of CQS function are biased, meaning their expected values are nonzero at the true parameter values, leading to the inconsistency of the CQML estimators. First let's make Assumption 1 below.

Assumption 1: For model (2.1), (i) the processes started m periods before the start of data collection, the 0th period, (ii) if $m \geq 1$, Δy_0 is independent of future disturbances $\{\epsilon_t, t \geq 1\}$; if $m = 0$, y_0 is independent of future disturbances $\{\epsilon_t, t \geq 1\}$.

Assumption 1 is the same as the Assumption A in Yang (2018). Compared with the assumptions in previous literature (e.g. Su & Yang (2015)), Assumption 1 requires minimum information about the past processes. It does not require the time-varying regressors to be trend-stationary or first-difference stationary. This is one of the advantages of M-estimation, i.e., some restrictive assumptions on the initial values and initial differences are removed. Denote $A_{21,0} = A_{20}e^{-\alpha_{10}W_1}$. The following lemma is necessary to compute the bias of CQS function.

Lemma 2.1. Under Assumption 1, $E(\Delta Y \Delta\epsilon') = -\sigma_{\epsilon 0}^2 \boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}} \boldsymbol{D}_0 \boldsymbol{e}^{-\boldsymbol{\alpha_{30} W_3}}$ and $E(\Delta Y_{-1}\Delta\epsilon') = -\sigma_{\epsilon 0}^2 \boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}} \boldsymbol{D}_{-1,0} \boldsymbol{e}^{-\boldsymbol{\alpha_{30} W_3}}$, where

$$\boldsymbol{D}_{-1,0} = \begin{pmatrix} I_n & 0 & \dots & \dots & 0 \\ A_{21,0}-2I_n & I_n & \ddots & \dots & \vdots \\ (A_{21,0}-I_n)^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ A_{21,0}^{T-4}(A_{21,0}-I_n)^2 & \dots & (A_{21,0}-I_n)^2 & A_{21,0}-2I_n & I_n \end{pmatrix} \text{ and}$$

$$\boldsymbol{D}_0 = \begin{pmatrix} A_{21,0}-2I_n & I_n & \dots & \dots & 0 \\ (A_{21,0}-I_n)^2 & A_{21,0}-2I_n & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & I_n \\ A_{21,0}^{T-3}(A_{21,0}-I_n)^2 & \dots & \dots & (A_{21,0}-I_n)^2 & A_{21,0}-2I_n \end{pmatrix}.$$

Here we used the fact that ϵ_{it} is i.i.d. across i and t, and that $e^{\alpha_{r0}W_r}$ is always invertible for $r = 1, 3$. The latter is an advantage of MESS over the traditional SAR models, i.e., the matrix exponential is always invertible so that the VC matrix of the dependent variable always exists and is positive definite. Thus no restrictions need to be imposed on the parameter space of α.

Utilizing Lemma 2.1, we have

$$E(\Delta u'\Sigma_0^{-1}\Delta Y_{-1}) = -\sigma_{\epsilon 0}^2 tr(\boldsymbol{D}_{-1,0}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}}), \tag{2.12}$$

$$E(\Delta u'\Sigma_0^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_{10} W_1}}\Delta Y) = -\sigma_{\epsilon 0}^2 tr(\boldsymbol{D}_0\boldsymbol{B}^{-1}\boldsymbol{W}_1), \tag{2.13}$$

$$E(\Delta u'\Sigma_0^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_{20} W_2}}\Delta Y_{-1}) = -\sigma_{\epsilon 0}^2 tr(\boldsymbol{D}_{-1,0}\boldsymbol{B}^{-1}\boldsymbol{W}_{21,0}), \tag{2.14}$$

where $\boldsymbol{W}_{21,0} = \boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_{20} W_2}}\boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}}$ and $\boldsymbol{B} = B \otimes I_n$. These equations imply that $E(\frac{\partial \ell(\theta)}{\partial \tau})$, $E(\frac{\partial \ell(\theta)}{\partial \alpha_1})$ and $E(\frac{\partial \ell(\theta)}{\partial \alpha_2})$ are nonzero, making τ, α_1 and α_2 elements of the CQS function biased. The set of CQS functions (2.11) are estimating functions for the CQML estimator. From previous literature (e.g., Van der Vaart (2000) Chapter 5), we know the consistency of an M-estimator requires the estimating function to have a probability limit of zero at the true parameter values, i.e., $plim_{n\to\infty}\frac{1}{nT}S(\theta_0) = 0$. However Lemma 2.1 implies that it does not hold for CQMLE. Typically $E(\frac{\partial \ell(\theta)}{\partial \tau})$, $E(\frac{\partial \ell(\theta)}{\partial \alpha_1})$ and $E(\frac{\partial \ell(\theta)}{\partial \alpha_2})$ are of order n, which implies $E[\sqrt{nT}(\tilde{\theta}-\theta_0)] = O(\sqrt{\frac{n}{T}})$. Even if $T \to \infty$ but grows at as fast as n, the asymptotic bias does not vanish. The bias vanishes when $\frac{n}{T} \to 0$, which refers to a large panel and is not of interest in our study. So the CQML estimation fails to produce consistent estimates.

To have a set of unbiased estimating functions, we modify them to get the adjusted quasi score (AQS) functions:

$$S^*(\theta) = \begin{cases} \frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma^{-1}\Delta u(\phi), \\ -\frac{n(T-1)}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}\Delta u(\phi)'\Sigma^{-1}\Delta u(\phi), \\ \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21}), \\ -\frac{1}{2\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi). \end{cases} \tag{2.15}$$

The M-estimator derived from the AQS functions are consistent and asymptotic normal, which will be shown in Theorem 3.1 and 3.2 respectively. It is interesting to compare the AQS functions with those in SDPD model in Yang (2018). First the bias term $tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_1 W_1}})$ in the τ elements has similar format with that for the τ element[3] in Yang (2018) (with MESS instead of SAR process as the multiplier). This means that while inherent spatial processes are different, the format of bias that come from the dynamic effect is not affected by the nature of the spatial structure. second thing to note is that, similar to SDPD model, the adjustments in AQS functions are free from MESS in the disturbance term, i.e., $e^{\alpha_3 W_3}$ is not involved in the trace terms. This implies that the AQS adjustments will not change if MESS in the disturbance term changes to other forms of spatial relationship, e.g., higher order MESS, autoregressive, moving average, etc. Third the adjustments modify the estimation of τ, α_1 and α_2 so that they become nonlinear, implying some hidden information is surfaced.

To derive the M-estimator, we first solve for the constrained M-estimators of β and σ_ϵ^2, given $\zeta = (\tau, \alpha')'$, as

$$\hat{\beta}_M(\zeta) = (\Delta X'\Sigma^{-1}\Delta X)^{-1}\Delta X'\Sigma^{-1}(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1}), \tag{2.16}$$

$$\hat{\sigma}^2_{\epsilon,M}(\zeta) = \frac{1}{n(T-1)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\Delta\hat{u}(\zeta), \tag{2.17}$$

where $\Delta\hat{u}(\zeta) = \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1} - \Delta X\hat{\beta}_M(\zeta)$. Then $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$ are substituted back into the other four elements of the AQS function $S^*(\theta)$ to get the concentrated AQS function:

$$S^{*c}(\zeta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21}), \\ -\frac{1}{2\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\hat{u}(\zeta). \end{cases} \tag{2.18}$$

The unconstrained M-estimator $\hat{\zeta}_M = (\hat{\tau}_M, \hat{\alpha}'_M)'$ can be solved by letting $S^{*c}(\zeta) = 0$. The unconstrained M-estimators $\hat{\beta}_M$ and $\hat{\sigma}^2_{\epsilon,M}$ are then derived by substituting $\hat{\zeta}_M$ into $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$. Note CQMLE and M-estimator use the same set of constrained estimators of β and σ^2_ϵ to derive unconstrained ones, i.e., $\hat{\beta}_M(\zeta) = \tilde{\beta}(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta) = \tilde{\sigma}^2_\epsilon(\zeta)$. The advantage of M-estimation comes from the AQS function (2.15). It adjusts the estimation function so that they become unbiased. For the CQML estimation, the unconstrained estimators $\tilde{\beta}$ and $\tilde{\sigma}^2$ are biased because of the spillover from the bias of the unconstrained estimators $\tilde{\zeta}$ when being substituted into (2.8) and (2.9).

After deriving the M-estimators $\hat{\theta}_M = (\hat{\beta}'_M, \hat{\sigma}^2_{\epsilon,M}, \hat{\tau}_M, \hat{\alpha}'_M)'$, the next task is to prove its consistency, establish its asymptotic distribution and provide a consistent estimator for the VC matrix. To get a consistent estimate of VC matrix is not as straightforward as it seems, i.e., substituting the consistent M-estimator into its asymptotic VC matrix. A consistent method is needed which is presented in section 3.

3 Asymptotic Properties of the M-estimator

In this section we explore the asymptotic properties of the M-estimator. We first prove it is consistent and then derive its asymptotic distribution. To facilitate valid inference, an OPMD

[3]Note the differences in the definition of matrix $\boldsymbol{D}$ and $\boldsymbol{D}_{-1}$ with those in Yang (2018).

estimator of the VC matrix is also proposed. Valid inference can thus be based on standard errors implied by the OPMD-estimator of the VC matrix.

3.1 Consistency of the M-estimator

To prove the consistency and to later derive the asymptotic distribution of the M-estimator, we first make some regularity assumptions. Let C_n be an $n \times n$ matrix. Then C'_n, $tr(C_n)$, $|C_n|$, $||C_n||$, $\gamma_{min}(C_n)$ and $\gamma_{max}(C_n)$ denote the transpose, trace, determinant, Euclidean norm, the smallest and largest eigenvalues of C_n respectively.

Assumption 2. Matrices $\{W_1\}$, $\{W_2\}$ and $\{W_3\}$ are bounded in both row and column sum norms. The diagonal elements of W_1, W_2 and W_3 are zeroes.

Assumption 3. $\{X_t\}$ is exogenous, with uniformly bounded elements, and has full column rank. Also $lim_{n\to\infty} \frac{1}{nT}\Delta X' \Delta X$ exists and is nonsingular.

Assumption 4. There exists a constant $\delta > 0$ such that $|\alpha_r| \leq \delta$ for $r = 1, 2$ and 3, and the true α_0 is in the interior of the parameter space $\mathcal{A}$. Also there exist a lower bound $\underline{c}_{\alpha_r}$ and an upper bound $\bar{c}_{\alpha_r}$ such that $0 < \underline{c}_{\alpha_r} \leq inf_{\alpha_r \in \mathcal{A}_r} \gamma_{min}(e^{\alpha_r W'_r} e^{\alpha_r W_r}) \leq sup_{\alpha_r \in \mathcal{A}_r} \gamma_{max}(e^{\alpha_r W'_r} e^{\alpha_r W_r}) \leq \bar{c}_{\alpha_r} < \infty$ for $r = 1, 2$ and 3.

Assumption 5. The $\{\epsilon_{it}\}$ are i.i.d. with mean zero and variance σ^2_ϵ, and $E|\epsilon_{it}|^{4+a}$ exists for some $a > 0$.

Assumption 6. For an $n \times n$ matrix C_n which is uniformly bounded in row and column sums, with elements of uniform order g_n^{-1}, and an $n \times 1$ vector c_n with elements of uniform order $g_n^{-1/2}$, (i) $\frac{g_n}{n}\Delta y'_1 C_n \Delta y_1 = O_p(1)$ and $\frac{g_n}{n}\Delta y'_1 C_n \Delta\epsilon_2 = O_p(1)$; (ii) $\frac{g_n}{n}[\Delta y_1 - E(\Delta y_1)]' c_n = o_p(1)$; (iii) $\frac{g_n}{n}[\Delta y'_1 C_n \Delta y_1 - E(\Delta y'_1 C_n \Delta y_1)] = o_p(1)$; (iv) $\frac{g_n}{n}[\Delta y'_1 C_n \Delta\epsilon_2 - E(\Delta y'_1 C_n \Delta\epsilon_2)] = o_p(1)$.

Assumptions 1-5 are standard in literature (e.g. Debarsy et al. (2015)). Assumption 6 is the same as Assumption F in Yang (2018). It imposes some mild conditions on the initial difference Δy_1 which will be used in the later proofs.

To prove the consistency of $\hat{\theta}_M$, we note that it follows from the consistency of $\hat{\zeta}_M$ since $\hat{\beta}_M = \hat{\beta}_M(\hat{\zeta}_M)$ and $\hat{\sigma}^2_{\epsilon,M} = \hat{\sigma}^2_{\epsilon,M}(\hat{\zeta}_M)$. To prove the consistency of $\hat{\zeta}_M$, we first define the population counterpart of the AQS function as:

$$\bar{S}^*(\theta) = E[S^*(\theta)] = \begin{cases} \frac{1}{\sigma^2_\epsilon} E[\Delta X' \Sigma^{-1} \Delta u(\phi)], \\ -\frac{n(T-1)}{2\sigma^2_\epsilon} + \frac{1}{2\sigma^4_\epsilon} E[\Delta u(\phi)' \Sigma^{-1} \Delta u(\phi)], \\ \frac{1}{\sigma^2_\epsilon} E[\Delta u(\phi)' \Sigma^{-1} \Delta Y_{-1}] + tr(\boldsymbol{D}_{-1} \boldsymbol{B}^{-1} \boldsymbol{e}^{-\alpha_1 \boldsymbol{W}_1}), \\ -\frac{1}{\sigma^2_\epsilon} E[\Delta u(\phi)' \Sigma^{-1} \boldsymbol{W}_1 \boldsymbol{e}^{\alpha_1 \boldsymbol{W}_1} \Delta Y] - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ \frac{1}{\sigma^2_\epsilon} E[\Delta u(\phi)' \Sigma^{-1} \boldsymbol{W}_2 \boldsymbol{e}^{\alpha_2 \boldsymbol{W}_2} \Delta Y_{-1}] + tr(\boldsymbol{D}_{-1} \boldsymbol{B}^{-1} \boldsymbol{W}_{21}), \\ -\frac{1}{2\sigma^2_\epsilon} E[\Delta u(\phi)' (B^{-1} \otimes E_3) \Delta u(\phi)]. \end{cases} \tag{3.1}$$

Similar to deriving the M-estimator, we can first solve for $\bar{\beta}_M(\zeta)$ and $\bar{\sigma}^2_{\epsilon,M}(\zeta)$ as:

$$\bar{\beta}_M(\zeta) = (\Delta X' \Sigma^{-1} \Delta X)^{-1} \Delta X' \Sigma^{-1} [\boldsymbol{e}^{\alpha_1 \boldsymbol{W}_1} E(\Delta Y) - \boldsymbol{A}_2 E(\Delta Y_{-1})], \tag{3.2}$$

$$\bar{\sigma}^2_{\epsilon,M}(\zeta) = \frac{1}{n(T-1)} E[\Delta \bar{u}(\zeta)' \Sigma^{-1} \Delta \bar{u}(\zeta)], \tag{3.3}$$

where $\Delta \bar{u}(\zeta) = \boldsymbol{e}^{\alpha_1 \boldsymbol{W}_1} \Delta Y - \boldsymbol{A}_2 \Delta Y_{-1} - \Delta X \bar{\beta}_M(\zeta)$. By substituting them into the last four elements of $\bar{S}^*(\theta)$, we get the population counterpart of the concentrated AQS function (2.18) as

$$\bar{S}^{*c}(\zeta) = \begin{cases} \frac{1}{\bar{\sigma}^2_{\epsilon,M}(\zeta)} E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1}] + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\bar{\sigma}^2_{\epsilon,M}(\zeta)} E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y] - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ \frac{1}{\bar{\sigma}^2_{\epsilon,M}(\zeta)} E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}] + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21}), \\ -\frac{1}{2\bar{\sigma}^2_{\epsilon,M}(\zeta)} E[\Delta\bar{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\bar{u}(\zeta)]. \end{cases} \tag{3.4}$$

Note ζ_0 is a zero of $\bar{S}^{*c}(\zeta)$. According to Theorem 5.9 of Van der Vaart (2000), if $\hat{\zeta}_M$ is a zero of $S^{*c}(\zeta)$ and ζ_0 is a zero of $\bar{S}^{*c}(\zeta)$, then $\hat{\zeta}_M$ is a consistent estimator of ζ_0 if $sup_{\zeta\in\mathcal{Z}} \frac{1}{n(T-1)}||S^{*c}(\zeta) - \bar{S}^{*c}(\zeta)|| \xrightarrow{p} 0$ and the following Assumption holds.

Assumption 7. $inf_{\zeta:d(\zeta,\zeta_0)\geq\nu}||\bar{S}^{*c}(\zeta)|| > 0$ for every $\nu > 0$, where $d(\zeta,\zeta_0)$ is a measure of distance between ζ and ζ_0.

Before we show $sup_{\zeta\in\mathcal{Z}} \frac{1}{n(T-1)}||S^{*c}(\zeta) - \bar{S}^{*c}(\zeta)|| \xrightarrow{p} 0$, let's first define some convenient expressions. Let $\Delta\bar{u}^*(\zeta) = \Sigma^{-\frac{1}{2}}\Delta\bar{u}(\zeta)$, $\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} = \Sigma^{-\frac{1}{2}}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}$, $\boldsymbol{A}_2^* = \Sigma^{-\frac{1}{2}}\boldsymbol{A}_2$, $\Delta Y^\dagger = \Delta Y - E(\Delta Y)$, $\Delta Y_{-1}^\dagger = \Delta Y_{-1} - E(\Delta Y_{-1})$, $P = \Sigma^{-\frac{1}{2}}\Delta X(\Delta X'\Sigma^{-1}\Delta X)^{-1}\Delta X'\Sigma^{-\frac{1}{2}}$ and $M = I_{n(T-1)} - P$. Then we have

$$\Delta\bar{u}^*(\zeta) = P(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y^\dagger - \boldsymbol{A}_2^*\Delta Y_{-1}^\dagger) + M(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y - \boldsymbol{A}_2^*\Delta Y_{-1}). \tag{3.5}$$

The expression will be useful in deriving $\bar{\sigma}^2_{\epsilon,M}(\zeta)$ in (3.3) in the proof for Theorem 3.1 below.

Theorem 3.1. Suppose Assumptions 1-7 hold and further the following condition $0 < \underline{c}_{\Delta Y} \leq inf_{\zeta\in\mathcal{Z}}\gamma_{min}[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1})] \leq sup_{\zeta\in\mathcal{Z}}\gamma_{max}[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1})] \leq \bar{c}_{\Delta Y} < \infty$, we have $\hat{\theta}_M \xrightarrow{p} \theta_0$ as $n \to \infty$.

3.2 Asymptotic Distribution of the M-estimator

To derive the asymptotic distribution of $\hat{\theta}_M$, we apply the mean value theorem (MVT) to $S^*(\hat{\theta}_M) = 0$ at the true θ_0 to get $\sqrt{n(T-1)}(\hat{\theta}_M - \theta_0) = -(\frac{1}{n(T-1)}\frac{\partial S^*(\bar{\theta})}{\partial\theta'})^{-1}\frac{1}{\sqrt{n(T-1)}}S^*(\theta_0)$ for some $\bar{\theta}$ between θ_0 and $\hat{\theta}_M$ elementwise. Then we show that $\frac{1}{n(T-1)}\frac{\partial S^*(\bar{\theta})}{\partial\theta'}$ carries appropriate asymptotic properties and that $\frac{1}{\sqrt{n(T-1)}}S^*(\theta_0)$ is asymptotically normal. One thing to note here is that Δy_1 might not be exogenous and is unspecified, so the regular law of large numbers (LLN) and central limit theorem (CLT) for linear-quadratic forms is not sufficient. Instead we use extended LLN and CLT for bilinear-quadratic forms from Yang (2018) and Su & Yang (2015), which is listed in Lemmata A.3 and A.4. The following lemma that expresses ΔY and ΔY_{-1} in a convenient format will be crucial in deriving the asymptotic distribution and later a consistent estimate of the VC matrix. Let $blkdiag(C_1, ..., C_n)$ be the block diagonal matrix with diagonal $n\times n$ matrices $C_1, ..., C_n$. Denote $A_{12,0} = e^{-\alpha_{10}W_1}A_{20}$.

Lemma 3.1. Under Assumptions 1, 3 and 5,

$$\Delta Y = G\Delta\boldsymbol{y}_1 + \boldsymbol{\delta} + K\Delta\epsilon, \tag{3.6}$$

$$\Delta Y_{-1} = G_{-1}\Delta\boldsymbol{y}_1 + \boldsymbol{\delta}_{-1} + K_{-1}\Delta\epsilon, \tag{3.7}$$

where $\Delta\boldsymbol{y}_1 = I_{T-1}\otimes\Delta y_1, G = blkdiag[A_{12,0}, (A_{12,0})^2, ..., (A_{12,0})^{T-1}]$, $G_{-1} = blkdiag[I_n, A_{12,0}, ..., (A_{12,0})^{T-2}]$, $\boldsymbol{\delta} = J\boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}}\Delta X\beta_0$, $\boldsymbol{\delta}_{-1} = J_{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}}\Delta X\beta_0$, $K = J\boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}}\boldsymbol{e}^{-\boldsymbol{\alpha_{30} W_3}}$, $K_{-1} = J_{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_{10} W_1}}\boldsymbol{e}^{-\boldsymbol{\alpha_{30} W_3}}$,

$$J = \begin{pmatrix} I_n & 0 & \dots & \dots & 0 \\ A_{12,0} & \ddots & \ddots & \ddots & \vdots \\ A_{12,0}^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ A_{12,0}^{T-2} & \dots & A_{12,0}^2 & A_{12,0} & I_n \end{pmatrix} \text{ and } \mathrm{J}_{-1} = \begin{pmatrix} 0 & 0 & \dots & \dots & 0 \\ I_n & \ddots & \ddots & \ddots & \vdots \\ A_{12,0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ A_{12,0}^{T-3} & \dots & A_{12,0} & I_n & 0 \end{pmatrix}.$$

By substituting (3.6) and (3.7) into τ, α_1 and α_2 elements and $\Delta u = \boldsymbol{e^{-\alpha_{30} W_3}} \Delta\epsilon$ into the β, σ_ϵ^2 and α_3 elements of the AQS function (2.15) at the true value θ_0, we get

$$S^*(\theta_0) = \begin{cases} R_1' \Delta\epsilon, \\ -\frac{n(T-1)}{2\sigma_{\epsilon 0}^2} + \Delta\epsilon' O_1 \Delta\epsilon, \\ \Delta\epsilon' F_1 \Delta \boldsymbol{y}_1 + R_2' \Delta\epsilon + \Delta\epsilon' O_2 \Delta\epsilon + tr(\boldsymbol{D}_{-1,0} \boldsymbol{B}^{-1} \boldsymbol{e^{-\alpha_{10} W_1}}), \\ -\Delta\epsilon' F_2 \Delta \boldsymbol{y}_1 - R_3' \Delta\epsilon - \Delta\epsilon' O_3 \Delta\epsilon - tr(\boldsymbol{D}_0 \boldsymbol{B}^{-1} \boldsymbol{W}_1), \\ \Delta\epsilon' F_3 \Delta \boldsymbol{y}_1 + R_4' \Delta\epsilon + \Delta\epsilon' O_4 \Delta\epsilon + tr(\boldsymbol{D}_{-1,0} \boldsymbol{B}^{-1} \boldsymbol{W}_{21,0}), \\ \Delta\epsilon' O_5 \Delta\epsilon, \end{cases} \tag{3.8}$$

where $R_1 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \Delta X$, $R_2 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{\delta}_{-1}$, $R_3 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{W}_1 \boldsymbol{e^{\alpha_{10} W_1}} \boldsymbol{\delta}$, $R_4 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{W}_2 \boldsymbol{e^{\alpha_{20} W_2}} \boldsymbol{\delta}_{-1}$, $O_1 = \frac{1}{2\sigma_{\epsilon 0}^4}(B^{-1} \otimes I_n)$, $O_2 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) K_{-1}$, $O_3 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{W}_1 \boldsymbol{e^{\alpha_{10} W_1}} K$, $O_4 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{W}_2 \boldsymbol{e^{\alpha_{20} W_2}} K_{-1}$, $O_5 = -\frac{1}{2\sigma_{\epsilon 0}^2}[B^{-1} \otimes (W_3 + W_3')]$, $F_1 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) G_{-1}$, $F_2 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{W}_1 \boldsymbol{e^{\alpha_{10} W_1}} G$ and $F_3 = \frac{1}{\sigma_{\epsilon 0}^2}(B^{-1} \otimes e^{\alpha_{30} W_3}) \boldsymbol{W}_2 \boldsymbol{e^{\alpha_{20} W_2}} G_{-1}$.

Using $S^*(\theta_0)$ in (3.8), we can derive the expected Hessian of loglikehood function and the expected variance of score function at the true value to get the the asymptotic distribution of the M-estimator.

Theorem 3.2. Suppose assumptions of Theorem 3.1 hold, we have

$$\sqrt{n(T-1)}(\hat{\theta}_M - \theta_0) \xrightarrow{d} N[0, lim_{n\to\infty} \Psi^{*-1}(\theta_0) \Omega^*(\theta_0) \Psi^{*-1}(\theta_0)], \tag{3.9}$$

where $\Psi^*(\theta_0) = -\frac{1}{n(T-1)} E\left[\frac{\partial S^*(\theta_0)}{\partial \theta'}\right]$ and $\Omega^*(\theta_0) = \frac{1}{n(T-1)} var\left[S^*(\theta_0)\right]$ are assumed to exist and $\Psi^*(\theta_0)$ is positive definite for sufficiently large n.

3.3 The OPMD Estimator of VC Matrix

In this section we derive a feasible estimator for the VC matrix $\Psi^{*-1}(\theta_0)\Omega^*(\theta_0)\Psi^{*-1}(\theta_0)$. Denote the Hessian matrix by $H^*(\theta) = \frac{\partial S^*(\theta)}{\partial \theta'}$. Then a consistent estimate of $\Psi^*(\theta_0)$ is easily derived by substituting the consistent M-estimates in, i.e., $\Psi^*(\hat{\theta}_M) = -\frac{1}{n(T-1)} H^*(\hat{\theta}_M)$. The detailed expression and the proof for the consistency of $\Psi^*(\hat{\theta}_M)$ are provided in the proof of Theorem 3.2 in Appendix C.

For $\Omega^*(\theta_0)$, however, this method does not work. This is because from (3.8) we know that τ, α_1 and α_2 element of $\Omega^*(\theta_0)$ contain the initial difference Δy_1, which is unspecified. So we need to design a method that is free from the initial condition. Following Yang (2018), we propose an outer product of martingale difference (OPMD) method to consistently estimate $\Omega^*(\theta_0)$. The OPMD method first transforms $S^*(\theta_0)$ in (3.8) into a sum of vector martingale difference sequence (MDS). To be specific, we will write $R_r' \Delta\epsilon$, $\Delta\epsilon' O_r \Delta\epsilon - E(\Delta\epsilon' O_r \Delta\epsilon)$ and $\Delta\epsilon' F_r \Delta \boldsymbol{y}_1 - E(\Delta\epsilon' F_r \Delta \boldsymbol{y}_1)$ for suitable r as sums of MDS. The transformation enables us to write $\Omega^*(\theta_0)$, which is the variance of the outer product of thh sum of elements of a vector MDS, as the expected outer product of the elements of MDS because MDS has mean zero and the terms in the sum are independent (See (3.14) below). Then the averaged sum of the outer product of elements of the estimated vector MDS can be derived to be a consistent estimate of $\Omega^*(\theta_0)$.

For a square matrix $A = A^u + A^l + A^d$, let A^u, A^l and A^d be the upper-triangular, lower-triangular and diagonal matrix of A respectively. In the following we suppress the subscripts in R_r, O_r and F_r for suitable r to simplify notations. Let R_t be the $n\times k$ submatrix or $n\times 1$ subvector of R, where R could be a $n(T-1)\times K$ matrix (R_1) or $n(T-1)\times 1$ vector (R_2, R_3, R_4). Let O_{ts} and F_{ts} be the $n\times n$ submatrix of $n(T-1)\times n(T-1)$ matrix O and F respectively. Note R_t, O_{ts} and F_{ts} are partitioned by the $t,s=2,...,T$. Define $F_t^+ = \Sigma_{s=2}^T F_{ts}$, for $t=2,...,T$, $F_2^{++} = F_2^+ e^{-\alpha_{10}W_1}e^{\alpha_{30}W_3}$, $\Delta y_1^\diamond = e^{\alpha_3 W_3}e^{\alpha_{10}W_1}\Delta y_1$, $\Delta\xi = (F_2^{++u}+F_2^{++l})\Delta y_1^\diamond$, $\Delta\eta_t = \Sigma_{s=2}^T(O_{st}^{u'}+O_{ts}^l)\Delta\epsilon_s$, $\Delta\epsilon_t^* = \Sigma_{t=2}^T O_{ts}^d\Delta\epsilon_s$, d_{it} is the itth diagonal element of $\boldsymbol{BO}$ and $\Delta y_{1t}^* = F_t^+\Delta y_1$. Let $\{\Pi_{n,i}\}$ be the increasing sequence of σ-fields generated by $\{\epsilon_{j1},...,\epsilon_{jT}; j=1,...,i\}, i=1,...,n, n\geq 1$. Let $\Phi_{n,0}$ be the σ-field generated by $\{\epsilon_0,\Delta y_0\}$. Define $\Phi_{n,i} = \Phi_{n,0}\otimes\Pi_{n,i}$ as the σ-field on the Cartesian product generated by subset of the form $\phi_{n,0}\times\pi_{n,i}$, where $\phi_{n,0}\in\Phi_{n,0}$ and $\pi_{n,i}\in\Pi_{n,i}$. We show in the following lemma that $S^*(\theta_0)$ can be written as sums of vector MDS.

Lemma 3.2. Suppose the assumptions of Lemma 3.1 hold, define $a_{1i} = \Sigma_{t=2}^T R_{it}'\Delta\epsilon_{it}$, $a_{2i} = \Sigma_{t=2}^T(\Delta\epsilon_{it}\Delta\eta_{it} + \Delta\epsilon_{it}\Delta\epsilon_{it}^* - \sigma_{\epsilon0}^2 d_{it})$ and $a_{3i} = \Delta\epsilon_{2i}\Delta\xi_i + F_{2,ii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^\diamond + \sigma_{\epsilon0}^2) + \Sigma_{t=3}^T\Delta\epsilon_{it}\Delta y_{1it}^*$. Then

$$R'\Delta\epsilon = \Sigma_{i=1}^n a_{1i}, \tag{3.10}$$

$$\Delta\epsilon' O\Delta\epsilon - E(\Delta\epsilon' O\Delta\epsilon) = \Sigma_{i=1}^n a_{2i}, \tag{3.11}$$

$$\Delta\epsilon' F\Delta\boldsymbol{y}_1 - E(\Delta\epsilon' F\Delta\boldsymbol{y}_1) = \Sigma_{i=1}^n a_{3i}. \tag{3.12}$$

and $\{(a_{1i}', a_{2i}, a_{3i})', \Phi_{n,i}\}_{i=1}^n$ forms a vector MDS.

Now using Lemma 3.3, for each R_r, define $a_{1ri} = \Sigma_{t=2}^T R_{rit}'\Delta\epsilon_{it}$ for $r=1,2,3$ and 4; for each O_r, define $a_{2ri} = \Sigma_{t=2}^T(\Delta\epsilon_{it}\Delta\eta_{rit} + \Delta\epsilon_{it}\Delta\epsilon_{rit}^* - \sigma_{\epsilon0}^2 d_{rit})$ for $r=1,2,3,4$ and 5; for each F_r, define $a_{3ri} = \Sigma_{t=2}^T[\Delta\epsilon_{2i}\Delta\xi_{ri} + F_{2,rii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^\diamond + \sigma_{\epsilon0}^2) + \Sigma_{t=3}^T\Delta\epsilon_{it}\Delta y_{r1it}^*]$ for $r=1,2$ and 3. Then we can construct a vector $a_i = (a_{11i}', a_{21i}, a_{31i}+a_{12i}+a_{22i}, -a_{32i}-a_{13i}-a_{23i}, a_{33i}+a_{14i}+a_{24i}, a_{25i})'$. Here for the first element $E(R_1'\Delta\epsilon)=0$. For the second element $E(\Delta\epsilon' O_1\Delta\epsilon) = \frac{n(T-1)}{2\sigma_{\epsilon0}^2}$. For the third element $E(\Delta\epsilon' F_1\boldsymbol{y}_1 + R_2'\Delta\epsilon + \Delta\epsilon' O_2\Delta\epsilon) = -tr(\boldsymbol{D}_{-1,0}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_{10}W_1}})$. For the fourth element $E(\Delta\epsilon' F_2\boldsymbol{y}_1 + R_3'\Delta\epsilon + \Delta\epsilon' O_3\Delta\epsilon) = -tr(\boldsymbol{D}_0\boldsymbol{B}^{-1}\boldsymbol{W}_1)$. For the fifth element $E(\Delta\epsilon' F_3\Delta\boldsymbol{y}_1 + R_4'\Delta\epsilon + \Delta\epsilon' O_4\Delta\epsilon) = -tr(\boldsymbol{D}_{-1,0}\boldsymbol{B}^{-1}\boldsymbol{W}_2\boldsymbol{e}^{-\boldsymbol{\alpha_{10}W_1}})$. For the sixth element $E(\Delta\epsilon' O_5\Delta\epsilon)=0$. So

$$S^*(\theta_0) = \Sigma_{i=1}^n a_i. \tag{3.13}$$

Since $E(a_i|\Phi_{n,i-1})=0$, $\{a_i,\Phi_{n,i}\}$ form a vector MDS. Together with (3.13), we thus have

$$var[S^*(\theta_0)] = E[(\Sigma_{i=1}^n a_i)(\Sigma_{i=1}^n a_i)'] - [E(\Sigma_{i=1}^n a_i)][E(\Sigma_{i=1}^n a_i)]' = \Sigma_{i=1}^n E(a_i a_i') \tag{3.14}$$

A consistent estimator of $\Omega^*(\theta_0)$ is then given by $\widehat{\Omega}^* = \frac{1}{n(T-1)}\Sigma_{i=1}^n\hat{a}_i\hat{a}_i'$, where $\hat{a}_i$ is derived by replacing θ_0 in a_i by the M-estimator $\hat{\theta}_M$. The consistency of $\widehat{\Omega}^*$ and thus of the VC matrix $\Psi^{*-1}(\hat{\theta}_M)\widehat{\Omega}^*\Psi^{*-1}(\hat{\theta}_M)$ follows in the theorem below.

Theorem 3.3. Under the assumptions of Theorem 3.1, as $n\to\infty$,

$$\widehat{\Omega}^* - \Omega^*(\theta_0) = \frac{1}{n(T-1)}\left[\Sigma_{i=1}^n\hat{a}_i\hat{a}_i' - \Sigma_{i=1}^n a_i a_i'\right] \xrightarrow{p} 0 \tag{3.15}$$

and thus $\left[\Psi^{*-1}(\hat{\theta}_M)\widehat{\Omega}^*\Psi^{*-1}(\hat{\theta}_M) - \Psi^{*-1}(\theta_0)\Omega^*(\theta_0)\Psi^{*-1}(\theta_0)\right] \xrightarrow{p} 0.$ (3.16)

The M-estimator and OPMD estimator of the VC matrix subsume submodels that contain MESS in dependent variable, lagged dependent variable and/or disturbances. Their formats are derived in Appendix D. Different submodels are also explored in the Monte Carlo simulations in next section.

4 Monte Carlo Simulation

To fully investigate the performance of M-estimator and OPMD-based standard error, we establish the following models in the Monte Carlo simulation.

MESDPS(0,1,0): $y_t = \tau y_{t-1} + \alpha_2 W_2 y_{t-1} + \beta_0 l_n + X_t\beta_1 + Z\gamma + \mu + \epsilon_t$,

MESDPS(1,0,0): $e^{\alpha_1 W_1} y_t = \tau y_{t-1} + \beta_0 l_n + X_t\beta_1 + Z\gamma + \mu + \epsilon_t$,

MESDPS(1,1,0): $e^{\alpha_1 W_1} y_t = \tau y_{t-1} + \alpha_2 W_2 y_{t-1} + \beta_0 l_n + X_t\beta_1 + Z\gamma + \mu + \epsilon_t$,

MESDPS(0,1,1): $y_t = \tau y_{t-1} + \alpha_2 W_2 y_{t-1} + \beta_0 l_n + X_t\beta_1 + Z\gamma + \mu + u_t$, $e^{\alpha_3 W_3} u_t = \epsilon_t$,

MESDPS(1,0,1): $e^{\alpha_1 W_1} y_t = \tau y_{t-1} + \beta_0 l_n + X_t\beta_1 + Z\gamma + \mu + u_t$, $e^{\alpha_3 W_3} u_t = \epsilon_t$,

MESDPS(1,1,1): $e^{\alpha_1 W_1} y_t = \tau y_{t-1} + \alpha_2 W_2 y_{t-1} + \beta_0 l_n + X_t\beta_1 + Z\gamma + \mu + u_t$, $e^{\alpha_3 W_3} u_t = \epsilon_t$.

The elements of X_n is drawn from $N(0,4)$. Elements of Z and μ are drawn from $U(0,1)$ and $N(0,1)$ respectively. The spatial weight matrices are based on rook and queen contiguity. Three specifications of the disturbances ϵ_t are generated: (i) normal, (ii) normal mixture (10% $N(0,5^2)$ and 90% $N(0,1)$), (iii) standardized $\gamma(2,1)$. Both (ii) and (iii) are standardized to have the same variance with (i). Four sample sizes are considered, corresponding to $n = (49, 100)$ and $T = (3, 7)$.

The values of parameters are $\beta_0 = 10$, $\beta_1 = 1$, $\gamma = 1$ and $\sigma_\epsilon = 1$ or 2. For ρ and α_r, $r = 1, 2, 3$, we select from a set of values $(-1.5, -1.1, -0.5, -0.1, 0, 0.5, 1.1, 1.5)$ in different submodels. Each experiment is replicated 1000 times. To compare the performance of the OPMD estimator, we report the empirical standard deviations (sd), OPMD-based standard errors (**se**), standard errors based on $\widehat{\Omega}^{*-1}$ ($\tilde{se}$) and standard errors based on $\Psi^{*-1}(\hat{\theta}_M)$ ($\hat{se}$). Better performance is represented by closer approximation to sd.

Table 1a presents partial results for for empirical means of CQMLE and M-estimator and Table 1b presents the empirical standard deviations and standard errors for MESDPS(0,1,0). All tables in the main paper are partial and full results can be found in the Appendix. The results show that both the M-estimator and OPMD-based standard error exhibit excellent finite sample properties. For the empirical means in Table 1a, M-estimator provides almost unbiased estimates to the true coefficients in all three disturbance specifications while the CQMLE is biased. The bias of CQMLE are significant in many cases. When n grows larger to 100, the bias does not vanish. But when T grows bigger, the bias of CQMLE nearly disappears. On the other hand, the M-estimator is nearly unbiased for all coefficients for all n and T. The rational choice of n and T means that the M-estimator is useful in many real-world applications. It brings nearly unbiased results for studies with relatively small sample sizes. For the standard errors in Table 1b, the OPMD estimator has superior performance, exhibiting much closer approximation to the empirical sd than the other two candidates in most cases, especially when disturbance follows a gamma distribution. The OPMD estimator stays close to the empirical sd for all parameters under all n, T and σ^2_ϵ. Paying specific attention to τ under disturbance that follows gamma distribution, we find that the OPMD estimator gives especially better performance than the other two candidates of ses. This highlights the importance of conducting inference using the proposed OPMD estimator when the normality of the disturbance is in doubt. Overall the M-estimator and OPMD estimator provides good estimates and exhibits perfect finite sample property.

Table 2a and 2b present results for MESDPS(1,0,0). Similar conclusions can be made, i.e., the M-estimator and the OPMD estimator exhibit perfect finite sample properties. From Table 2a, we observe that he M-estimation gives nearly unbiased estimates while CQMLE is quite biased in many cases. The CQMLE again does not converge to the true value when n increases to 100 while the M-estimator remains nearly unbiased regardless of the size of n. From Table 2b we see the OPMD-based ses also has good performance. It is closer to the empirical sd in nearly most parameter combinations and provides especially better results for τ when the disturbance follows gamma distribution.

Table 3a and 3b contain partial results for MESDPS(1,1,0). We observe similar results. The M-estimator performs much better than the CQMLE when $T = 3$ in most cases. For α_1, CQMLE is almost as unbiased as M-estimator even when $T = 3$. For other coefficients, the M-estimator remains nearly unbiased for small and large T. For Table 3b, the OPMD method gives good estimates for ses, especially when disturbances are non-normal.

Table 4a and 4b present the results for MESDPS(1,0,1). It again shows that the proposed M-estimator and OPMD-based ses perform well. The M-estimator has smaller bias than CQMLE for all parameters, in many cases nearly unbiased. The OPMD-based ses better approximate the empirical standard deviations in most cases.

Table 5a and 5b present the estimation results fro MESDPS(0,1,1). The CQMLE has much smaller bias when T is increased from 3 to 7, but it remains biased when n is increased from 49 to 100. For α_3, the CQMLE and M-estimator are both biased by a similar magnitude when $T = 3$. Larger n and T erase the biases for both. For other parameters, M-estimator remains nearly perfect, even when n and T are small. The OPMD-based ses approximates the empirical sd well, especially when disturbance follows gamma distribution.

Table 6a and 6b present the results for the full model MESDPS(1,1,1). The M-estimator also provides better results than CQMLE in most cases. The bias is relatively big for α_3 when T is small even for M-estimator, but vanishes when n grows to 100 or T grows to 7. The OPMD-based se remains reliable and provides quite good estimates to the empirical sd.

5 Empirical Application to US Outward FDI

Foreign direct investment (FDI) has been discussed a lot in the literature. The FDI stock grows at a much faster rate than export and GDP. The multinational enterprises (MNEs) play important role in the process. Formal MNS theory was developed in the 1980s. Markusen (1984) and Helpman (1984) establish general equilibrium models with different motives for MNEs and coined the term "*horizontal*" and "*vertical*" FDI respectively. Horizontal FDI is for market access while vertical FDI is for access to cheaper factor inputs. A weakness of the two country framework is that it ignores third markets. Later work relax this assumption to develop more complicated models. Ekholm et al. (2007) and Yeaple (2003) propose a model in which the parent country invests in the host country to serve third market and is named "*export-platform*" FDI. Baltagi et al. (2007), on the other hand, argue MNEs can exploit local comparative advantages and set up plants in third markets. This motive is called "*complex vertical*".

Recent literature explore third market as a determinant of bilateral FDI. Coughlin et al. (1999) is the first paper to study FDI using spatial econometrics. They find a positive spatial lag (SL) and spatial error (SE) effect for China's inward FDI for neighboring regions. Baltagi et al. (2007) use the industries and countries FDI data to explore the knowledge-capital model of US outbound FDI using generalized moments (GM) estimators. They find that the spatial coefficients are significant while evidence of various modes of FDI emerges. Blonigen et al. (2007) study the US outward FDI by including spatial lag in the model and find that the estimates of the traditional determinants of FDI are robust to the inclusion of spatial lag. They find a positive and significant spatial lag using the whole sample which suggests complex-vertical motivations for MNE activity. Garretsen & Peeters (2009) apply a spatial lag model (SLM) and spatial error model (SEM) for Dutch FDI and find positive and significant spatial effects in both. Debarsy et al. (2015) utilizes a cross-sectional MESS model on Belgium's outward FDI and find evidence of pure vertical FDI. They argue that this is because Belgium has high production costs such as labor. In our study, the focus will be placed on the spatial coefficients since the dynamic nature of the model changes the situation in a significant way.

We explore the US outward FDI using MESDPS. Our data contains 40 countries from both developed and developing world over 7 years (2011-2017). The list of countries are listed in Table 7.

Table 7. List of Countries

Argentina	Australia	Belgium	Brazil	Canada	Chile	China
Cyprus	Czech	Denmark	Estonia	Finland	France	Germany
Hungary	India	Ireland	Italy	Japan	South Korea	Luxembourg
Malaysia	Mexico	Netherlands	New Zealand	Norway	Poland	Portugal
Romania	Russia	Singapore	South Africa	Spain	Sweden	Switzerland
Thailand	Turkey	Ukraine	United Kingdom	Vietnam		

The model to be estimated is a dynamic panel version of a modified gravity-type framework,

$$LFDI_{it} = \tau LFDI_{i,t-1} + \beta_1 LGDP_{it} + \beta_2 LPOP_{it} + \beta_3 LRISK_{it} + \beta_4 MP_{it} + u_{it}, \tag{3.17}$$

with corresponding spatial effects built in different models. Here $LFDI_{it}$ is the log of stock of outward FDI from US to host country i in year t. FDI are US outward positions (stocks) from International Direct Investment Statistics. The independent variables are a set of host country variables which includes log of GDP (LGDP), log of population (LPOP), log of an investment risk variable (LRISK), which is found to be important in the International Finance literature, and a surrounding-market potential variable (MP). As shown in Table 7, MP is an important characteristic to distinguish between export-platform and pure vertical FDI. We follow Garretsen & Peeters (2009) and compute it as the distance-weighted sums of other countries' GDP in the sample where the distance is the bilateral distance between capitals from Mayer & Zignago (2011). GDP and population data are extracted from the World Bank's World Development Indicators (WDI). Risk is the inverse of an investment profile index from International Country Risk Guide. Table 8 contains the summary statistics of these variables. The spatial weight matrix is an inverse arc-distance between capitals of host countries. Similar to Blonigen et al. (2007), we multiply the weights by the shortest distance between capitals (80.98 km between capitals of Estonia and Finland). The same spatial weight matrix will be applied to all spatial processes.

As discussed in section 2.2, there exists a relation between the spatial coefficients in STLE model and MESDPS[4], i.e., $\lambda_1 = 1 - e^{\alpha_1}$, $\rho = \tau + 1$ and $\lambda_2 = e^{\alpha_2} - 1$. The M-estimation results of corresponding models in Yang (2018) are thus also reported to highlight the relation in

[4] STLE specification is the comprehensive model which contains the spatial lag effect, dynamic effect, space-time effect and spatial error effect. It corresponds to our MESDPS(1,1,1). See section 2.2 for the detailed model specification.

Table 8. Descriptive Statistics

Variable	Mean	Std	Min	Max
Log of FDI ($millions)	10.09	1.97	4.09	13.75
Log of host country GDP(2010 constant dollars)	27.02	1.34	23.77	29.95
Log of host country population	17.05	1.64	13.16	21.05
Log of investment risk	-2.22	0.2	-2.48	-1.73
Surrounding market potential	25.66	2.07	22.5	27.16

interpretations of the two methods.

Table 10a and 10b summarize the estimation results. We run four specifications: SL model, MESDPS(1,0,0), STL model and MESDPS(1,1,0) for the full sample and the last 5 years. The SL and STL models are based on Yang (2018). All specifications contain an CQMLE and an M estimator. Table 10a contains results for SL and MESDPS(1,0,0) and 10b for STL and MESDPS(1,1,0).

We make three important observations. First we would like to emphasize the fact that the results perfectly capture the expected relation between coefficient estimates. For spatial coefficients, they satisfy the relation discussed above. In table 10a, the coefficient estimates for dynamic effects of CQMLE for SL is 0.3890 and for MESDPS(1,0,0) is -0.6074 for the full sample. For the M-estimator they are 0.6979 and -0.2978 respectively. They satisfy the relation $\rho = \tau + 1$. The similar situation is also found for the last-5-year sample. For W_1, which represents the spatial lag in SL model and MESS in MESDPS(1,0,0) for the dependent variables, we find that the signs of CQMLE and M-estimator of coefficients are positive and negative respectively. For CQMLE, the SL model has a coefficient of -0.3058 and MESDPS(1,0,0) has a coefficient of 0.2044. For M-estimators they are -0.3415 and 0.2154 respectively. These are in line with the relation $\lambda_1 = 1 - e^{\alpha_1}$. For table 10b, the similar situation is also found. The estimates for the dynamic effect and W_1 have the similar signs and magnitudes with those in table 10a and thus satisfy the expected relation. For W_2, we find that the coefficients are both positive. For the full sample the CQMLE are 0.1684 for STL and 0.1195 for MESDPS(1,1,0). Combined with their magnitudes, the expected relation $\lambda_2 = e^{\alpha_2} - 1$ holds. The similar situation also applies to the last-5-year sample. On the other hand, for the host variables, their coefficient estimates of CQMLE are similar for those in SL/STL and MESPDS(1,0,0)/(1,1,0) in the full sample and last-5-years sample. The similarity is also found for the M-estimator. Thus the results confirm our proposed relation between the coefficient estimates in the theory.

The second observation is that the inclusion of dynamic effects makes the coefficients of host country variables insignificant compared with the panel data case. In Blonigen et al. (2007) where the data from 1983 to 1998 are used, the signs for LGDP is positive and for LPOP and RISK are negative (see table 3 on p1315). The estimates are mostly significant in their study except MP variable. The insignificance of MP, combined with a positive coefficient estimate for spatially lag of LFDI, point to complex vertical FDI. In our study, however, adding in a lagged dependent variable changes the model estimates extensively. Although the estimates (except LPOP) have the same signs with those in Blonigen et al. (2007), they are no longer significant. The sign for spatial lag of LFDI stays significant but becomes negative. The significance of the coefficient estimates for $LFDI_{t-1}$ tells us that the inclusion of dynamic effect is an relatively important variable in explaining the variation in LFDI. The spatial terms are also significant in most cases.

The third observation is the difference between of the CQMLE and M-estimator. While in most cases they have same signs in respective groups, their magnitudes differ. For example, in table

Table 10a. Estimation results of US outbound log(FDI) for SL and MESDPS(1,0,0)

	Full sample				Last 5 years			
	(1) SL		(2) MEDPS(1,0,0)		(3) SL		(4) MEDPS(1,0,0)	
Variables	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
LGDP	0.7953	0.2510	0.7799	0.2325	0.2651	−0.2861	0.2240	−0.3909
		(0.445)		(0.673)		(0.441)		(0.995)
LPOP	1.4047	1.1872	1.4003	1.1824	1.6485	1.1407	1.6142	1.0465
		(1.077)		(0.950)		(1.1)		(1.301)
RISK	−0.1007	−0.1121	−0.0927	−0.1014	−0.0138	−0.0453	−0.0003	−0.0339
		(0.101)		(0.116)		(0.156)		(0.168)
MP	−0.1953	0.0187	−0.2877	−0.1075	−0.0608	0.3485	−0.0684	0.3729
		(1.004)		(0.850)		(0.940)		(1.029)
LFDI_{t-1}	0.3890	0.6979	−0.6074	−0.2978	0.2285	0.5371	−0.7579	−0.4140
		(0.246)***		(0.306)		(0.116)***		(0.509)
W_1	−0.3058	−0.3415	0.2044	0.2154	−0.4574	−0.4573	0.2898	0.2801
		(0.278)		(0.152)*		(0.166)***		(0.104)***

Table 10b. Estimation results of US outbound log(FDI) for STL and MESDPS(1,1,0)

	Full sample				Last 5 years			
	(5) STL		(6) MESDPS(1,1,0)		(7) STL		(8) MESDPS(1,1,0)	
Variables	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est	CQMLE	M-Est
LGDP	0.7843	0.2602	0.7767	0.2381	0.2336	−0.3787	0.2087	−0.4591
		(0.346)		(0.275)		(0.541)		(0.511)
LPOP	1.4665	1.2709	1.4572	1.2526	1.609	1.0292	1.5947	0.9613
		(0.972)*		(0.980)		(1.130)		(1.129)
RISK	−0.0965	−0.1032	−0.0895	−0.0959	−0.0034	−0.0191	0.0052	−0.0123
		(0.108)		(0.111)		(0.167)		(0.195)
MP	−0.607	−0.5238	−0.6047	−0.4934	−0.1416	0.1522	−0.1101	0.2051
		(0.623)		(0.559)		(0.908)		(0.961)
LFDI_{t-1}	0.4063	0.7049	−0.5977	−0.292	0.2433	0.5815	−0.7507	−0.3804
		(0.147)***		(0.077)***		(0.171)***		(0.161)***
W_1	−0.3508	−0.3674	0.2322	0.2378	−0.4650	−0.4597	0.2917	0.2796
		(0.104)***		(0.083)***		(0.101)***		(0.066)***
W_2	0.1684	0.1991	0.1195	0.1333	0.0695	0.1679	0.0341	0.1213
		(0.077)***		(0.056)***		(0.298)		(0.222)

OPMD standard errors are in parenthesis. W_1 and W_2 are spatial weight matrix in terms of SAR in SL and STL models and MESS in MESDPS(1,0,0) and MESDPS(1,1,0).

* Correspond to significance at 10%.

** Correspond to significance at 5%.

*** Correspond to significance at 1%.

10b, the estimate for $LGDP$ is 0.7843 for CQMLE in STL and 0.2602 for M-estimator in the full sample. This tells us that the M-estimator might correctly captures the impact of LGDP on LFDI. Although we do not have a reference in this field to examine its validity, the difference do tell us that we need to be careful in using the CQMLE which provide biased results.

6 Conclusion

In this paper we proposed a consistent way to estimate the matrix exponential spatial dynamic panel specification. To the best of our knowledge, this is the first paper to tackle this problem. The comprehensive model includes matrix exponential in the dependent variable, spatial effect in the lagged dependent variable and disturbance. We also propose an OPMD estimator for the VC matrix. Valid inference can be based on the standard error derived from the OPMD estimator, especially when the normality of the disturbance is in doubt. The method can be applied to submodels and works perfectly. The method is free from the initial condition specification and simple to use. It provides scholars a reliable way to conduct empirical research. Future research might focus on modifying the type of spatial processes in the model, for example to moving average.

References

Anselin, L. (2001), 'Spatial econometrics', *A companion to theoretical econometrics* **310330**.

Anselin, L. (2013), *Spatial econometrics: methods and models*, Vol. 4, Springer Science & Business Media.

Baltagi, B. H., Egger, P. & Pfaffermayr, M. (2007), 'Estimating models of complex fdi: Are there third-country effects?', *Journal of econometrics* **140**(1), 260–281.

Baltagi, B. H., Song, S. H. & Koh, W. (2003), 'Testing panel data regression models with spatial error correlation', *Journal of econometrics* **117**(1), 123–150.

Blonigen, B. A., Davies, R. B., Waddell, G. R. & Naughton, H. T. (2007), 'Fdi in space: Spatial autoregressive relationships in foreign direct investment', *European economic review* **51**(5), 1303–1325.

Chiu, T. Y., Leonard, T. & Tsui, K.-W. (1996), 'The matrix-logarithmic covariance model', *Journal of the American Statistical Association* **91**(433), 198–210.

Coughlin, C. C., Segev, E. et al. (1999), *Foreign direct investment in China: a spatial econometric study*, Research Department, Federal Reserve Bank of St. Louis.

Debarsy, N., Jin, F. & Lee, L.-F. (2015), 'Large sample properties of the matrix exponential spatial specification with an application to fdi', *Journal of Econometrics* **188**(1), 1–21.

Ekholm, K., Forslid, R. & Markusen, J. R. (2007), 'Export-platform foreign direct investment', *Journal of the European Economic Association* **5**(4), 776–795.

Elhorst, J. P. (2010), 'Dynamic panels with endogenous interaction effects when t is small', *Regional Science and Urban Economics* **40**(5), 272–282.

Figueiredo, C. & Da Silva, A. R. (2015), 'A matrix exponential spatial specification approach to panel data models', *Empirical Economics* **49**(1), 115–129.

Garretsen, H. & Peeters, J. (2009), 'Fdi and the relevance of spatial linkages: do third-country effects matter for dutch fdi?', *Review of World Economics* **145**(2), 319–338.

Helpman, E. (1984), 'A simple theory of international trade with multinational corporations', *Journal of political economy* **92**(3), 451–471.

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

Lee, L.-f. & Yu, J. (2010*b*), 'Some recent developments in spatial panel data models', *Regional Science and Urban Economics* **40**(5), 255–271.

Lee, L.-f. & Yu, J. (2010*c*), 'A spatial dynamic panel data model with both time and individual fixed effects', *Econometric Theory* **26**(2), 564–597.

Lee, L.-f. & Yu, J. (2013), Spatial panel data models, *in* 'The Oxford handbook of panel data'.

LeSage, J. & Chih, Y.-Y. (2018), 'A matrix exponential spatial panel model with heterogeneous coefficients', *Geographical Analysis* **50**(4), 422–453.

LeSage, J. & Pace, R. K. (2007), ‘A matrix exponential spatial specification’, *Journal of Econometrics* **140**(1), 190–214.

Markusen, J. R. (1984), ‘Multinationals, multi-plant economies, and the gains from trade’, *Journal of international economics* **16**(3-4), 205–226.

Mayer, T. & Zignago, S. (2011), Notes on cepii?s distances measures: The geodist database, Working Papers 2011-25, CEPII.

Mutl, J. (2006), Dynamic panel data models with spatially correlated disturbances, PhD thesis.

Su, L. & Yang, Z. (2015), ‘Qml estimation of dynamic panel data models with spatial errors’, *Journal of Econometrics* **185**(1), 230–258.

Van der Vaart, A. W. (2000), *Asymptotic statistics*, Vol. 3, Cambridge university press.

Xu, X. & Lee, L.-f. (2019), ‘Theoretical foundations for spatial econometric research’, *Regional Science and Urban Economics* **76**, 2–12.

Yang, Z. (2018), ‘Unified m-estimation of fixed-effects spatial dynamic models with short panels’, *Journal of econometrics* **205**(2), 423–447.

Yeaple, S. R. (2003), ‘The complex integration strategies of multinationals and cross country dependencies in the structure of foreign direct investment’, *Journal of international economics* **60**(2), 293–314.

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’, *Journal of Econometrics* **146**(1), 118–134.

Yu, J. & Lee, L.-f. (2010), ‘Estimation of unit root spatial dynamic panel data models’, *Econometric Theory* **26**(5), 1332–1362.

Zhang, Y., Feng, S. & Jin, F. (2019), ‘Qml estimation of the matrix exponential spatial specification panel data model with fixed effects and heteroskedasticity’, *Economics Letters* **180**, 1–5.

Table 1a. Empirical mean of CQMLE and M-estimator, MESDPS(0,1,0)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9537	0.9987	0.9565	1.0007	0.9934	1.0000	0.9936	1.0001
	1	0.9163	0.9607	0.9442	0.9883	0.9873	0.9942	0.9872	0.9937
	0.5	0.4245	0.5000	0.4269	0.5011	0.4948	0.5000	0.4951	0.5001
	−0.1	−0.0217	−0.1011	−0.0236	−0.1016	−0.0944	−0.1000	−0.0946	−0.1000
2	1	0.9529	0.9988	0.9560	1.0004	0.9947	1.0012	0.9941	1.0004
	1	0.9302	0.9756	0.9393	0.9828	0.9825	0.9892	0.9865	0.9930
	0.5	0.4263	0.5022	0.4265	0.5002	0.4951	0.5002	0.4950	0.5000
	−0.1	−0.0237	−0.1038	−0.0231	−0.1005	−0.0947	−0.1003	−0.0946	−0.1000
3	1	0.9565	1.0023	0.9558	0.9996	0.9916	0.9981	0.9950	1.0014
	1	0.9261	0.9731	0.9362	0.9801	0.9828	0.9896	0.9852	0.9917
	0.5	0.4267	0.5037	0.4266	0.5005	0.4951	0.5002	0.4951	0.5001
	−0.1	−0.0239	−0.1052	−0.0230	−0.1007	−0.0946	−0.1002	−0.0947	−0.1001
1	1	0.9469	0.9994	0.9465	0.9992	0.9837	0.9994	0.9848	1.0002
	1	0.9174	0.9700	0.9333	0.9853	0.9754	0.9912	0.9813	0.9968
	0	−0.1022	0.0010	−0.1026	0.0001	−0.0283	0.0008	−0.0286	0.0001
	−0.1	0.0044	−0.1029	0.0057	−0.1008	−0.0699	−0.1011	−0.0694	−0.1001
2	1	0.9481	1.0009	0.9473	0.9994	0.9835	0.9992	0.9851	1.0004
	1	0.9213	0.9743	0.9376	0.9896	0.9740	0.9897	0.9779	0.9931
	0	−0.1014	0.0023	−0.1033	−0.0010	−0.0286	0.0005	−0.0285	−0.0001
	−0.1	0.0036	−0.1044	0.0064	−0.0995	−0.0694	−0.1007	−0.0695	−0.1000
3	1	0.9490	1.0011	0.9460	0.9977	0.9853	1.0009	0.9842	0.9993
	1	0.9148	0.9675	0.9346	0.9870	0.9709	0.9866	0.9761	0.9913
	0	−0.1015	0.0009	−0.1013	0.0008	−0.0281	0.0008	−0.0272	0.0009
	−0.1	0.0044	−0.1019	0.0044	−0.1016	−0.0699	−0.1009	−0.0708	−0.1011
1	1	0.9523	0.9960	0.9555	0.9994	0.9898	0.9995	0.9908	1.0006
	1	0.9298	0.9733	0.9511	0.9952	0.9808	0.9904	0.9846	0.9942
	−0.5	−0.6050	−0.5026	−0.6011	−0.4982	−0.5390	−0.5011	−0.5384	−0.5002
	−0.1	0.0074	−0.0982	0.0040	−0.1023	−0.0594	−0.0991	−0.0595	−0.0996
2	1	0.9565	1.0002	0.9570	1.0007	0.9895	0.9994	0.9895	0.9992
	1	0.9345	0.9780	0.9450	0.9880	0.9781	0.9877	0.9855	0.9951
	−0.5	−0.6006	−0.4987	−0.6032	−0.5017	−0.5390	−0.5010	−0.5394	−0.5012
	−0.1	0.0027	−0.1028	0.0056	−0.0990	−0.0602	−0.1000	−0.0589	−0.0989
3	1	0.9523	0.9964	0.9570	1.0003	0.9907	1.0006	0.9914	1.0011
	1	0.9359	0.9811	0.9411	0.9841	0.9828	0.9926	0.9836	0.9932
	−0.5	−0.6035	−0.5005	−0.6008	−0.5000	−0.5389	−0.5007	−0.5374	−0.4996
	−0.1	0.0056	−0.1010	0.0030	−0.1012	−0.0589	−0.0989	−0.0611	−0.1008

Note: Disturbance 1=normal, 2=normal-mixture and 3=gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_2)'$. W_2 is generated by rook contiguity.

Table 1b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(0,1,0)

dis	par	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.056	.055	.060	.055	.039	.039	.040	.039	.030	.029	.032	.030	.021	.020	.021	.021
	1	.148	.143	.159	.144	.103	.103	.109	.103	.082	.082	.088	.083	.059	.057	.060	.058
	0.5	.040	.040	.041	.036	.027	.028	.027	.025	.004	.004	.004	.004	.003	.003	.003	.003
	−0.1	.045	.049	.042	.041	.030	.034	.027	.028	.004	.005	.004	.004	.003	.003	.003	.003
2	1	.058	.056	.060	.055	.040	.039	.040	.038	.031	.029	.032	.030	.021	.021	.021	.021
	1	.149	.145	.162	.146	.106	.103	.108	.103	.083	.081	.088	.082	.058	.057	.060	.058
	0.5	.040	.040	.041	.036	.027	.028	.027	.025	.004	.004	.004	.004	.003	.003	.003	.003
	−0.1	.044	.049	.042	.041	.030	.034	.027	.028	.004	.005	.004	.004	.003	.003	.003	.003
3	1	.055	.055	.062	.055	.040	.038	.041	.038	.029	.029	.032	.029	.021	.020	.022	.021
	1	.209	.188	.131	.146	.142	.137	.084	.103	.123	.118	.063	.082	.086	.085	.041	.058
	0.5	.042	.039	.042	.037	.028	.026	.027	.025	.004	.004	.004	.004	.003	.003	.003	.003
	−0.1	.045	.047	.044	.041	.031	.031	.029	.028	.004	.004	.004	.004	.003	.003	.003	.003
1	1	.057	.057	.061	.056	.041	.040	.041	.039	.030	.030	.032	.030	.022	.021	.022	.021
	1	.153	.148	.161	.147	.103	.105	.109	.104	.084	.082	.089	.083	.059	.058	.060	.059
	0	.052	.052	.052	.046	.037	.036	.035	.032	.015	.014	.015	.013	.010	.010	.010	.009
	−0.1	.059	.064	.052	.052	.041	.044	.035	.036	.017	.017	.015	.015	.011	.012	.010	.011
2	1	.057	.057	.061	.056	.041	.040	.041	.039	.031	.030	.032	.030	.022	.021	.022	.021
	1	.150	.147	.164	.147	.109	.105	.110	.105	.083	.081	.090	.083	.059	.058	.060	.058
	0	.052	.052	.052	.046	.036	.035	.035	.032	.015	.014	.015	.014	.010	.010	.010	.009
	−0.1	.058	.064	.053	.052	.040	.043	.035	.036	.017	.018	.015	.015	.011	.012	.010	.011
3	1	.059	.056	.063	.056	.040	.039	.042	.039	.031	.030	.033	.030	.021	.021	.022	.021
	1	.195	.188	.133	.146	.146	.138	.087	.104	.117	.117	.065	.083	.086	.085	.042	.058
	0	.053	.049	.053	.046	.038	.034	.035	.032	.014	.014	.015	.013	.010	.010	.010	.009
	−0.1	.058	.059	.055	.051	.040	.040	.036	.036	.015	.017	.016	.015	.011	.012	.010	.011
1	1	.058	.055	.060	.055	.042	.039	.041	.039	.031	.029	.032	.030	.021	.021	.022	.021
	1	.149	.143	.161	.146	.104	.104	.110	.104	.084	.081	.089	.083	.056	.057	.060	.058
	−0.5	.055	.052	.057	.050	.040	.037	.038	.035	.024	.022	.026	.023	.016	.015	.017	.016
	−0.1	.062	.065	.058	.056	.044	.046	.039	.040	.027	.027	.026	.026	.018	.020	.018	.018
2	1	.057	.055	.060	.055	.039	.039	.040	.038	.030	.029	.032	.030	.020	.021	.021	.021
	1	.147	.144	.162	.146	.107	.103	.108	.103	.081	.081	.089	.082	.057	.058	.060	.058
	−0.5	.053	.052	.057	.050	.037	.036	.038	.035	.023	.022	.026	.023	.017	.015	.017	.016
	−0.1	.060	.065	.058	.056	.042	.045	.039	.039	.027	.028	.027	.026	.019	.020	.018	.018
3	1	.058	.055	.062	.055	.041	.038	.041	.038	.030	.029	.033	.030	.021	.021	.022	.021
	1	.212	.191	.132	.147	.149	.137	.085	.103	.129	.118	.064	.083	.085	.086	.042	.058
	−0.5	.057	.053	.058	.050	.040	.035	.038	.035	.025	.022	.026	.023	.017	.015	.018	.016
	−0.1	.063	.064	.060	.056	.043	.043	.040	.039	.028	.027	.027	.026	.019	.019	.018	.018

Note: Same configuration as Table 1a. Here sd is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 2a. Empirical mean of CQMLE and M-estimator, MESDPS(1,0,0)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9659	0.9995	0.9676	1.0006	0.9963	1.0001	0.9963	1.0000
	1	0.9279	0.9619	0.9549	0.9886	0.9901	0.9941	0.9897	0.9934
	0.5	0.4566	0.5009	0.4571	0.5011	0.4966	0.4999	0.4970	0.4999
	1.1	1.0785	1.1019	1.0776	1.1006	1.1047	1.1003	1.1039	1.1002
2	1	0.9649	0.9995	0.9677	1.0009	0.9971	1.0009	0.9968	1.0004
	1	0.9419	0.9776	0.9505	0.9837	0.9845	0.9885	0.9896	0.9933
	0.5	0.4574	0.5029	0.4574	0.5009	0.4966	0.5000	0.4970	0.4999
	1.1	1.0795	1.1037	1.0778	1.1005	1.1046	1.1002	1.1040	1.1002
3	1	0.9684	1.0024	0.9674	1.0002	0.9943	0.9980	0.9978	1.0013
	1	0.9372	0.9733	0.9480	0.9814	0.9852	0.9892	0.9880	0.9916
	0.5	0.4579	0.5032	0.4578	0.5013	0.4967	0.5001	0.4971	0.5000
	1.1	1.0789	1.1030	1.0775	1.1002	1.1044	1.1000	1.1037	1.1000
1	1	0.9539	1.0004	0.9528	0.9995	0.9924	0.9992	0.9938	1.0004
	1	0.9231	0.9723	0.9389	0.9864	0.9837	0.9910	0.9904	0.9973
	0	−0.0710	0.0026	−0.0717	0.0009	−0.0101	0.0000	−0.0099	0.0001
	1.1	1.0390	1.1041	1.0371	1.1011	1.1100	1.1022	1.1098	1.1008
2	1	0.9541	1.0012	0.9539	1.0004	0.9922	0.9990	0.9939	1.0006
	1	0.9250	0.9749	0.9436	0.9914	0.9828	0.9899	0.9867	0.9935
	0	−0.0711	0.0029	−0.0718	0.0008	−0.0099	0.0001	−0.0100	0.0000
	1.1	1.0386	1.1041	1.0367	1.1007	1.1093	1.1013	1.1099	1.1010
3	1	0.9548	1.0019	0.9522	0.9985	0.9941	1.0008	0.9926	0.9992
	1	0.9198	0.9693	0.9403	0.9885	0.9795	0.9865	0.9840	0.9909
	0	−0.0708	0.0028	−0.0708	0.0018	−0.0094	0.0006	−0.0096	0.0004
	1.1	1.0359	1.1011	1.0378	1.1019	1.1083	1.1003	1.1091	1.1005
1	1	0.9573	0.9982	0.9598	1.0003	0.9884	0.9996	0.9896	1.0005
	1	0.9351	0.9782	0.9547	0.9973	0.9793	0.9905	0.9832	0.9941
	−0.5	−0.5783	−0.4976	−0.5772	−0.4964	−0.5341	−0.5003	−0.5336	−0.5001
	1.1	0.9868	1.1042	0.9884	1.1061	1.0683	1.1011	1.0666	1.0994
2	1	0.9604	1.0014	0.9611	1.0012	0.9880	0.9995	0.9881	0.9990
	1	0.9382	0.9820	0.9481	0.9898	0.9763	0.9875	0.9839	0.9949
	−0.5	−0.5775	−0.4958	−0.5797	−0.4997	−0.5346	−0.5009	−0.5347	−0.5012
	1.1	0.9892	1.1086	0.9848	1.1016	1.0709	1.1036	1.0660	1.0988
3	1	0.9565	0.9972	0.9605	1.0000	0.9896	1.0009	0.9901	1.0010
	1	0.9389	0.9824	0.9427	0.9839	0.9819	0.9933	0.9821	0.9931
	−0.5	−0.5785	−0.4983	−0.5787	−0.5000	−0.5332	−0.4996	−0.5333	−0.5000
	1.1	0.9869	1.1037	0.9878	1.1024	1.0672	1.0998	1.0692	1.1016

Note: Disturbance 1=normal, 2=normal-mixture and 3=Gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1)'$. W_1 is generated by queen contiguity.

Table 2b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,0,0)

dis	par	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.055	.068	.059	.054	.037	.047	.040	.038	.030	.030	.032	.029	.020	.021	.021	.021
	1	.147	.153	.157	.141	.101	.112	.108	.102	.081	.112	.088	.091	.058	.072	.060	.062
	0.5	.028	.034	.027	.025	.019	.024	.018	.018	.004	.005	.004	.004	.002	.003	.002	.002
	1.1	.026	.027	.026	.024	.018	.019	.018	.017	.006	.008	.006	.006	.004	.004	.004	.004
2	1	.056	.069	.059	.054	.039	.048	.039	.038	.030	.031	.031	.029	.021	.021	.021	.021
	1	.150	.156	.160	.143	.105	.111	.107	.101	.084	.111	.088	.090	.057	.072	.060	.062
	0.5	.028	.034	.028	.025	.019	.024	.018	.018	.004	.005	.004	.004	.002	.003	.002	.002
	1.1	.026	.027	.026	.025	.018	.019	.018	.017	.006	.008	.006	.006	.004	.005	.004	.004
3	1	.054	.069	.060	.054	.039	.048	.040	.038	.029	.030	.032	.029	.021	.021	.021	.021
	1	.210	.205	.128	.143	.142	.150	.083	.101	.123	.156	.063	.090	.086	.104	.041	.062
	0.5	.031	.036	.027	.025	.020	.026	.018	.018	.004	.006	.004	.004	.002	.003	.002	.002
	1.1	.024	.026	.029	.025	.017	.018	.019	.017	.006	.009	.006	.006	.004	.005	.004	.004
1	1	.056	.075	.060	.055	.039	.053	.040	.039	.030	.031	.032	.030	.022	.022	.021	.021
	1	.153	.164	.160	.144	.104	.117	.109	.102	.083	.093	.088	.085	.059	.065	.060	.060
	0	.042	.051	.038	.035	.028	.036	.026	.025	.009	.010	.009	.009	.006	.007	.006	.006
	1.1	.051	.055	.049	.046	.036	.039	.033	.032	.020	.023	.020	.020	.014	.015	.013	.014
2	1	.058	.076	.060	.055	.040	.053	.040	.039	.030	.031	.032	.030	.021	.022	.021	.021
	1	.152	.164	.163	.144	.108	.117	.109	.102	.083	.092	.089	.085	.058	.065	.060	.059
	0	.043	.052	.038	.035	.029	.036	.026	.025	.008	.010	.009	.009	.006	.007	.006	.006
	1.1	.053	.056	.049	.046	.035	.039	.033	.032	.019	.023	.020	.020	.013	.015	.013	.014
3	1	.058	.077	.062	.055	.041	.053	.041	.039	.031	.031	.032	.029	.021	.022	.021	.021
	1	.195	.212	.131	.143	.147	.157	.085	.102	.116	.131	.063	.085	.085	.095	.041	.059
	0	.043	.056	.038	.035	.030	.039	.025	.025	.009	.011	.009	.009	.006	.008	.006	.006
	1.1	.046	.054	.054	.045	.033	.038	.036	.032	.020	.024	.020	.020	.014	.016	.013	.014
1	1	.058	.073	.060	.055	.042	.051	.040	.039	.031	.032	.032	.030	.021	.023	.022	.021
	1	.150	.162	.162	.144	.105	.118	.110	.103	.084	.085	.089	.083	.056	.060	.060	.058
	-0.5	.049	.059	.045	.042	.034	.042	.031	.029	.021	.021	.021	.019	.014	.015	.014	.013
	1.1	.101	.107	.097	.090	.069	.076	.065	.063	.052	.051	.052	.049	.035	.036	.035	.035
2	1	.056	.073	.060	.055	.040	.051	.040	.038	.030	.032	.032	.030	.020	.023	.022	.021
	1	.150	.163	.162	.145	.107	.117	.108	.102	.081	.084	.089	.082	.057	.060	.060	.058
	-0.5	.048	.059	.046	.042	.033	.042	.031	.029	.021	.021	.021	.019	.014	.015	.014	.013
	1.1	.099	.107	.097	.090	.070	.076	.065	.063	.052	.051	.051	.049	.036	.036	.035	.035
3	1	.058	.075	.062	.055	.040	.051	.041	.038	.030	.032	.033	.030	.021	.023	.022	.021
	1	.212	.216	.131	.145	.149	.155	.084	.102	.129	.123	.064	.083	.085	.089	.042	.058
	-0.5	.051	.064	.045	.042	.035	.045	.029	.029	.021	.022	.021	.019	.014	.016	.014	.013
	1.1	.100	.107	.103	.090	.066	.075	.068	.063	.049	.050	.053	.049	.035	.036	.035	.035

Note: Same configuration as Table 2a. Here sd is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 3a. Empirical mean of CQMLE and M-estimator, MESDPS(1,1,0)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9865	0.9977	0.9878	0.9991	0.9981	0.9995	0.9990	1.0005
	1	0.9544	0.9661	0.9784	0.9900	0.9860	0.9875	0.9913	0.9928
	−1.5	−1.5307	−1.5001	−1.5295	−1.4988	−1.5098	−1.4998	−1.5095	−1.4994
	1.1	1.1007	1.1031	1.1013	1.1027	1.0953	1.0986	1.0982	1.1017
	−0.5	−0.4636	−0.5023	−0.4662	−0.5038	−0.4863	−0.4982	−0.4879	−0.5002
2	1	0.9892	1.0003	0.9902	1.0013	0.9980	0.9995	0.9978	0.9993
	1	0.9568	0.9685	0.9728	0.9840	0.9830	0.9845	0.9923	0.9939
	−1.5	−1.5285	−1.4978	−1.5302	−1.5001	−1.5096	−1.4995	−1.5102	−1.5001
	1.1	1.1012	1.1027	1.0985	1.1002	1.0958	1.0992	1.0958	1.0991
	−0.5	−0.4664	−0.5038	−0.4613	−0.4981	−0.4870	−0.4989	−0.4869	−0.4990
3	1	0.9854	0.9964	0.9896	1.0006	0.9990	1.0006	0.9995	1.0010
	1	0.9597	0.9718	0.9694	0.9808	0.9878	0.9894	0.9900	0.9915
	−1.5	−1.5313	−1.5007	−1.5301	−1.5000	−1.5089	−1.4987	−1.5093	−1.4993
	1.1	1.1006	1.1028	1.0993	1.1010	1.0991	1.1025	1.0962	1.0996
	−0.5	−0.4651	−0.5036	−0.4668	−0.5041	−0.4863	−0.4983	−0.4892	−0.5013

Note: Disturbance 1=normal, 2=normal-mixture and 3=Gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1, \alpha_2)'$. W_1 and W_2 are generated by rook and queen contiguity respectively.

Table 3b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,1,0)

dis	par	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.054	.051	.057	.052	.038	.036	.038	.036	.030	.029	.032	.029	.021	.020	.021	.021
	1	.141	.135	.158	.140	.100	.099	.106	.100	.084	.080	.089	.082	.056	.057	.060	.057
	−1.5	.037	.034	.040	.035	.026	.024	.027	.025	.019	.018	.021	.019	.013	.013	.014	.013
	1.1	.073	.071	.079	.072	.050	.051	.053	.051	.043	.043	.047	.043	.030	.031	.032	.031
	−0.5	.079	.077	.084	.077	.056	.055	.057	.055	.041	.040	.044	.041	.028	.029	.030	.029
2	1	.052	.051	.057	.052	.037	.036	.038	.036	.029	.029	.032	.029	.020	.020	.021	.021
	1	.141	.136	.157	.140	.103	.098	.105	.100	.080	.080	.089	.081	.057	.057	.060	.058
	−1.5	.036	.034	.040	.035	.025	.024	.027	.025	.019	.018	.021	.019	.014	.013	.014	.013
	1.1	.073	.071	.079	.072	.052	.050	.053	.051	.043	.043	.047	.043	.031	.030	.032	.031
	−0.5	.077	.075	.083	.076	.056	.055	.056	.054	.042	.041	.044	.041	.029	.029	.030	.029
3	1	.054	.051	.060	.052	.037	.036	.039	.036	.029	.029	.033	.029	.021	.020	.021	.020
	1	.206	.183	.127	.141	.145	.134	.081	.099	.128	.117	.064	.082	.085	.085	.041	.057
	−1.5	.036	.034	.042	.035	.025	.024	.027	.025	.019	.018	.021	.019	.013	.013	.014	.013
	1.1	.074	.071	.082	.072	.051	.050	.054	.051	.045	.043	.048	.043	.031	.030	.032	.031
	−0.5	.080	.076	.086	.076	.054	.054	.058	.054	.040	.040	.045	.041	.029	.029	.030	.029

Note: Same configuration as Table 3a. Here *sd* is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 4a. Empirical mean of CQMLE and M-estimator, MESDPS(1,0,1)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9769	0.9971	0.9793	0.9997	0.9941	0.9992	0.9956	1.0005
	1	0.9374	0.9634	0.9629	0.9898	0.9809	0.9868	0.9866	0.9926
	−0.5	−0.5608	−0.5012	−0.5589	−0.4982	−0.5241	−0.5010	−0.5235	−0.5000
	0.5	0.4398	0.5050	0.4382	0.5067	0.4915	0.5027	0.4866	0.4998
	−1.1	−1.0764	−1.1122	−1.0687	−1.1082	−1.0999	−1.1071	−1.0901	−1.0987
2	1	0.9796	0.9996	0.9802	1.0003	0.9933	0.9984	0.9948	0.9997
	1	0.9419	0.9682	0.9569	0.9832	0.9783	0.9842	0.9875	0.9935
	−0.5	−0.5580	−0.4983	−0.5611	−0.5012	−0.5243	−0.5011	−0.5245	−0.5010
	0.5	0.4455	0.5117	0.4356	0.5035	0.4960	0.5077	0.4856	0.4986
	−1.1	−1.0694	−1.1073	−1.0672	−1.1070	−1.0990	−1.1069	−1.0923	−1.1008
3	1	0.9760	0.9963	0.9788	0.9989	0.9956	1.0007	0.9956	1.0004
	1	0.9436	0.9707	0.9527	0.9790	0.9836	0.9897	0.9853	0.9913
	−0.5	−0.5606	−0.5004	−0.5600	−0.5004	−0.5233	−0.5000	−0.5233	−0.5000
	0.5	0.4389	0.5067	0.4389	0.5070	0.4900	0.5016	0.4900	0.5026
	−1.1	−1.0599	−1.0995	−1.0665	−1.1079	−1.0978	−1.1050	−1.0938	−1.1020

Note: Disturbance 1=normal, 2=normal-mixture and 3=Gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1, \alpha_3)'$. W_1 and W_3 are generated by rook and queen contiguity respectively.

Table 4b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,0,1)

dis	par	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.044	.042	.046	.042	.031	.029	.031	.029	.024	.024	.026	.024	.017	.017	.017	.017
	1	.144	.137	.159	.141	.102	.100	.108	.102	.084	.081	.089	.082	.056	.057	.060	.058
	−0.5	.041	.037	.041	.037	.029	.026	.028	.027	.019	.017	.019	.018	.013	.012	.013	.013
	0.5	.121	.116	.123	.115	.084	.082	.083	.081	.068	.069	.071	.068	.048	.049	.049	.048
	−1.1	.159	.153	.166	.152	.110	.105	.110	.105	.087	.085	.094	.086	.059	.060	.063	.061
2	1	.044	.042	.046	.042	.029	.029	.031	.029	.025	.024	.026	.024	.016	.017	.017	.017
	1	.144	.138	.159	.142	.105	.100	.107	.101	.080	.080	.089	.082	.057	.058	.060	.058
	−0.5	.040	.037	.042	.037	.028	.026	.028	.026	.018	.017	.020	.018	.013	.012	.013	.013
	0.5	.116	.116	.123	.114	.085	.082	.083	.081	.069	.068	.072	.068	.048	.049	.049	.048
	−1.1	.158	.154	.166	.152	.109	.105	.110	.105	.085	.085	.093	.086	.061	.060	.063	.060
3	1	.044	.043	.048	.042	.029	.030	.031	.029	.024	.023	.026	.024	.017	.017	.017	.017
	1	.207	.184	.129	.142	.147	.134	.083	.100	.129	.118	.064	.082	.085	.085	.042	.058
	−0.5	.042	.039	.041	.037	.030	.027	.027	.026	.019	.017	.020	.018	.013	.012	.013	.013
	0.5	.122	.112	.131	.114	.083	.080	.086	.080	.067	.067	.074	.067	.048	.049	.049	.048
	−1.1	.152	.152	.174	.152	.106	.103	.115	.105	.085	.084	.097	.086	.062	.059	.064	.060

Note: Same configuration as Table 4a. Here *sd* is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 5a. Empirical mean of CQMLE and M-estimator, MESDPS(0,1,1)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9915	0.9983	0.9930	0.9999	0.9985	0.9994	0.9996	1.0005
	1	0.9585	0.9650	0.9821	0.9888	0.9862	0.9872	0.9916	0.9926
	−1.5	−1.5282	−1.5000	−1.5264	−1.4979	−1.5097	−1.5005	−1.5090	−1.4996
	0.5	0.5108	0.5011	0.5095	0.4994	0.4993	0.4999	0.5000	0.5002
	−1.1	−1.0753	−1.0769	−1.0883	−1.0891	−1.0945	−1.0947	−1.0931	−1.0932
2	1	0.9936	1.0003	0.9945	1.0012	0.9988	0.9997	0.9985	0.9993
	1	0.9599	0.9665	0.9775	0.9840	0.9830	0.9840	0.9928	0.9938
	−1.5	−1.5253	−1.4967	−1.5284	−1.5004	−1.5098	−1.5006	−1.5095	−1.5002
	0.5	0.5085	0.4983	0.5100	0.5002	0.4983	0.4987	0.4993	0.4996
	−1.1	−1.0554	−1.0563	−1.0884	−1.0889	−1.0886	−1.0889	−1.0988	−1.0990
3	1	0.9894	0.9961	0.9934	1.0001	0.9996	1.0005	0.9999	1.0008
	1	0.9624	0.9691	0.9731	0.9797	0.9884	0.9894	0.9904	0.9914
	−1.5	−1.5299	−1.5014	−1.5275	−1.4994	−1.5090	−1.4998	−1.5087	−1.4994
	0.5	0.5111	0.5009	0.5087	0.4987	0.5007	0.5011	0.4985	0.4988
	−1.1	−1.0554	−1.0563	−1.0844	−1.0848	−1.0892	−1.0894	−1.0954	−1.0956

Note: Disturbance 1=normal, 2=normal-mixture and 3=Gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_2, \alpha_3)'$. W_2 and W_3 are generated by rook and queen contiguity respectively.

Table 5b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(0,1,1)

dis	par	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.047	.044	.050	.045	.033	.032	.034	.032	.027	.025	.028	.026	.018	.018	.019	.018
	1	.140	.134	.157	.139	.099	.098	.106	.100	.084	.080	.088	.082	.056	.057	.060	.057
	−1.5	.037	.035	.040	.036	.027	.025	.027	.025	.018	.018	.020	.018	.013	.013	.013	.013
	0.5	.033	.032	.035	.032	.023	.023	.024	.023	.022	.021	.023	.022	.015	.015	.016	.015
	−1.1	.198	.199	.214	.197	.142	.139	.143	.138	.111	.109	.119	.110	.076	.078	.081	.078
2	1	.046	.045	.050	.045	.032	.031	.033	.032	.026	.025	.028	.026	.017	.018	.019	.018
	1	.141	.135	.156	.139	.102	.098	.105	.099	.080	.080	.089	.081	.057	.057	.060	.057
	−1.5	.038	.035	.041	.036	.025	.024	.027	.025	.018	.018	.020	.018	.013	.013	.013	.013
	0.5	.033	.032	.035	.032	.023	.022	.024	.023	.022	.021	.023	.021	.015	.015	.016	.015
	−1.1	.202	.197	.213	.196	.140	.138	.144	.138	.111	.109	.119	.110	.078	.078	.081	.078
3	1	.048	.044	.052	.045	.032	.031	.034	.032	.025	.025	.029	.026	.018	.018	.019	.018
	1	.204	.182	.125	.139	.145	.133	.080	.099	.128	.117	.064	.082	.085	.085	.041	.057
	−1.5	.037	.035	.042	.036	.025	.025	.027	.025	.018	.018	.020	.018	.013	.013	.013	.013
	0.5	.034	.032	.036	.032	.023	.023	.024	.023	.022	.021	.024	.021	.015	.015	.016	.015
	−1.1	.204	.194	.224	.196	.136	.136	.148	.137	.111	.107	.122	.110	.077	.077	.083	.078

Note: Same configuration as Table 5a. Here *sd* is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 6a. Empirical mean of CQMLE and M-estimator, MESDPS(1,1,1)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9360	0.9979	0.9392	1.0001	0.9839	1.0003	0.9840	1.0004
	1	0.8965	0.9568	0.9269	0.9868	0.9702	0.9845	0.9767	0.9909
	−0.5	−0.6370	−0.4984	−0.6328	−0.4961	−0.5502	−0.4997	−0.5506	−0.5002
	1.1	1.0286	1.1000	1.0337	1.1031	1.0815	1.1013	1.0805	1.1000
	1.1	1.0893	1.1006	1.0913	1.1019	1.1016	1.1012	1.1009	1.1002
	1.1	1.1984	1.1643	1.1591	1.1287	1.1277	1.1169	1.1235	1.1131
2	1	0.9392	1.0009	0.9409	1.0013	0.9826	0.9994	0.9831	0.9994
	1	0.8995	0.9604	0.9209	0.9798	0.9666	0.9808	0.9778	0.9919
	−0.5	−0.6339	−0.4945	−0.6354	−0.4998	−0.5509	−0.5004	−0.5510	−0.5007
	1.1	1.0290	1.1001	1.0333	1.1024	1.0796	1.0996	1.0802	1.0997
	1.1	1.0882	1.0990	1.0916	1.1022	1.1002	1.0998	1.1007	1.1001
	1.1	1.2135	1.1844	1.1568	1.1275	1.1332	1.1226	1.1175	1.1074
3	1	0.9357	0.9969	0.9400	1.0001	0.9841	1.0010	0.9851	1.0013
	1	0.8995	0.9607	0.9172	0.9759	0.9717	0.9862	0.9764	0.9905
	−0.5	−0.6364	−0.4982	−0.6329	−0.4985	−0.5509	−0.5001	−0.5489	−0.4988
	1.1	1.0241	1.0944	1.0302	1.0985	1.0807	1.1009	1.0815	1.1007
	1.1	1.0856	1.0957	1.0883	1.0986	1.1012	1.1010	1.1010	1.1003
	1.1	1.2181	1.1889	1.1647	1.1360	1.1320	1.1209	1.1202	1.1102

Note: Disturbance 1=normal, 2=normal-mixture and 3=Gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1, \alpha_2, \alpha_3)'$. W_1, W_2 and W_3 are generated by queen, rook and queen contiguity respectively.

Table 6b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,1,1)

dis	par	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.064	.063	.067	.062	.045	.044	.044	.043	.033	.032	.034	.032	.022	.022	.023	.022
	1	.152	.147	.167	.148	.106	.107	.113	.107	.085	.082	.091	.083	.056	.058	.061	.058
	−0.5	.070	.062	.068	.061	.048	.043	.045	.043	.027	.026	.028	.026	.019	.018	.019	.018
	1.1	.095	.100	.099	.094	.066	.069	.065	.065	.048	.048	.049	.046	.033	.034	.033	.033
	1.1	.075	.080	.080	.076	.053	.056	.053	.053	.041	.041	.042	.040	.028	.029	.028	.028
	1.1	.231	.229	.240	.221	.161	.158	.158	.153	.121	.121	.130	.120	.084	.086	.088	.085
2	1	.063	.063	.067	.061	.043	.044	.044	.043	.033	.032	.034	.032	.022	.022	.023	.022
	1	.153	.147	.168	.149	.108	.106	.112	.106	.082	.082	.091	.083	.057	.059	.061	.058
	−0.5	.070	.062	.068	.061	.047	.043	.044	.042	.028	.025	.029	.026	.019	.018	.019	.018
	1.1	.093	.100	.099	.094	.063	.068	.065	.065	.048	.048	.049	.046	.033	.034	.033	.033
	1.1	.075	.081	.080	.076	.051	.055	.052	.052	.041	.041	.042	.040	.028	.029	.028	.028
	1.1	.235	.228	.239	.220	.156	.157	.158	.153	.125	.121	.130	.120	.085	.086	.088	.085
3	1	.064	.062	.070	.061	.044	.044	.045	.043	.032	.032	.035	.032	.023	.022	.023	.022
	1	.209	.192	.141	.149	.151	.140	.090	.105	.129	.118	.067	.083	.086	.086	.043	.058
	−0.5	.072	.063	.070	.061	.049	.043	.045	.042	.029	.025	.029	.026	.019	.018	.019	.018
	1.1	.097	.100	.104	.095	.063	.068	.067	.065	.047	.048	.050	.046	.032	.034	.033	.033
	1.1	.078	.081	.084	.076	.050	.055	.054	.052	.040	.041	.043	.040	.027	.029	.029	.028
	1.1	.236	.227	.252	.221	.151	.155	.163	.153	.121	.119	.134	.120	.085	.085	.090	.085

Note: Same configuration as Table 6a. Here sd is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Appendix

Ye Yang

1 Appendix A. Some Lemmas

In the following, Lemmas A.1 can be found in Kelejian & Prucha (1999). Lemma A.2 can be found in, e.g., Debarsy et al. (2015), Lee (2004). Lemma A.3 can be found in, e.g., Yang (2015) and Yang (2018). Lemma A.4, a central limit theorem for bilinear quadratic forms, can be found in Yang (2018). The proofs are contained in these papers and thus are omitted. Let UB stand for "bounded in both row and column sum norms".

Lemma A.1. Suppose that $n \times n$ matrices $\{A_n\}$ and $\{B_n\}$ are UB and C_n is a sequence of conformable matrices whose elements are uniformly $O(g_n^{-1})$. Then
(i) the sequence $\{A_n B_n\}$ are UB.
(ii) the elements of A_n are uniformly bounded and $tr(A_n) = O(n)$, and
(iii) the elements of $A_n C_n$ and $C_n A_n$ are uniformly $O(g_n^{-1})$.

Lemma A.2. Suppose that elements of $n \times k$ matrix X_n are uniformly bounded and $lim_{n\to\infty} n^{-1} X_n' X_n$ exists and is nonsingular, then $P_n = X_n(X_n'X_n)^{-1}X_n'$ and $M_n = I_n - P_n$ are UB.

Lemma A.3. Suppose that $n \times n$ matrices $\{A_n\}$ are uniformly bounded in either row or column sums norm and the elements $a_{n,ij}$ of A_n are $O(g_n^{-1})$ uniformly in all i and j. Also suppose that ϵ_n is an $n \times 1$ random vector of i.i.d. elements with mean zero, variance σ^2 and finite 4th moment and b_n is an $n \times 1$ vector with constant elements of uniform order $O(g_n^{-1/2})$. Then

(i) $E(\epsilon_n' A_n \epsilon_n) = O(\frac{n}{g_n})$; (ii) $Var(\epsilon_n' A_n \epsilon_n) = O(\frac{n}{g_n})$; (iii) $Var(\epsilon_n' A_n \epsilon_n + b_n'\epsilon_n) = O(\frac{n}{g_n})$; (iv) $\epsilon_n' A_n \epsilon_n = O_p(\frac{n}{g_n})$; (v) $\epsilon_n' A_n \epsilon_n - E(\epsilon_n' A_n \epsilon_n) = O_p[(\frac{n}{g_n})^{1/2}]$; (vi) $\epsilon_n' A_n b_n = O_p[(\frac{n}{g_n})^{1/2}]$; (vii) The results in (iii) and (vi) remain valid if b_n is an $n \times 1$ random vector independent of ϵ_n such that $\{E(\epsilon_n^2)\}$ are of uniform order $O(g_n^{-1})$.

Lemma A.4. Suppose that $n \times n$ matrices $\{A_n\}$ is UB with elements of uniform order $O(g_n^{-1})$. Suppose $\{\epsilon_n\}$ is a $n \times 1$ random vector of i.i.d. elements with mean zero, variance σ_ϵ^2 and finite $(4+2\nu_0)$th moment for some $\nu_0 > 0$. Suppose an $n \times 1$ random vector $b_n = \{b_{ni}\}$ is independent of ϵ_n and satisfies the following conditions (i) $\{E(b_{ni}^2)\}$ are of uniform order $O(g_n^{-1})$, (ii) $sup_i E|b_{ni}|^{2+\nu_0} < \infty$, (iii) $\frac{g_n}{n}\Sigma_{i=1}^n [A_{n,ii}(b_{ni} - Eb_{ni})] = o_p(1)$ where $\{A_{n,ii}\}$ are the diagonal elements of A_n, (iv) $\frac{g_n}{n}\Sigma_{i=1}^n [b_{ni} - E(b_{ni}^2)] = o_p(1)$. Define the bilinear-quadratic form as $C_n = b_n'\epsilon_n + \epsilon_n' A_n \epsilon_n - \sigma_\epsilon^2 tr(A_n)$ with variance $\sigma_{C_n}^2$. If $lim_{n\to\infty} g_n^{1+2/\nu_0}/n = 0$ and $\{\frac{g_n}{n}\sigma_{C_n}^2\}$ are bounded away from zero, then $C_n/\sigma_{C_n}^2 \xrightarrow{d} N(0,1)$.

2 Appendix B. Proofs of Lemmas 2.1, 3.1 and 3.2

Proof of Lemma 2.1. First note the reduced form of ΔY is given by $\Delta Y = \boldsymbol{e^{-\alpha_{10} W_1}} \boldsymbol{A}_{20} \Delta Y_{-1} + \boldsymbol{e^{-\alpha_{10} W_1}} \Delta X \beta_0 + \boldsymbol{e^{-\alpha_{10} W_1}} \boldsymbol{e^{-\alpha_{30} W_3}} \Delta\epsilon$. For each element of ΔY, we have:

(1)$E(\Delta y_{t-1}\Delta\epsilon_t') = E[(e^{-\alpha_{10}W_1}A_{20}\Delta y_{t-2} + e^{-\alpha_{10}W_1}\Delta X_{t-1}\beta_0 + e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}\Delta\epsilon_{t-1})\Delta\epsilon_t']$
$=-\sigma_{\epsilon 0}^2 e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}$;

(2) $E(\Delta y_t\Delta\epsilon_t') = E[(e^{-\alpha_{10}W_1}A_{20}\Delta y_{t-1} + e^{-\alpha_{10}W_1}\Delta X_t\beta_0 + e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}\Delta\epsilon_t)]\Delta\epsilon_t'$
$= -\sigma_{\epsilon 0}^2 e^{-\alpha_{10}W_1}(A_{20}e^{-\alpha_{10}W_1} - 2I_n)e^{\alpha_{30}W_3}$;

(3) $E(\Delta y_{t+1}\Delta\epsilon_t') = E[(e^{-\alpha_{10}W_1}A_{20}\Delta y_t + e^{-\alpha_{10}W_1}\Delta X_{t+1}\beta_0 + e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}\Delta\epsilon_{t+1})\Delta\epsilon_t']$
$=-\sigma_{\epsilon 0}^2 e^{-\alpha_{10}W_1}(A_{20}e^{-\alpha_{10}W_1} - I_n)^2 e^{-\alpha_{30}W_3}$;

(4) For $t \geq s+1$ and $s \geq 2$, $E(\Delta y_t\Delta\epsilon_s') = -\sigma_{\epsilon 0}^2 e^{-\alpha_{10}W_1}(A_{20}e^{-\alpha_{10}W_1})^{t-(s+1)}(A_{20}e^{-\alpha_{10}W_1} - \mathrm{I}_n)^2 e^{-\alpha_{30}W_3}$;

(5) All the remaining terms $E(\Delta y_t\Delta\epsilon_{t+2}') = 0$ for $t \geq 1$.

Combining all elements, we have

$$E(\Delta Y_{-1}\Delta\epsilon') = E\left[\begin{pmatrix}\Delta y_1\\ \vdots\\ \Delta y_{T-1}\end{pmatrix} \times \begin{pmatrix}\Delta\epsilon_2 & \dots & \Delta\epsilon_T\end{pmatrix}\right] = -\sigma_{\epsilon 0}^2 \boldsymbol{e^{-\alpha_{10}W_1}D}_{-1,0}\boldsymbol{e^{-\alpha_{30}W_3}}$$

$$\text{and } \mathrm{E}(\Delta Y\Delta\epsilon') = E\left[\begin{pmatrix}\Delta y_2\\ \vdots\\ \Delta y_T\end{pmatrix} \times \begin{pmatrix}\Delta\epsilon_2 & \dots & \Delta\epsilon_T\end{pmatrix}\right] = -\sigma_{\epsilon 0}^2 \boldsymbol{e^{-\alpha_{10}W_1}D}_0\boldsymbol{e^{-\alpha_{30}W_3}}.$$

Proof of Lemma 3.1. Recall $A_{12,0} = e^{-\alpha_{10}W_1}A_{20}$. By the reduced form of Δy_t and continuous substitution, we have:

$$\begin{aligned}\Delta y_t &= A_{12,0}\Delta y_{t-1} + e^{-\alpha_{10}W_1}\Delta X_t\beta_0 + e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}\Delta\epsilon_t\\ &= \mathrm{A}_{12,0}^{t-1}\Delta y_1 + \Sigma_{i=0}^{t-2}A_{12,0}^i e^{-\alpha_{10}W_1}\Delta X_{t-i}\beta_0 + \Sigma_{i=0}^{t-2}A_{12,0}^i e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}\Delta\epsilon_{t-i}\\ &= \mathrm{A}_{12,0}^{t-1}\Delta y_1 + [A_{12,0}^{t-2}\ \ A_{12,0}^{t-3}\ \ \dots\ \ I_n\ \ 0\ \ \dots\ \ 0]\boldsymbol{e^{-\alpha_{10}W_1}}\Delta X\beta_0\\ &\quad + [A_{12,0}^{t-2}\ \ A_{12,0}^{t-3}\ \ \dots\ \ I_n\ \ 0\ \ \dots\ \ 0]\boldsymbol{e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}}\Delta\epsilon.\end{aligned}$$

Stacking them in one column we have:

$$\Delta Y = G\Delta\boldsymbol{y}_1 + J\boldsymbol{e^{-\alpha_{10}W_1}}\Delta X\beta_0 + J\boldsymbol{e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}}\Delta\epsilon = G\Delta\boldsymbol{y}_1 + \boldsymbol{\delta} + K\Delta\epsilon.$$

Similarly for ΔY_{-1} we have:

$$\Delta Y_{-1} = G_{-1}\Delta\boldsymbol{y}_1 + J_{-1}\boldsymbol{e^{-\alpha_{10}W_1}}\Delta X\beta_0 + J_{-1}\boldsymbol{e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}}\Delta\epsilon = G_{-1}\Delta\boldsymbol{y}_1 + \boldsymbol{\delta}_{-1} + K_{-1}\Delta\epsilon.$$

Proof of Lemma 3.2. First note $R'\Delta\epsilon = \Sigma_{i=1}^T R_t'\Delta\epsilon_t = \Sigma_{i=1}^n\Sigma_{t=2}^T R_{it}'\Delta\epsilon_{it} = \Sigma_{i=1}^n a_{1i}$. Here we partition $R'\Delta\epsilon$ by time periods in the first equality and then by individuals in the second equality.

Second $E(\Delta\epsilon' O\Delta\epsilon) = E[tr(\Delta\epsilon' O\Delta\epsilon)] = tr[E(\Delta\epsilon'\Delta\epsilon)O] = \sigma_{\epsilon 0}^2 tr(\boldsymbol{B}O) = \sigma_{\epsilon 0}^2\Sigma_{i=1}^n\Sigma_{t=2}^T d_{it}$, where d_{it} is the itth diagonal element of $\boldsymbol{B}O$. So we have:

$$\begin{aligned}\Delta\epsilon' O\Delta\epsilon - E(\Delta\epsilon' O\Delta\epsilon) &= \Sigma_{t=2}^T\Sigma_{s=2}^T\Delta\epsilon_t'(O_{ts}^u + O_{ts}^l + O_{ts}^d)\Delta\epsilon_s - \sigma_{\epsilon 0}^2\Sigma_{i=1}^n\Sigma_{t=2}^T d_{it}\\ &= \Sigma_{t=2}^T\Sigma_{s=2}^T[\Delta\epsilon_s' O_{ts}^u{}'\Delta\epsilon_t + \Delta\epsilon_t'(O_{ts}^l + O_{ts}^d)\Delta\epsilon_s] - \sigma_{\epsilon 0}^2\Sigma_{t=2}^T\Sigma_{s=2}^T d_{it}\\ &= \Sigma_{i=1}^n\Sigma_{t=2}^T(\Delta\epsilon_{it}\Delta\eta_{it} + \Delta\epsilon_{it}\Delta\epsilon_{it}^* - \sigma_{\epsilon 0}^2 d_{it})\\ &= \Sigma_{i=1}^n a_{2i},\end{aligned}$$

where $\Delta\eta_t = \Sigma_{s=2}^T(O_{st}^u{}' + O_{ts}^l)\Delta\epsilon_s$ and $\Delta\epsilon_t^* = \Sigma_{s=2}^T O_{ts}^d\Delta\epsilon_s$. Here we change the way $\Delta\epsilon' O\Delta\epsilon$ is partitioned from by time periods to by individual and time periods.

Third we first write $E(\Delta\epsilon' F\Delta\boldsymbol{y}_1)$ as following:

$$\begin{aligned}E(\Delta\epsilon' F\Delta\boldsymbol{y}_1) &= \Sigma_{t=2}^T\Delta\epsilon_t' F_t^+\Delta y_1 = \Delta\epsilon_2' F_2^+\Delta y_1 + \Sigma_{t=3}^T\Delta\epsilon_t' F_t^+\Delta y_1\\ &= \Delta\epsilon_2' F_2^+ e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}e^{\alpha_{30}W_3}e^{\alpha_{10}W_1}\Delta y_1 + \Sigma_{t=3}^T\Delta\epsilon_t' F_t^+\Delta y_1\end{aligned}$$

$$=\Delta\epsilon_2' F_2^{++}\Delta y_1^{\diamond} + \Sigma_{t=3}^T \Delta\epsilon_t' \Delta y_{1t}^*,$$

where $F_2^{++} = F_2^{+} e^{-\alpha_{10}W_1} e^{-\alpha_{30}W_3}$, $\Delta y_1^{\diamond} = e^{\alpha_{30}W_3} e^{\alpha_{10}W_1}\Delta y_1$ and $\Delta y_{1t}^* = F_t^{+}\Delta y_1$. Also note $\Delta y_1^{\diamond} = e^{\alpha_{30}W_3} e^{\alpha_{20}W_2}\Delta y_0 + e^{\alpha_{30}W_3}\Delta X_1\beta_0 + \Delta\epsilon_1$. By Assumption 1, Δy_0 is independent of ϵ_t for $t \geq 1$. So $E(\Delta\epsilon_2' F_2^{++}\Delta y_1^{\diamond}) = E[\epsilon_2' F_2^{++}\Delta\epsilon_1] = -\sigma_{\epsilon 0}^2 tr(F_2^{++})$, which leads to the following:

$$\begin{aligned}\Delta\epsilon_2' F_2^{++}\Delta y_1^{\diamond} - E(\Delta\epsilon_2' F_2^{++}\Delta y_1^{\diamond}) &= \Delta\epsilon_2'(F_2^{++u} + F_2^{++l})\Delta y_1^{\diamond} + \Delta\epsilon_2' F_2^{++d}\Delta y_1^{\diamond} + \sigma_{\epsilon 0}^2 tr(F_2^{++}) \\ &= \Sigma_{i=1}^n \Delta\epsilon_{2i}\Delta\xi_i + \Sigma_{i=1}^n F_{2,ii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond} + \sigma_{\epsilon 0}^2),\end{aligned}$$

where $\Delta\xi = (F_2^{++u} + F_2^{++l})\Delta y_1^{\diamond}$. Combining the equations above, we get the following:

$$\begin{aligned}\Delta\epsilon' F\Delta\boldsymbol{y}_1 - E(\Delta\epsilon' F\Delta\boldsymbol{y}_1) &= \Delta\epsilon_2' F_2^{++}\Delta y_1^{\diamond} - E(\Delta\epsilon_2' F_2^{++}\Delta y_1^{\diamond}) + \Sigma_{t=3}^T\Delta\epsilon_t'\Delta y_{1t}^* - E(\Sigma_{t=3}^T\Delta\epsilon_t'\Delta y_{1t}^*) \\ &= \Sigma_{i=1}^n \Delta\epsilon_{2i}\Delta\xi_i + \Sigma_{i=1}^n F_{2,ii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond} + \sigma_{\epsilon 0}^2) + \Sigma_{t=3}^T\Delta\epsilon_t'\Delta y_{1t}^* \\ &= \Sigma_{i=1}^n a_{3i},\end{aligned}$$

where $a_{3i} = \Delta\epsilon_{2i}\Delta\xi_i + F_{2,ii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond} + \sigma_{\epsilon 0}^2) + \Sigma_{t=3}^T\Delta\epsilon_{it}'\Delta y_{1it}^*$. Note $E(\Sigma_{t=3}^T\Delta\epsilon_t'\Delta y_{1t}^*) = 0$ according to Assumption 1.

Because $E[(a_{1i}', a_{2i}, a_{3i})|\Phi_{n,i-1}] = 0$, where $\Phi_{n,i-1} = \Phi_{n,0}\otimes\Pi_{n,i-1}$ is the Cartesian product generated by subsets of $X_1 \times X_2$, where $X_1 \in \Phi_{n,0}$ and $X_2 \in \Pi_{n,i-1}$, $\{(a_{1i}', a_{2i}, a_{3i}), \Phi_{n,i}\}$ form a vector MDS.

3 Appendix C. Proofs of Theorems 3.1-3.3

Proof of Theorem 3.1. Given Assumption 7, we need to prove $sup_{\zeta\in\mathcal{A}}||S^{*c}(\zeta) - \bar{S}^{*c}(\zeta)|| \xrightarrow{p} 0$. Note $S^{*c}(\zeta)$ has four elements by (2.18). So

$$S^{*c}(\zeta) - \bar{S}^{*c}(\zeta) = \begin{cases} \frac{1}{\hat{\sigma}_{\epsilon,M}^2(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1} - \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\zeta)}E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1}], \\ -\frac{1}{\hat{\sigma}_{\epsilon,M}^2(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y + \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\zeta)}E\big[\Delta\bar{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y\big], \\ \frac{1}{\hat{\sigma}_{\epsilon,M}^2(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} - \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\zeta)}E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}], \\ -\frac{1}{2}\big\{\frac{1}{\hat{\sigma}_{\epsilon,M}^2(\zeta)}\Delta\hat{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\hat{u}(\zeta) - \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\zeta)}E\big[\Delta\bar{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\bar{u}(\zeta)\big]\big\} \end{cases}$$

Now each function can be written as, while neglecting $\frac{1}{2}$, $[\frac{\pm(\bar{\sigma}_{\epsilon,M}^2(\zeta) - \hat{\sigma}_{\epsilon,M}^2(\zeta))}{\bar{\sigma}_{\epsilon,M}^2(\zeta)\hat{\sigma}_{\epsilon,M}^2(\zeta)}]f[\Delta\hat{u}(\zeta)] + \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\zeta)}\{f[\Delta\hat{u}(\zeta)] - Ef[\Delta\bar{u}(\zeta)]\}$, where $f[\Delta\hat{u}(\zeta)]$ and $f[\Delta\bar{u}(\zeta)]$ are the functions of $\Delta\hat{u}(\zeta)$ and $\Delta\bar{u}(\zeta)$ respectively. To show that these functions are $o_p(1)$, we need to prove the following:

(i) $inf_{\zeta\in\mathcal{Z}}\bar{\sigma}_{\epsilon,M}^2(\zeta) > c > 0$ for some positive number c;

(ii) $sup_{\zeta\in\mathcal{Z}}|\hat{\sigma}_{\epsilon,M}^2(\zeta) - \bar{\sigma}_{\epsilon,M}^2(\zeta)| = o_p(1)$;

(iii) $sup_{\zeta\in\mathcal{Z}}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1} - E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1}]\big| = o_p(1)$;

(iv) $sup_{\zeta\in\mathcal{Z}}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y - E\big[\Delta\bar{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y\big]\big| = o_p(1)$;

(v) $sup_{\zeta\in\mathcal{Z}}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} - E[\Delta\bar{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}]\big| = o_p(1)$;

(vi) $sup_{\zeta\in\mathcal{Z}}\frac{1}{n(T-1)}\big|\Delta\hat{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\hat{u}(\zeta) - \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\zeta)}E\big[\Delta\bar{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\bar{u}(\zeta)\big]\big| = o_p(1)$.

Proof of (i): Utilizing (3.5) $\bar{\sigma}_{\epsilon,M}^2(\zeta)$ can be expressed as:

$$\begin{aligned}\bar{\sigma}_{\epsilon,M}^2(\zeta) &= \frac{1}{n(T-1)}E\big[\Delta\bar{u}^*(\zeta)'\Delta\bar{u}^*(\zeta)\big] \\ &= \frac{1}{n(T-1)}E\big[(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y^{\dagger} - \boldsymbol{A}_2^*\Delta Y_{-1}^{\dagger})'P(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y^{\dagger} - \boldsymbol{A}_2^*\Delta Y_{-1}^{\dagger}) \\ &\quad + (e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y - \boldsymbol{A}_2^*\Delta Y_{-1})'M(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y - \boldsymbol{A}_2^*\Delta Y_{-1})\big] \\ &= \frac{1}{n(T-1)}E\big\{tr[(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y^{\dagger} - \boldsymbol{A}_2^*\Delta Y_{-1}^{\dagger})'(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y^{\dagger} - \boldsymbol{A}_2^*\Delta Y_{-1}^{\dagger})]\big\} \\ &\quad + \frac{1}{n(T-1)}E\{[e^{\boldsymbol{\alpha_1 W_1^*}}E(\Delta Y) - \boldsymbol{A}_2^*E(\Delta Y_{-1})]'M(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y - \boldsymbol{A}_2^*\Delta Y_{-1})\} \\ &\quad + \frac{1}{n(T-1)}E\{(e^{\boldsymbol{\alpha_1 W_1^*}}\Delta Y^{\dagger} - \boldsymbol{A}_2^*\Delta Y_{-1}^{\dagger})'M[e^{\boldsymbol{\alpha_1 W_1^*}}E(\Delta Y^{\dagger}) - \boldsymbol{A}_2^*E(\Delta Y_{-1}^{\dagger})]\}\end{aligned}$$

$$= \frac{1}{n(T-1)} tr[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})]$$
$$+ \frac{1}{n(T-1)} [\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} E(\Delta Y) - \boldsymbol{A}_2^* E(\Delta Y_{-1})]' M [\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} E(\Delta Y) - \boldsymbol{A}_2^* E(\Delta Y_{-1})],$$

where we used $E(\Delta Y^\dagger) = E(\Delta Y_{-1}^\dagger) = 0$ in the last equality.

For the first term, note

$$\begin{aligned}
&\frac{1}{n(T-1)} tr[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})] \\
&= \frac{1}{n(T-1)} tr[\Sigma_3^{-1} var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}_2 \Delta Y_{-1})] \\
&= \frac{1}{n(T-1)} tr[(B^{-1} \otimes e^{\alpha_3 W_3'} e^{\alpha_3 W_3}) var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}_2 \Delta Y_{-1})] \\
&\geq \frac{1}{n(T-1)} \gamma_{min}(B^{-1}) \gamma_{min}(e^{\alpha_3 W_3'} e^{\alpha_3 W_3}) tr[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}_2 \Delta Y_{-1})],
\end{aligned}$$

where $\gamma_{min}(B^{-1}) > 0$ given the structure of B, $\gamma_{min}(e^{\alpha_3 W_3'} e^{\alpha_3 W_3}) > 0$ by Assumption 4 and $tr[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}_2 \Delta Y_{-1})] > 0$ by the assumptions of the theorem. So $\frac{1}{n(T-1)} tr[var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})] > 0$.

For the second term, note M is positive semi-definite, so $\frac{1}{n(T-1)} [\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} E(\Delta Y) - \boldsymbol{A}_2^* E(\Delta Y_{-1})]' M [\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} E(\Delta Y) - \boldsymbol{A}_2^* E(\Delta Y_{-1})] \geq 0$ uniformly in $\zeta \in \mathcal{Z}$. So (i) holds.

Proof of (ii): We first express $\Delta \hat{u}^*(\zeta)$ as $\Delta \hat{u}^*(\zeta) = \Sigma^{-\frac{1}{2}} \Delta \hat{u}(\zeta) = \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1} - P(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1}) = M(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})$. So $\hat{\sigma}^2_{\epsilon,M}(\zeta) = \frac{1}{n(T-1)} \Delta \hat{u}^{*'}(\zeta) \Delta \hat{u}^*(\zeta) = \frac{1}{n(T-1)} (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})' M (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})$. Utilizing the expression in the proof of (i) for $\overline{\sigma}^2_{\epsilon,M}(\zeta)$ leads to the following:

$$\begin{aligned}
\hat{\sigma}^2_{\epsilon,M}(\zeta) - \overline{\sigma}^2_{\epsilon,M}(\zeta) &= \frac{1}{n(T-1)} (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})' M (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1}) \\
&\quad - \frac{1}{n(T-1)} E\big[(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y^\dagger - \boldsymbol{A}_2^* \Delta Y_{-1}^\dagger)' P (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y^\dagger - \boldsymbol{A}_2^* \Delta Y_{-1}^\dagger) \\
&\quad + (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})' M (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y - \boldsymbol{A}_2^* \Delta Y_{-1})\big] \\
&= \frac{1}{n(T-1)} [N_1 - E(N_1)] - \frac{2}{n(T-1)} [N_2 - E(N_2)] + \frac{1}{n(T-1)} [N_3 - E(N_3)] + E(N_4),
\end{aligned}$$

where $N_1 = \Delta Y' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y$, $N_2 = \Delta Y' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* \Delta Y_{-1}$, $N_3 = \Delta Y_{-1}' \boldsymbol{A}_2^{*\prime} M \boldsymbol{A}_2^* \Delta Y_{-1}$ and $N_4 = (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y^\dagger - \boldsymbol{A}_2^* \Delta Y_{-1}^\dagger)' P (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \Delta Y^\dagger - \boldsymbol{A}_2^* \Delta Y_{-1}^\dagger)$. We need to prove $N_r - E(N_r) \xrightarrow{p} 0$ uniformly in $\zeta \in \mathcal{Z}$ for $r = 1, 2, 3$ and $E[N_4(\zeta)] \longrightarrow 0$ uniformly in $\zeta \in \mathcal{Z}$.

To prove $N_r - E(N_r) \xrightarrow{p} 0$ uniformly in $\zeta \in \mathcal{Z}$ for $r = 1, 2, 3$, we need to prove the pointwise convergence of $N_r - E(N_r)$ in each $\zeta \in \mathcal{Z}$ and the stochastic equicontinuity of N_r.

Proof of pointwise convergence: By Lemma 3.1, we can express N_r's for $r = 1, 2$ and 3 as a function of $\Delta \boldsymbol{y}_1$, $\boldsymbol{\delta}$ and $\Delta\epsilon$ as follows:

$$\begin{aligned}
N_1 &= \Delta \boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} G \Delta \boldsymbol{y}_1 + \boldsymbol{\delta}' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \boldsymbol{\delta} + \Delta\epsilon' K' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} K \Delta\epsilon \\
&\quad + 2\Delta \boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} \boldsymbol{\delta} + 2\boldsymbol{\delta}' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} K \Delta\epsilon + 2\Delta \boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}} K \Delta\epsilon,
\end{aligned}$$

$$\begin{aligned}
N_2 &= \Delta \boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* G_{-1} \Delta \boldsymbol{y}_1 + \Delta \boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* \boldsymbol{\delta}_{-1} + \Delta \boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* K_{-1} \Delta\epsilon \\
&\quad + \boldsymbol{\delta}' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* G_{-1} \Delta \boldsymbol{y}_1 + \boldsymbol{\delta}' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* \boldsymbol{\delta}_{-1} + \boldsymbol{\delta}' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* K_{-1} \Delta\epsilon \\
&\quad + \Delta\epsilon' K' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* G_{-1} \Delta \boldsymbol{y}_1 + \Delta\epsilon' K' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* \boldsymbol{\delta}_{-1} + \Delta\epsilon' K' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} M \boldsymbol{A}_2^* K_{-1} \Delta\epsilon,
\end{aligned}$$

$$\begin{aligned}
N_3 &= \Delta \boldsymbol{y}_1' G_{-1}' \boldsymbol{A}_2^* M \boldsymbol{A}_2^* G_{-1} \Delta \boldsymbol{y}_1 + \boldsymbol{\delta}_{-1}' \boldsymbol{A}_2^{*\prime} M \boldsymbol{A}_2^* \boldsymbol{\delta}_{-1} + \Delta\epsilon' K_{-1}' \boldsymbol{A}_2^{*\prime} M \boldsymbol{A}_2^* K_{-1} \Delta\epsilon \\
&\quad + 2\Delta \boldsymbol{y}_1' G_{-1}' \boldsymbol{A}_2^{*\prime} M \boldsymbol{A}_2^* \boldsymbol{\delta}_{-1} + 2\Delta \boldsymbol{y}_1' G_{-1}' \boldsymbol{A}_2^* M \boldsymbol{A}_2^* K_{-1} \Delta\epsilon + 2\boldsymbol{\delta}_{-1}' \boldsymbol{A}_2^* M \boldsymbol{A}_2^* K_{-1} \Delta\epsilon.
\end{aligned}$$

Denote $N_1 = \Sigma_{q=1}^6 N_{1,q}$, $N_2 = \Sigma_{q=1}^9 N_{2,q}$ and $N_3 = \Sigma_{q=1}^6 N_{3,q}$, where each q denotes the corresponding terms in N_1, N_2 and N_3. We can prove that each element satisfies $N_{r,q} - E[N_{r,q}] = o_p(1)$ for all r and q. First note that $N_{1,2} - E(N_{1,2}) = 0$, $N_{2,5} - E(N_{2,5}) = 0$ and $N_{3,2} - E(N_{3,2}) = 0$ because they are nonstochastic. For the rest of the terms, we group them into five categories:

(A) $\Delta \boldsymbol{y}_1' C_1 \Delta \boldsymbol{y}_1 : N_{1,1}, N_{2,1}$ and $N_{3,1}$;
(B) $\Delta\epsilon' C_2 \Delta\epsilon : N_{1,3}, N_{2,9}$ and $N_{3,3}$;
(C) $\Delta \boldsymbol{y}_1' c_3 : N_{1,4}, N_{2,2}, N_{2,4}$ and $N_{3,4}$;
(D) $\Delta \boldsymbol{y}_1' C_4 \Delta\epsilon : N_{1,6}, N_{2,3}, N_{2,7}$ and $N_{3,5}$;

(E) $\Delta\epsilon' c_5 : N_{1,5}, N_{2,6}, N_{2,8}$ and $N_{3,6}$,

where C_1, C_2 and C_4 are $n(T-1)\times n(T-1)$ nonstochastic matrices and c_3 and c_5 are $n(T-1)\times 1$ nonstochastic vectors.

For (A), we can write $\frac{1}{n(T-1)}\Delta\boldsymbol{y}_1' C_1 \Delta\boldsymbol{y}_1 = \frac{1}{n}\Delta\boldsymbol{y}_1' C_1^* \Delta\boldsymbol{y}_1$, where $C_1^* = \frac{1}{T-1}\Sigma_s\Sigma_t C_{1,st}$, where $C_{1,st}$ are functions of the true parameters α_{10}, α_{30} and unknown parameters α_1, α_3. By Assumption 4, Lemma A.1 and Lemma A.2 it is uniformly bounded in row or column sums. Hence $\frac{1}{n(T-1)}[\Delta\boldsymbol{y}_1' C_1\Delta\boldsymbol{y}_1 - E(\Delta\boldsymbol{y}_1' C_1 \Delta\boldsymbol{y}_1)] = \frac{1}{n}[\Delta\boldsymbol{y}_1' C_1^*\Delta\boldsymbol{y}_1 - E(\Delta\boldsymbol{y}_1' C_1^*\Delta\boldsymbol{y}_1)]$ is pointwise convergent by Assumption 6(iii).

For (B), we can write $\frac{1}{n(T-1)}\Delta\epsilon' C_2\Delta\epsilon = \frac{1}{T-1}\Sigma_s\Sigma_t \frac{1}{n}\epsilon' C_{2,st}\epsilon$. By Lemma A.3(v), $\frac{1}{n}[\epsilon' C_{2,st}\epsilon - E(\epsilon' C_{2,st}\epsilon)]$ is pointwise convergent for each s and t.

For (C), the pointwise convergence of $\frac{1}{n(T-1)}[\Delta\boldsymbol{y}_1' c_3 - E(\Delta\boldsymbol{y}_1' c_3)]$ follows from Assumption 6(ii).

For (D), we can write $\Delta\boldsymbol{y}_1' C_4\Delta\epsilon = \Sigma_s \Delta y_1 C_{4,s}^*\Delta\epsilon_s$ and the pointwise convergence follows from Lemma A.3(vii) and Assumption 6(iv).

For (E), we can write $\Delta\epsilon' c_5 = \Sigma_s\Delta\epsilon_s c_{5,s}$. Note $E(\Delta\epsilon_s c_{5,s}) = 0$. By Chebyshev's inequality, $\Delta\epsilon_s c_{5,s}$ is pointwise convergent for each s.

Proof of stochastic equicontinuity: Denote each $N_{r,q}$ for $r = 1, 2$ and 3 by $N_{r,q}(\zeta)$. Then for any two parameter vectors $\zeta_1 \in \mathcal{Z}$ and $\zeta_2 \in \mathcal{Z}$, we have by mean value theorem: $N_{r,q}(\zeta_1) - N_{r,q}(\zeta_2) = \frac{\partial}{\partial\zeta'}N_{r,q}(\bar{\zeta})(\zeta_1 - \zeta_2)$, where $\bar{\zeta}$ is between ζ_1 and ζ_2 elementwise. We can prove each of $\frac{\partial}{\partial\zeta'}N_{r,q}(\zeta)$ is $O_p(1)$ for the five categories above. For example for $N_{1,1}(\zeta)$ we have:

$$
\begin{aligned}
sup_{\zeta\in\mathcal{Z}}\left|\frac{1}{n(T-1)}\frac{\partial N_{1,1}(\zeta)}{\partial\alpha_1}\right| &= sup_{\zeta\in\mathcal{A}}\left|\frac{2}{n(T-1)}\Delta\boldsymbol{y}_1' G' \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1'}}\boldsymbol{W}_1'\Sigma^{-\frac{1}{2}}M\Sigma^{-\frac{1}{2}}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}G\Delta\boldsymbol{y}_1\right| \\
&\leq \gamma_{max}(\boldsymbol{W}_1\Sigma^{-1})\gamma_{max}(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1'}}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}})\frac{2}{n(T-1)}|\Delta\boldsymbol{y}_1' G' G_1\Delta\boldsymbol{y}_1| \\
&= \mathrm{O}_p(1)
\end{aligned}
$$

where we used $\gamma_{max}(M) = 1$ and Assumption 6(i). So $\frac{\partial}{\partial\zeta'}N_{1,1}(\zeta) = O_p(1)$ and $N_{1,1}(\zeta)$ is stochastic equicontinuous. The proofs for stochastic equicontinuity of each of the remaining $N_{r,q}(\zeta)$ follow similarly. By Corollary 2.2 in Newey (1991), $N_{r,q}(\zeta) - E[N_{r,q}(\zeta)] \xrightarrow{p} 0$ uniformly in $\zeta\in\mathcal{Z}$ for all r and q. Hence $N_r(\zeta) - E[N_r(\zeta)] \xrightarrow{p} 0$ uniformly in $\zeta\in\mathcal{Z}$ for $r = 1, 2$ and 3.

To prove $E[N_4(\zeta)] \longrightarrow 0$ uniformly in $\zeta\in\mathcal{Z}$, first note that

$$
\begin{aligned}
\frac{1}{n(T-1)}E[N_4(\zeta)] &= \frac{1}{n(T-1)}E[(Y^{\dagger\prime}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}} - Y_{-1}^{\dagger'}\boldsymbol{A}_2^{*\prime})\Sigma^{-\frac{1}{2}}P\Sigma^{-\frac{1}{2}}(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^*}}Y^{\dagger} - \boldsymbol{A}_2^* Y_{-1}^{\dagger})] \\
&= \frac{1}{n(T-1)}tr[\Sigma^{-1}\Delta X(\Delta X'\Sigma^{-1}\Delta X)^{-1}\Delta X'\Sigma^{-1}var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1})] \\
&\leq \frac{\gamma_{max}(\Sigma^{-2})}{n(T-1)}\gamma_{min}^{-1}(\Delta X'\Sigma^{-1}\Delta X)tr[\Delta X' var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1})\Delta X] \\
&\leq \frac{\gamma_{max}(\Sigma^{-2})}{n(T-1)}\gamma_{min}^{-1}[\frac{\Delta X'\Sigma^{-1}\Delta X}{n(T-1)}]\frac{1}{n(T-1)}tr[\Delta X' var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1})\Delta X].
\end{aligned}
$$

By Assumption 4, there exists two positive constants C_{α_3} and $\bar{C}_{\alpha_3}$ such that $0 < C_{\alpha_3} \leq inf_{\alpha_3\in\mathcal{A}_3}$ $\gamma_{min}(\Sigma^{-1}) \leq sup_{\alpha_3\in\mathcal{A}_3}\gamma_{max}(\Sigma^{-1}) \leq \bar{C}_{\alpha_3} < \infty$. So there exists another two constants $c_{\Delta X}$ and $\bar{c}_{\Delta X}$ such that $0 < c_{\Delta X} \leq inf_{\alpha_3\in\mathcal{A}_3}\gamma_{min}(\Sigma^{-1})\gamma_{min}[\frac{\Delta X'\Delta X}{n(T-1)}] \leq \gamma_{min}[\frac{\Delta X'\Sigma^{-1}\Delta X}{n(T-1)}] \leq \gamma_{max}[\frac{\Delta X'\Sigma^{-1}\Delta X}{n(T-1)}]$ $\leq sup_{\alpha_3\in\mathcal{A}_3}\gamma_{max}(\Sigma^{-1})\gamma_{max}[\frac{\Delta X'\Delta X}{n(T-1)}] \leq \bar{c}_{\Delta X} < \infty$, which can be used in the inequality above and leads to

$$
\begin{aligned}
\frac{1}{n(T-1)}E[N_4(\zeta)] &\leq \frac{1}{n(T-1)}\bar{C}_{\alpha_3}^2 c_{\Delta X}\frac{1}{n(T-1)}tr[\Delta X' var(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1})\Delta X] \\
&\leq \frac{1}{n(T-1)}\bar{C}_{\alpha_3}^2 c_{\Delta X}\bar{c}_{\Delta Y}\frac{1}{n(T-1)}tr(\Delta X'\Delta X) \\
&= O(\tfrac{1}{n})
\end{aligned}
$$

by assumption of the theorem and bounds on Rayleigh quotient. Hence $\hat{\sigma}_{\epsilon,M}^2(\zeta) - \overline{\sigma}_{\epsilon,M}^2(\zeta) = o_p(1)$ uniformly in $\zeta\in\mathcal{Z}$ and (ii) holds.

Proof of (iii)-(vi): Using the similar transformations in the proof of (ii), by letting $\widetilde{\boldsymbol{W}}_r = \Sigma^{-\frac{1}{2}}\boldsymbol{W}_r\Sigma^{\frac{1}{2}}$ for $r = 1$ and 2, we can express the functions in (iii)-(vi) as follows:

$\Delta\hat{u}(\alpha)'\Sigma^{-1}\Delta Y_{-1} - E\big[\Delta\bar{u}(\alpha)'\Sigma^{-1}\Delta Y_{-1}\big]$
$= \Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M\Sigma^{-\frac{1}{2}}\Delta Y_{-1} - E(\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M\Sigma^{-\frac{1}{2}}\Delta Y_{-1})$
$\quad -\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M\Sigma^{-\frac{1}{2}}\Delta Y_{-1} + E(\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M\Sigma^{-\frac{1}{2}}\Delta Y_{-1})$
$\quad -E(\Delta Y^{\dagger\prime}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}P\Sigma^{-\frac{1}{2}}\Delta Y_{-1}) + E(\Delta Y_{-1}^{\dagger}{}'\boldsymbol{A}_2^{*\prime}P\Sigma^{-\frac{1}{2}}\Delta Y_{-1})$

$\Delta\hat{u}(\alpha)'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - E\big[\Delta\bar{u}(\alpha)'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y\big]$
$= \Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M\widetilde{\boldsymbol{W}}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y - E(\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M\widetilde{\boldsymbol{W}}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y)$
$\quad -\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M\widetilde{\boldsymbol{W}}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y + E(\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M\widetilde{\boldsymbol{W}}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y)$
$\quad -E(\Delta Y^{\dagger\prime}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}P\widetilde{\boldsymbol{W}}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y) + E(\Delta Y_{-1}^{\dagger}{}'\boldsymbol{A}_2^{*\prime}P\widetilde{\boldsymbol{W}}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y)$

$\Delta\hat{u}(\alpha)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} - E\big[\Delta\bar{u}(\alpha)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}\big]$
$= \Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M\widetilde{\boldsymbol{W}}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2^{*}}}\Delta Y_{-1} - E(\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M\widetilde{\boldsymbol{W}}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2^{*}}}\Delta Y_{-1})$
$\quad -\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M\widetilde{\boldsymbol{W}}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2^{*}}}\Delta Y_{-1} + E(\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M\widetilde{\boldsymbol{W}}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2^{*}}}\Delta Y_{-1})$
$\quad -E(\Delta Y^{\dagger\prime}\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}P\widetilde{\boldsymbol{W}}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2^{*}}}\Delta Y_{-1}) + E(\Delta Y_{-1}^{\dagger}{}'\boldsymbol{A}_2^{*\prime}P\widetilde{\boldsymbol{W}}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2^{*}}}\Delta Y_{-1})$

$\Delta\hat{u}(\alpha)'(B^{-1}\otimes E_3)\Delta\hat{u}(\alpha) - \frac{1}{\bar{\sigma}_{\epsilon,M}^2(\alpha)}E\big[\Delta\bar{u}(\alpha)'(B^{-1}\otimes E_3)\Delta\bar{u}(\alpha)\big]$
$= \Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M(B^{-1}\otimes E_3)M\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y - E[\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M(B^{-1}\otimes E_3)M\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y]$
$\quad +\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M(B^{-1}\otimes E_3)M\boldsymbol{A}_2^{*}\Delta Y_{-1} - E[\Delta Y_{-1}'\boldsymbol{A}_2^{*\prime}M(B^{-1}\otimes E_3)M\boldsymbol{A}_2^{*}\Delta Y_{-1}]$
$\quad -2\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M(B^{-1}\otimes E_3)M\boldsymbol{A}_2^{*}\Delta Y_{-1} - 2E[\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*\prime}}}M(B^{-1}\otimes E_3)M\boldsymbol{A}_2^{*}\Delta Y_{-1}]$
$\quad -2E[(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y^{\dagger} - \boldsymbol{A}_2^{*}\Delta Y_{-1}^{\dagger})'P(B^{-1}\otimes E_3)P(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y^{\dagger} - \boldsymbol{A}_2^{*}\Delta Y_{-1}^{\dagger})]$
$\quad -2E[(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y^{\dagger} - \boldsymbol{A}_2^{*}\Delta Y_{-1}^{\dagger})'P(B^{-1}\otimes E_3)M(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1^{*}}}\Delta Y^{\dagger} - \boldsymbol{A}_2^{*}\Delta Y_{-1}^{\dagger})]$

Using Lemma 3.1, we can express these terms as functions of $\Delta\boldsymbol{y}_1, \boldsymbol{\delta}$ and $\Delta\epsilon$. Similar proofs follow from those for (ii) and thus are omitted.

Proof of Theorem 3.2. After applying the mean value theorem, to obtain the asymptotic distribution of $\sqrt{n(T-1)}(\hat{\theta}_M - \theta_0) = -[\frac{1}{n(T-1)}H^*(\bar{\theta})]^{-1}\frac{1}{\sqrt{n(T-1)}}S^*(\theta_0)$, where $H^*(\bar{\theta}) = \frac{\partial S^*(\bar{\theta})}{\partial\theta'}$ and $\bar{\theta}$ is between $\hat{\theta}_M$ and θ_0 elementwise, we will first prove that $\frac{1}{n(T-1)}H^*(\bar{\theta}) = \frac{1}{n(T-1)}H^*(\theta_0) + o_p(1) = \frac{1}{n(T-1)}E[H^*(\theta_0)] + o_p(1)$ and then $\frac{1}{\sqrt{n(T-1)}}S^*(\theta_0) \xrightarrow{d} N[0, lim_{n\to\infty}\Omega^*(\theta_0)]$.

The generic form $H^*(\theta) = \frac{\partial S^*(\theta)}{\partial\theta'}$ are comprised of the following elements:

$H_{\beta\beta}^*(\theta) = -\frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma^{-1}\Delta X,$
$H_{\beta\sigma_\epsilon^2}(\theta) = -\frac{1}{\sigma_\epsilon^4}\Delta X'\Sigma^{-1}\Delta u(\phi),$
$H_{\beta\tau}^*(\theta) = -\frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma^{-1}\Delta Y_{-1},$
$H_{\beta\alpha_1}^*(\theta) = \frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y,$
$H_{\beta\alpha_2}^*(\theta) = -\frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1},$
$H_{\beta\alpha_3}^*(\theta) = \frac{1}{\sigma_\epsilon^2}\Delta X'(B^{-1}\otimes E_3)\Delta u(\phi),$
$H_{\sigma_\epsilon^2\sigma_\epsilon^2}^*(\theta) = \frac{n(T-1)}{2\sigma_\epsilon^4} - \frac{1}{\sigma_\epsilon^6}\Delta u(\phi)'\Sigma^{-1}\Delta u(\phi),$
$H_{\sigma_\epsilon^2\tau}^*(\theta) = -\frac{1}{\sigma_\epsilon^4}\Delta Y_{-1}'\Sigma^{-1}\Delta u(\phi),$
$H_{\sigma_\epsilon^2\alpha_1}^*(\theta) = -\frac{1}{\sigma_\epsilon^4}\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1'}}\boldsymbol{W}_1'\Sigma^{-1}\Delta u(\phi),$
$H_{\sigma_\epsilon^2\alpha_2}^*(\theta) = -\frac{1}{\sigma_\epsilon^4}\Delta Y_{-1}'\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2'}}\boldsymbol{W}_2'\Sigma^{-1}\Delta u(\phi)$
$H_{\sigma_\epsilon^2\alpha_3}^*(\theta) = -\frac{1}{2\sigma_\epsilon^4}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi)$
$H_{\tau\tau}^*(\theta) = -\frac{1}{\sigma_\epsilon^2}\Delta Y_{-1}'\Sigma^{-1}\Delta Y_{-1}$
$H_{\tau\alpha_1}^*(\theta) = \frac{1}{\sigma_\epsilon^2}\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1'}}\boldsymbol{W}_1'\Sigma^{-1}\Delta Y_{-1},$
$H_{\tau\alpha_2}^*(\theta) = -\frac{1}{\sigma_\epsilon^2}\Delta Y_{-1}'\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2'}}\boldsymbol{W}_2'\Sigma^{-1}\Delta Y_{-1},$
$H_{\tau\alpha_3}^*(\theta) = \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta Y_{-1}$
$H_{\alpha_1\alpha_1}^*(\theta) = -\frac{1}{\sigma_\epsilon^2}[\Delta Y'\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1'}}\boldsymbol{W}_1'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y + \Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_1^2\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y] - tr(\boldsymbol{D}_{\alpha_1}\boldsymbol{B}^{-1}\boldsymbol{W}_1)$

$$H^*_{\alpha_1\alpha_2}(\theta) = \frac{1}{\sigma_\epsilon^2}\Delta Y'_{-1}\boldsymbol{e^{\alpha_2 W_2'}}\boldsymbol{W}_2'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e^{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}_{\alpha_2}\boldsymbol{B}^{-1}\boldsymbol{W}_1)$$
$$H^*_{\alpha_1\alpha_3}(\theta) = -\frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_3)\boldsymbol{W}_1\boldsymbol{e^{\alpha_1 W_1}}\Delta Y$$
$$H^*_{\alpha_2\alpha_2}(\theta) = \frac{1}{\sigma_\epsilon^2}[-\Delta Y'_{-1}\boldsymbol{e^{\alpha_2 W_2'}}\boldsymbol{W}_2'\Sigma_3^{-1}\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}\Delta Y_{-1} + \Delta u(\phi)'\Sigma_3^{-1}\boldsymbol{W}_2^2\boldsymbol{e^{\alpha_2 W_2}}\Delta Y_{-1}]$$
$$+\mathrm{tr}(\boldsymbol{D}_{-1,\alpha_2}\boldsymbol{B}^{-1}\boldsymbol{W}_{21}) + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2})$$
$$H^*_{\alpha_2\alpha_3}(\theta) = \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_3)\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}\Delta Y_{-1}$$
$$H^*_{\alpha_3\alpha_3}(\theta) = -\frac{1}{2\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_{33})\Delta u(\phi),$$

where $\boldsymbol{D}_{\alpha_1} = \frac{\partial \boldsymbol{D}}{\partial \alpha_1}$, $\boldsymbol{D}_{\alpha_2} = \frac{\partial \boldsymbol{D}}{\partial \alpha_2}$, $\boldsymbol{D}_{-1,\alpha_2} = \frac{\partial \boldsymbol{D}_{-1}}{\partial \alpha_2}$, $\boldsymbol{W}_{21,\alpha_2} = \frac{\partial \boldsymbol{W}_{21}}{\partial \alpha_2}$ and $E_{33} = \frac{\partial E_3}{\partial \alpha_3}$.

We will first prove $\frac{1}{n(T-1)}[H^*(\bar\theta) - H^*(\theta_0)] = o_p(1)$. Note there are stochastic and nonstochastic elements in $H^*(\theta)$. The stochastic elements are comprised of all the terms other than the trace terms and $\frac{n(T-1)}{2\sigma_\epsilon^4}$ in $H^*_{\sigma_\epsilon^2\sigma_\epsilon^2}(\theta)$. By the model assumptions and Lemma A.1, all elements in $H^*(\theta_0)$ are uniformly bounded in both row and column sums and thus $\frac{1}{n(T-1)}H^*(\theta_0) = O_p(1)$. Note $\hat\theta_M \xrightarrow{p} \theta_0$ by Theorem 3.1. So $\bar\theta \xrightarrow{p} \theta_0$ as well because $\bar\theta$ is between $\hat\theta_M$ and θ_0. It follows that $\frac{1}{n(T-1)}H^*(\bar\theta) = O_p(1)$. For $\bar\sigma_\epsilon^{-2}$, $\bar\sigma_\epsilon^{-4}$ and $\bar\sigma_\epsilon^{-6}$ in $H^*(\bar\theta)$, note $\bar\sigma_\epsilon^{-2} \xrightarrow{p} \sigma_{\epsilon 0}^{-2}$, $\bar\sigma_\epsilon^{-4} \xrightarrow{p} \sigma_{\epsilon 0}^{-4}$ and $\bar\sigma_\epsilon^{-6} \xrightarrow{p} \sigma_{\epsilon 0}^{-6}$ which is implied by $\bar\theta \xrightarrow{p} \theta_0$. So they can be replaced by $\sigma_{\epsilon 0}^{-2}$, $\sigma_{\epsilon 0}^{-4}$ and $\sigma_{\epsilon 0}^{-6}$ respectively during the proof, i.e., we need to show $\frac{1}{n(T-1)}[H^*(\bar\beta, \sigma_{\epsilon 0}^2, \bar\tau, \bar\alpha) - H^*(\beta_0, \sigma_{\epsilon 0}^2, \tau_0, \alpha_0)] = o_p(1)$. Noting that $\Delta u(\phi) = \boldsymbol{e^{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}_2\Delta Y_{-1} - \Delta X\beta$ and $\Delta u = \boldsymbol{e^{\alpha_{10} W_1}}\Delta Y - \boldsymbol{A}_{20}\Delta Y_{-1} - \Delta X\beta_0$, we get the expression $\Delta u(\phi) = \Delta u + (\boldsymbol{e^{\alpha_1 W_1}} - \boldsymbol{e^{\alpha_{10} W_1}})\Delta Y - (\boldsymbol{A}_2 - \boldsymbol{A}_{20})\Delta Y_{-1} - \Delta X(\beta - \beta_0)$. Similarly $\Sigma_3^{-1} = \Sigma_{30}^{-1} + B^{-1}\otimes(e^{\alpha_3 W_3'}e^{\alpha_3 W_3} - e^{\alpha_{30} W_3'}e^{\alpha_{30} W_3})$. These two expressions being substituted into $H^*(\theta)$ implies that the stochastic elements are linear, bilinear or quadratic in ΔY, ΔY_{-1} or Δu, and linear or quadratic in β, τ and α. Hence the stochastic elements in $\frac{1}{n(T-1)}[H^*(\bar\beta, \sigma_{\epsilon 0}^2, \bar\tau, \bar\alpha) - H^*(\beta_0, \sigma_{\epsilon 0}^2, \tau_0, \alpha_0)]$ are linear, bilinear or quadratic in ΔY, ΔY_{-1} or Δu, and linear or quadratic in $\bar\beta - \beta_0$, $\bar\tau - \tau_0$ and $\bar\alpha - \alpha_0$. By Lemma 3.1, we can express these elements in terms of $\Delta \boldsymbol{y}_1$ and $\Delta\epsilon$. Because $\bar\theta \xrightarrow{p} \theta_0$, we can prove all the stochastic elements are $o_p(1)$ using the similar proof to Theorem 3.1.

For the nonstochastic elements, we will prove that all the trace terms are $o_p(1)$. There are two types of trace terms, the first being $tr(\boldsymbol{D}_{\alpha_1}\boldsymbol{B}^{-1}\boldsymbol{W}_1)$, $tr(\boldsymbol{D}_{\alpha_2}\boldsymbol{B}^{-1}\boldsymbol{W}_1)$ and $tr(\boldsymbol{D}_{-1,\alpha_2}\boldsymbol{B}^{-1}\boldsymbol{W}_{21})$ and the second being $tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2})$. For $tr(\boldsymbol{D}_{\alpha_1}\boldsymbol{B}^{-1}\boldsymbol{W}_1)$, assume $(\breve\alpha_1, \breve\alpha_2)$ is between $(\bar\alpha_1, \bar\alpha_2)$ and $(\alpha_{10}, \alpha_{20})$ elementwise. By the mean value theorem:

$$\frac{1}{n(T-1)}\{tr[\boldsymbol{D}_{\alpha_1}(\bar\alpha_1, \bar\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_1] - tr[\boldsymbol{D}_{\alpha_1}(\alpha_{10}, \alpha_{20})\boldsymbol{B}^{-1}\boldsymbol{W}_1]\}$$
$$= \frac{1}{n(T-1)}\Big\{(\bar\alpha_1 - \alpha_{10})\{tr[\boldsymbol{D}_{\alpha_1\alpha_1}(\breve\alpha_1, \breve\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_1]\} + (\bar\alpha_2 - \alpha_{20})\{tr[\boldsymbol{D}_{\alpha_1\alpha_2}(\breve\alpha_1, \breve\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_1]\}\Big\},$$

where $\boldsymbol{D}_{\alpha_1\alpha_1}(\breve\alpha_1, \breve\alpha_2)$ and $\boldsymbol{D}_{\alpha_1\alpha_2}(\breve\alpha_1, \breve\alpha_2)$ are the derivatives of $\boldsymbol{D}_{\alpha_1}$ with respect to α_1 and α_2 respectively evaluated at $(\breve\alpha_1, \breve\alpha_2)$. WLOG we assume $T = 3$, then

$$\boldsymbol{D}_{\alpha_1\alpha_1} = \begin{pmatrix} A_2W_1^2e^{-\alpha_1 W_1} & \boldsymbol{0}_{n\times n} \\ 2[A_2W_1e^{-\alpha_1 W_1}A_2W_1e^{-\alpha_1 W_1} + (A_2e^{-\alpha_1 W_1} - I_n)A_2W_1^2e^{-\alpha_1 W_1}] & A_2W_1^2e^{-\alpha_1 W_1} \end{pmatrix}$$

and

$$\boldsymbol{D}_{\alpha_1\alpha_2} = \begin{pmatrix} -W_2A_2W_1e^{-\alpha_1 W_1} & \boldsymbol{0}_{n\times n} \\ -2[W_2A_2e^{-\alpha_1 W_1}A_2W_1e^{-\alpha_1 W_1} + (A_2e^{-\alpha_1 W_1} - I_n)W_2A_2W_1e^{-\alpha_1 W_1}] & -W_2A_2W_1e^{-\alpha_1 W_1} \end{pmatrix},$$

By Lemma A.1, $\boldsymbol{D}_{\alpha_1\alpha_1}$ and $\boldsymbol{D}_{\alpha_1\alpha_2}$ are uniformly bounded in a matrix norm in the neighborhood of $(\alpha_{10}, \alpha_{20})$, leading to $\frac{1}{n(T-1)}\{tr[\boldsymbol{D}_{\alpha_1}(\bar\alpha_1, \bar\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_1] - tr[\boldsymbol{D}_{\alpha_1}(\alpha_{10}, \alpha_{20})\boldsymbol{B}^{-1}\boldsymbol{W}_1]\} = o_p(1)$. The rest of the first type are proved similarly. For $tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2})$, we similarly apply the mean value theorem and get the following:

$$\frac{1}{n(T-1)}\{tr[\boldsymbol{D}_{-1}(\bar\alpha_1, \bar\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2}(\bar\alpha_1, \bar\alpha_2)] - tr[\boldsymbol{D}_{-1}(\alpha_{10}, \alpha_{20})\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2}(\alpha_{10}, \alpha_{20})]\}$$
$$= \frac{1}{n(T-1)}\Big\{(\bar\alpha_1 - \alpha_{10})\{tr[\boldsymbol{D}_{-1,\alpha_1}(\breve\alpha_1, \breve\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2}(\breve\alpha_1, \breve\alpha_2)] + tr[\boldsymbol{D}(\breve\alpha_1, \breve\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2\alpha_1}(\breve\alpha_1, \breve\alpha_2)]\}$$
$$+(\bar\alpha_2 - \alpha_{20})\{tr[\boldsymbol{D}_{-1,\alpha_2}(\breve\alpha_1, \breve\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2}(\breve\alpha_1, \breve\alpha_2)] + tr[\boldsymbol{D}(\breve\alpha_1, \breve\alpha_2)\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2\alpha_2}(\breve\alpha_1, \breve\alpha_2)]\}\Big\},$$

where $\boldsymbol{D}_{-1,\alpha_1}(\breve{\alpha}_1, \breve{\alpha}_2)$ and $\boldsymbol{D}_{-1,\alpha_2}(\breve{\alpha}_1, \breve{\alpha}_2)$ are the derivatives of $\boldsymbol{D}_{-1}$ with respect to α_1 and α_2 respectively and $\boldsymbol{W}_{21,\alpha_2\alpha_1}(\breve{\alpha}_1, \breve{\alpha}_2)$ and $\boldsymbol{W}_{21,\alpha_2\alpha_2}(\breve{\alpha}_1, \breve{\alpha}_2)$ are derivatives of $\boldsymbol{W}_{21,\alpha_2}$ with respect to α_1 and α_2 respectively, all evaluated at $(\breve{\alpha}_1, \breve{\alpha}_2)$. Again WLOG assuming T=3, we have

$$\boldsymbol{D}_{-1,\alpha_1} = \begin{pmatrix} I_n & \mathbf{0}_{n\times n} \\ -e^{\alpha_2 W_2} W_1 e^{-\alpha_1 W_1} & I_n \end{pmatrix}, \boldsymbol{D}_{-1,\alpha_2} = \begin{pmatrix} I_n & \mathbf{0}_{n\times n} \\ W_2 e^{\alpha_2 W_2} e^{-\alpha_1 W_1} & I_n \end{pmatrix},$$

$$\boldsymbol{W}_{21,\alpha_2\alpha_1} = \begin{pmatrix} -W_2^2 e^{\alpha_2 W_2} W_1 e^{-\alpha_1 W_1} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & -W_2^2 e^{\alpha_2 W_2} W_1 e^{-\alpha_1 W_1} \end{pmatrix} \text{ and}$$

$$\boldsymbol{W}_{21,\alpha_2\alpha_2} = \begin{pmatrix} W_2^3 e^{\alpha_2 W_2} e^{-\alpha_1 W_1} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & W_2^3 e^{\alpha_2 W_2} e^{-\alpha_1 W_1} \end{pmatrix}.$$

Here $\boldsymbol{D}_{-1,\alpha_1}, \boldsymbol{D}_{-1,\alpha_2}, \boldsymbol{W}_{21,\alpha_2}$ and $\boldsymbol{W}_{21,\alpha_2\alpha_1}$ are uniformly bounded in a matrix norm by Lemma A.1. So $\frac{1}{n(T-1)}\{tr[\boldsymbol{D}_{-1}(\bar{\alpha}_1, \bar{\alpha}_2)\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2}(\bar{\alpha}_1, \bar{\alpha}_2)] - tr[\boldsymbol{D}_{-1}(\alpha_{10}, \alpha_{20})\boldsymbol{B}^{-1}\boldsymbol{W}_{21,\alpha_2}(\alpha_{10}, \alpha_{20})]\} = o_p(1)$. It follows that $\frac{1}{n(T-1)}[H^*(\bar{\theta}) - H^*(\theta_0)] = o_p(1)$.

Next let's prove $\frac{1}{n(T-1)}\{H^*(\theta_0) - E[H^*(\theta_0)]\} = o_p(1)$. The term is comprised of differences of linear, bilinear or quadratic forms in ΔY, ΔY_{-1} or Δu and their expected values at the true values. For terms involving ΔY and ΔY_{-1}, using Lemma 3.1, they can be expressed as formulas of sums of terms linear in $\Delta \boldsymbol{y}_1$, quadratic in $\Delta \boldsymbol{y}_1$, bilinear in $\Delta \boldsymbol{y}_1$ and $\Delta\epsilon$ and quadratic in $\Delta\epsilon$. Using Lemma A.1, Lemma A.4 and Assumption 6, these terms are $o_p(1)$. For terms involving Δu, note $\Delta u = e^{-\alpha_{30}W_3}\Delta\epsilon = e^{-\alpha_{30}W_3}C\epsilon$, where C is an $n(T-1)\times nT$ matrix:

$$C = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}.$$

So we have, for example, $H^*_{\sigma_\epsilon^2\alpha_3}(\theta_0) - E[H^*_{\sigma_\epsilon^2\alpha_3}(\theta_0)] = \frac{1}{2\sigma_{\epsilon 0}^4}\{\epsilon' C' \boldsymbol{e}^{\boldsymbol{\alpha_{30}W_3}'}(B^{-1}\otimes E_{30})\boldsymbol{e}^{\boldsymbol{\alpha_{30}W_3}}C\epsilon - E[\epsilon' C' \boldsymbol{e}^{\boldsymbol{\alpha_{30}W_3}'}(B^{-1}\otimes E_{30})\boldsymbol{e}^{\boldsymbol{\alpha_{30}W_3}}C\epsilon]\}$. By Lemma A.4(v), $\frac{1}{n(T-1)}\{H^*_{\sigma_\epsilon^2\alpha_3}(\theta_0) - E[H^*_{\sigma_\epsilon^2\alpha_3}(\theta_0)]\} = o_p(1)$. Similar proofs can be done for all other terms involving Δu. So $\frac{1}{n(T-1)}[H^*(\theta_0) - E(H^*(\theta_0))] = o_p(1)$.

Finally let's prove $\frac{1}{\sqrt{n(T-1)}}S^*(\theta_0) \xrightarrow{d} N[0, lim_{n\to\infty}\Omega^*(\theta_0)]$. From (3.8) we know $S^*(\theta_0)$ consists of three types of components: $R'\Delta\epsilon$, $\Delta\epsilon' F \Delta\boldsymbol{y}_1$ and $\Delta\epsilon' O \Delta\epsilon$ where subscripts r for R_r, F_r and O_r are suppressed for simplicity. Partitioning them using matrix C above gives us the following: $R'\Delta\epsilon = \Sigma_{t=1}^T R_t^{*\prime}\epsilon_t$, $\Delta\epsilon' F\Delta\boldsymbol{y}_1 = \Sigma_{t=1}^T \epsilon_t' F_t^* \Delta y_1$, $\Delta\epsilon' O\Delta\epsilon = \Sigma_{s=1}^T\Sigma_{t=1}^T \epsilon_t' O_t^* \epsilon_t$, where $R_t^* = R_t' C_t$, $F_t^* = C_t' F_t$ and $O_t^* = C_t' O_t C_t$ are $n\times k$, $n\times n$ and $n\times n$ partitioned matrices of $R'C$, $C'T$ and $C'OC$ respectively. After substituting $\Delta y_1 = (e^{-\alpha_{10}W_1}A_2 - I_n)y_0 + c_1 + e^{-\alpha_{10}W_1}e^{\alpha_{30}W_3}\epsilon_1$ into $\epsilon_t' F_t^* \Delta y_1$, where c_1 is a non-stochastic term, we get $\Sigma_{t=1}^T\epsilon_t' F_t^*\Delta y_1 = \Sigma_{t=1}^T \epsilon_t' F_{t1}^* y_0 + \Sigma_{t=1}^T \epsilon_t' F_t^* c_1 + \Sigma_{t=1}^T \epsilon_t' F_{t2}^* \epsilon_1$. So for an $(k+5)\times 1$ vector of constants a, $a'S^*(\theta_0) = \Sigma_{s=1}^T\Sigma_{t=1}^T \epsilon_t' A_{ts}\epsilon_s + \Sigma_{t=1}^T \epsilon_t' B_t \epsilon_1 + \Sigma_{t=1}^T \epsilon_t' f(y_0) + a'd$ for nonstochastic matrices A_{ts}, B_t, vector d and $f(y_0)$ as a function of y_0. By Assumption 1, y_0 is independent of ϵ_t for $t = 1, ..., T$. Also $\epsilon_1, ..., \epsilon_T$ are independent of each other by Assumption 5. Hence $\frac{1}{n(T-1)}a'S^*(\theta_0)$ is asymptotically normal by Lemma A.5. Since every fixed linear combination of elements of $S^*(\theta_0)$ converges in distribution, by Cramer-Wold device, $\frac{1}{n(T-1)}S^*(\theta_0) \xrightarrow{d} N[0, lim_{n\to\infty}\Omega^*(\theta_0)]$.

Proof of Theorem 3.3. To prove $\hat{\Omega}^* = \frac{1}{n(T-1)}\Sigma_{i=1}^n \hat{a}_i\hat{a}_i' \xrightarrow{p} \Omega^*(\theta_0) = \frac{1}{n(T-1)}\Sigma_{i=1}^n E(a_i a_i')$, we need to prove the following:

(i) $\frac{1}{n(T-1)}\Sigma_{i=1}^n \hat{a}_i\hat{a}_i' \xrightarrow{p} \frac{1}{n(T-1)}\Sigma_{i=1}^n a_i a_i'$;
(ii) $\frac{1}{n(T-1)}\Sigma_{i=1}^n a_i a_i' \xrightarrow{p} \frac{1}{n(T-1)}\Sigma_{i=1}^n E(a_i a_i')$.

Proof of (i): For $\bar{\theta}$ between $\hat{\theta}_M$ and θ_0 elementwise, we can utilize mean value theorem to each of the elements in $\frac{1}{n(T-1)}\Sigma_{i=1}^n(\hat{a}_{li}\hat{a}_{mi}' - a_{li}a_{mi}')$ for $l,m = 1,2,3$ and prove each of them is $o_p(1)$. For example, for the first element when $l = m = 1$, $a_{11i}a_{11i}'$ is an $k \times k$ matrix where k is the number of regressors in ΔX, and $\frac{1}{n(T-1)}(\hat{a}_{11i}\hat{a}_{11i}' - a_{11i}a_{11i}') = -\frac{2}{n(T-1)}\bar{a}_{11i}\Sigma_{j=1}^k\Sigma_{t=2}^T\bar{R}_{it}'(\boldsymbol{e^{\bar{\alpha}_3 W_3}}\Delta X_j)_{it}(\hat{\beta}_{jM} - \beta_{j0}) - \frac{2}{n(T-1)}\bar{a}_{11i}(\Sigma_{t=2}^T\frac{1}{\bar{\sigma}_\epsilon^4}\bar{R}_{it}'\Delta\bar{\epsilon}_{it})'(\hat{\sigma}_{\epsilon,M}^2 - \sigma_{\epsilon0}^2) - \frac{2}{n(T-1)}\bar{a}_{11i}\ [\Sigma_{t=2}^T\bar{R}_{it}'(\boldsymbol{e^{\bar{\alpha}_3 W_3}}\Delta Y_{-1})_{it}]'(\hat{\tau}_M - \tau_0) + \frac{2}{n(T-1)}\bar{a}_{11i}\ [\Sigma_{t=2}^T\bar{R}_{it}'(\boldsymbol{e^{\bar{\alpha}_3 W_3}}\boldsymbol{W}_1\boldsymbol{e^{\bar{\alpha}_1 W_1}}\Delta Y)_{it}]'(\hat{\alpha}_{1M} - \alpha_{10}) - \frac{2}{n(T-1)}\bar{a}_{11i}[\Sigma_{t=2}^T\bar{R}_{it}'(\boldsymbol{e^{\bar{\alpha}_3 W_3}}\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}\Delta Y_{-1})_{it}]'(\hat{\alpha}_{2M} - \alpha_{20}) + \frac{2}{n(T-1)}\bar{a}_{11i}\big\{\Sigma_{t=2}^T\{[\frac{1}{\bar{\sigma}_\epsilon^2}(B^{-1}\otimes W_3 e^{\bar{\alpha}_3 W_3})\Delta X]_{it}'\Delta\bar{\epsilon}_{it} + \bar{R}_{1it}'(\boldsymbol{W}_3\Delta\bar{\epsilon})_{it}\}\big\}'(\hat{\alpha}_{3M} - \alpha_{30})$, where the terms with bars on top denote the values implied by $\bar{\theta}$ which is between θ_M and θ_0. By model assumptions and Lemma A.1, all the multipliers before the differences of parameters $\hat{\theta}_M - \theta_0$ are $O_p(1)$. Since $\hat{\theta}_M - \theta_0 = o_p(1)$ by Theorem 1, $\frac{1}{n(T-1)}(\hat{a}_{11i}\hat{a}_{11i}' - a_{11i}a_{11i}') = o_p(1)$. The proofs for other terms follow similarly.

Proof of (ii): We need to prove $\frac{1}{n(T-1)}\Sigma_{i=1}^n[a_{li}a_{mi}' - E(a_{li}a_{mi}')] \xrightarrow{p} 0$ for $l,m = 1,2,3$. We will prove it for $l = m = 1$, $l = m = 2$ and $l = m = 3$ and the cross multiplied cases are done in a similar way.

Before proceeding with the proof we define the following notations.

(1) For $n(T-1) \times 1$ vector $\Delta\epsilon$, we denote $\Delta\epsilon_{\cdot t}$ as the $n \times 1$ vector that selects all elements corresponding to period t and denote $\Delta\epsilon_{i\cdot}$ as the $(T-1) \times 1$ vector that selects all elements corresponding to individual i.

(2) For $n(T-1) \times n(T-1)$ matrix O, we denote $O_{\cdot t,\cdot s}$ as the $n \times n$ matrix that selects all elements corresponding to period (t,s), denote $O_{i\cdot,j\cdot}$ as the $(T-1) \times (T-1)$ matrix that selects all elements corresponding to individual (i,j) and denote $O_{it,j\cdot}$ as the $(T-1) \times 1$ vector that is the tth column of $O_{i\cdot,j\cdot}$.

Then we can express a_{1i}, a_{2i} and a_{3i} as $a_{1i} = \Sigma_{t=2}^T R_{it}'\epsilon_{it} = R_{i\cdot}\Delta\epsilon_{i\cdot}$, $a_{2i} = \Sigma_{t=2}^T(\Delta\epsilon_{it}\Delta\eta_{it} + \Delta\epsilon_{it}\Delta\epsilon_{it}^* - \sigma_{\epsilon0}^2 d_{it}) = \Delta\epsilon_{i\cdot}'\Delta\eta_{i\cdot} + \Delta\epsilon_{i\cdot}'\Delta\epsilon_{i\cdot}^* - \sigma_{\epsilon0}^2 l_{T-1}'d_{i\cdot}$ and $a_{3i} = \Delta\epsilon_{2i}\Delta\xi_i + F_{2,ii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^\diamond + \sigma_{\epsilon0}^2) + \Sigma_{t=3}^T\Delta\epsilon_{it}\Delta y_{1it}^* = \Delta\epsilon_{2i}\Delta\xi_i + F_{2,ii}^{++}(\Delta\epsilon_{2i}\Delta y_{1i}^\diamond + \sigma_{\epsilon0}^2) + \Delta\epsilon_{i-}'\Delta y_{1i-}^*$, where $-$ denotes the selection of all element from $t = 3$ to T. These expressions will be convenient to use in the proof below.

For a_{1i}, $\frac{1}{n(T-1)}\Sigma_{i=1}^n[a_{1i}a_{1i}' - E(a_{1i}a_{1i}')] = \frac{1}{n(T-1)}\Sigma_{i=1}^n R_{i\cdot}'(\Delta\epsilon_{i\cdot}\Delta\epsilon_{i\cdot}' - \sigma_{\epsilon0}^2 B)R_i = \frac{1}{n(T-1)}\Sigma_{i=1}^n z_{n,i}$. Note $z_{n,i}$ is a MDS since $\{z_{n,i}\}$ are independent and $E(z_{n,i}) = 0$. Given Assumption 6, we know from Lemma A.1 that the elements of $R_{i\cdot}$ are uniformly bounded in row and column sums. Then $E|z_{n,i}|^{1+\zeta}$ is bounded above by some constant for $\zeta > 0$ which implies $z_{n,i}$ is uniformly integrable. Also for the multiplying coefficient, $limsup_{n\to\infty}\Sigma_{i=1}^n\frac{1}{n(T-1)} = \frac{1}{T-1} < \infty$ and $lim_{n\to\infty}\Sigma_{i=1}^n[\frac{1}{n(T-1)}]^2 = lim_{n\to\infty}\frac{1}{n(T-1)^2} = 0$. By Theorem 19.7 on Weak Law of Large Numbers for MD array (p299 Davidson (1994)), $\frac{1}{n(T-1)}\Sigma_{i=1}^n z_{n,i} \xrightarrow{p} 0$.

For a_{2i}, first note $E(\Delta\epsilon_{i\cdot}'\Delta\eta_{i\cdot}) = 0$ for $i = 1,...,n$ because each multiplying group of elements in $\Delta\epsilon_{i\cdot}$ and $\Delta\eta_{i\cdot}$ are from different individuals. Then we have $\frac{1}{n(T-1)}\Sigma_{i=1}^n[a_{2i}^2 - E(a_{2i}^2)] = \frac{1}{n(T-1)}\Sigma_{i=1}^n\{[(\Delta\epsilon_{i\cdot}'\Delta\eta_{i\cdot})^2 - E(\Delta\epsilon_{i\cdot}'\Delta\eta_{i\cdot})^2] + [(\Delta\epsilon_{i\cdot}'\Delta\epsilon_{i\cdot}^*)^2 - E(\Delta\epsilon_{i\cdot}'\Delta\epsilon_{i\cdot}^*)^2] + 2(\Delta\epsilon_{i\cdot}'\Delta\eta_{i\cdot})(\Delta\epsilon_{i\cdot}'\Delta\epsilon_{i\cdot}^*) - 2\sigma_{\epsilon0}^2 l_{T-1}'d_{i\cdot}(\Delta\epsilon_{i\cdot}'\Delta\eta_{i\cdot}) - 2[\sigma_{\epsilon0}^2 l_{T-1}'d_{i\cdot}(\Delta\epsilon_{i\cdot}'\Delta\epsilon_{i\cdot}^* - E(\Delta\epsilon_{i\cdot}'\Delta\epsilon_{i\cdot}^*))]\}$. We can prove that each of the five terms is $o_p(1)$. For example, for the first term, subtracting and adding a same term and noticing $E(\Delta\epsilon_{i\cdot}\Delta\epsilon_{i\cdot}') = \sigma_{\epsilon0}^2 B$, it equals $\frac{1}{n(T-1)}\Sigma_{i=1}^n\Delta\eta_{i\cdot}'(\Delta\epsilon_{i\cdot}\Delta\epsilon_{i\cdot}' - \sigma_{\epsilon0}^2 B)\Delta\eta_{i\cdot}] + \frac{\sigma_{\epsilon0}^2}{n(T-1)}\Sigma_{i=1}^n[\Delta\eta_{i\cdot}'B\Delta\eta_{i\cdot} - E(\Delta\eta_{i\cdot}'B\Delta\eta_{i\cdot})]$. Let $N_{n,i} = \Delta\eta_{i\cdot}'(\Delta\epsilon_{i\cdot}\Delta\epsilon_{i\cdot}' - \sigma_{\epsilon0}^2 B)\Delta\eta_{i\cdot}$. Since $\Delta\eta_{i\cdot}$ is $\Pi_{n,i-1}$ measurable, $E(N_{n,i}|\Pi_{n,i-1}) = 0$. To be a MD array, it is also necessary that $E(N_{n,i}) < \infty$ (e.g. Davidson p232), which is obviously satisfied. Thus $\{N_{n,i}, \Pi_{n,i-1}\}$ is a MD array. Also $E|N_{n,i}^{1+\zeta}|$ is bounded above by some positive constant for some $\zeta > 0$. So $\{N_{n,i}\}$ is uniformly integrable. The multiplier $\frac{1}{n(T-1)}$ was shown in proof of a_{1i} to satisfy the other two conditions of Theorem 19.7 in Davidson (1994). So $\frac{1}{n(T-1)}\Sigma_{i=1}^{n}N_{n,i}=o_p(1)$.

For the second term, we can express $\Delta\eta_{i\cdot}'B\Delta\eta_{i\cdot}=\Sigma_{t=1}^{T-1}\Sigma_{s=1}^{T-1}\Delta\eta_{it}B_{ts}\Delta\eta_{it}$, where $\Delta\eta_{it}$ is the ith element of the $n\times 1$ vector $\Delta\eta_t=\Sigma_{s=2}^{T}(O_{st}^{u'}+O_{ts}^{l})\Delta\epsilon_s$. Here O is an $n(T-1)\times n(T-1)$ matrix and O_{st} is its stth $n\times n$ block matrix. So $\Delta\eta_{it}=\Sigma_{s=2}^{T}\Sigma_{j=1}^{i-1}(O_{js,it}+O_{it,js})\Delta\epsilon_{js}=\Sigma_{j=1}^{i-1}\Sigma_{s=2}^{T}(O_{js,it}+O_{it,js})\Delta\epsilon_{js}=\Sigma_{j=1}^{i-1}O_{ijt}'\Delta\epsilon_{j\cdot}$, where $O_{ijt}=O_{j\cdot,it}+O_{it,j\cdot}$. Then for $\Delta\eta_{is}=\Delta\eta_{it}$, we have $(\Delta\eta_{it})^2-E[(\Delta\eta_{it})^2]=\Sigma_{j=1}^{i-1}O_{ijt}'(\Delta\epsilon_{j\cdot}\Delta\epsilon_{j\cdot}'-\sigma_{\epsilon 0}^2B)O_{ijt}+2\Sigma_{j=1}^{i-1}\Sigma_{k=1}^{j-1}O_{ijt}'\Delta\epsilon_{j\cdot}'\Delta\epsilon_{k\cdot}'O_{ikt}$, which implies

$$
\begin{aligned}
&\frac{1}{n(T-1)}\Sigma_{i=1}^{n}\{(\Delta\eta_{it})^2-E[(\Delta\eta_{it})^2]\}\\
=&\frac{1}{n(T-1)}\Sigma_{i=1}^{n}\{\Sigma_{j=1}^{i-1}O_{ijt}'(\Delta\epsilon_{j\cdot}\Delta\epsilon_{j\cdot}'-\sigma_{\epsilon 0}^2B)O_{ijt}+2\Sigma_{j=1}^{i-1}\Sigma_{k=1}^{j-1}\Delta\epsilon_{j\cdot}'O_{ijt}O_{ikt}'\Delta\epsilon_{k\cdot}\}\\
=&\frac{1}{n(T-1)}\Sigma_{j=1}^{n-1}\{\Sigma_{i=j+1}^{n}[O_{ijt}'(\Delta\epsilon_{j\cdot}\Delta\epsilon_{j\cdot}'-\sigma_{\epsilon 0}^2B)O_{ijt}]\}+\frac{2}{n(T-1)}\Sigma_{j=1}^{n-1}\Delta\epsilon_{j\cdot}'\Sigma_{i=j+1}^{n}\Sigma_{k=1}^{j-1}O_{ijt}O_{ikt}'\Delta\epsilon_{k\cdot}
\end{aligned}
$$

Now the terms in the summation in the first element are independent, and $\Sigma_{i=j+1}^{n}\Sigma_{k=1}^{j-1}O_{ijt}O_{ikt}'\Delta\epsilon_{k\cdot}$ is $\Pi_{n,j-1}$-measurable. By Theorem 19.7 in Davidson (1994), $\frac{1}{n(T-1)}\Sigma_{i=1}^{n}\{(\Delta\eta_{it})^2-E[(\Delta\eta_{it})^2]\}=o_p(1)$. Similar proofs can be done for $\Delta\eta_{is}\neq\Delta\eta_{it}$. Thus $\frac{1}{n(T-1)}\Sigma_{i=1}^{n}[\Delta\eta_{i\cdot}'B\Delta\eta_{i\cdot}-E(\Delta\eta_{i\cdot}'B\Delta\eta_{i\cdot})]=o_p(1)$. It follows that the first term in $\frac{1}{n(T-1)}\Sigma_{i=1}^{n}[a_{2i}^2-E(a_{2i}^2)]$ is $o_p(1)$. The proofs for the second and the fifth term are similar to that of the first element of the first term, the proofs for the third and fourth terms are similar to that of the second element of the first term and thus they are omitted.

For a_{3i}, we have:

$$
\begin{aligned}
&\frac{1}{n(T-1)}\Sigma_{i=1}^{n}[a_{3i}^2-E(a_{3i}^2)]\\
=&\frac{1}{n(T-1)}\Sigma_{i=1}^{n}(\Delta\epsilon_{2i}^2\Delta\xi_i^2-2\sigma_{\epsilon 0}^2\Delta\xi_i^2)+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}[\sigma_{\epsilon 0}^2(\Delta\xi_i^2-E(\Delta\xi_i)^2)]\\
&+\frac{1}{n(T-1)}\Sigma_{i=1}^{n}(F_{2,ii}^{++})^2[\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond}-E(\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond})^2]+\frac{1}{n(T-1)}\Sigma_{i=1}^{n}[(\Delta\epsilon_{i-}'\Delta y_{1i-}^{*})^2-E(\Delta\epsilon_{i-}'\Delta y_{1i-}^{*})^2]\\
&+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}[F_{2,ii}^{++}\Delta\epsilon_{2i}^2\Delta\xi_i\Delta y_{1i}^{\diamond}-E(F_{2,ii}^{++}\Delta\epsilon_{2i}^2\Delta\xi_i\Delta y_{1i}^{\diamond})]+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}[\sigma_{\epsilon 0}^2(F_{2,ii}^{++}\Delta\epsilon_{2i}\Delta\xi_i)]\\
&+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}[(\Delta\epsilon_{2i}\Delta\xi_i\Delta\epsilon_{i-}'\Delta y_{1i-}^{*})-E(\Delta\epsilon_{2i}\Delta\xi_i\Delta\epsilon_{i-}'\Delta y_{1i-}^{*})]\\
&+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}\{F_{2,ii}^{++}\sigma_{\epsilon 0}^2[\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond}-E(\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond})]\}\\
&+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}\{F_{2,ii}^{++}[\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond}\Delta\epsilon_{i-}'\Delta y_{1i-}^{*}-E(\Delta\epsilon_{2i}\Delta y_{1i}^{\diamond}\Delta\epsilon_{i-}'\Delta y_{1i-}^{*})]\}\\
&+\frac{2}{n(T-1)}\Sigma_{i=1}^{n}[F_{2,ii}^{++}\sigma_{\epsilon 0}^2(\Delta\epsilon_{i-}'\Delta y_{1i-}^{*}-E(\Delta\epsilon_{i-}'\Delta y_{1i-}^{*}))],
\end{aligned}
$$

where we subtracted and added $\frac{2}{n(T-1)}\Sigma_{i=1}^{n}\sigma_{\epsilon 0}^2\Delta\xi_i^2$ and used the fact that $F_{2,ii}^{++}$ is nonstochastic.

Note $\Delta\xi_i^2$ is $\Phi_{n,i-1}$-measurable, which implies that the first term is the average of a MD array. By Theorem 19.7, the first term is $o_p(1)$. The sixth term is thus also convergent. Note $\Delta\xi=(F_2^{++u}+F_2^{++l})\Delta y_1^{\diamond}=(F_2^{++u}+F_2^{++l})e^{\alpha_{30}W_3}e^{\alpha_{10}W_1}\Delta y_1$, so the second term equals $\frac{2}{n(T-1)}\Sigma_{i=1}^{n}\{\sigma_{\epsilon 0}^2$ $[\Delta y_1'e^{\alpha_{10}W_1'}e^{\alpha_{30}W_3'}(F_2^{++u}+F_2^{++l})'(F_2^{++u}+F_2^{++l})e^{\alpha_{30}W_3}e^{\alpha_{10}W_1}\Delta y_1-E(\Delta y_1'e^{\alpha_{10}W_1'}e^{\alpha_{30}W_3'}(F_2^{++u}+F_2^{++l})'(F_2^{++u}+F_2^{++l})e^{\alpha_{30}W_3}e^{\alpha_{10}W_1}\Delta y_1)]\}$. Since $e^{\alpha_{10}W_1'}e^{\alpha_{30}W_3'}(F_2^{++u}+F_2^{++l})'(F_2^{++u}+F_2^{++l})e^{\alpha_{30}W_3}$ $e^{\alpha_{10}W_1}$ is uniformly bounded in row and column sums by Lemma A.1, the convergence of the second term follows from Assumption F. For the third, fifth and eighth term, we can substitute $\Delta y_1^{\diamond}=e^{\alpha_{30}W_3}e^{\alpha_{10}W_1}\Delta y_0+e^{\alpha_{30}W_3}\Delta X_1\beta_0+\Delta\epsilon_1$ in them and prove they are convergent. For the fourth and tenth term, we can prove they are convergent using Assumption 6 since $\Delta\epsilon_{i-}$ are from $t=3$ to T and Δy_{1i-}^{*} is constructed based on Δy_1 which implies they are independent. For the seventh and ninth term, note $\Delta y_{1t}^{*}=F_t^{+}\Delta y_1=F_t^{+}\Delta y_0+F_t^{+}e^{-\alpha_{10}W_1}\Delta X_1\beta_0+F_t^{+}e^{-\alpha_{10}W_1}e^{-\alpha_{30}W_3}\Delta\epsilon_1$. The convergence follows.

4 Appendix D. Estimation of Submodels

The M-estimation proposed in the main paper can be modified to incorporate different submodels by getting rid of matrix exponential in dependent variable, lagged dependent variable and/or disturbance. In this part of the appendix we describe the estimation of submodels used in the Monte Carlo simulation.

MESDPS(1,0,0). By setting $\alpha_2 = 0$ and $\alpha_3 = 0$ we get MESDPS(1,0,0). Let $\boldsymbol{A}_0 = I_{t-1} \otimes A_0$ with $A_0 = \tau_0 + 1$. The first differenced model is given by

$$\boldsymbol{e^{\alpha_{10} W_1}} \Delta Y = \boldsymbol{A}_0 \Delta Y_{-1} + \Delta X \beta_0 + \Delta\epsilon, \tag{D.1}$$

and the conditional quasi loglikelihood is thus given by

$$\ell_{(1,0,0)}(\theta) = -\frac{n(T-1)}{2} log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \Delta\epsilon(\phi)' \boldsymbol{B}^{-1} \Delta\epsilon(\phi), \tag{D.2}$$

where $\theta = (\beta', \sigma_\epsilon^2, \tau, \alpha_1)'$, $\phi = (\beta', \tau, \alpha_1)'$, $\boldsymbol{B} = B \otimes I_n$ and $\Delta\epsilon(\phi) = \boldsymbol{e^{\alpha_1 W_1}} \Delta Y - \boldsymbol{A} \Delta Y_{-1} - \Delta X \beta$. Given $\zeta = (\tau, \alpha_1)'$, the constrained estimators of β and σ_ϵ^2 are given by

$$\tilde{\beta}(\zeta) = (\Delta X' \boldsymbol{B}^{-1} \Delta X)^{-1} \Delta X' \boldsymbol{B}^{-1} (\boldsymbol{e^{\alpha_1 W_1}} \Delta Y - \boldsymbol{A} \Delta Y_{-1}), \tag{D.3}$$

$$\tilde{\sigma}_\epsilon^2(\zeta) = \frac{1}{n(T-1)} \Delta\tilde{\epsilon}(\zeta)' \boldsymbol{B}^{-1} \Delta\tilde{\epsilon}(\zeta), \tag{D.4}$$

where $\Delta\tilde{\epsilon}(\zeta) = \boldsymbol{e^{\alpha_1 W_1}} \Delta Y - \boldsymbol{A} \Delta Y_{-1} - \Delta X \tilde{\beta}(\alpha_1)$. Substituting them back into (D.2), ignoring constants, the concentrated log-likelihood function is derived as:

$$l^c_{(1,0,0)}(\zeta) = -log[\Delta\tilde{\epsilon}(\zeta)' \boldsymbol{B}^{-1} \Delta\tilde{\epsilon}(\zeta)] \tag{D.5}$$

Maximizing (D.5) gives us CQMLE $\tilde{\zeta}$ and then CQMLEs $\tilde{\beta} = \tilde{\beta}(\tilde{\zeta})$ and $\tilde{\sigma}_\epsilon^2 = \tilde{\sigma}_\epsilon^2(\tilde{\zeta})$.

The conditional quasi score (CQS) function corresponding to (2.11) in the paper is given by

$$S_{(1,0,0)}(\theta) = \begin{cases} \frac{1}{\sigma_\epsilon^2} \Delta X' \boldsymbol{B}^{-1} \Delta\epsilon(\phi), \\ -\frac{n(T-1)}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \Delta\epsilon(\phi)' \boldsymbol{B}^{-1} \Delta\epsilon(\phi), \\ \frac{1}{\sigma_\epsilon^2} \Delta\epsilon(\phi)' \boldsymbol{B}^{-1} \Delta Y_{-1}, \\ -\frac{1}{\sigma_\epsilon^2} \Delta\epsilon(\phi)' \boldsymbol{B}^{-1} \boldsymbol{W}_1 \boldsymbol{e^{\alpha_1 W_1}} \Delta Y. \end{cases} \tag{$D.6$}$$

Note here the expectations in Lemma 2.1 reduce to $E(\Delta Y \Delta\epsilon') = -\sigma_{\epsilon 0}^2 \boldsymbol{e^{-\alpha_{10} W_1}} \boldsymbol{D}_0$ and $E(\Delta Y_{-1} \Delta\epsilon') = -\sigma_{\epsilon 0}^2 \boldsymbol{e^{-\alpha_{10} W_1}} \boldsymbol{D}_{-1,0}$, where

$$\boldsymbol{D}_0 = \begin{pmatrix} A_0 e^{-\alpha_{10} W_1} - 2I_n & I_n & \dots & \dots & 0 \\ (A_0 e^{-\alpha_{10} W_1} - I_n)^2 & A_0 e^{-\alpha_{10} W_1} - 2I_n & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & I_n \\ (A_0 e^{-\alpha_{10} W_1})^{T-3} (A_0 e^{-\alpha_{10} W_1} - I_n)^2 & \dots & \dots & (A_0 e^{-\alpha_{10} W_1} - I_n)^2 & A_0 e^{-\alpha_{10} W_1} - 2I_n \end{pmatrix},$$

and $$\boldsymbol{D}_{-1,0} = \begin{pmatrix} I_n & 0 & \dots & \dots & 0 \\ A_0 e^{-\alpha_{10} W_1} - 2I_n & I_n & \ddots & \dots & \vdots \\ (A_0 e^{-\alpha_{10} W_1} - I_n)^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ (A_0 e^{-\alpha_{10} W_1})^{T-4} (A_0 e^{-\alpha_{10} W_1} - I_n)^2 & \dots & (A_0 e^{-\alpha_{10} W_1} - I_n)^2 & A_0 e^{-\alpha_{10} W_1} - 2I_n & I_n \end{pmatrix}.$$

The adjusted quasi score (AQS) corresponding to (2.15) in the paper is thus given by

$$S^*_{(1,0,0)}(\theta) = \begin{cases} \frac{1}{\sigma_\epsilon^2}\Delta X'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ -\frac{n(T-1)}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ \frac{1}{\sigma_\epsilon^2}\Delta\epsilon(\phi)'\Sigma^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\sigma_\epsilon^2}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \end{cases} \qquad (D.7)$$

To derive the M-estimator, the constrained M-estimators of β and σ_ϵ^2 are first solved as

$$\hat{\beta}_M(\zeta) = (\Delta X'\boldsymbol{B}^{-1}\Delta X)^{-1}\Delta X'\boldsymbol{B}^{-1}(\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1}), \qquad (D.8)$$

$$\hat{\sigma}^2_{\epsilon,M}(\zeta) = \frac{1}{n(T-1)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta\hat{\epsilon}(\zeta), \qquad (D.9)$$

where $\Delta\hat{\epsilon}(\zeta) = \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\hat{\beta}_M(\zeta)$. Then $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$ are substituted back into the third and fourth elements of the AQS function to get the concentrated AQS function:

$$S^{*c}_{(1,0,0)}(\zeta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1) \end{cases}$$

The unconstrained M-estimators $\hat{\tau}_M$ and $\hat{\alpha}_{1M}$ can be solved by letting $S^{*c}_{(1,0,0)}(\zeta) = 0$ and consequently the unconstrained M-estimators $\hat{\beta}_M = \hat{\beta}_M(\hat{\zeta}_M)$ and $\hat{\sigma}^2_{\epsilon,M} = \hat{\sigma}^2_{\epsilon,M}(\hat{\zeta}_M)$.

MESDPS(0,1,0). By setting $\alpha_1 = 0$ and $\alpha_3 = 0$, MESDPS(0,1,0) appears. Let $\boldsymbol{A}_0 = I_{T-1} \otimes A_0$ with $A_0 = \tau_0 I_n + e^{\alpha_{20}W_2}$. The first differenced model is given by

$$\Delta Y = \boldsymbol{A}_0\Delta Y_{-1} + \Delta X\beta_0 + \Delta\epsilon, \qquad (D.10)$$

and the conditional quasi loglikelihood is subsequently given by

$$\ell_{(0,1,0)}(\theta) = -\frac{n(T-1)}{2}log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \qquad (D.11)$$

where $\theta = (\beta', \sigma_\epsilon^2, \tau, \alpha_2)'$, $\phi = (\beta', \tau, \alpha_2)'$, $\boldsymbol{B} = B \otimes I_n$ and $\Delta\epsilon(\phi) = \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\beta$. Given $\zeta = (\tau, \alpha_2)'$, the constrained estimators of β and σ_ϵ^2 are given by

$$\tilde{\beta}(\zeta) = (\Delta X'\boldsymbol{B}^{-1}\Delta X)^{-1}\Delta X'\boldsymbol{B}^{-1}(\Delta Y - \boldsymbol{A}\Delta Y_{-1}), \qquad (D.12)$$

$$\tilde{\sigma}^2_\epsilon(\zeta) = \frac{1}{n(T-1)}\Delta\tilde{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta\tilde{\epsilon}(\zeta), \qquad (D.13)$$

where $\Delta\tilde{\epsilon}(\zeta) = \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\tilde{\beta}(\zeta)$. Substituting them back into (D.11), ignoring constants, the concentrated log-likelihood function is:

$$l^c_{(0,1,0)}(\zeta) = -log[\Delta\tilde{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta\tilde{\epsilon}(\zeta)] \qquad (D.14)$$

Maximizing (D.14) gives us CQMLE $\tilde{\zeta}$ and then CQMLEs $\tilde{\beta} = \tilde{\beta}(\tilde{\zeta})$ and $\tilde{\sigma}^2_\epsilon = \tilde{\sigma}^2_\epsilon(\tilde{\zeta})$.

The conditional quasi score (CQS) function corresponding to (2.11) in the paper is given by

$$S_{(0,1,0)}(\theta) = \begin{cases} \frac{1}{\sigma_\epsilon^2}\Delta X'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ -\frac{n(T-1)}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ \frac{1}{\sigma_\epsilon^2}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma_\epsilon^2}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}. \end{cases} \qquad (D.15)$$

In this case the expectation in Lemma 2.1 reduces to $E(\Delta Y_{-1}\Delta\epsilon') = -\sigma^2_{\epsilon 0}\boldsymbol{D}_{-1,0}$, where

$$\boldsymbol{D}_{-1,0} = \begin{pmatrix} I_n & 0 & \dots & \dots & 0 \\ A_0 - 2I_n & I_n & \ddots & \dots & \vdots \\ (A_0 - I_n)^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ (A_0)^{T-4}(A_0 - I_n)^2 & \dots & (A_0 - I_n)^2 & A_0 - 2I_n & I_n \end{pmatrix},$$

$$\boldsymbol{D}_0 = \begin{pmatrix} A_0 - 2I_n & I_n & \dots & \dots & 0 \\ (A_0 - I_n)^2 & A_0 - 2I_n & \ddots & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & I_n \\ A_0^{T-3}(A_0 - I_n)^2 & \dots & \dots & (A_0 - I_n)^2 & A_0 - 2I_n \end{pmatrix}.$$

The adjusted quasi score (AQS) corresponding to (2.15) in the paper is then given by

$$S^*_{(0,1,0)}(\theta) = \begin{cases} \frac{1}{\sigma_\epsilon^2} \Delta X' \boldsymbol{B}^{-1} \Delta\epsilon(\phi), \\ -\frac{n(T-1)}{2\sigma_\epsilon^2} + \frac{1}{2\sigma_\epsilon^4} \Delta\epsilon(\phi)' \boldsymbol{B}^{-1} \Delta\epsilon(\phi), \\ \frac{1}{\sigma_\epsilon^2} \Delta\epsilon(\phi)' \Sigma^{-1} \Delta Y_{-1} + tr(\boldsymbol{D}_{-1} \boldsymbol{B}^{-1}), \\ \frac{1}{\sigma_\epsilon^2} \Delta\epsilon(\phi)' \Sigma^{-1} \boldsymbol{W}_2 \boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}} \Delta Y_{-1} + tr(\boldsymbol{D}_{-1} \boldsymbol{B}^{-1} \boldsymbol{W}_2 \boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}). \end{cases} \tag{D.16}$$

To derive the M-estimator, the constrained M-estimators of β and σ_ϵ^2 are first solved as

$$\hat{\beta}_M(\zeta) = (\Delta X' \boldsymbol{B}^{-1} \Delta X)^{-1} \Delta X' \boldsymbol{B}^{-1} (\Delta Y - \boldsymbol{A} \Delta Y_{-1}), \tag{D.17}$$

$$\hat{\sigma}^2_{\epsilon,M}(\zeta) = \frac{1}{n(T-1)} \Delta\hat{\epsilon}(\zeta)' \boldsymbol{B}^{-1} \Delta\hat{\epsilon}(\zeta), \tag{D.18}$$

where $\Delta\hat{\epsilon}(\zeta) = \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X \hat{\beta}_M(\zeta)$. Then $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$ are substituted back into the third and fourth elements of the AQS function:

$$S^{*c}_{(0,1,0)}(\zeta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)} \Delta\hat{\epsilon}(\zeta)' \Sigma^{-1} \Delta Y_{-1} + tr(\boldsymbol{D}_{-1} \boldsymbol{B}^{-1}), \\ \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)} \Delta\hat{\epsilon}(\zeta)' \Sigma^{-1} \boldsymbol{W}_2 \boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}} \Delta Y_{-1} + tr(\boldsymbol{D}_{-1} \boldsymbol{B}^{-1} \boldsymbol{W}_2 \boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}), \end{cases}$$

which is the concentrated AQS function. The unconstrained M-estimators $\hat{\zeta}_M$ can be solved by letting $S^{*c}_{(0,1,0)}(\zeta) = 0$. The unconstrained M-estimators are then derived as $\hat{\beta}_M = \hat{\beta}_M(\hat{\zeta}_M)$ and $\hat{\sigma}^2_{\epsilon,M} = \hat{\sigma}^2_{\epsilon,M}(\hat{\zeta}_M)$.

MESDPS(1,1,0). By setting $\alpha_3 = 0$, MESDPS(1,1,0) appears. Again let $\boldsymbol{A}_0 = I_{T-1} \otimes A_0$ with $A_0 = \tau_0 I_n + e^{\alpha_{20} W_2}$. The first differenced model is given by

$$\boldsymbol{e}^{\boldsymbol{\alpha_{10} W_1}} \Delta Y = \boldsymbol{A}_0 \Delta Y_{-1} + \Delta X \beta_0 + \Delta\epsilon, \tag{D.19}$$

and the conditional quasi loglikelihood is thus given by

$$\ell_{(1,1,0)}(\theta) = -\frac{n(T-1)}{2} log(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} \Delta\epsilon(\phi)' \boldsymbol{B}^{-1} \Delta\epsilon(\phi), \tag{D.20}$$

where $\theta = (\beta', \sigma_\epsilon^2, \tau, \alpha_1, \alpha_2)'$, $\phi = (\beta', \tau, \alpha_1, \alpha_2)'$, $\boldsymbol{B} = B \otimes I_n$ and $\Delta\epsilon(\phi) = \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\beta$. Given $\zeta = (\tau, \alpha_1, \alpha_2)'$, the constrained estimators of β and σ_ϵ^2 are given by

$$\tilde{\beta}(\zeta) = (\Delta X' \boldsymbol{B}^{-1} \Delta X)^{-1} \Delta X' \boldsymbol{B}^{-1} (\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}\Delta Y_{-1}), \tag{D.21}$$

$$\tilde{\sigma}^2_\epsilon(\zeta) = \frac{1}{n(T-1)} \Delta\tilde{\epsilon}(\zeta)' \boldsymbol{B}^{-1} \Delta\tilde{\epsilon}(\zeta), \tag{D.22}$$

where $\Delta\tilde{\epsilon}(\zeta) = \boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}} \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\tilde{\beta}(\zeta)$. Substituting them back into (D.20), ignoring constants, the concentrated log-likelihood function is given by:

$$l^c_{(1,1,0)}(\zeta) = -log[\Delta\tilde{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta\tilde{\epsilon}(\zeta)] \tag{D.23}$$

Maximizing (D.23) gives us CQMLE $\tilde{\zeta}$, with the implied CQMLEs $\tilde{\beta} = \tilde{\beta}(\tilde{\zeta})$ and $\tilde{\sigma}^2_\epsilon = \tilde{\sigma}^2_\epsilon(\tilde{\zeta})$.

Correspondingly, the conditional quasi score (CQS) function (2.11) in the paper becomes

$$S_{(1,1,0)}(\theta) = \begin{cases} \frac{1}{\sigma^2_\epsilon}\Delta X'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ -\frac{n(T-1)}{2\sigma^2_\epsilon} + \frac{1}{2\sigma^4_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ \frac{1}{\sigma^2_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta Y_{-1}, \\ -\frac{1}{\sigma^2_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y, \\ \frac{1}{\sigma^2_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}. \end{cases} \tag{D.24}$$

Here the expectations in Lemma 2.1 are simplified to $E(\Delta Y\Delta\epsilon') = -\sigma^2_{\epsilon 0}e^{-\boldsymbol{\alpha_{10}W_1}}\boldsymbol{D}_0$ and $E(\Delta Y_{-1}\Delta\epsilon')$ $= -\sigma^2_{\epsilon 0}e^{-\boldsymbol{\alpha_{10}W_1}}\boldsymbol{D}_{-1,0}$, where $\boldsymbol{D}_0$ and $\boldsymbol{D}_{-1,0}$ have the same expression as those in Lemma 2.1.

The adjusted quasi score (AQS) in (2.15) in the main paper is then reduced to

$$S^*_{(1,1,0)}(\theta) = \begin{cases} \frac{1}{\sigma^2_\epsilon}\Delta X'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ -\frac{n(T-1)}{2\sigma^2_\epsilon} + \frac{1}{2\sigma^4_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\Delta\epsilon(\phi), \\ \frac{1}{\sigma^2_\epsilon}\Delta\epsilon(\phi)'\Sigma^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}e^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\sigma^2_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ \frac{1}{\sigma^2_\epsilon}\Delta\epsilon(\phi)'\boldsymbol{B}^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21}), \end{cases} \tag{$D.25$}$$

where $\boldsymbol{W}_{21} = \boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}e^{-\boldsymbol{\alpha_1 W_1}}$.To derive the M-estimator, the constrained M-estimators of β and σ^2_ϵ are first solved as

$$\hat{\beta}_M(\zeta) = (\Delta X'\boldsymbol{B}^{-1}\Delta X)^{-1}\Delta X'\boldsymbol{B}^{-1}(e^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1}), \tag{D.26}$$

$$\hat{\sigma}^2_{\epsilon,M}(\zeta) = \tfrac{1}{n(T-1)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta\hat{\epsilon}(\zeta), \tag{D.27}$$

where $\Delta\hat{\epsilon}(\zeta) = e^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\hat{\beta}_M(\zeta)$. Then $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$ are substituted back into the rest of the AQS function to get the concentrated AQS function:

$$S^{*c}_{(1,1,0)}(\zeta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}e^{-\boldsymbol{\alpha_1 W_1}}), \\ -\frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\boldsymbol{W}_1 e^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{\epsilon}(\zeta)'\boldsymbol{B}^{-1}\boldsymbol{W}_2 e^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_{21}). \end{cases}$$

The unconstrained M-estimators $\hat{\zeta}_M$ can be solved by letting $S^{*c}_{(1,1,0)}(\zeta) = 0$ and then $\hat{\beta}_M = \hat{\beta}_M(\hat{\zeta}_M)$ and $\hat{\sigma}^2_{\epsilon,M} = \hat{\sigma}^2_{\epsilon,M}(\hat{\zeta}_M)$.

MESDPS(1,0,1). By setting $\alpha_2 = 0$, we have MESDPS(1,0,1). Here let $\boldsymbol{A}_0 = I_{t-1}\otimes A_0$ with $A_0 = \tau_0 + 1$. The first differenced model is given by

$$e^{\boldsymbol{\alpha_{10}W_1}}\Delta Y = \boldsymbol{A}_0\Delta Y_{-1} + \Delta X\beta_0 + \Delta u,\ e^{\boldsymbol{\alpha_{30}W_3}}\Delta u = \Delta\epsilon, \tag{D.28}$$

and the conditional quasi loglikelihood is thus given by

$$\ell_{(1,0,1)}(\theta) = -\tfrac{n(T-1)}{2}log(\sigma^2_\epsilon) - \tfrac{1}{2\sigma^2_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \tag{D.29}$$

where $\theta = (\beta', \sigma^2_\epsilon, \tau, \alpha_1, \alpha_3)'$, $\phi = (\beta', \tau, \alpha_1)'$, $\Sigma(\alpha_3) = B\otimes e^{-\alpha_3 W_3}e^{-\alpha_3 W_3'}$, and $\Delta u(\phi) = e^{\boldsymbol{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\beta$. Given $\zeta = (\tau, \alpha_1, \alpha_3)'$, the constrained estimators of β and σ^2_ϵ are given by

$$\tilde{\beta}(\zeta) = (\Delta X'\Sigma(\alpha_3)^{-1}\Delta X)^{-1}\Delta X'\Sigma(\alpha_3)^{-1}(\boldsymbol{e^{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1}), \tag{D.30}$$

$$\tilde{\sigma}^2_\epsilon(\zeta) = \tfrac{1}{n(T-1)}\Delta\tilde{\epsilon}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\tilde{\epsilon}(\zeta), \tag{D.31}$$

where $\Delta\tilde{\epsilon}(\zeta) = \boldsymbol{e^{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\tilde{\beta}(\zeta)$. Substituting them back into (D.29), ignoring constants, the concentrated log-likelihood function is given by:

$$l^c_{(1,0,1)}(\zeta) = -log[\Delta\tilde{u}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\tilde{u}(\zeta)] \tag{D.32}$$

Maximizing (D.32) gives us CQMLE $\tilde{\zeta}$ and then the implied CQMLEs $\tilde{\beta} = \tilde{\beta}(\tilde{\zeta})$ and $\tilde{\sigma}^2_\epsilon = \tilde{\sigma}^2_\epsilon(\tilde{\zeta})$.

Correspondingly, the conditional quasi score (CQS) function (2.11) in the paper becomes

$$S_{(1,0,1)}(\theta) = \begin{cases} \frac{1}{\sigma^2_\epsilon}\Delta X'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \\ -\frac{n(T-1)}{2\sigma^2_\epsilon} + \frac{1}{2\sigma^4_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \\ \frac{1}{\sigma^2_\epsilon}\Delta u(\phi)'\Sigma^{-1}\Delta Y_{-1}, \\ -\frac{1}{\sigma^2_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\boldsymbol{W}_1\boldsymbol{e^{\alpha_1 W_1}}\Delta Y, \\ -\frac{1}{2\sigma^2_\epsilon}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi). \end{cases} \tag{D.33}$$

Now the expectations in Lemma 2.1 become $E(\Delta Y\Delta\epsilon') = -\sigma^2_{\epsilon 0}\boldsymbol{e^{-\alpha_{10}W_1}}\boldsymbol{D}_0\boldsymbol{e^{-\alpha_{30}W_3}}$ and $E(\Delta Y_{-1}\Delta\epsilon')$ $= -\sigma^2_{\epsilon 0}\boldsymbol{e^{-\alpha_{10}W_1}}\boldsymbol{D}_{-1,0}\boldsymbol{e^{-\alpha_{30}W_3}}$, where

$$\boldsymbol{D}_0 = \begin{pmatrix} A_0e^{-\alpha_{10}W_1} - 2I_n & I_n & \cdots & \cdots & 0 \\ (A_0e^{-\alpha_{10}W_1} - I_n)^2 & A_0e^{-\alpha_{10}W_1} - 2I_n & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & I_n \\ (A_0e^{-\alpha_{10}W_1})^{T-3}(A_0e^{-\alpha_{10}W_1} - I_n)^2 & \cdots & \cdots & (A_0e^{-\alpha_{10}W_1} - I_n)^2 & A_0e^{-\alpha_{10}W_1} - 2I_n \end{pmatrix},$$

and $$\boldsymbol{D}_{-1,0} = \begin{pmatrix} I_n & 0 & \cdots & \cdots & 0 \\ A_0e^{-\alpha_{10}W_1} - 2I_n & I_n & \ddots & \cdots & \vdots \\ (A_0e^{-\alpha_{10}W_1} - I_n)^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ (A_0e^{-\alpha_{10}W_1})^{T-4}(A_0e^{-\alpha_{10}W_1} - I_n)^2 & \cdots & (A_0e^{-\alpha_{10}W_1} - I_n)^2 & A_0e^{-\alpha_{10}W_1} - 2I_n & I_n \end{pmatrix}.$$

The adjusted quasi score (AQS) in (2.15) in the main paper is then reduced to

$$S^*_{(1,0,1)}(\theta) = \begin{cases} \frac{1}{\sigma^2_\epsilon}\Delta X'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \\ -\frac{n(T-1)}{2\sigma^2_\epsilon} + \frac{1}{2\sigma^4_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \\ \frac{1}{\sigma^2_\epsilon}\Delta u(\phi)'\Sigma^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}), \\ -\frac{1}{\sigma^2_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\boldsymbol{W}_1\boldsymbol{e^{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{W}_1), \\ -\frac{1}{2\sigma^2_\epsilon}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi). \end{cases} \tag{$D.34$}$$

To derive the M-estimator, the constrained M-estimators of β and σ^2_ϵ are first solved as

$$\hat{\beta}_M(\zeta) = (\Delta X'\Sigma(\alpha_3)^{-1}\Delta X)^{-1}\Delta X'\Sigma(\alpha_3)^{-1}(\boldsymbol{e^{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1}), \tag{D.35}$$

$$\hat{\sigma}^2_{\epsilon,M}(\zeta) = \tfrac{1}{n(T-1)}\Delta\hat{\epsilon}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\hat{\epsilon}(\zeta), \tag{D.36}$$

where $\Delta\hat{\epsilon}(\zeta) = \boldsymbol{e^{\alpha_1 W_1}}\Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\hat{\beta}_M(\zeta)$. Then $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$ are substituted back into the rest of the AQS function to get the concentrated AQS function:

$$S^{*c}_{(1,0,1)}(\zeta) = \begin{cases} \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)} \Delta\hat{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1} + tr(\boldsymbol{D}_{-1,0}\boldsymbol{B}^{-1}\boldsymbol{e}^{-\boldsymbol{\alpha_{10}W_1}}), \\ -\frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)} \Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_1\boldsymbol{e}^{\boldsymbol{\alpha_1 W_1}}\Delta Y - tr(\boldsymbol{DB}^{-1}\boldsymbol{W}_1), \\ -\frac{1}{2\hat{\sigma}^2_{\epsilon,M}(\zeta)} \Delta\hat{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\hat{u}(\zeta). \end{cases}$$

The unconstrained M-estimators $\hat{\zeta}_M$ can be solved by letting $S^{*c}_{(1,0,1)}(\zeta) = 0$ and then $\hat{\beta}_M = \hat{\beta}_M(\hat{\zeta}_M)$ and $\hat{\sigma}^2_{\epsilon,M} = \hat{\sigma}^2_{\epsilon,M}(\hat{\zeta}_M)$.

MESDPS(0,1,1). By setting $\alpha_1 = 0$, we have MESDPS(0,1,1). Let $\boldsymbol{A}_0 = I_{T-1}\otimes A_0$ with $A_0 = \tau_0 I_n + e^{\alpha_{20}W_2}$. The first differenced model is given by

$$\Delta Y = \boldsymbol{A}_0\Delta Y_{-1} + \Delta X\beta_0 + \Delta u,\ \boldsymbol{e}^{\boldsymbol{\alpha_{30}W_3}}\Delta u = \Delta\epsilon, \tag{D.37}$$

and the conditional quasi loglikelihood is thus given by

$$\ell_{(1,0,1)}(\theta) = -\tfrac{n(T-1)}{2} log(\sigma^2_\epsilon) - \tfrac{1}{2\sigma^2_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \tag{D.38}$$

where $\theta = (\beta', \sigma^2_\epsilon, \tau, \alpha_2, \alpha_3)'$, $\phi = (\beta', \tau, \alpha_2)'$, $\Sigma(\alpha_3) = B\otimes e^{-\alpha_3 W_3}e^{-\alpha_3 W_3'}$, and $\Delta u(\phi) = \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\beta$. Given $\zeta = (\tau, \alpha_2, \alpha_3)'$, the constrained estimators of β and σ^2_ϵ are given by

$$\tilde{\beta}(\zeta) = (\Delta X'\Sigma(\alpha_3)^{-1}\Delta X)^{-1}\Delta X'\Sigma(\alpha_3)^{-1}(\Delta Y - \boldsymbol{A}\Delta Y_{-1}), \tag{D.39}$$

$$\tilde{\sigma}^2_\epsilon(\zeta) = \tfrac{1}{n(T-1)}\Delta\tilde{u}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\tilde{u}(\zeta), \tag{D.40}$$

where $\Delta\tilde{u}(\zeta) = \Delta Y - \boldsymbol{A}\Delta Y_{-1} - \Delta X\tilde{\beta}(\zeta)$. Substituting them back into (D.38), ignoring constants, the concentrated log-likelihood function is given by:

$$l^c_{(0,1,1)}(\zeta) = -log[\Delta\tilde{u}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\tilde{u}(\zeta)] \tag{D.41}$$

Maximizing (D.41) gives us CQMLE $\tilde{\zeta}$ and then the implied CQMLEs $\tilde{\beta} = \tilde{\beta}(\tilde{\zeta})$ and $\tilde{\sigma}^2_\epsilon = \tilde{\sigma}^2_\epsilon(\tilde{\zeta})$.

Correspondingly, the conditional quasi score (CQS) function (2.11) in the paper becomes

$$S_{(0,1,1)}(\theta) = \begin{cases} \frac{1}{\sigma^2_\epsilon}\Delta X'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \\ -\frac{n(T-1)}{2\sigma^2_\epsilon} + \frac{1}{2\sigma^4_\epsilon}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi), \\ \frac{1}{\sigma^2_\epsilon}\Delta u(\phi)'\Sigma^{-1}\Delta Y_{-1}, \\ \frac{1}{\sigma^2_\epsilon}\Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e}^{\boldsymbol{\alpha_2 W_2}}\Delta Y_{-1}, \\ -\frac{1}{2\sigma^2_\epsilon}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi). \end{cases} \tag{D.42}$$

Now the expectations in Lemma 2.1 become $E(\Delta Y\Delta\epsilon') = -\sigma^2_{\epsilon 0}\boldsymbol{D}_0\boldsymbol{e}^{-\boldsymbol{\alpha_{30}W_3}}$ and $E(\Delta Y_{-1}\Delta\epsilon')$ $= -\sigma^2_{\epsilon 0}\boldsymbol{D}_{-1,0}\boldsymbol{e}^{-\boldsymbol{\alpha_{30}W_3}}$, where

$$\boldsymbol{D}_{-1,0} = \begin{pmatrix} I_n & 0 & \cdots & \cdots & 0 \\ A_0 - 2I_n & I_n & \ddots & \cdots & \vdots \\ (A_0 - I_n)^2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ (A_0)^{T-4}(A_0 - I_n)^2 & \cdots & (A_0 - I_n)^2 & A_0 - 2I_n & I_n \end{pmatrix},$$

$$\boldsymbol{D}_0 = \begin{pmatrix} A_0 - 2I_n & I_n & \cdots & \cdots & 0 \\ (A_0 - I_n)^2 & A_0 - 2I_n & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & I_n \\ A_0^{T-3}(A_0 - I_n)^2 & \cdots & \cdots & (A_0 - I_n)^2 & A_0 - 2I_n \end{pmatrix}.$$

The adjusted quasi score (AQS) in (2.15) in the main paper is then reduced to

$$S^{*}_{(0,1,1)}(\theta)=\begin{cases}\frac{1}{\sigma_\epsilon^2}\Delta X'\Sigma(\alpha_3)^{-1}\Delta u(\phi),\\ -\frac{n(T-1)}{2\sigma_\epsilon^2}+\frac{1}{2\sigma_\epsilon^4}\Delta u(\phi)'\Sigma(\alpha_3)^{-1}\Delta u(\phi),\\ \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\Delta Y_{-1}+tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{e^{-\alpha_1 W_1}}),\\ \frac{1}{\sigma_\epsilon^2}\Delta u(\phi)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}\Delta Y_{-1}+tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}),\\ -\frac{1}{2\sigma_\epsilon^2}\Delta u(\phi)'(B^{-1}\otimes E_3)\Delta u(\phi).\end{cases} \quad (D.43)$$

To derive the M-estimator, the constrained M-estimators of β and σ_ϵ^2 are first solved as

$$\hat{\beta}_M(\zeta)=(\Delta X'\Sigma(\alpha_3)^{-1}\Delta X)^{-1}\Delta X'\Sigma(\alpha_3)^{-1}(\Delta Y-\boldsymbol{A}\Delta Y_{-1}), \quad (D.44)$$

$$\hat{\sigma}^2_{\epsilon,M}(\zeta)=\frac{1}{n(T-1)}\Delta\hat{u}(\zeta)'\Sigma(\alpha_3)^{-1}\Delta\hat{u}(\zeta), \quad (D.45)$$

where $\Delta\hat{u}(\zeta)=\Delta Y-\boldsymbol{A}\Delta Y_{-1}-\Delta X\hat{\beta}_M(\zeta)$. Then $\hat{\beta}_M(\zeta)$ and $\hat{\sigma}^2_{\epsilon,M}(\zeta)$ are substituted back into the rest of the AQS function to get the concentrated AQS function:

$$S^{*c}_{(0,1,1)}(\zeta)=\begin{cases}\frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\Delta Y_{-1}+tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}),\\ \frac{1}{\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'\Sigma^{-1}\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}\Delta Y_{-1}+tr(\boldsymbol{D}_{-1}\boldsymbol{B}^{-1}\boldsymbol{W}_2\boldsymbol{e^{\alpha_2 W_2}}),\\ -\frac{1}{2\hat{\sigma}^2_{\epsilon,M}(\zeta)}\Delta\hat{u}(\zeta)'(B^{-1}\otimes E_3)\Delta\hat{u}(\zeta).\end{cases}$$

The unconstrained M-estimators $\hat{\rho}_M$, $\hat{\alpha}_{2M}$ and $\hat{\alpha}_{3M}$ can be solved by letting $S^{*c}_{(0,1,1)}(\zeta)=0$ and consequently the unconstrained M-estimators $\hat{\beta}_M=\hat{\beta}_M(\hat{\zeta}_M)$ and $\hat{\sigma}^2_{\epsilon,M}=\hat{\sigma}^2_{\epsilon,M}(\hat{\zeta}_M)$.

Table 1a. Empirical mean of CQMLE and M-estimator, MESDPS(1,1,0)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9848	0.9990	0.9862	1.0003	0.9992	1.0000	0.9993	1.0001
	1	0.9363	0.9511	0.9679	0.9825	0.9898	0.9907	0.9910	0.9918
	1.5	1.4853	1.5002	1.4858	1.5006	1.4999	1.5000	1.4999	1.5000
	1.1	1.0986	1.1005	1.0982	1.1000	1.1001	1.1000	1.1001	1.1000
	−0.5	−0.4987	−0.5005	−0.4993	−0.5006	−0.5000	−0.5000	−0.5000	−0.5000
2	1	0.9832	0.9980	0.9862	1.0004	1.0001	1.0009	0.9997	1.0004
	1	0.9502	0.9654	0.9640	0.9783	0.9843	0.9852	0.9907	0.9915
	1.5	1.4854	1.5007	1.4856	1.5003	1.4999	1.5000	1.4999	1.5000
	1.1	1.0993	1.1014	1.0987	1.1003	1.1001	1.1000	1.1001	1.1000
	−0.5	−0.4986	−0.4998	−0.4984	−0.4998	−0.5000	−0.5000	−0.5000	−0.5000
3	1	0.9871	1.0015	0.9859	0.9998	0.9971	0.9978	1.0006	1.0013
	1	0.9452	0.9608	0.9617	0.9762	0.9848	0.9857	0.9890	0.9898
	1.5	1.4856	1.5008	1.4861	1.5008	1.4999	1.5000	1.4999	1.5000
	1.1	1.0987	1.1006	1.0983	1.1000	1.1001	1.1000	1.1001	1.1000
	−0.5	−0.4993	−0.5014	−0.4987	−0.5003	−0.5000	−0.5000	−0.5000	−0.5000
1	1	0.9593	1.0004	0.9583	0.9996	0.9933	0.9992	0.9945	1.0003
	1	0.9184	0.9623	0.9400	0.9822	0.9813	0.9879	0.9895	0.9957
	0	−0.0544	0.0022	−0.0549	0.0011	−0.0104	−0.0003	−0.0096	0.0000
	1.1	1.0513	1.1063	1.0486	1.1032	1.1135	1.1017	1.1120	1.1007
	−0.5	−0.4860	−0.5004	−0.4861	−0.4996	−0.5036	−0.5007	−0.5030	−0.5002
2	1	0.9601	1.0014	0.9596	1.0007	0.9932	0.9991	0.9946	1.0005
	1	0.9221	0.9662	0.9446	0.9870	0.9801	0.9865	0.9857	0.9918
	0	−0.0542	0.0025	−0.0546	0.0013	−0.0098	0.0002	−0.0097	−0.0002
	1.1	1.0479	1.1041	1.0484	1.1030	1.1121	1.1002	1.1120	1.1009
	−0.5	−0.4909	−0.5039	−0.4860	−0.4994	−0.5031	−0.5003	−0.5029	−0.5001
3	1	0.9610	1.0025	0.9575	0.9984	0.9950	1.0010	0.9932	0.9991
	1	0.9165	0.9608	0.9409	0.9838	0.9772	0.9837	0.9828	0.9890
	0	−0.0531	0.0035	−0.0542	0.0018	−0.0094	0.0005	−0.0095	0.0001
	1.1	1.0470	1.1029	1.0481	1.1029	1.1115	1.0998	1.1119	1.1007
	−0.5	−0.4883	−0.5013	−0.4870	−0.5005	−0.5031	−0.5003	−0.5029	−0.5000
1	1	0.9865	0.9977	0.9878	0.9991	0.9981	0.9995	0.9990	1.0005
	1	0.9544	0.9661	0.9784	0.9900	0.9860	0.9875	0.9913	0.9928
	−1.5	−1.5307	−1.5001	−1.5295	−1.4988	−1.5098	−1.4998	−1.5095	−1.4994
	1.1	1.1007	1.1031	1.1013	1.1027	1.0953	1.0986	1.0982	1.1017
	−0.5	−0.4636	−0.5023	−0.4662	−0.5038	−0.4863	−0.4982	−0.4879	−0.5002
2	1	0.9892	1.0003	0.9902	1.0013	0.9980	0.9995	0.9978	0.9993
	1	0.9568	0.9685	0.9728	0.9840	0.9830	0.9845	0.9923	0.9939
	−1.5	−1.5285	−1.4978	−1.5302	−1.5001	−1.5096	−1.4995	−1.5102	−1.5001
	1.1	1.1012	1.1027	1.0985	1.1002	1.0958	1.0992	1.0958	1.0991
	−0.5	−0.4664	−0.5038	−0.4613	−0.4981	−0.4870	−0.4989	−0.4869	−0.4990
3	1	0.9854	0.9964	0.9896	1.0006	0.9990	1.0006	0.9995	1.0010
	1	0.9597	0.9718	0.9694	0.9808	0.9878	0.9894	0.9900	0.9915
	−1.5	−1.5313	−1.5007	−1.5301	−1.5000	−1.5089	−1.4987	−1.5093	−1.4993
	1.1	1.1006	1.1028	1.0993	1.1010	1.0991	1.1025	1.0962	1.0996
	−0.5	−0.4651	−0.5036	−0.4668	−0.5041	−0.4863	−0.4983	−0.4892	−0.5013

Note: Disturbance 1=normal, 2=normal-mixture and 3=gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1, \alpha_2)'$. W_1 and W_2 are generated by rook and queen contiguity respectively.

Table 1b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,1,0)

dis	par	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.052	.051	.057	.052	.036	.036	.038	.036	.029	.029	.032	.029	.020	.020	.021	.020
	1	.143	.140	.155	.140	.098	.102	.106	.101	.081	.088	.089	.085	.058	.060	.060	.059
	1.5	.015	.014	.016	.014	.010	.010	.011	.010	.000	.000	.000	.000	.000	.000	.000	.000
	1.1	.014	.014	.015	.014	.010	.010	.010	.010	.000	.000	.000	.000	.000	.000	.000	.000
	−0.5	.028	.027	.030	.027	.020	.019	.020	.019	.000	.000	.000	.000	.000	.000	.000	.000
2	1	.054	.051	.057	.052	.037	.036	.038	.036	.030	.029	.032	.029	.021	.020	.021	.020
	1	.146	.142	.158	.142	.103	.102	.105	.101	.084	.088	.088	.084	.057	.060	.060	.059
	1.5	.015	.014	.016	.015	.010	.010	.010	.010	.000	.000	.000	.000	.000	.000	.000	.000
	1.1	.015	.015	.016	.014	.009	.010	.010	.010	.000	.000	.000	.000	.000	.000	.000	.000
	−0.5	.028	.027	.030	.027	.019	.019	.020	.019	.000	.000	.000	.000	.000	.000	.000	.000
3	1	.052	.051	.059	.052	.037	.036	.039	.036	.029	.029	.032	.029	.020	.020	.021	.020
	1	.205	.187	.126	.141	.140	.138	.081	.100	.123	.127	.063	.084	.085	.089	.041	.058
	1.5	.016	.015	.015	.014	.011	.011	.010	.010	.000	.000	.000	.000	.000	.000	.000	.000
	1.1	.014	.015	.016	.014	.010	.010	.011	.010	.000	.000	.000	.000	.000	.000	.000	.000
	−0.5	.029	.028	.031	.027	.019	.019	.020	.019	.000	.000	.000	.000	.000	.000	.000	.000
1	1	.056	.054	.060	.055	.039	.038	.040	.038	.030	.029	.032	.030	.022	.021	.021	.021
	1	.151	.150	.161	.146	.102	.107	.109	.103	.083	.105	.089	.090	.059	.070	.060	.062
	0	.036	.030	.034	.031	.024	.021	.022	.021	.009	.013	.010	.011	.007	.008	.006	.007
	1.1	.087	.086	.083	.080	.059	.060	.056	.056	.018	.025	.019	.020	.013	.015	.012	.013
	−0.5	.066	.064	.063	.060	.043	.045	.042	.042	.007	.007	.006	.006	.004	.005	.004	.004
2	1	.057	.054	.060	.055	.040	.038	.040	.038	.030	.029	.032	.030	.021	.021	.021	.021
	1	.150	.149	.163	.146	.106	.107	.109	.104	.083	.104	.089	.090	.058	.069	.060	.062
	0	.037	.030	.034	.031	.025	.021	.022	.021	.009	.013	.010	.011	.006	.008	.006	.007
	1.1	.089	.086	.083	.080	.060	.061	.056	.056	.018	.025	.019	.020	.012	.015	.012	.013
	−0.5	.061	.063	.063	.060	.044	.045	.043	.042	.007	.007	.006	.006	.004	.005	.004	.004
3	1	.058	.054	.062	.055	.040	.038	.041	.038	.031	.029	.032	.029	.021	.021	.022	.021
	1	.193	.194	.133	.146	.145	.143	.086	.104	.116	.144	.064	.090	.085	.099	.042	.062
	0	.038	.033	.033	.031	.027	.023	.021	.021	.010	.013	.010	.011	.006	.008	.006	.007
	1.1	.085	.087	.088	.080	.057	.060	.058	.056	.019	.026	.019	.020	.012	.016	.012	.013
	−0.5	.065	.064	.065	.060	.044	.045	.043	.042	.006	.007	.006	.006	.004	.005	.004	.004
1	1	.054	.051	.057	.052	.038	.036	.038	.036	.030	.029	.032	.029	.021	.020	.021	.021
	1	.141	.135	.158	.140	.100	.099	.106	.100	.084	.080	.089	.082	.056	.057	.060	.057
	−1.5	.037	.034	.040	.035	.026	.024	.027	.025	.019	.018	.021	.019	.013	.013	.014	.013
	1.1	.073	.071	.079	.072	.050	.051	.053	.051	.043	.043	.047	.043	.030	.031	.032	.031
	−0.5	.079	.077	.084	.077	.056	.055	.057	.055	.041	.040	.044	.041	.028	.029	.030	.029
2	1	.052	.051	.057	.052	.037	.036	.038	.036	.029	.029	.032	.029	.020	.020	.021	.021
	1	.141	.136	.157	.140	.103	.098	.105	.100	.080	.080	.089	.081	.057	.057	.060	.058
	−1.5	.036	.034	.040	.035	.025	.024	.027	.025	.019	.018	.021	.019	.014	.013	.014	.013
	1.1	.073	.071	.079	.072	.052	.050	.053	.051	.043	.043	.047	.043	.031	.030	.032	.031
	−0.5	.077	.075	.083	.076	.056	.055	.056	.054	.042	.041	.044	.041	.029	.029	.030	.029
3	1	.054	.051	.060	.052	.037	.036	.039	.036	.029	.029	.033	.029	.021	.020	.021	.020
	1	.206	.183	.127	.141	.145	.134	.081	.099	.128	.117	.064	.082	.085	.085	.041	.057
	−1.5	.036	.034	.042	.035	.025	.024	.027	.025	.019	.018	.021	.019	.013	.013	.014	.013
	1.1	.074	.071	.082	.072	.051	.050	.054	.051	.045	.043	.048	.043	.031	.030	.032	.031
	−0.5	.080	.076	.086	.076	.054	.054	.058	.054	.040	.040	.045	.041	.029	.029	.030	.029

Note: Same configuration as Table 1a. Here *sd* is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 2a. Empirical mean of CQMLE and M-estimator, MESDPS(1,0,1)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9734	0.9996	0.9746	1.0005	0.9956	0.9994	0.9962	1.0000
	1	0.9263	0.9510	0.9576	0.9825	0.9871	0.9905	0.9885	0.9918
	0.5	0.4604	0.5002	0.4610	0.5006	0.4972	0.4999	0.4973	0.5000
	0.5	0.4846	0.5014	0.4835	0.5003	0.5022	0.5002	0.5019	0.5000
	−1.1	−1.0778	−1.0916	−1.0861	−1.1000	−1.1049	−1.1039	−1.0990	−1.0986
2	1	0.9734	1.0004	0.9746	1.0005	0.9968	1.0005	0.9965	1.0002
	1	0.9386	0.9639	0.9536	0.9781	0.9820	0.9853	0.9881	0.9914
	1.5	0.4617	0.5020	0.4613	0.5003	0.4972	0.5000	0.4973	0.5000
	0.5	0.4862	0.5030	0.4837	0.5002	0.5024	0.5004	0.5022	0.5002
	−1.1	−1.0851	−1.0984	−1.0900	−1.1042	−1.1044	−1.1041	−1.1034	−1.1026
3	1	0.9747	1.0011	0.9748	1.0005	0.9955	0.9992	0.9969	1.0006
	1	0.9341	0.9600	0.9503	0.9752	0.9823	0.9856	0.9866	0.9899
	1.5	0.4617	0.5019	0.4614	0.5008	0.4973	0.5001	0.4973	0.5000
	0.5	0.4855	0.5024	0.4830	0.4997	0.5021	0.5001	0.5021	0.5001
	−1.1	−1.0899	−1.1027	−1.0867	−1.1012	−1.1041	−1.1034	−1.0996	−1.0992
1	1	0.9700	1.0000	0.9686	0.9990	0.9903	0.9992	0.9908	0.9996
	1	0.9276	0.9591	0.9487	0.9804	0.9793	0.9875	0.9870	0.9951
	0	−0.0574	0.0013	−0.0589	0.0002	−0.0140	0.0002	−0.0141	0.0000
	0.5	0.4618	0.5040	0.4576	0.5010	0.5013	0.5025	0.4995	0.5007
	−1.1	−1.0711	−1.1008	−1.0657	−1.0957	−1.0997	−1.1013	−1.0998	−1.1014
2	1	0.9692	0.9995	0.9703	1.0003	0.9901	0.9991	0.9911	0.9999
	1	0.9314	0.9631	0.9538	0.9855	0.9784	0.9865	0.9832	0.9913
	0	−0.0578	0.0012	−0.0589	−0.0001	−0.0139	0.0003	−0.0142	−0.0001
	0.5	0.4619	0.5044	0.4568	0.4998	0.5001	0.5010	0.4995	0.5007
	−1.1	−1.0743	−1.1046	−1.0702	−1.1000	−1.0989	−1.1002	−1.0997	−1.1011
3	1	0.9715	1.0017	0.9681	0.9980	0.9915	1.0003	0.9907	0.9994
	1	0.9255	0.9572	0.9493	0.9811	0.9750	0.9832	0.9812	0.9892
	0	−0.0569	0.0015	−0.0580	0.0007	−0.0136	0.0004	−0.0135	0.0005
	0.5	0.4580	0.5000	0.4580	0.5010	0.4998	0.5009	0.4998	0.5012
	−1.1	−1.0660	−1.0945	−1.0688	−1.0987	−1.1034	−1.1047	−1.1014	−1.1029
1	1	0.9769	0.9971	0.9793	0.9997	0.9941	0.9992	0.9956	1.0005
	1	0.9374	0.9634	0.9629	0.9898	0.9809	0.9868	0.9866	0.9926
	−0.5	−0.5608	−0.5012	−0.5589	−0.4982	−0.5241	−0.5010	−0.5235	−0.5000
	0.5	0.4398	0.5050	0.4382	0.5067	0.4915	0.5027	0.4866	0.4998
	−1.1	−1.0764	−1.1122	−1.0687	−1.1082	−1.0999	−1.1071	−1.0901	−1.0987
2	1	0.9796	0.9996	0.9802	1.0003	0.9933	0.9984	0.9948	0.9997
	1	0.9419	0.9682	0.9569	0.9832	0.9783	0.9842	0.9875	0.9935
	−0.5	−0.5580	−0.4983	−0.5611	−0.5012	−0.5243	−0.5011	−0.5245	−0.5010
	0.5	0.4455	0.5117	0.4356	0.5035	0.4960	0.5077	0.4856	0.4986
	−1.1	−1.0694	−1.1073	−1.0672	−1.1070	−1.0990	−1.1069	−1.0923	−1.1008
3	1	0.9760	0.9963	0.9788	0.9989	0.9956	1.0007	0.9956	1.0004
	1	0.9436	0.9707	0.9527	0.9790	0.9836	0.9897	0.9853	0.9913
	−0.5	−0.5606	−0.5004	−0.5600	−0.5004	−0.5233	−0.5000	−0.5233	−0.5000
	0.5	0.4389	0.5067	0.4389	0.5070	0.4900	0.5016	0.4900	0.5026
	−1.1	−1.0599	−1.0995	−1.0665	−1.1079	−1.0978	−1.1050	−1.0938	−1.1020

Note: Disturbance 1=normal, 2=normal-mixture and 3=gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1, \alpha_3)'$. W_1 and W_3 is generated by queen contiguity.

Table 2b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,0,1)

dis	par	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$	sd	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.041	.041	.045	.041	.029	.029	.030	.029	.022	.021	.024	.022	.015	.015	.016	.015
	1	.144	.135	.156	.139	.099	.099	.107	.101	.081	.083	.089	.083	.058	.058	.060	.058
	0.5	.027	.025	.029	.026	.019	.018	.019	.018	.003	.003	.004	.003	.002	.002	.002	.002
	0.5	.031	.030	.033	.030	.022	.022	.023	.022	.006	.006	.006	.006	.004	.004	.004	.004
	−1.1	.145	.143	.155	.142	.100	.099	.103	.099	.081	.080	.086	.080	.058	.057	.058	.056
2	1	.044	.041	.045	.041	.029	.029	.030	.029	.022	.022	.024	.022	.015	.015	.016	.015
	1	.146	.137	.159	.141	.104	.099	.106	.100	.083	.082	.089	.082	.057	.058	.060	.058
	1.5	.028	.025	.029	.026	.018	.018	.019	.018	.003	.003	.004	.003	.002	.002	.002	.002
	0.5	.032	.031	.034	.031	.022	.022	.022	.022	.006	.006	.006	.006	.004	.004	.004	.004
	−1.1	.151	.144	.154	.142	.098	.099	.103	.099	.082	.079	.086	.080	.058	.056	.058	.056
3	1	.043	.041	.046	.041	.030	.029	.031	.029	.022	.021	.024	.022	.015	.015	.016	.015
	1	.206	.181	.128	.141	.140	.134	.082	.100	.123	.120	.063	.083	.085	.086	.042	.058
	1.5	.030	.027	.029	.026	.020	.019	.018	.018	.004	.003	.004	.003	.002	.002	.002	.002
	0.5	.030	.029	.036	.031	.022	.021	.024	.022	.006	.006	.007	.006	.004	.005	.004	.004
	−1.1	.150	.142	.163	.143	.101	.097	.107	.099	.081	.079	.088	.080	.056	.056	.060	.056
1	1	.042	.041	.046	.042	.029	.029	.031	.029	.022	.022	.025	.022	.017	.016	.016	.016
	1	.148	.138	.158	.141	.100	.099	.107	.101	.083	.082	.089	.083	.059	.058	.060	.058
	0	.038	.034	.037	.034	.026	.024	.025	.024	.010	.010	.011	.010	.007	.007	.007	.007
	0.5	.059	.057	.061	.056	.041	.040	.041	.040	.025	.025	.025	.024	.018	.018	.017	.017
	−1.1	.154	.144	.156	.144	.101	.099	.105	.100	.081	.080	.086	.080	.056	.056	.059	.057
2	1	.042	.041	.046	.042	.030	.029	.030	.029	.022	.022	.024	.022	.016	.016	.016	.016
	1	.145	.137	.161	.142	.104	.100	.108	.101	.082	.081	.090	.082	.058	.058	.060	.058
	0	.037	.033	.037	.034	.026	.024	.025	.024	.010	.010	.011	.010	.007	.007	.007	.007
	0.5	.059	.057	.061	.056	.041	.040	.041	.040	.024	.025	.025	.024	.017	.018	.017	.017
	−1.1	.145	.144	.156	.143	.098	.099	.105	.100	.081	.080	.086	.080	.059	.056	.059	.057
3	1	.044	.042	.047	.042	.031	.029	.031	.029	.022	.022	.024	.022	.016	.016	.017	.016
	1	.190	.179	.130	.141	.143	.134	.084	.101	.116	.117	.065	.082	.085	.086	.042	.058
	0	.038	.035	.037	.034	.027	.025	.024	.024	.010	.010	.011	.010	.007	.007	.007	.007
	0.5	.054	.053	.065	.056	.039	.038	.044	.040	.025	.025	.026	.024	.017	.018	.017	.017
	−1.1	.148	.142	.164	.143	.103	.098	.108	.100	.082	.079	.089	.080	.057	.056	.060	.057
1	1	.044	.042	.046	.042	.031	.029	.031	.029	.024	.024	.026	.024	.017	.017	.017	.017
	1	.144	.137	.159	.141	.102	.100	.108	.102	.084	.081	.089	.082	.056	.057	.060	.058
	−0.5	.041	.037	.041	.037	.029	.026	.028	.027	.019	.017	.019	.018	.013	.012	.013	.013
	0.5	.121	.116	.123	.115	.084	.082	.083	.081	.068	.069	.071	.068	.048	.049	.049	.048
	−1.1	.159	.153	.166	.152	.110	.105	.110	.105	.087	.085	.094	.086	.059	.060	.063	.061
2	1	.044	.042	.046	.042	.029	.029	.031	.029	.025	.024	.026	.024	.016	.017	.017	.017
	1	.144	.138	.159	.142	.105	.100	.107	.101	.080	.080	.089	.082	.057	.058	.060	.058
	−0.5	.040	.037	.042	.037	.028	.026	.028	.026	.018	.017	.020	.018	.013	.012	.013	.013
	0.5	.116	.116	.123	.114	.085	.082	.083	.081	.069	.068	.072	.068	.048	.049	.049	.048
	−1.1	.158	.154	.166	.152	.109	.105	.110	.105	.085	.085	.093	.086	.061	.060	.063	.060
3	1	.044	.043	.048	.042	.029	.030	.031	.029	.024	.023	.026	.024	.017	.017	.017	.017
	1	.207	.184	.129	.142	.147	.134	.083	.100	.129	.118	.064	.082	.085	.085	.042	.058
	−0.5	.042	.039	.041	.037	.030	.027	.027	.026	.019	.017	.020	.018	.013	.012	.013	.013
	0.5	.122	.112	.131	.114	.083	.080	.086	.080	.067	.067	.074	.067	.048	.049	.049	.048
	−1.1	.152	.152	.174	.152	.106	.103	.115	.105	.085	.084	.097	.086	.062	.059	.064	.060

Note: Same configuration as Table 2a. Here sd is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 3a. Empirical mean of CQMLE and M-estimator, MESDPS(0,1,1)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9771	0.9999	0.9781	1.0006	0.9979	0.9994	0.9985	1.0001
	1	0.9304	0.9503	0.9624	0.9826	0.9889	0.9905	0.9903	0.9918
	1.5	1.4704	1.5004	1.4708	1.5006	1.4997	1.5000	1.4997	1.5000
	0.5	0.5167	0.4997	0.5165	0.4998	0.5002	0.5000	0.5002	0.5000
	−1.1	−1.0526	−1.0531	−1.0850	−1.0848	−1.0883	−1.0886	−1.0879	−1.0882
2	1	0.9757	0.9993	0.9772	0.9999	0.9992	1.0008	0.9991	1.0006
	1	0.9429	0.9634	0.9571	0.9769	0.9837	0.9852	0.9899	0.9914
	1.5	1.4711	1.5017	1.4703	1.4996	1.4997	1.5000	1.4997	1.5000
	0.5	0.5160	0.4987	0.5168	0.5003	0.5002	0.5000	0.5002	0.5000
	−1.1	−1.0574	−1.0574	−1.0863	−1.0876	−1.0886	−1.0891	−1.0940	−1.0944
3	1	0.9784	1.0016	0.9771	0.9996	0.9971	0.9987	0.9997	1.0012
	1	0.9389	0.9598	0.9551	0.9751	0.9845	0.9860	0.9885	0.9900
	1.5	1.4708	1.5014	1.4709	1.5004	1.4997	1.5000	1.4997	1.5000
	0.5	0.5165	0.4992	0.5166	0.5000	0.5002	0.5000	0.5002	0.5000
	−1.1	−1.0582	−1.0582	−1.0868	−1.0874	−1.0873	−1.0876	−1.0906	−1.0909
1	1	0.9515	1.0001	0.9499	0.9995	0.9840	0.9992	0.9846	0.9997
	1	0.9152	0.9581	0.9367	0.9802	0.9732	0.9876	0.9807	0.9950
	0	−0.0960	0.0011	−0.0972	0.0003	−0.0320	0.0006	−0.0324	−0.0001
	0.5	0.5540	0.4989	0.5550	0.4998	0.5192	0.4996	0.5194	0.5001
	−1.1	−1.0597	−1.0602	−1.0748	−1.0753	−1.0842	−1.0872	−1.0910	−1.0948
2	1	0.9504	0.9998	0.9508	0.9998	0.9844	0.9996	0.9853	1.0004
	1	0.9188	0.9623	0.9411	0.9845	0.9715	0.9858	0.9773	0.9915
	0	−0.0963	0.0017	−0.0979	−0.0011	−0.0322	0.0003	−0.0323	0.0000
	0.5	0.5541	0.4984	0.5555	0.5007	0.5193	0.4998	0.5193	0.5000
	−1.1	−1.0674	−1.0671	−1.0834	−1.0837	−1.0822	−1.0862	−1.0912	−1.0949
3	1	0.9520	1.0010	0.9492	0.9980	0.9853	1.0006	0.9844	0.9994
	1	0.9136	0.9571	0.9374	0.9812	0.9682	0.9825	0.9752	0.9896
	0	−0.0968	0.0004	−0.0966	0.0004	−0.0321	0.0004	−0.0313	0.0011
	0.5	0.5552	0.5003	0.5548	0.4999	0.5192	0.4997	0.5188	0.4993
	−1.1	−1.0621	−1.0611	−1.0790	−1.0788	−1.0893	−1.0927	−1.0928	−1.0971
1	1	0.9915	0.9983	0.9930	0.9999	0.9985	0.9994	0.9996	1.0005
	1	0.9585	0.9650	0.9821	0.9888	0.9862	0.9872	0.9916	0.9926
	−1.5	−1.5282	−1.5000	−1.5264	−1.4979	−1.5097	−1.5005	−1.5090	−1.4996
	0.5	0.5108	0.5011	0.5095	0.4994	0.4993	0.4999	0.5000	0.5002
	−1.1	−1.0753	−1.0769	−1.0883	−1.0891	−1.0945	−1.0947	−1.0931	−1.0932
2	1	0.9936	1.0003	0.9945	1.0012	0.9988	0.9997	0.9985	0.9993
	1	0.9599	0.9665	0.9775	0.9840	0.9830	0.9840	0.9928	0.9938
	−1.5	−1.5253	−1.4967	−1.5284	−1.5004	−1.5098	−1.5006	−1.5095	−1.5002
	0.5	0.5085	0.4983	0.5100	0.5002	0.4983	0.4987	0.4993	0.4996
	−1.1	−1.0554	−1.0563	−1.0884	−1.0889	−1.0886	−1.0889	−1.0988	−1.0990
3	1	0.9894	0.9961	0.9934	1.0001	0.9996	1.0005	0.9999	1.0008
	1	0.9624	0.9691	0.9731	0.9797	0.9884	0.9894	0.9904	0.9914
	−1.5	−1.5299	−1.5014	−1.5275	−1.4994	−1.5090	−1.4998	−1.5087	−1.4994
	0.5	0.5111	0.5009	0.5087	0.4987	0.5007	0.5011	0.4985	0.4988
	−1.1	−1.0554	−1.0563	−1.0844	−1.0848	−1.0892	−1.0894	−1.0954	−1.0956

Note: Disturbance 1=normal, 2=normal-mixture and 3=gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_2, \alpha_3)'$. W_2 and W_3 are generated by rook and queen contiguity respectively.

Table 3b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(0,1,1)

dis	par	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.047	.047	.052	.047	.033	.033	.035	.033	.026	.025	.028	.026	.018	.018	.019	.018
	1	.143	.135	.155	.139	.099	.099	.106	.100	.081	.081	.089	.082	.058	.057	.060	.057
	1.5	.022	.021	.024	.021	.016	.015	.016	.015	.000	.000	.001	.000	.000	.000	.000	.000
	0.5	.016	.015	.017	.015	.011	.011	.011	.011	.000	.000	.000	.000	.000	.000	.000	.000
	−1.1	.205	.197	.214	.196	.138	.138	.144	.138	.113	.109	.118	.110	.078	.078	.081	.078
2	1	.050	.047	.052	.048	.034	.033	.035	.033	.027	.025	.028	.026	.018	.018	.019	.018
	1	.144	.137	.159	.141	.103	.098	.106	.100	.083	.080	.089	.081	.057	.057	.060	.057
	1.5	.023	.021	.024	.022	.015	.015	.016	.015	.000	.000	.001	.000	.000	.000	.000	.000
	0.5	.016	.015	.017	.015	.011	.010	.011	.011	.000	.000	.000	.000	.000	.000	.000	.000
	−1.1	.208	.199	.212	.196	.141	.137	.144	.137	.108	.108	.119	.110	.079	.077	.081	.078
3	1	.048	.047	.055	.048	.034	.033	.036	.033	.025	.025	.029	.026	.018	.018	.019	.018
	1	.205	.181	.127	.140	.139	.134	.082	.100	.123	.117	.063	.082	.085	.085	.041	.057
	1.5	.023	.022	.025	.022	.016	.015	.016	.015	.000	.000	.001	.000	.000	.000	.000	.000
	0.5	.015	.014	.018	.015	.011	.010	.012	.011	.000	.000	.000	.000	.000	.000	.000	.000
	−1.1	.204	.194	.223	.197	.138	.137	.147	.138	.109	.108	.121	.110	.077	.077	.083	.078
1	1	.052	.050	.056	.051	.037	.036	.038	.036	.027	.026	.029	.027	.020	.019	.019	.019
	1	.148	.141	.160	.144	.102	.102	.109	.103	.083	.081	.090	.083	.059	.058	.061	.058
	0	.054	.049	.053	.049	.039	.035	.035	.034	.017	.016	.018	.016	.012	.011	.012	.011
	0.5	.035	.032	.035	.032	.025	.023	.023	.022	.010	.010	.011	.010	.007	.007	.007	.007
	−1.1	.212	.198	.214	.197	.140	.138	.143	.138	.112	.109	.119	.110	.077	.078	.081	.078
2	1	.052	.050	.056	.051	.037	.036	.037	.036	.027	.026	.029	.027	.019	.018	.020	.019
	1	.148	.140	.163	.144	.107	.102	.109	.103	.083	.081	.090	.083	.059	.058	.060	.058
	0	.055	.049	.054	.049	.037	.035	.036	.034	.017	.016	.018	.016	.012	.011	.012	.011
	0.5	.035	.032	.035	.032	.024	.023	.023	.022	.011	.010	.011	.010	.007	.007	.007	.007
	−1.1	.204	.196	.214	.196	.141	.139	.143	.138	.111	.110	.118	.110	.077	.077	.081	.078
3	1	.052	.050	.058	.051	.037	.036	.038	.036	.027	.026	.029	.026	.019	.019	.020	.019
	1	.191	.182	.133	.143	.144	.136	.086	.103	.116	.116	.066	.082	.085	.085	.043	.058
	0	.054	.050	.055	.049	.037	.035	.036	.034	.017	.016	.018	.016	.011	.011	.012	.011
	0.5	.032	.031	.038	.032	.022	.022	.025	.022	.010	.009	.011	.010	.007	.007	.007	.007
	−1.1	.201	.195	.224	.197	.145	.136	.148	.137	.114	.108	.122	.110	.080	.077	.082	.078
1	1	.047	.044	.050	.045	.033	.032	.034	.032	.027	.025	.028	.026	.018	.018	.019	.018
	1	.140	.134	.157	.139	.099	.098	.106	.100	.084	.080	.088	.082	.056	.057	.060	.057
	−1.5	.037	.035	.040	.036	.027	.025	.027	.025	.018	.018	.020	.018	.013	.013	.013	.013
	0.5	.033	.032	.035	.032	.023	.023	.024	.023	.022	.021	.023	.022	.015	.015	.016	.015
	−1.1	.198	.199	.214	.197	.142	.139	.143	.138	.111	.109	.119	.110	.076	.078	.081	.078
2	1	.046	.045	.050	.045	.032	.031	.033	.032	.026	.025	.028	.026	.017	.018	.019	.018
	1	.141	.135	.156	.139	.102	.098	.105	.099	.080	.080	.089	.081	.057	.057	.060	.057
	−1.5	.038	.035	.041	.036	.025	.024	.027	.025	.018	.018	.020	.018	.013	.013	.013	.013
	0.5	.033	.032	.035	.032	.023	.022	.024	.023	.022	.021	.023	.021	.015	.015	.016	.015
	−1.1	.202	.197	.213	.196	.140	.138	.144	.138	.111	.109	.119	.110	.078	.078	.081	.078
3	1	.048	.044	.052	.045	.032	.031	.034	.032	.025	.025	.029	.026	.018	.018	.019	.018
	1	.204	.182	.125	.139	.145	.133	.080	.099	.128	.117	.064	.082	.085	.085	.041	.057
	−1.5	.037	.035	.042	.036	.025	.025	.027	.025	.018	.018	.020	.018	.013	.013	.013	.013
	0.5	.034	.032	.036	.032	.023	.023	.024	.023	.022	.021	.024	.021	.015	.015	.016	.015
	−1.1	.204	.194	.224	.196	.136	.136	.148	.137	.111	.107	.122	.110	.077	.077	.083	.078

Note: Same configuration as Table 3a. Here *sd* is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

Table 4a. Empirical mean of CQMLE and M-estimator, MESDPS(1,1,1)

dis	par	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est	CQMLE	M-est
		n=49, T=3		n=100, T=3		n=49, T=7		n=100, T=7	
1	1	0.9606	1.0016	0.9629	1.0013	1.0002	1.0009	0.9993	1.0001
	1	0.8833	0.9446	0.9214	0.9812	0.9824	0.9872	0.9858	0.9903
	0.5	0.4102	0.5024	0.4136	0.5028	0.4966	0.5000	0.4967	0.5000
	1.1	1.0856	1.1060	1.0828	1.1013	1.1034	1.1004	1.1032	1.1001
	1.1	1.1123	1.1053	1.1085	1.1001	1.1051	1.1005	1.1048	1.1002
	1.1	1.1516	1.1726	1.1069	1.1301	1.1125	1.1208	1.1084	1.1168
2	1	0.9602	1.0030	0.9625	1.0015	1.0000	1.0009	0.9994	1.0002
	1	0.8958	0.9605	0.9165	0.9754	0.9778	0.9826	0.9856	0.9900
	0.5	0.4124	0.5069	0.4133	0.5013	0.4968	0.5001	0.4968	0.5000
	1.1	1.0839	1.1075	1.0838	1.1039	1.1031	1.1001	1.1032	1.1001
	1.1	1.1095	1.1054	1.1098	1.1037	1.1046	1.1001	1.1047	1.1002
	1.1	1.1531	1.1749	1.1062	1.1254	1.1129	1.1210	1.1018	1.1099
3	1	0.9629	1.0047	0.9628	1.0009	0.9976	0.9984	1.0002	1.0010
	1	0.8934	0.9593	0.9131	0.9720	0.9783	0.9831	0.9840	0.9884
	0.5	0.4132	0.5075	0.4137	0.5015	0.4967	0.5001	0.4967	0.5000
	1.1	1.0855	1.1066	1.0841	1.1023	1.1033	1.1003	1.1033	1.1001
	1.1	1.1112	1.1044	1.1100	1.1017	1.1049	1.1004	1.1049	1.1002
	1.1	1.1514	1.1752	1.1047	1.1255	1.1149	1.1230	1.1054	1.1135
1	1	0.9383	1.0016	0.9379	1.0005	1.0006	1.0004	1.0018	1.0012
	1	0.8768	0.9537	0.9039	0.9779	0.9725	0.9856	0.9821	0.9943
	0	−0.1288	0.0062	−0.1289	0.0022	−0.0203	0.0017	−0.0207	0.0005
	1.1	1.0445	1.1037	1.0424	1.1013	1.1155	1.0999	1.1176	1.1011
	1.1	1.0890	1.1017	1.0869	1.1007	1.1220	1.0992	1.1243	1.1009
	1.1	1.1804	1.1768	1.1490	1.1408	1.0971	1.1238	1.0824	1.1095
2	1	0.9378	1.0008	0.9398	1.0016	0.9998	0.9998	1.0015	1.0009
	1	0.8801	0.9566	0.9084	0.9822	0.9710	0.9838	0.9785	0.9907
	0	−0.1286	0.0058	−0.1295	0.0005	−0.0211	0.0006	−0.0209	0.0004
	1.1	1.0398	1.0979	1.0473	1.1049	1.1171	1.1017	1.1178	1.1010
	1.1	1.0844	1.0960	1.0918	1.1047	1.1238	1.1014	1.1245	1.1008
	1.1	1.1786	1.1742	1.1351	1.1284	1.0965	1.1218	1.0818	1.1089
3	1	0.9396	1.0039	0.9377	0.9994	1.0013	1.0014	1.0001	0.9996
	1	0.8749	0.9531	0.9052	0.9798	0.9672	0.9799	0.9760	0.9884
	0	−0.1299	0.0054	−0.1270	0.0035	−0.0209	0.0006	−0.0204	0.0011
	1.1	1.0446	1.1050	1.0433	1.1007	1.1167	1.1014	1.1170	1.1004
	1.1	1.0895	1.1034	1.0872	1.0996	1.1234	1.1011	1.1236	1.1000
	1.1	1.1750	1.1704	1.1442	1.1373	1.0906	1.1158	1.0807	1.1079
1	1	0.9360	0.9979	0.9392	1.0001	0.9839	1.0003	0.9840	1.0004
	1	0.8965	0.9568	0.9269	0.9868	0.9702	0.9845	0.9767	0.9909
	−0.5	−0.6370	−0.4984	−0.6328	−0.4961	−0.5502	−0.4997	−0.5506	−0.5002
	1.1	1.0286	1.1000	1.0337	1.1031	1.0815	1.1013	1.0805	1.1000
	1.1	1.0893	1.1006	1.0913	1.1019	1.1016	1.1012	1.1009	1.1002
	1.1	1.1984	1.1643	1.1591	1.1287	1.1277	1.1169	1.1235	1.1131
2	1	0.9392	1.0009	0.9409	1.0013	0.9826	0.9994	0.9831	0.9994
	1	0.8995	0.9604	0.9209	0.9798	0.9666	0.9808	0.9778	0.9919
	−0.5	−0.6339	−0.4945	−0.6354	−0.4998	−0.5509	−0.5004	−0.5510	−0.5007
	1.1	1.0290	1.1001	1.0333	1.1024	1.0796	1.0996	1.0802	1.0997
	1.1	1.0882	1.0990	1.0916	1.1022	1.1002	1.0998	1.1007	1.1001
	1.1	1.2135	1.1844	1.1568	1.1275	1.1332	1.1226	1.1175	1.1074
3	1	0.9357	0.9969	0.9400	1.0001	0.9841	1.0010	0.9851	1.0013
	1	0.8995	0.9607	0.9172	0.9759	0.9717	0.9862	0.9764	0.9905
	−0.5	−0.6364	−0.4982	−0.6329	−0.4985	−0.5509	−0.5001	−0.5489	−0.4988
	1.1	1.0241	1.0944	1.0302	1.0985	1.0807	1.1009	1.0815	1.1007
	1.1	1.0856	1.0957	1.0883	1.0986	1.1012	1.1010	1.1010	1.1003
	1.1	1.2181	1.1889	1.1647	1.1360	1.1320	1.1209	1.1202	1.1102

Note: Disturbance 1=normal, 2=normal-mixture and 3=gamma. Parameters $\theta = (\beta, \sigma_\epsilon^2, \tau, \alpha_1, \alpha_2, \alpha_3)'$. W_1, W_2 and W_3 are generated by queen, rook and queen contiguity respectively.

Table 4b. Empirical sd and asymptotic standard errors of M-estimator, MESDPS(1,1,1)

dis	par	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$	*sd*	**se**	$\tilde{se}$	$\hat{se}$
		n=49, T=3				n=100, T=3				n=49, T=7				n=100, T=7			
1	1	.054	.054	.057	.052	.037	.037	.038	.037	.024	.023	.026	.024	.017	.017	.017	.017
	1	.152	.148	.163	.146	.105	.107	.112	.106	.081	.089	.090	.085	.058	.062	.060	.059
	0.5	.049	.049	.047	.045	.032	.034	.030	.031	.004	.004	.004	.004	.003	.003	.002	.003
	1.1	.072	.082	.074	.073	.050	.057	.050	.052	.006	.006	.005	.005	.004	.004	.004	.004
	1.1	.088	.104	.092	.092	.062	.072	.061	.065	.008	.009	.007	.007	.005	.006	.005	.005
	1.1	.224	.226	.231	.214	.149	.154	.153	.149	.116	.113	.120	.111	.080	.080	.082	.079
2	1	.057	.054	.058	.053	.037	.037	.038	.037	.025	.024	.026	.024	.017	.017	.017	.017
	1	.154	.150	.168	.149	.109	.107	.111	.105	.083	.087	.090	.084	.057	.062	.060	.059
	0.5	.049	.049	.047	.045	.032	.033	.030	.031	.004	.004	.004	.004	.003	.003	.002	.002
	1.1	.077	.083	.074	.074	.049	.057	.049	.051	.005	.006	.005	.005	.004	.004	.004	.004
	1.1	.094	.104	.092	.092	.061	.072	.061	.064	.007	.008	.007	.007	.005	.006	.005	.005
	1.1	.228	.227	.229	.214	.151	.153	.153	.148	.112	.111	.121	.111	.080	.079	.082	.079
3	1	.057	.054	.060	.053	.038	.038	.039	.036	.024	.023	.027	.024	.017	.017	.018	.017
	1	.218	.196	.139	.149	.144	.141	.088	.105	.123	.126	.065	.084	.085	.090	.042	.059
	0.5	.053	.052	.049	.046	.034	.034	.031	.031	.004	.004	.004	.004	.003	.003	.002	.003
	1.1	.074	.085	.077	.074	.051	.057	.051	.051	.005	.006	.006	.005	.004	.004	.004	.004
	1.1	.089	.107	.096	.093	.064	.072	.063	.064	.007	.008	.008	.007	.005	.006	.005	.005
	1.1	.225	.227	.240	.215	.149	.153	.157	.148	.111	.111	.124	.111	.078	.078	.084	.079
1	1	.062	.060	.064	.059	.042	.042	.042	.041	.026	.026	.028	.026	.019	.019	.019	.018
	1	.162	.148	.168	.149	.109	.106	.113	.106	.083	.084	.091	.083	.059	.059	.061	.059
	0	.067	.059	.059	.056	.045	.040	.038	.038	.014	.016	.014	.014	.010	.011	.009	.010
	1.1	.091	.093	.090	.087	.058	.065	.059	.061	.028	.032	.026	.027	.020	.023	.018	.019
	1.1	.090	.096	.092	.089	.057	.066	.060	.062	.031	.037	.030	.031	.022	.026	.020	.022
	1.1	.238	.229	.235	.218	.153	.155	.156	.151	.117	.118	.124	.115	.081	.084	.084	.082
2	1	.061	.060	.064	.059	.043	.042	.042	.041	.026	.027	.028	.026	.018	.019	.018	.018
	1	.159	.147	.170	.149	.114	.106	.113	.106	.083	.083	.091	.083	.058	.059	.061	.059
	0	.066	.058	.059	.055	.044	.040	.038	.038	.014	.016	.014	.014	.010	.011	.009	.010
	1.1	.088	.093	.090	.087	.060	.066	.060	.061	.027	.032	.026	.028	.019	.023	.018	.019
	1.1	.087	.096	.092	.089	.060	.067	.061	.062	.031	.037	.030	.031	.021	.026	.020	.022
	1.1	.233	.227	.236	.218	.155	.156	.155	.151	.118	.119	.123	.115	.081	.083	.084	.082
3	1	.064	.060	.067	.059	.044	.042	.043	.041	.027	.027	.028	.026	.018	.019	.019	.018
	1	.203	.188	.143	.149	.152	.140	.091	.106	.116	.118	.066	.083	.085	.087	.043	.058
	0	.068	.060	.061	.056	.046	.042	.039	.038	.014	.016	.014	.014	.010	.011	.009	.010
	1.1	.088	.094	.094	.087	.063	.065	.062	.061	.026	.033	.027	.028	.019	.023	.018	.019
	1.1	.088	.097	.096	.089	.063	.067	.062	.062	.029	.037	.031	.031	.022	.026	.021	.022
	1.1	.227	.226	.248	.219	.163	.153	.162	.151	.120	.117	.128	.115	.083	.083	.085	.082
1	1	.064	.063	.067	.062	.045	.044	.044	.043	.033	.032	.034	.032	.022	.022	.023	.022
	1	.152	.147	.167	.148	.106	.107	.113	.107	.085	.082	.091	.083	.056	.058	.061	.058
	−0.5	.070	.062	.068	.061	.048	.043	.045	.043	.027	.026	.028	.026	.019	.018	.019	.018
	1.1	.095	.100	.099	.094	.066	.069	.065	.065	.048	.048	.049	.046	.033	.034	.033	.033
	1.1	.075	.080	.080	.076	.053	.056	.053	.053	.041	.041	.042	.040	.028	.029	.028	.028
	1.1	.231	.229	.240	.221	.161	.158	.158	.153	.121	.121	.130	.120	.084	.086	.088	.085
2	1	.063	.063	.067	.061	.043	.044	.044	.043	.033	.032	.034	.032	.022	.022	.023	.022
	1	.153	.147	.168	.149	.108	.106	.112	.106	.082	.082	.091	.083	.057	.059	.061	.058
	−0.5	.070	.062	.068	.061	.047	.043	.044	.042	.028	.025	.029	.026	.019	.018	.019	.018
	1.1	.093	.100	.099	.094	.063	.068	.065	.065	.048	.048	.049	.046	.033	.034	.033	.033
	1.1	.075	.081	.080	.076	.051	.055	.052	.052	.041	.041	.042	.040	.028	.029	.028	.028
	1.1	.235	.228	.239	.220	.156	.157	.158	.153	.125	.121	.130	.120	.085	.086	.088	.085
3	1	.064	.062	.070	.061	.044	.044	.045	.043	.032	.032	.035	.032	.023	.022	.023	.022
	1	.209	.192	.141	.149	.151	.140	.090	.105	.129	.118	.067	.083	.086	.086	.043	.058
	−0.5	.072	.063	.070	.061	.049	.043	.045	.042	.029	.025	.029	.026	.019	.018	.019	.018
	1.1	.097	.100	.104	.095	.063	.068	.067	.065	.047	.048	.050	.046	.032	.034	.033	.033
	1.1	.078	.081	.084	.076	.050	.055	.054	.052	.040	.041	.043	.040	.027	.029	.029	.028
	1.1	.236	.227	.252	.221	.151	.155	.163	.153	.121	.119	.134	.120	.085	.085	.090	.085

Note: Same configuration as Table 4a. Here *sd* is empirical standard deviation, **se** is OPMD estimator, $\tilde{se}$ is standard error based on $\widehat{\Omega}^{*-1}$ and $\hat{se}$ based on $\Psi^{*-1}(\hat{\theta}_M)$.

References

Debarsy, N., Jin, F. & Lee, L.-F. (2015), 'Large sample properties of the matrix exponential spatial specification with an application to fdi', *Journal of Econometrics* **188**(1), 1–21.

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

Lee, L.-F. (2004), 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models', *Econometrica* **72**(6), 1899–1925.

Newey, W. K. (1991), 'Uniform convergence in probability and stochastic equicontinuity', *Econometrica* **59**(4), 1161–1167.

Yang, Z. (2015), 'A general method for third-order bias and variance corrections on a nonlinear estimator', *Journal of Econometrics* **186**(1), 178–200.

Yang, Z. (2018), 'Unified m-estimation of fixed-effects spatial dynamic models with short panels', *Journal of econometrics* **205**(2), 423–447.