# QML Estimation of the Matrix Exponential Spatial Specification Panel Data Model with Fixed Effects and Heteroskedasticity

February 17, 2019

Yuanqing Zhang[a,b], Shuhui Feng[c], Fei Jin[d,1]

[a]School of Business, Shanghai University of International Business and Economics, Shanghai, 201620, China

[b]School of Economics, Peking University, Beijing, 100871, China

[c]Institute of Economics, Shanghai Academy of Social Sciences, Shanghai, 200020, China

[d]School of Economics, Fudan University, Shanghai, 200433, China

**Abstract**

This paper studies a spatial panel data model with fixed effects and heteroskedasticity, where the spatial effects in the dependent variable and disturbances are in the form of matrix exponential spatial specification (MESS). The asymptotic properties of quasi maximum likelihood (QML) estimators with large $n$ and finite or large $T$ are established. We show that the QML estimator (QMLE) can be consistent and asymptotically normal under unknown heteroskedasticity when the spatial weights matrices in the two MESS processes are commutative. We provide a consistent estimator for the standard deviation of the QMLE under regularity conditions, which can be used for inference.

*Keywords:* MESS, Panel data, QMLE, Heteroskedasticity, Spatial dependence

*JEL classification:* C12, C13, C21, C23

# 1 Introduction

LeSage and Pace (2007) propose the matrix exponential spatial specification (MESS) as a substitute for the spatial autoregressive (SAR) specification, based on two advantages of the MESS model: one is that the specification always yields positive definite estimated covariance matrices, and the other is its computational simplicity because its quasi maximum likelihood (QML) function does not involve any Jacobian. Debarsy et al. (2015) consider a general MESS model that has MESS processes in both the dependent variable and the disturbances. They study the consistency of QML estimators (QMLE) for the model with homoskedastic and heteroskedastic disturbances.

[1]Corresponding author (E-mail: jin.fei@live.com).

Under unknown heteroskedasticity, the QMLE of the MESS model is consistent when the spatial weights matrices in the two MESS processes are commutative, but the QMLE of the SAR model is not in general.

The extension of cross sectional SAR models to panel data has a long history. The MESS can also be extended to panel data. Figueiredo and Silva (2015) employ a direct approach to establish asymptotic properties of maximum likelihood estimators for a MESS panel model with fixed effects. However, we notice that even for the SAR specification, only a few papers in the literature have considered heteroskedasticity for panel data. Moscone and Tosetti (2011) employ the generalized method of moments (GMM) to study a SAR panel data model with fixed effects and heteroskedasticity. However, the GMM needs special quadratic moment conditions.

We consider a fixed effects spatial panel data model with unknown heteroskedasticity, which has MESS processes in both the dependent variable and disturbances. With individual effects being concentrated out from the QML function, we find that the QMLE can still be consistent when the two spatial weights matrices for the MESS processes are commutative. We derive the asymptotic distribution of the QMLE and provide a consistent estimator for its standard deviation. In practice, the spatial weights matrices in the two processes are often specified to be the same, so they are commutative. If there is no spatial dependence in disturbances, the commutativity property is automatically satisfied. Some regularity assumptions, lemmas and proofs are provided in a supplementary file.

## 2 Model and estimation

We consider the following MESS panel data model (MESSPD(1,1)):

$$e^{\alpha W_n} Y_{nt} = X_{nt}\beta + \mathbf{c}_n + U_{nt}, \quad e^{\tau M_n} U_{nt} = V_{nt}, \quad t = 1, 2, \ldots, T, \tag{1}$$

where $Y_{nt} = (y_{1t}, y_{2t}, \cdots, y_{nt})'$ and $V_{nt} = (v_{1t}, \cdots, v_{nt})'$ are $n \times 1$ vectors of observations on the dependent variable and disturbances at time $t$. The $v_{it}$'s are independent $(0, \sigma_{it}^2)$. $X_{nt}$ is an $n \times k$ matrix of exogenous time-varying regressors with coefficient vector $\beta$. The $n \times n$ nonstochastic spatial weights matrices $W_n$ and $M_n$, which may or may not be different, capture the spatial dependence on, respectively, $y_{it}$ and $v_{it}$ among cross sectional units. $\alpha$ and $\tau$ are scalar spatial dependence parameters, and $\mathbf{c}_n$ is an $n \times 1$ vector of fixed effects.

Denote $\zeta = (\beta', \alpha, \tau)'$ and $\theta = (\zeta, \sigma^2)'$, with $\zeta_0 = (\beta_0', \alpha_0, \tau_0)'$ and $\theta_0 = (\zeta_0, \sigma_0^2)'$ being their true values. As $(e^{\alpha D})^{-1} = e^{-\alpha D}$ for any square matrix $D$, the reduced form of $Y_{nt}$ is

$$Y_{nt} = e^{-\alpha_0 W_n}(X_{nt}\beta_0 + \mathbf{c}_{n0} + e^{-\tau_0 M_n} V_{nt}). \tag{2}$$

It follows that the variance-covariance (VC) matrix of $Y_{nt}$ is $e^{-\alpha_0 W_n} e^{-\tau_0 M_n} E(V_{nt}V_{nt}') e^{-\tau_0 M_n'} e^{-\alpha_0 W_n'}$, which is always positive definite. It is unnecessary to impose any constraint on the parameter spaces of $\alpha$ and $\tau$. By contrast, for the SAR model with SAR disturbances (SARAR model), if the eigenvalues of a spatial weights matrix $W_n$ are all real, the parameter space for the associated spatial dependence parameter is $(1/\mu_{\min}, 1/\mu_{\max})$, where $\mu_{\min}$ and $\mu_{\max}$ are, respectively, the smallest and largest eigenvalues of $W_n$.[2]

[2] If $W_n$ is row-normalized, as $\mu_{\max} = 1$, the parameter space becomes $(1/\mu_{\min}, 1)$. See, e.g., Kelejian and Prucha (2010) and Elhorst et al. (2012) on discussions of parameter spaces for SAR models.

For any square matrix $D$, $|e^{\alpha D}| = e^{\text{trace}(\alpha D)} = e^{\alpha\,\text{trace}(D)}$. Then as $W_n$ and $M_n$ have zero diagonals, the quasi log likelihood function of the model, as if the disturbances $v_{it}$'s were normal, is

$$\ln L_{nT}(\theta, \mathbf{c}_n) = -\frac{nT}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} V'_{nt}(\zeta, \mathbf{c}_n)V_{nt}(\zeta, \mathbf{c}_n). \tag{3}$$

where $V_{nt}(\zeta, \mathbf{c}_n) = e^{\tau M_n}(e^{\alpha W_n}Y_{nt} - X_{nt}\beta - \mathbf{c}_n)$. Let $\tilde{Y}_{nt} = Y_{nt} - \bar{Y}_{nT}$ for $t = 1, ..., T$, where $\bar{Y}_{nT} = \frac{1}{T}\sum_{t=1}^{T} Y_{nt}$. Define $\tilde{X}_{nt}$ and $\tilde{V}_{nt}$ similarly. With a large $n$, our analysis is based on the following concentrated log likelihood function with individual effects $\mathbf{c}_n$ being concentrated out:

$$\ln L_{nT}(\theta) = -\frac{nT}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta), \tag{4}$$

where $\tilde{V}_{nt}(\zeta) = e^{\tau M_n}(e^{\alpha W_n}\tilde{Y}_{nt} - \tilde{X}_{nt}\beta)$. Under unknown heteroskedasticity, the parameters we are interested in do not cover variance parameters, as the number of different variances increases with the number of cross sections. By (4), for given $\sigma^2$, other parameter estimates are derived from the function

$$\Gamma_{nT}(\zeta) = \sum_{t=1}^{T}\tilde{V}'_{nt}(\zeta)\tilde{V}_{nt}(\zeta), \tag{5}$$

which does not involve any variance parameter. The first order derivative of (5) is

$$\frac{\partial \Gamma_{nT}(\zeta)}{\partial \zeta} = \begin{pmatrix} -2\sum_{t=1}^{T}(e^{\tau M_n}\tilde{X}_{nt})'\tilde{V}_{nt}(\zeta) \\ 2\sum_{t=1}^{T}(e^{\tau M_n}W_n e^{\alpha W_n}\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta) \\ 2\sum_{t=1}^{T}(M_n\tilde{V}_{nt}(\zeta))'\tilde{V}_{nt}(\zeta) \end{pmatrix}. \tag{6}$$

The VC matrix of disturbances $\Sigma_{nT} = diag(\Sigma_{nT,1}, \cdots, \Sigma_{nT,T})$ is a diagonal matrix, where each block $\Sigma_{nT,t} = diag(\sigma_{1t}^2, \cdots, \sigma_{nt}^2)$ is a diagonal matrix formed by $\sigma_{it}^2 = E(v_{it}^2)$ for $i = 1, \ldots, n$. Note that

$$E\Big(\sum_{t=1}^{T}(M_n\tilde{V}_{nt}(\zeta_0))'\tilde{V}_{nt}(\zeta_0)\Big) = tr\big[(J_T \otimes M_n)\Sigma_{nT}\big],$$

where $J_T = I_T - \frac{1}{T}l_T l'_T$ is an operator due to the elimination of individual effects. As the diagonal elements of $J_T$ are the same and the diagonal elements of $M_n$ are all zero, $tr\big[(J_T \otimes M_n)\Sigma_{nT}\big] = 0$. In addition,

$$E\Big(\sum_{t=1}^{T}(e^{\tau_0 M_n}W_n e^{\alpha_0 W_n}\tilde{Y}_{nt})'\tilde{V}_{nt}(\zeta_0)\Big) = tr\big[(J_T \otimes e^{-\tau_0 M'_n}W'_n e^{\tau_0 M'_n})\Sigma_{nT}\big],$$

which may not be equal to zero in general. But if the spatial weights matrices $W_n$ and $M_n$ are commutative, then $W'_n e^{\tau_0 M'_n} = e^{\tau_0 M'_n}W'_n$ and $tr\big[(J_T \otimes e^{-\tau_0 M'_n}W'_n e^{\tau_0 M'_n})\Sigma_{nT}\big] = tr\big[(J_T \otimes W'_n)\Sigma_{nT}\big] = 0$. Thus, with commutative $W_n$ and $M_n$, the QMLE $\hat{\zeta}_{nT}$ of $\zeta$ derived from the minimization of (5) can be consistent under heteroskedasticity. Without the commutativity of $W_n$ and $M_n$, the QMLE is generally inconsistent, so this condition may not be extended and we maintain the following assumption.

**Assumption 1.** *$W_n$ and $M_n$ are constant spatial weights matrices with zero diagonals, and they are commutative.*

Note that for given $\eta = (\alpha, \tau)'$, minimizing $\Gamma_{nT}(\zeta)$ yields a closed form estimate of $\beta$:

$$\hat{\beta}_{nT}(\eta) = \Big(\sum_{t=1}^{T} (e^{\tau M_n} \tilde{X}_{nt})' e^{\tau M_n} \tilde{X}_{nt}\Big)^{-1} \sum_{t=1}^{T} (e^{\tau M_n} \tilde{X}_{nt})' e^{\tau M_n} e^{\alpha W_n} \tilde{Y}_{nt}. \tag{7}$$

This estimate can be substituted into (5) to derive a function of only $\eta$.

**Theorem 1.** *Under Assumption 1 and other standard regularity conditions in the supplementary file, the QMLE $\hat{\zeta}_{nT}$ is consistent for $\zeta_0$, i.e., $\hat{\zeta}_{nT} \xrightarrow{p} \zeta_0$.*

For the asymptotic distribution of $\hat{\zeta}_{nT}$, let $\mathbf{X}_{nT} = (X'_{n1}, X'_{n2}, \cdots, X'_{nT})'$,

$$\begin{aligned} \mathcal{H}_{\zeta_0,nT} &= \frac{1}{nT} E\Big(\frac{\partial^2 \Gamma_{nT}(\zeta_0)}{\partial\zeta\partial\zeta'}\Big) \\ &= \frac{2}{nT} \begin{pmatrix} \sum_{t=1}^{T} (e^{\tau_0 M_n} \tilde{X}_{nt})' e^{\tau_0 M_n} \tilde{X}_{nt} & * & * \\ -\sum_{t=1}^{T} (e^{\tau_0 M_n} W_n \tilde{X}_{nt} \beta_0)' e^{\tau_0 M_n} \tilde{X}_{nt} & \mathcal{H}_{22} & * \\ 0 & tr\big[(J_T \otimes M_n^s W_n)\Sigma_{nT}\big] & tr\big[(J_T \otimes M_n^s M_n)\Sigma_{nT}\big] \end{pmatrix}, \end{aligned} \tag{8}$$

where $\mathcal{H}_{22} = \sum_{t=1}^{T} (e^{\tau_0 M_n} W_n \tilde{X}_{nt} \beta_0)' (e^{\tau_0 M_n} W_n \tilde{X}_{nt} \beta_0) + tr\big[(J_T \otimes W_n^s W_n)\Sigma_{nT}\big]$, and

$$\begin{aligned} \Delta_{\zeta_0,nT} &= \frac{1}{nT} E\Big(\frac{\partial \Gamma_{nT}(\zeta_0)}{\partial\zeta} \cdot \frac{\partial \Gamma_{nT}(\zeta_0)}{\partial\zeta'}\Big) \\ &= \frac{2}{nT} \begin{pmatrix} 2\mathbf{X}'_{nT}(J_T \otimes e^{\tau_0 M'_n})\Sigma_{nT}(J_T \otimes e^{\tau_0 M_n})\mathbf{X}_{nT} & * & * \\ -2\beta'_0 \mathbf{X}'_{nT}(J_T \otimes W'_n e^{\tau_0 M'_n})\Sigma_{nT}(J_T \otimes e^{\tau_0 M_n})\mathbf{X}_{nT} & \Delta_{22} & * \\ 0 & \Delta_{32} & \Delta_{33} \end{pmatrix}, \end{aligned} \tag{9}$$

where $\Delta_{22} = 2\beta'_0 \mathbf{X}'_{nT}(J_T \otimes W'_n e^{\tau_0 M'_n})\Sigma_{nT}(J_T \otimes e^{\tau_0 M_n} W_n)\mathbf{X}_{nT}\beta_0 + tr\big[\Sigma_{nT}(J_T \otimes W_n^s)\Sigma_{nT}(J_T \otimes W_n^s)\big]$, $\Delta_{32} = tr\big[\Sigma_{nT}(J_T \otimes M_n^s)\Sigma_{nT}(J_T \otimes W_n^s)\big]$, and $\Delta_{33} = tr\big[\Sigma_{nT}(J_T \otimes M_n^s)\Sigma_{nT}(J_T \otimes M_n^s)\big]$.[3] For any square matrix $A_n$, let $vec(A_n)$ denote its vectorization and $A_n^s = A_n + A'_n$. As $M_n$ and $W_n$ are commutative, we can write $\Delta_{\zeta_0,nT}$ as $\Delta_{\zeta_0,nT} = \frac{1}{nT}\Delta'_{\zeta_0,1nT}\Delta_{\zeta_0,1nT}$, where

$$\Delta_{\zeta_0,1nT} = \begin{pmatrix} -2\Sigma_{nT}^{\frac{1}{2}}(J_T \otimes e^{\tau_0 M_n})\mathbf{X}_{nT} & 2\Sigma_{nT}^{\frac{1}{2}}(J_T \otimes e^{\tau_0 M_n} W_n)\mathbf{X}_{nT}\beta_0 & 0 \\ 0 & vec\big(\Sigma_{nT}^{\frac{1}{2}}(J_T \otimes W_n^s)\Sigma_{nT}^{\frac{1}{2}}\big) & vec\big(\Sigma_{nT}^{\frac{1}{2}}(J_T \otimes M_n^s)\Sigma_{nT}^{\frac{1}{2}}\big) \end{pmatrix}. \tag{10}$$

Thus, $\Delta_{\zeta_0,nT}$ is positive semi-definite. The following assumption can ensure the nonsingularity of $\mathcal{H}_{\zeta_0,nT}$ in the limit.

**Assumption 2.** $\lim \frac{1}{nT} tr[(J_T \otimes M_n^s M_n)\Sigma_{nT}] \neq 0$ *and* $\lim \frac{1}{nT}\big[(\tilde{\mathbf{X}}_{nT}\beta_0)'(I_T \otimes W'_n e^{\tau_0 M'_n})H_{nT}(\tau_0)(I_T \otimes e^{\tau_0 M_n} W_n)\tilde{\mathbf{X}}_{nT}\beta_0 + tr[(J_T \otimes W_n^s W_n)\Sigma_{nT}] - \frac{tr^2[(J_T \otimes M_n^s W_n)\Sigma_{nT}]}{tr[(J_T \otimes M_n^s M_n)\Sigma_{nT}]}\big] \neq 0$*, where* $\tilde{\mathbf{X}}_{nT} = [\tilde{X}'_{n1}, \ldots, \tilde{X}'_{nT}]'$ *and* $H_{nT}(\tau_0) = I_{nT} - (I_T \otimes e^{\tau_0 M_n})\tilde{\mathbf{X}}_{nT}[\tilde{\mathbf{X}}'_{nT}(I_T \otimes e^{\tau_0 M'_n} e^{\tau_0 M_n})\tilde{\mathbf{X}}_{nT}]^{-1}\tilde{\mathbf{X}}'_{nT}(I_T \otimes e^{\tau_0 M'_n})$.

[3] We might expect that the variance matrix $\Delta_{\zeta_0,nT}$ involves the third and fourth moments of disturbances. However, due to the commutativity of $W_n$ and $M_n$, relevant terms involving those higher order moments will be multiplied by the diagonal elements of $W_n$ and $M_n$, which are zero. Thus, $\Delta_{\zeta_0,nT}$ does not involve the third and fourth moments of disturbances.

**Theorem 2.** *Under Assumptions 1–2 and other regularity conditions in the supplementary file,*

$$\sqrt{nT}(\hat{\zeta}_{nT}-\zeta_0) \xrightarrow{d} N\big(0, \lim(\mathcal{H}_{\zeta_0,nT}^{-1}\Delta_{\zeta_0,nT}\mathcal{H}_{\zeta_0,nT}^{-1})\big),$$

*where the limit is taken under a large $n$ and a large or finite $T$.*

As the number of individual effects increases with $n$, one might expect the incidental parameter problem with a finite $T$ (Neyman and Scott, 1948) and there can still be an inconsistency problem for a large $T$ (Lee and Yu, 2010). However, for the SAR panel data model, with individual effects being concentrated out from the quasi log likelihood function, those problems only occur to the variance parameter as shown in Lee and Yu (2010). As our MESSPD(1,1) is similar to the SAR panel data model but we do not estimate any variance parameter for our model with heteroskedasticity, there are no such problems and the convergence rate of the QMLE is $\sqrt{nT}$.

We can define White (1980) type consistent estimators of $\mathcal{H}_{\zeta_0,nT}$ and $\Delta_{\zeta_0,nT}$ to make valid inference using the QMLE $\hat{\zeta}_{nT}$ under heteroskedasticity. Due to the use of the deviation from the mean operator to eliminate individual effects, we can only obtain estimates of $\tilde{v}_{it} = v_{it} - \frac{1}{T}\sum_{t=1}^{T} v_{it}$, namely the residuals $\hat{\tilde{v}}_{it}$ from the QML estimation. As $E(\sum_{t=1}^{T}\tilde{v}_{it}^2) = \frac{T-1}{T}\sum_{t=1}^{T}\sigma_{it}^2$, $\Sigma_{nT}$ in $\mathcal{H}_{\zeta_0,nT}$ and $\Delta_{\zeta_0,nT}$ can be replaced by $\hat{\Sigma}_{nT} = \frac{T}{T-1}diag(\hat{\tilde{v}}_{11}^2,\cdots,\hat{\tilde{v}}_{n1}^2,\cdots,\hat{\tilde{v}}_{1T}^2,\cdots,\hat{\tilde{v}}_{nT}^2)$ to obtain $\hat{\mathcal{H}}_{\zeta_0,nT}$ and $\hat{\Delta}_{\zeta_0,nT}$. When $\sigma_{it}^2$ is allowed to depend on $t$, even if $\Sigma_{nT}$ is replaced by $\hat{\Sigma}_{nT}$, we may not have a consistent covariance matrix estimator when $T$ is finite.[4] A large $T$ ensures the consistency of $\hat{\mathcal{H}}_{\zeta_0,nT}$ and $\hat{\Delta}_{\zeta_0,nT}$. When the heteroskedasticity is set as $v_{it} \sim (0,\sigma_i^2)$, i.e., $\sigma_{it}^2$ only depends on $i$, $E(\tilde{v}_{it}^2) = \frac{T-1}{T}\sigma_i^2$. Then as long as $n$ is large, we will have a consistent covariance matrix estimator regardless of whether $T$ is fixed or tends to infinity.

**Theorem 3.** *Suppose that Assumptions 1–2 and regularity assumptions in the supplementary file hold. Under either (a) both $n$ and $T$ are large or (b) $n$ is large, $T$ is finite and $v_{it} \sim (0,\sigma_i^2)$, $\hat{\mathcal{H}}_{\zeta_0,nT} = \mathcal{H}_{\zeta_0,nT} + o_p(1)$ and $\hat{\Delta}_{\zeta_0,nT} = \Delta_{\zeta_0,nT} + o_p(1)$.*

# 3 Monte Carlo

We conduct some Monte Carlo simulations to evaluate the finite sample performance of our estimator. Samples are generated from model (1), where $v_{it}$'s are normally distributed with mean zero and standard deviation $1+\frac{k}{nT}$ for $k = 1,\ldots,nT$, and $X_{1,nt}$ and $X_{2,nt}$ are independently drawn from $N(0,4)$ and $U(1,5)$ respectively. The true value of $\zeta$ is either $(1,1,-2,1)'$ or $(1,1,-2,-1)'$. We use the latitudes and longitudes of 30 provinces and autonomous regions of Chinese mainland to generate a geographical weights matrices based on 10 nearest neighbors and then row-normalize it. This matrix is used to construct block-diagonal spatial weights matrix $W_n = M_n$ for $n = 90$, 120, or 150. The number of time periods $T$ is 5, 10, or 20. The number of Monte Carlo repetitions is 1000. We report Bias, E-SD, T-SD and RMSE, where Bias is the average of estimation biases from 1000 repetitions, E-SD is the empirical standard deviation of the estimates, T-SD is the estimated theoretical standard

[4]When $v_{it} \sim (0,\sigma_{it}^2)$ and $T$ is finite, this replacement generally does not generate consistent estimators for terms in $\Delta_{\zeta_0,nT}$ that involve two $\Sigma_{nT}$, i.e., terms with the form $\frac{1}{nT}tr[\Sigma_{nT}(J_T\otimes W_n^s)\Sigma_{nT}(J_T\otimes M_n^s)]$. See the proof of Theorem 3 in the supplementary file for details.

deviation and RMSE is the root mean squared error.[5] Monte Carlo results are summarized in Table 1. We observe that the biases of the QMLEs are very small, so E-SDs and RMSEs are close. The E-SDs and T-SDs are similar. For given $n$, Biases, E-SDs, T-SDs and RMSEs decrease as $T$ increases.

Table 1: QML estimation results for $\zeta$ under heteroskedasticity

| | | | $\zeta_0 = (1, 1, -2, 1)'$ | | | | | $\zeta_0 = (1, 1, -2, -1)'$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $T$ | | $\beta_1$ | $\beta_2$ | $\alpha$ | $\tau$ | | $\beta_1$ | $\beta_2$ | $\alpha$ | $\tau$ |
| 90 | 5 | Bias | 0.0005 | -0.0022 | -0.0013 | 0.0365 | Bias | -0.0007 | -0.0022 | 0.0061 | 0.0126 |
| | | E-SD | 0.0427 | 0.0439 | 0.0504 | 0.1507 | E-SD | 0.0459 | 0.0458 | 0.1538 | 0.2116 |
| | | T-SD | 0.0416 | 0.0416 | 0.0494 | 0.1362 | T-SD | 0.0440 | 0.0440 | 0.1447 | 0.1928 |
| | | RMSE | 0.0427 | 0.0439 | 0.0504 | 0.1550 | RMSE | 0.0459 | 0.0459 | 0.1540 | 0.2120 |
| | 10 | Bias | -0.0003 | 0.0020 | 0.0021 | 0.0115 | Bias | -0.0015 | 0.0014 | 0.0096 | -0.0034 |
| | | E-SD | 0.0282 | 0.0301 | 0.0330 | 0.0938 | E-SD | 0.0299 | 0.0316 | 0.0976 | 0.1289 |
| | | T-SD | 0.0278 | 0.0278 | 0.0328 | 0.0912 | T-SD | 0.0294 | 0.0293 | 0.0969 | 0.1291 |
| | | RMSE | 0.0282 | 0.0302 | 0.0331 | 0.0945 | RMSE | 0.0299 | 0.0316 | 0.0981 | 0.1289 |
| | 20 | Bias | 0.0007 | -0.0000 | 0.0012 | 0.0060 | Bias | 0.0007 | -0.0004 | 0.0029 | 0.0011 |
| | | E-SD | 0.0190 | 0.0196 | 0.0219 | 0.0649 | E-SD | 0.0199 | 0.0212 | 0.0665 | 0.0905 |
| | | T-SD | 0.0192 | 0.0192 | 0.0225 | 0.0629 | T-SD | 0.0202 | 0.0202 | 0.0667 | 0.0889 |
| | | RMSE | 0.0190 | 0.0196 | 0.0219 | 0.0652 | RMSE | 0.0199 | 0.0212 | 0.0665 | 0.0905 |
| 120 | 5 | Bias | -0.0001 | -0.0004 | -0.0008 | 0.0236 | Bias | -0.0004 | -0.0004 | -0.0016 | 0.0112 |
| | | E-SD | 0.0367 | 0.0370 | 0.0442 | 0.1231 | E-SD | 0.0386 | 0.0385 | 0.1268 | 0.1710 |
| | | T-SD | 0.0360 | 0.0359 | 0.0428 | 0.1181 | T-SD | 0.0381 | 0.0380 | 0.1259 | 0.1674 |
| | | RMSE | 0.0367 | 0.0370 | 0.0442 | 0.1254 | RMSE | 0.0386 | 0.0385 | 0.1268 | 0.1713 |
| | 10 | Bias | 0.0002 | 0.0010 | 0.0015 | 0.0075 | Bias | 0.0002 | 0.0007 | 0.0007 | 0.0030 |
| | | E-SD | 0.0244 | 0.0246 | 0.0281 | 0.0811 | E-SD | 0.0259 | 0.0262 | 0.0856 | 0.1140 |
| | | T-SD | 0.0241 | 0.0241 | 0.0285 | 0.0791 | T-SD | 0.0254 | 0.0254 | 0.0837 | 0.1117 |
| | | RMSE | 0.0244 | 0.0246 | 0.0281 | 0.0815 | RMSE | 0.0259 | 0.0262 | 0.0856 | 0.1140 |
| | 20 | Bias | -0.0001 | -0.0006 | 0.0007 | 0.0012 | Bias | -0.0002 | -0.0010 | 0.0019 | -0.0025 |
| | | E-SD | 0.0166 | 0.0176 | 0.0200 | 0.0556 | E-SD | 0.0176 | 0.0187 | 0.0586 | 0.0762 |
| | | T-SD | 0.0166 | 0.0166 | 0.0195 | 0.0546 | T-SD | 0.0175 | 0.0175 | 0.0578 | 0.0771 |
| | | RMSE | 0.0166 | 0.0176 | 0.0200 | 0.0556 | RMSE | 0.0176 | 0.0187 | 0.0586 | 0.0763 |
| 150 | 5 | Bias | 0.0004 | -0.0023 | -0.0017 | 0.0215 | Bias | 0.0002 | -0.0026 | 0.0008 | 0.0097 |
| | | E-SD | 0.0340 | 0.0333 | 0.0387 | 0.1151 | E-SD | 0.0353 | 0.0347 | 0.1104 | 0.1528 |
| | | T-SD | 0.0323 | 0.0323 | 0.0382 | 0.1057 | T-SD | 0.0342 | 0.0341 | 0.1124 | 0.1497 |
| | | RMSE | 0.0340 | 0.0334 | 0.0388 | 0.1171 | RMSE | 0.0353 | 0.0348 | 0.1104 | 0.1531 |
| | 10 | Bias | 0.0006 | 0.0002 | -0.0000 | 0.0068 | Bias | 0.0004 | -0.0002 | 0.0020 | 0.0006 |
| | | E-SD | 0.0215 | 0.0215 | 0.0249 | 0.0717 | E-SD | 0.0232 | 0.0229 | 0.0751 | 0.1016 |
| | | T-SD | 0.0216 | 0.0215 | 0.0254 | 0.0708 | T-SD | 0.0228 | 0.0227 | 0.0751 | 0.1001 |
| | | RMSE | 0.0215 | 0.0215 | 0.0249 | 0.0720 | RMSE | 0.0232 | 0.0229 | 0.0751 | 0.1016 |
| | 20 | Bias | -0.0007 | -0.0004 | -0.0003 | 0.0037 | Bias | -0.0008 | -0.0005 | 0.0005 | 0.0008 |
| | | E-SD | 0.0150 | 0.0152 | 0.0169 | 0.0487 | E-SD | 0.0158 | 0.0160 | 0.0505 | 0.0681 |
| | | T-SD | 0.0148 | 0.0149 | 0.0174 | 0.0488 | T-SD | 0.0157 | 0.0156 | 0.0516 | 0.0689 |
| | | RMSE | 0.0150 | 0.0152 | 0.0169 | 0.0488 | RMSE | 0.0158 | 0.0160 | 0.0505 | 0.0681 |

[5]The T-SD is obtained from the diagonal elements of the estimated $\mathcal{H}^{-1}_{\zeta_0,nT}\Delta_{\zeta_0,nT}\mathcal{H}^{-1}_{\zeta_0,nT}$.

# 4 Real data analysis

Ertur and Musolesi (2017) use the spatial error model to estimate both trade-unrelated and trade-related geographic spillovers among 24 countries. Due to differences in the regional environment of different countries, there might be heteroskedasticity in $V_{nt}$. We use our MESSPD(1,1) to analyze the geographic R&D spillovers among 24 countries over the period 1971–2004.[6]

We consider a MESSPD(1,1) model:

$$\begin{aligned} e^{\alpha W_n} \log F_{nt} &= \beta_1 \log GS_{nt}^d + \beta_2 \log NS_{nt}^d + \beta_3 \log GS_{nt}^f + \beta_4 \log NS_{nt}^f + \beta_5 \log GS_{nt}^{tf} \\ &\quad + \beta_6 \log NS_{nt}^{tf} + \beta_7 \log GH_{nt} + \beta_8 \log NH_{nt} + a_n + U_{nt}, \\ e^{\tau W_n} U_{nt} &= V_{nt}, \end{aligned} \tag{11}$$

where $F_{nt}$ is TFP of 24 countries at $t$, $GS_{nt}^d$ and $NS_{nt}^d$ are domestic R&D capital stocks of G7 and non-G7 countries respectively, $GS_{nt}^f$ and $NS_{nt}^f$ are foreign capital stocks containing trade-unrelated geographic spillovers of G7 and non-G7 countries respectively, $GS_{nt}^{tf}$ and $NS_{nt}^{tf}$ are foreign capital stocks containing trade-related geographic spillovers of G7 and non-G7 countries respectively, $GH_{nt}$ and $NH_{nt}$ are human capital of G7 and non-G7 countries respectively, $W_n = [w_{ij}]$ is an $n \times n$ matrix with $w_{ij} = e^{-d_{ij}} / \sum_{j=1}^n e^{-d_{ij}}$, where $d_{ij}$ is the geographic distance between country $i$ and country $j$, and $a_n$ is a vector of individual fixed effects. We also report estimation results from MESSPD(0,1), where $\alpha$ in (11) is restricted to zero.

Table 2 summarizes estimation results. We note that there is no significant spatial dependence in $\log F_{nt}$, thus the results of the two models are similar and they are close to those in Ertur and Musolesi (2017). Domestic R&D significantly affects both G7 and non-G7 countries; Non-G7 countries benefit more from trade-unrelated geographic spillovers than G7 countries; The two groups significantly benefit from trade-related geographic spillovers; and human capital of non-G7 countries does not significantly affect the two groups of countries.

Table 2: Estimation results for the real data analysis

| | $\log GS_{nt}^d$ | $\log NS_{nt}^d$ | $\log GS_{nt}^f$ | $\log NS_{nt}^f$ | $\log GS_{nt}^{tf}$ | $\log NS_{nt}^{tf}$ | $\log GH_{nt}$ | $\log NH_{nt}$ | Spat auto in $\log F_{nt}$ | Spat auto in $U_{nt}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| (i) | $0.10^{***}$ | $0.034^{***}$ | $-0.029$ | $0.16^{***}$ | $0.14^{***}$ | $0.13^{***}$ | $0.21^{***}$ | 0.076 | 0.026 | $-0.43^{**}$ |
| | (2.90) | (4.95) | (-0.63) | (2.33) | (10.97) | (2.94) | (4.37) | (0.45) | (0.099) | (-1.99) |
| (ii) | $0.099^{***}$ | $0.035^{***}$ | $-0.031$ | $0.15^{***}$ | $0.14^{***}$ | $0.13^{***}$ | $0.21^{**}$ | 0.077 | | $-0.41^{*}$ |
| | (2.92) | (5.04) | (-0.70) | (2.36) | (10.86) | (2.93) | (4.71) | (0.46) | | (-1.95) |

1. Asymptotic t-statistics in brackets.
2. (i): MESSPD(1,1); (ii): MESSPD(0,1).
3. $^{***}$, $^{**}$, $^{*}$: Significant at 1%, 5%, and 10%, respectively.

[6] The data set is from Ertur and Musolesi (2017).

# 5 Conclusion

In this article, we show that, for the MESS panel data model (1), when the spatial weights matrices are commutative, the QMLEs can be consistent and asymptotically normal. This is a desirable feature that the SAR counterpart does not have. The MESS panel model does not impose additional restrictions on the parameter spaces. As the QML function of the model does not involve any Jacobian, it is computationally more attractive than the SAR counterpart. Monte Carlo simulations demonstrate that QMLEs perform satisfactorily in finite samples and a real data analysis shows the practicability of our model.

## Acknowledgements

We are grateful to the Editor and a referee for helpful comments that lead to improvements of this paper. This work is partially supported by grants from the National Natural Science Foundation of China (No. 71874106 and 71833004). Fei Jin also acknowledges financial support from Shanghai Pujiang Program (No. 15PJC043) and Program for Innovative Research Team of Shanghai University of Finance and Economics.

## References

[1] Debarsy N, Jin F, Lee L F. Large sample properties of the matrix exponential spatial specification with an application to FDI [J]. *Journal of Econometrics*, 188 (1): 1–21 (2015).

[2] Elhorst J P, Lacombe D J, Piras G. On model specification and parameter space definitions in higher order spatial econometric models [J]. *Regional Science & Urban Economics*, 42 (1–2): 211–220 (2012).

[3] Ertur, Cem, and Antonio Musolesi. Weak and Strong Cross-Sectional Dependence: A Panel Data Analysis of International Technology Diffusion [J]. *Journal of Applied Econometrics*, 32 (3): 477–503 (2017).

[4] Figueiredo C, Silva A R D. A matrix exponential spatial specification approach to panel data models [J]. *Empirical Economics*, 49 (1): 115–129 (2015).

[5] Kelejian H H, Prucha I R. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances [J]. *Journal of Econometrics*, 157 (1): 53–67 (2010).

[6] Lesage J P, Pace R K. A matrix exponential spatial specification [J]. *Journal of Econometrics*, 140(1): 190–214 (2007).

[7] Lee L F, Yu J. Estimation of spatial autoregressive panel data models with fixed effects [J]. *Journal of Econometrics*, 154 (2): 165–185 (2010).

[8] Moscone F, Tosetti E. GMM estimation of spatial panels with fixed effects and unknown heteroskedasticity [J]. *Regional Science & Urban Economics*, 41 (5): 487–497 (2011).

[9] Neyman, J., Scott, E L. Consistent estimates based on partially consistent observations [J]. *Econometrica*, 16: 1–32 (1948).

[10] White H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity [J]. *Econometrica*, 48 (4): 817–838 (1980).