Statistica Sinica Preprint No: SS-2021-0280	
Title	Spatial-temporal Model with Heterogeneous Random Effects
Manuscript ID	SS-2021-0280
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0280
Complete List of Authors	Xingdong Feng, Wenyu Li and Qianqian Zhu
Corresponding Author	Qianqian Zhu
E-mail	zhu.qianqian@mail.shufe.edu.cn

doi:10.5705/ss.202021.0280

Spatial-temporal Model with Heterogeneous Random Effects

Xingdong Feng, Wenyu Li and Qianqian Zhu

School of Statistics and Management

Shanghai University of Finance and Economics

Abstract: In this paper, we propose a novel spatial-temporal model with individual random effects characterized by a location-scale structure, which allows us to flexibly capture the pure influence of space-specific factors in a quantile regression framework. A hybrid two-stage estimation procedure is introduced for this model. The first stage proposes a Gaussian quasi-maximum likelihood estimator for the spatial-temporal effects, and the second constructs a weighted conditional quantile estimator, which we use to study the conditional quantiles of the random effects related to space-specific attributes. We verify the validity of the two-stage hybrid estimation, and establish the asymptotic properties of our estimators. The results of our simulation study indicate that the proposed estimation procedure performs well in different scenarios with finite-samples. Lastly, we apply the proposed method to data from a real case study on the air quality of China.

Keywords: Dynamic spatial autoregressive models; Hybrid estimation; Quantile regression; Quasi-maximum likelihood estimation; Random effects.

1. Introduction

With the rapid development of the world economy and the gradual deterioration of the global ecological environment, many countries have established and improved environmental protection mechanisms to maintain the ecological equilibrium (Streck, 2004; Busch and Jörgens, 2005). At the same time, urban air quality has attracted much attention (Wang and

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

Hao, 2012; Gulia et al., 2015), because poor air quality has adverse effects on human health and survival (Cakmak et al., 2011; Bowatte et al., 2015). Researchers have increasingly focused on quantitative analyses of the impact of social and anthropogenic driving forces on urban air quality. For example, De Bruyn et al. (1998) and Chen and Xu (2017) studied the relationship between air quality and economic development; Cramer (1998) and Chen et al. (2020) investigated the influence of population density on air quality; and Li et al. (2018) and Yu and Liu (2020) considered the impact of changes to industrial structure on air quality. However, as a complex and large system, the urban ambient air system is not only affected by human activities, but also by spatial-temporal effects, including historical patterns, geographical location, meteorological conditions, and so on (Yang et al., 2007; Giri et al., 2008; Zhou et al., 2022). As a result, it has become necessary to explore the social forces affecting urban air quality pressure after removing spatial-temporal effects.

In the literature, spatial effects that depict a geographical correlation are usually modeled in three forms: (i) spatial autoregressive (SAR) models, with the spatial effect built at the response variable (Lee, 2004; Yu et al., 2008); (ii) spatial error models, with the spatial effect modeled in the disturbance term (Kapoor et al., 2007; Kelejian and Prucha, 2010; Su and Yang, 2015); and (iii) spatial panel data models with the spatial effect simultaneously reflected in both the responses and the errors (Kelejian and Prucha, 1998, 1999). Existing models include temporal effects that capture autocorrelation in the form of a lagged response in the same way as in time series models. As the practical need to analyze spatial-temporal data has increased, numerous spatial panel data models that include spatial and temporal effects simultaneously have been studied. For example, Yu et al. (2008) considered a Gaussian quasi-maximum likelihood estimator (QMLE) for the spatial dynamic panel data (SDPD) model with fixed effects, when both the individual dimension N and the time dimension T are large. Lee and Yu (2014) proposed generalized method of moments estimators for the SDPD model with fixed effects when N is large and T can be large, but small relative to N. Su and Yang (2015) studied the QMLE for dynamic panel models with spatial errors and random/fixed effects when N is large, but T is fixed. In this paper, we propose a dynamic spatial autoregressive (DSAR) model to capture the spatial-temporal effects and investigate its estimation.

For spatial panel data models, individual effects are important, and are widely investigated in the form of fixed effects (Yu et al., 2008; Lee and Yu, 2010; Shi and Lee, 2017) or random effects (Baltagi et al., 2013; Su and Yang, 2015). Fixed and random effects both have advantages and disadvantages. The random-effect model avoids the "incidental parameter" problem (Neyman and Scott, 1948) in fixed-effect models and has more efficient estimators, but it neglects possible correlations between individual effects and regressors. The estimators of fixed-effect models are robust in the sense that the fixed effect can be arbitrarily related to the regressors; see page 624 of Hansen (2021), Mundlak (1978) and Li and Yang (2021). Note that the random effects in existing studies are assumed to have zero means and constant variances, which implies that the random effects are homogeneous for individuals. However, individual random effects can be influenced by space-specific covariates, which may make the random effects heterogeneous. For example, the influence of socio-economic factors on air quality is heterogeneous between cities. In general, heterogeneity across individuals can be described by fixed effects or heteroskedastic random effects (Baltagi et al., 2010), both of which emphasize the differences between individuals, but cannot be used to explore the heterogeneity information. Motivated by this fact, we consider a space-specific, covariate-dependent location-scale structure for the individual effects. Specifically, the location of an individual effect depicts the differences between individuals on the conditional mean, and can be arbitrarily related to time-variant regressors. Thus it can be viewed as a fixed effect. In contrast, the scale of an individual effect is random and allows for conditional heteroscedasticity. Hence, the location-scale structure combines the advantages of fixed and random effects. Moreover, it can flexibly extract the influence of space-specific covariates on the whole distribution of the individual random effect.

Quantile regressions (Koenker and Bassett, 1978) are popular tools for capturing heterogeneity and exploiting the distributional information in data, and provide a complete picture for studying the influence of the covariates on the response. In the example of air quality, we are interested in exploring the influence of socio-economic factors on air quality after removing spatial-temporal effects, not only on the average level, but also on the different quantile levels, especially the lower and upper ones. This motivates us to adopt a quantile regression for the proposed individual effects in the location-scale structure. In summary, we propose a new framework that captures individual heterogeneity in the presence of spatial-temporal effects, and explore the influence of space-specific covariates on the response. This study contributes to the literature in three ways:

(a) We propose a DSAR model with possibly heterogeneous random effects. To the best of our knowledge, this is the first study to assume that the random effects of spatial models have a location-scale structure that can be affected by space-specific covariates. As a result, this model can capture the pure influence of space-specific factors on different quantile levels of the response, after the spatial-temporal effects have been removed.

(b) We introduce a two-stage hybrid estimation procedure for the proposed model. Specifically, a QMLE is proposed in the first stage to estimate the spatial-temporal effects, and a weighted conditional quantile regression estimator (WCQE) is used in the second stage by replacing unobserved random effects with observed pseudo ones based on the QMLE.

(c) From a theoretical viewpoint, we derive the consistency and asymptotic normality for the QMLE, establish the weak convergence of the quantile regression process and the asymptotic distribution of the WCQE, and propose two estimators for the conditional scale coefficients of the random effects and establish their asymptotic results.

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 proposes a two-stage hybrid estimation procedure for this model, and establishes all related asymptotic results. A simulation study is conducted in Section 4 to assess the finite-sample performance of the estimation procedure, and an empirical analysis of air quality is provided in Section 5 to illustrate the usefulness of the proposed model and its inference tools. Section 6 concludes the paper. A bootstrapping procedure for the QMLE and all technical details are relegated to the Supplementary Material. Throughout this paper, $\otimes$ denotes the Kronecker product of two matrices, $|\cdot|$ denotes the absolute value of a scalar/vector or the determinant of a matrix, $\|\cdot\|$ denotes the Euclidean norm of a vector, and $tr(\cdot)$ denotes the trace of a matrix. Let $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ be the smallest and largest eigenvalues of a matrix, respectively. Denote N and T as the total numbers of spatial units and time periods, respectively. Define $I(\cdot)$ as the indicator function. For a positive integer m, denote I_m as an $m \times m$ identity matrix, $\mathbf{0}_m$ as an $m \times 1$ vector of zeros and $\boldsymbol{\iota}_m$ as an $m \times 1$ vector of ones, and let $J_m = \boldsymbol{\iota}_m \boldsymbol{\iota}_m'$. The operator E denotes the expectation with respect to the probability measure, $\mathbb{E}_n$ denotes the expectation with respect to the empirical measure, and $\mathbb{G}_n = \sqrt{n}(\mathbb{E}_n - E)$. In addition, $\ell^\infty(\mathcal{T})$ denotes the space of all uniformly bounded functions on $\mathcal{T}$. Moreover, $\to_p$ denotes convergence in probability, $\to_d$ denotes convergence in distribution, and $\rightsquigarrow$ denotes weak convergence. An R package implementing the proposed two-stage hybrid estimation procedure and the data set from Section 5 are available at *https://github.com/wyLI2020/QuantileDSAR*.

2. Model Specification

Consider the following DSAR model with random effects:

$$y_{it} = \theta_i + \alpha y_{i,t-1} + \lambda \sum_{j=1}^{N} w_{N,ij} y_{jt} + \sum_{\ell=1}^{q} \gamma_\ell z_{\ell it} + \varepsilon_{it},\ i = 1, \ldots, N, t = 1, \ldots, T, \tag{2.1}$$

where $\{\theta_i\}$ are space-specific or individual random effects, the scalar parameters α and λ characterize the dynamic effect and the spatial autoregressive effect, respectively, $\{\gamma_\ell\}$ are the coefficients of the time-varying exogenous variables $\{z_{\ell it}\}$, $\{w_{N,ij}\}$ are specified constant spatial weights that generate the spatial dependence, and $\{\varepsilon_{it}\}$ are independent and identically distributed (*i.i.d.*) across i and t with mean zero and finite variance σ_ε^2. To flexibly depict the possible influence of space-specific covariates on an individual random effect, we assume a location-scale structure for θ_i, that is,

$$\theta_i = \psi_0 + \sum_{\ell=1}^{p} \psi_\ell x_{\ell i} + \left(\beta_0 + \sum_{\ell=1}^{p} \beta_\ell |x_{\ell i}|\right) \eta_i, \tag{2.2}$$

where $\{x_{\ell i}\}$ are time-invariant constant regressors, $\{\psi_\ell\}$ and $\{\beta_\ell\}$ are the coefficients of $\{x_{\ell i}\}$ and $\{|x_{\ell i}|\}$, respectively, $\beta_0 > 0$ and $\beta_\ell \geq 0$ for $1 \leq \ell \leq p$, and $\{\eta_i\}$ are *i.i.d.* disturbances with mean zero and finite variance σ_η^2; see also Koenker and Zhao (1994) and Zhao and Xiao (2014). Let $\mu_i = \psi_0 + \sum_{\ell=1}^{p} \psi_\ell x_{\ell i}$ and $\vartheta_i = (\beta_0 + \sum_{\ell=1}^{p} \beta_\ell |x_{\ell i}|)\, \eta_i$. It then follows that

$$\theta_i = \mu_i + \vartheta_i.$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

The location-scale structure combines the advantages of fixed and random effects. Specifically, the location μ_i depicts the difference between individuals on the conditional mean and can be arbitrarily related to time-variant regressors $\{z_{\ell it}\}$, whereas the scale effect ϑ_i allows for conditional heteroscedasticity and reflects the influence of $\{x_{\ell i}\}$ on the fluctuation. Note that if $\psi_\ell = \beta_\ell = 0$ for all $\ell = 1, \ldots, p$, then θ_i reduces to the commonly used individual random effect. Moreover, if $\beta_0 = \beta_1 = \cdots = \beta_p = 0$, then θ_i is the individual fixed effect captured by space-specific covariates with $p+1$ unknown parameters, which avoids the "incidental parameters" problem due to N unknown parameters in commonly used individual fixed-effect models.

Denote $\boldsymbol{Y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_T')'$ and $\boldsymbol{Y}_{-1} = (\boldsymbol{y}_0', \ldots, \boldsymbol{y}_{T-1}')'$ with $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{Nt})'$, $Z = (Z_1', \ldots, Z_T')'$ with $Z_t = (\boldsymbol{z}_{1t}, \ldots, \boldsymbol{z}_{Nt})'$ and $\boldsymbol{z}_{it} = (z_{1it}, \ldots, z_{qit})'$, $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)'$ with $\boldsymbol{x}_i = (1, x_{1i}, \ldots, x_{pi})'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \ldots, \boldsymbol{\varepsilon}_T')'$ with $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$. Let $S_{NT}(\lambda) = I_T \otimes B_N(\lambda)$ with $B_N(\lambda) = I_N - \lambda W_N$ and $W_N = \{w_{N,ij}, 1 \le i, j \le N\}$ be the $N \times N$ specified constant spatial weights matrix, and denote $\widetilde{Z} = (\boldsymbol{Y}_{-1}, Z, (\boldsymbol{\iota}_T \otimes I_N)X)$. Then, model (2.1) with (2.2) can be rewritten in the following matrix form:

$$S_{NT}(\lambda)\boldsymbol{Y} = \widetilde{Z}\boldsymbol{\phi} + (\boldsymbol{\iota}_T \otimes I_N)\boldsymbol{\vartheta} + \boldsymbol{\varepsilon}, \tag{2.3}$$

where $\boldsymbol{\phi} = (\alpha, \boldsymbol{\gamma}', \boldsymbol{\psi}')'$ with $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ and $\boldsymbol{\psi} = (\psi_0, \psi_1, \ldots, \psi_p)'$, and $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_N)'$ with $\vartheta_i = \boldsymbol{x}_{ai}'\boldsymbol{\beta}\eta_i$, $\boldsymbol{x}_{ai} = (1, |x_{1i}|, \ldots, |x_{pi}|)'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$. The following proposition provides a sufficient condition for the covariance stationarity of $\boldsymbol{y}_t$.

Proposition 1. *Suppose that η_i and ε_{it} are mutually independent. If $\left|\alpha B_N^{-1}(\lambda)\right| < 1$ and the process $\{\boldsymbol{z}_{it}\}$ is covariance stationary, then there exists a unique covariance stationary* *solution to models* (2.1) *and* (2.2)*. The solution takes the form*

$$\boldsymbol{y}_t = \sum_{m=0}^{\infty} \alpha^m \left[B_N^{-1}(\lambda)\right]^{m+1} (\boldsymbol{\theta} + Z_{t-m}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_{t-m}).$$

Proposition 1 indicates that the stationarity of $\{\boldsymbol{y}_t\}$ restricts the spatial and temporal effects to some extent by imposing conditions on the weights matrix W_N, spatial effect parameter λ, and autoregressive coefficient α. Specifically, $|\alpha| + |\lambda| < 1$ implies $\left|\alpha B_N^{-1}(\lambda)\right| < 1$ if W_N is row-normalized. Moreover, the stationarity of $\{\boldsymbol{y}_t\}$ depends on that of the exogenous covariates $\{\boldsymbol{z}_{it}\}$. However, assuming that both $\{\boldsymbol{y}_t\}$ and $\{\boldsymbol{z}_{it}\}$ are stationary limits the application of the proposed model. This motivates us to consider a theoretically valid estimation method without any stationarity assumption on $\{\boldsymbol{y}_t\}$.

We are interested in estimating $\boldsymbol{\phi}$, $\boldsymbol{\beta}$, λ, σ_ε^2, and σ_η^2, and providing a conditional quantile estimation of the individual random effect θ_i. Let $Q_\eta(\tau)$ be the τth quantile of η_i. Denote $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)'$ and $X_a = (\boldsymbol{x}_{a1}, \ldots, \boldsymbol{x}_{aN})'$. The conditional quantile of $\boldsymbol{\theta}$, given the space-specific covariates X, has the form of

$$Q_{\boldsymbol{\theta}}(\tau|X) = X\boldsymbol{\psi} + X_a\boldsymbol{\varphi}(\tau), \tag{2.4}$$

where $\boldsymbol{\varphi}(\tau) = (\varphi_0(\tau), \varphi_1(\tau), \ldots, \varphi_p(\tau))' = Q_\eta(\tau)\boldsymbol{\beta}$. Note that the individual effect $\boldsymbol{\theta}$ are unobservable. Thus, to estimate $Q_{\boldsymbol{\theta}}(\tau|X)$ in the framework of model (2.3), we propose a two-stage hybrid estimation procedure.

3. Two-stage Hybrid Estimation

The space-specific random effect θ_i is latent. Thus, it is natural for the estimation under model (2.4) to obtain its observable approximate using (2.1) based on reasonable estimates of λ, α, and $\boldsymbol{\gamma}$, and then to apply the quantile regression estimation method to model (2.4), with θ_i replaced by its observable approximate, to obtain estimates of $\boldsymbol{\psi}$ and $\boldsymbol{\varphi}(\tau)$. Note that $\boldsymbol{\psi}$ is independent of τ and it is not necessary to estimate it using a conditional quantile estimation in our model setting, whereas $\boldsymbol{\varphi}(\tau)$ is τ-dependent and can be estimated using the quantile estimation method. Moreover, λ and $\boldsymbol{\phi}$ can be estimated directly, and thus the observable approximate of ϑ_i is obtained easily from by model (2.3). Together with the fact that $Q_{\boldsymbol{\theta}}(\tau|X) = X\boldsymbol{\psi} + Q_{\boldsymbol{\vartheta}}(\tau|X)$, this motivates us to obtain observable approximates of ϑ_i, and then to apply a quantile regression, given by

$$Q_{\boldsymbol{\vartheta}}(\tau|X) = X_a\boldsymbol{\varphi}(\tau). \tag{3.1}$$

Specifically, we adopt the QMLE for model (2.3), denoted by $\widehat{\boldsymbol{\zeta}} = (\widehat{\boldsymbol{\phi}}', \widehat{\sigma}_\varepsilon^2, \widehat{\lambda}, \widehat{\boldsymbol{\beta}}^{\star\prime})'$ for $\boldsymbol{\zeta} = (\boldsymbol{\phi}', \sigma_\varepsilon^2, \lambda, \boldsymbol{\beta}^{\star\prime})'$ with $\boldsymbol{\beta}^\star = \boldsymbol{\beta}\sigma_\eta/\sigma_\varepsilon$. Then, the observable approximate of ϑ_i called the pseudo scale effect and denoted by $\widehat{\vartheta}_i$, can be constructed based on $\widehat{\lambda}$ and $\widehat{\boldsymbol{\phi}}$; see (3.11). We next build a quantile regression model for $\widehat{\boldsymbol{\vartheta}} = (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_N)'$ using X_a as the regressor, and construct a WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$ for $\boldsymbol{\varphi}(\tau)$ and a weighted quantile average estimator (WQAE) $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$. We propose the following two-stage hybrid estimation procedure:

Stage 1: Obtain the Gaussian QMLE $\widehat{\boldsymbol{\zeta}}$ based on model (2.3).

Stage 2: Define the pseudo scale effects $\{\widehat{\vartheta}_i\}$ using (3.11) based on $\widehat{\boldsymbol{\zeta}}$. Then, obtain the WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$ for $\boldsymbol{\varphi}(\tau)$ based on model (3.1), with $\boldsymbol{\vartheta}$ replaced by $\widehat{\boldsymbol{\vartheta}}$, and propose the WQAE $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ by combining the information on $\widehat{\boldsymbol{\varphi}}(\cdot)$ at multiple quantile levels τ.

As a result, if $S_{NT}(\widehat{\lambda})$ is invertible, then the conditional mean of the data $\boldsymbol{Y}$ can be estimated by $\widehat{E}(\boldsymbol{Y}|X) = S_{NT}^{-1}(\widehat{\lambda})\widetilde{Z}\widehat{\boldsymbol{\phi}}$ based on model (2.3) and $\widehat{\boldsymbol{\zeta}}$. Moreover, the conditional quantile of the random effect $\boldsymbol{\theta}$ can be estimated by $\widehat{Q}_{\boldsymbol{\theta}}(\tau|X) = X\widehat{\boldsymbol{\psi}} + X_a'\widehat{\boldsymbol{\varphi}}(\tau)$ based on model (2.4), $\widehat{\boldsymbol{\psi}}$, and $\widehat{\boldsymbol{\varphi}}(\tau)$. Statistical inference tools, such as significance tests for $\boldsymbol{\phi}$, $\boldsymbol{\varphi}(\tau)$, and $\boldsymbol{\beta}$, can

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

also be constructed based on $\widehat{\boldsymbol{\phi}}$, $\widehat{\boldsymbol{\varphi}}(\tau)$, and $\widehat{\boldsymbol{\beta}}$. Further details of the proposed estimation procedure and its asymptotics are provided in Sections 3.1 and 3.2.

To establish the asymptotics for the proposed two-stage hybrid estimation procedure, we assume the following regularity conditions for N and T:

Assumption 3.1. $T = cN^r$, *for some constants* $r \geq 0$ *and* $c > 0$.

Assumption 3.2. $T = cN^r$, *for some constants* $0 \leq r \leq 1$ *and* $c > 0$.

Assumption 3.3. $T = cN^r$, *for some constants* $0 < r \leq 1$ *and* $c > 0$.

Assumptions 3.1 and 3.2 allow T to be fixed or go to infinity as N goes to infinity, and Assumption 3.2 restricts T to a rate not greater than N when it goes to infinity. Assumption 3.3 requires that both T and N go to infinity, and is a special case of Assumption 3.2. Specifically, the consistency of the QMLE in the first stage is established under Assumption 3.1, and the asymptotic normality of the QMLE is obtained under Assumption 3.2. For the asymptotic normality of the QMLE, we need to restrict the rate of T (i.e., $T/N \to c'$, for some constant $c' \geq 0$) to prove the asymptotic normality of $(NT)^{-1/2}\partial \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)/\partial \boldsymbol{\zeta}$ using Theorem A.1 of Kelejian and Prucha (2001); see the proof of Theorem 2 in the Supplementary Material. Moreover, because $T \to \infty$ is necessary for the validity of replacing $\boldsymbol{\vartheta}$ with $\widehat{\boldsymbol{\vartheta}}$ in the quantile estimation, Assumption 3.3 is used to establish the asymptotic results of the WCQE and WQAE in the second stage; see Section 3.2.

3.1 The Quasi-maximum Likelihood Estimation

Recall that the parameter vector of model (2.3) is $\boldsymbol{\zeta} = (\boldsymbol{\phi}', \sigma_\varepsilon^2, \lambda, \boldsymbol{\beta}^{\star\prime})'$ with $\boldsymbol{\beta}^\star = \boldsymbol{\beta}\sigma_\eta/\sigma_\varepsilon$, and let $\boldsymbol{\zeta}_0 = (\boldsymbol{\phi}_0', \sigma_{\varepsilon 0}^2, \lambda_0, \boldsymbol{\beta}_0^{\star\prime})'$ be the true value of $\boldsymbol{\zeta}$. Denote $\boldsymbol{V}_{NT}(\lambda, \boldsymbol{\phi}) = S_{NT}(\lambda)\boldsymbol{Y} - \widetilde{Z}\boldsymbol{\phi}$,

$\boldsymbol{V}_{NT} = (\boldsymbol{\iota}_T \otimes I_N)\boldsymbol{\vartheta} + \boldsymbol{\varepsilon}$, $B_{0N} = B_N(\lambda_0)$, and $S_{0NT} = S_{NT}(\lambda_0)$. Then it follows that $\boldsymbol{V}_{NT}(\lambda_0, \boldsymbol{\phi}_0) = \boldsymbol{V}_{NT} = S_{0NT}\boldsymbol{Y} - \widetilde{Z}\boldsymbol{\phi}_0$. We consider the case that initial observations are generated exogenously, that is, $\boldsymbol{y}_0$ is exogenous such that it can be treated as a constant vector; see also Su and Yang (2015).

Assume that ε_{it} and η_i are distributed with a zero mean and variances σ_ε^2 and σ_η^2, respectively, which implies that $\text{cov}\{\boldsymbol{V}_{NT}(\lambda, \boldsymbol{\phi})\} = \sigma_\varepsilon^2 \Omega_{NT}(\boldsymbol{\beta}^\star)$, where $\Omega_{NT}(\boldsymbol{\beta}^\star) = J_T \otimes A_N(\boldsymbol{\beta}^\star) + I_{NT}$ with $A_N(\boldsymbol{\beta}^\star) = \text{diag}\,\{(\boldsymbol{x}'_{a1}\boldsymbol{\beta}^\star)^2, \ldots, (\boldsymbol{x}'_{aN}\boldsymbol{\beta}^\star)^2\}$. The Gaussian quasi-likelihood function of $\boldsymbol{\zeta}$ is then given by

$$\mathcal{L}_{NT}(\boldsymbol{\zeta}) = (2\pi\sigma_\varepsilon^2)^{-\frac{NT}{2}} \left|S'_{NT}(\lambda)\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)S_{NT}(\lambda)\right|^{\frac{1}{2}} \exp\left\{-\frac{\boldsymbol{V}'_{NT}(\lambda,\boldsymbol{\phi})\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)\boldsymbol{V}_{NT}(\lambda,\boldsymbol{\phi})}{2\sigma_\varepsilon^2}\right\},$$

and the corresponding quasi-log-likelihood function of $\boldsymbol{\zeta}$ is

$$\begin{aligned}\ln \mathcal{L}_{NT}(\boldsymbol{\zeta}) = &-\frac{NT}{2}\ln(2\pi) - \frac{NT}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2}\ln|\Omega_{NT}(\boldsymbol{\beta}^\star)| + \ln|S_{NT}(\lambda)| \\ &- \frac{1}{2\sigma_\varepsilon^2}\boldsymbol{V}'_{NT}(\lambda,\boldsymbol{\phi})\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)\boldsymbol{V}_{NT}(\lambda,\boldsymbol{\phi}).\end{aligned} \tag{3.2}$$

Maximizing (3.2) gives the Gaussian QMLE of $\boldsymbol{\zeta}$, denoted by $\widehat{\boldsymbol{\zeta}}$. If ε_{it} and η_i are normally distributed, then $\widehat{\boldsymbol{\zeta}}$ is the MLE.

For ease of computation, we consider the concentrated log-likelihood by concentrating out $\boldsymbol{\phi}$ and σ_ε^2. Given $\boldsymbol{\delta} = (\lambda, \boldsymbol{\beta}^{\star\prime})'$, maximizing (3.2) leads to the following QMLEs of $\boldsymbol{\phi}$ and σ_ε^2:

$$\widehat{\boldsymbol{\phi}}(\boldsymbol{\delta}) = \left(\widetilde{Z}'\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)\widetilde{Z}\right)^{-1}\widetilde{Z}'\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)S_{NT}(\lambda)\boldsymbol{Y}, \tag{3.3}$$

and

$$\widehat{\sigma}_\varepsilon^2(\boldsymbol{\delta}) = \frac{1}{NT}\widehat{\boldsymbol{V}}'_{NT}(\boldsymbol{\delta})\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)\widehat{\boldsymbol{V}}_{NT}(\boldsymbol{\delta}), \tag{3.4}$$

where $\widehat{\boldsymbol{V}}_{NT}(\boldsymbol{\delta}) = S_{NT}(\lambda)\boldsymbol{Y} - \widetilde{Z}\widehat{\boldsymbol{\phi}}(\boldsymbol{\delta})$. Then, by plugging (3.3) and (3.4) into (3.2), the concentrated log-likelihood function of $\boldsymbol{\delta}$ is

$$\ln \ell_{NT}(\boldsymbol{\delta}) = -\frac{NT}{2}\left(\ln(2\pi)+1\right) - \frac{NT}{2}\ln\widehat{\sigma}_{\varepsilon}^{2}(\boldsymbol{\delta}) - \frac{1}{2}\ln|\Omega_{NT}(\boldsymbol{\beta}^{\star})| + \ln|S_{NT}(\lambda)|. \tag{3.5}$$

The QMLE $\widehat{\boldsymbol{\delta}}$ that maximizes the function (3.5) is given by

$$\widehat{\boldsymbol{\delta}} = \arg\max_{\boldsymbol{\delta}\in\Delta} \ln \ell_{NT}(\boldsymbol{\delta}),$$

where Δ is the parameter space of $\boldsymbol{\delta}$, which is assumed to be compact and the true value $\boldsymbol{\delta}_0 = (\lambda_0, \boldsymbol{\beta}_0^{\star\prime})'$ is in the interior of Δ. Then, the QMLEs of $\boldsymbol{\phi}$ and σ_{ε}^2 are $\widehat{\boldsymbol{\phi}} = \widehat{\boldsymbol{\phi}}(\widehat{\boldsymbol{\delta}})$ and $\widehat{\sigma}_{\varepsilon}^2 = \widehat{\sigma}_{\varepsilon}^2(\widehat{\boldsymbol{\delta}})$, respectively. As a result, the QMLE of $\boldsymbol{\zeta}$ is given by $\widehat{\boldsymbol{\zeta}} = (\widehat{\boldsymbol{\phi}}', \widehat{\sigma}_{\varepsilon}^2, \widehat{\boldsymbol{\delta}}')' = (\widehat{\boldsymbol{\phi}}', \widehat{\sigma}_{\varepsilon}^2, \widehat{\lambda}, \widehat{\boldsymbol{\beta}}^{\star\prime})'$.

To establish the consistency and asymptotic normality for the QMLE, we introduce the following assumptions.

Assumption 3.4. *(i) $E|\varepsilon_{it}|^{4+\epsilon_0} < \infty$ and $E|\eta_i|^{4+\epsilon_0} < \infty$, for some $\epsilon_0 > 0$; (ii) ε_{it} and η_i are mutually independent.*

Assumption 3.5. *(i) The elements $w_{N,ij}$ of W_N are uniformly bounded constants for all N in all i and j; as a normalization, $w_{N,ii} = 0$ for all i; (ii) the matrix B_{0N} is nonsingular; (iii) the sequences of matrices $\{W_N\}$ and $\{B_{0N}^{-1}\}$ are bounded in both row and column sums uniformly for all N; (iv) $\{B_N^{-1}(\lambda)\}$ are bounded in either row or column sums uniformly for all N, and uniformly in $\lambda \in \Delta(\lambda)$; and (v) the sequence of matrices $\{\sum_{t=1}^{T}\alpha_0^{t-1}(B_{0N}^{-1})^t\}$ are uniformly bounded for all N and T in both row and column sums.*

Assumption 3.6. *(i) $\{x_{\ell i}\}$ are uniformly bounded constants for all N; (ii) $E(z_{\ell it}) = 0$ and $E|z_{\ell it}|^{4+\epsilon_0} < \infty$ for some $\epsilon_0 > 0$; (iii) for any $1 \le \ell \le q$, the covariance matrix of $\{|z_{\ell it}|, 1 \le i \le N, 1 \le t \le T\}$ is uniformly bounded in row sums, and it holds that*

$$\sum_{1\le i_3,i_4\le N}\ \sum_{1\le t_3,t_4\le T} \operatorname{cov}\left(z_{\ell i_1 t_1} z_{\ell i_2 t_2}, z_{\ell i_3 t_3} z_{\ell i_4 t_4}\right) = O(1) \tag{3.6}$$

for any $1 \le i_1, i_2 \le N$ *and* $1 \le t_1, t_2 \le T$*; and (iv)* $\frac{1}{NT}\widetilde{Z}\widetilde{Z}'$ *is positive definite for sufficiently large* N *and* $\lambda_{\min}\left(\frac{1}{NT}\widetilde{Z}'\Omega_{NT}^{-1}(\boldsymbol{\beta}^{\star})\widetilde{Z}\right)$ *is uniformly bounded away from zero almost surely (a.s.).*

Assumption 3.4 provides basic conditions on the disturbances that are standard in the random effect panel data literature; see Assumptions R(i) and R(ii) of Su and Yang (2015). Note that Assumption 3.4(i) does not require $\{\varepsilon_{it}\}$ to be identically distributed. Because the heterogeneity among individuals is captured by the random effect θ_i in model (2.1), we assume ε_{it} has a common variance instead of i-dependent variance.

Assumption 3.5 provides essential features of the weights matrix that are commonly assumed for spatial models; see also Assumption G2 of Su and Yang (2015) and Assumptions 2–5 of Lee (2004). In empirical applications, W_N is usually row normalized such that $B_N(\lambda)$ is nonsingular for $\lambda \in (-1, 1)$; see also Kelejian and Prucha (2010). The invertibility of B_{0N} in Assumption 3.5(ii) guarantees that (2.3) has an equilibrium and the disturbance term is well defined. Assumption 3.5(iii) limits the spatial correlation to some degree, but facilitates the study of the asymptotic properties of the spatial parameter estimators. Assumption 3.5(iv) is required to establish the consistency. Assumption 3.5(v) is implied by Assumption 3.5(iii) when T is fixed. When both N and T go to infinity, Assumption 3.5(v) is necessary to guarantee that the variance of y_{it} is bounded, and is an extension of Assumption 3.5(iii), which guarantees that the variance of y_i in the SAR model is bounded as N goes to infinity (Lee, 2004). Note that Assumption 3.5(iii) implies that both the row and the column sums of $(B_{0N}^{-1})^t$ are bounded for an arbitrary t. This, together with $|\alpha_0| < 1$, implies that Assumption 3.5(v) holds, as long as α_0^{t-1} decreases faster than the row or column sum of $(B_{0N}^{-1})^t$ increases as t increases, which is usually satisfied in practice. Moreover, if W_N

is a row-normalized matrix, then both $\left\{B_N^{-1}(\lambda)\right\}$ and $\{\sum_{t=1}^{T}\alpha^{t-1}[B_N^{-1}(\lambda)]^t\}$ are uniformly bounded in row sums for $|\alpha|+|\lambda|<1$; see also Lee (2004). In addition, Assumption 3.5 implies some restrictions on the process $\{\boldsymbol{y}_t\}$ by imposing conditions on the weights matrix W_N, spatial effect parameter λ, and autoregressive coefficient α. However, both stationary and nonstationary cases of $\{\boldsymbol{y}_t\}$ are allowable under Assumption 3.5.

The conditions on the regressors in Assumptions 3.6(i), (ii), and (iv) are general and parallel Assumption 6 in Lee (2004), Assumption 4 in Yu et al. (2008), and Assumptions G1(iii) and R($iv^\star$) in Su and Yang (2015). The regressors in Lee (2004) are time-invariant constants, in Yu et al. (2008), they are time-variant constants, and in Su and Yang (2015), they are time-invariant or time-variant random variables. Note that Theorems 1 and 2 still hold if $x_{\ell i}$ (or $z_{\ell it}$) is assumed to be either constant or random variable. Here, we assume time-variant random $z_{\ell it}$ but time-invariant constant $x_{\ell i}$ to emphasize the randomness by time t. This also allows $x_{\ell i}$ to be arbitrarily related to $z_{\ell it}$. Assumption 3.6(iii) guarantees that the variances associated with $\{z_{\ell it}, 1\le i\le N, 1\le t\le T\}$ are bounded; see Lemmas 4–6. In particularly, (3.6) implies that given (i_1,t_1) and (i_2,t_2), $z_{\ell i_1 t_1}z_{\ell i_2 t_2}$ is correlated with $z_{\ell i_3 t_3}z_{\ell i_4 t_4}$ for only finite terms of (i_3,t_3) and (i_4,t_4). Note that Assumption 3.6(iii) allows for individual dependence among finite neighbors and time dependence up to finite lags, which makes it weaker than the independence assumption of regressors across individuals in Su and Yang (2015).

Recall that $\boldsymbol{\zeta}=(\boldsymbol{\phi}',\sigma_\varepsilon^2,\boldsymbol{\delta}')'$ and $\widehat{\boldsymbol{\zeta}}=(\widehat{\boldsymbol{\phi}}'(\widehat{\boldsymbol{\delta}}),\widehat{\sigma}_\varepsilon^2(\widehat{\boldsymbol{\delta}}),\widehat{\boldsymbol{\delta}}')'$. To show the consistency of the QMLE $\widehat{\boldsymbol{\zeta}}$, we first verify that $\boldsymbol{\zeta}_0=(\boldsymbol{\phi}_0',\sigma_{\varepsilon 0}^2,\boldsymbol{\delta}_0')'$ is identifiable. For the quasi-log-likelihood function at (3.2), given $\boldsymbol{\delta}$, its expectation $E[\ln\mathcal{L}_{NT}(\boldsymbol{\zeta})]$ is maximized at

$$\widetilde{\boldsymbol{\phi}}(\boldsymbol{\delta})=\left[E\left(\widetilde{Z}'\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)\widetilde{Z}\right)\right]^{-1}E\left(\widetilde{Z}'\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)S_{NT}(\lambda)\boldsymbol{Y}\right), \tag{3.7}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

and

$$\widetilde{\sigma}_{\varepsilon}^2(\boldsymbol{\delta}) = \frac{1}{NT} E\left(\widetilde{\boldsymbol{V}}'_{NT}(\boldsymbol{\delta})\Omega_{NT}^{-1}(\boldsymbol{\beta}^{\star})\widetilde{\boldsymbol{V}}_{NT}(\boldsymbol{\delta})\right), \tag{3.8}$$

where $\widetilde{\boldsymbol{V}}_{NT}(\boldsymbol{\delta}) = S_{NT}(\lambda)\boldsymbol{Y} - \widetilde{Z}\widetilde{\boldsymbol{\phi}}(\boldsymbol{\delta})$. Define $\ln \ell^{\star}_{NT}(\boldsymbol{\delta}) = \max_{\boldsymbol{\phi},\sigma_{\varepsilon}^2} E[\ln \mathcal{L}_{NT}(\boldsymbol{\zeta})]$. Then, by (3.7) and (3.8), we have

$$\ln \ell^{\star}_{NT}(\boldsymbol{\delta}) = -\frac{NT}{2}(\ln(2\pi) + 1) - \frac{NT}{2}\ln \widetilde{\sigma}_{\varepsilon}^2(\boldsymbol{\delta}) - \frac{1}{2}\ln|\Omega_{NT}(\boldsymbol{\beta}^{\star})| + \ln|S_{NT}(\lambda)|.$$

Denote $\Omega_{0NT} = \Omega_{NT}(\boldsymbol{\beta}^{\star}_0)$. Because $E\left(\widetilde{Z}'\Omega_{0NT}^{-1}\boldsymbol{V}_{NT}\right) = 0$, by Lemma 8 in the Supplementary Material, it can be shown that $\widetilde{\boldsymbol{\phi}}(\boldsymbol{\delta}_0) = \boldsymbol{\phi}_0 + [E(\widetilde{Z}'\Omega_{0NT}^{-1}\widetilde{Z})]^{-1}\, E(\widetilde{Z}'\Omega_{0NT}^{-1}\boldsymbol{V}_{NT}) = \boldsymbol{\phi}_0$. Then, it follows that $\widetilde{\boldsymbol{V}}_{NT}(\boldsymbol{\delta}_0) = \boldsymbol{V}_{NT}$ and $\widetilde{\sigma}_{\varepsilon}^2(\boldsymbol{\delta}_0) = \sigma_{\varepsilon 0}^2$. The identification of $\boldsymbol{\delta}_0$ can be based on the maximum values of $\frac{1}{NT}\ln \ell^{\star}_{NT}(\boldsymbol{\delta})$, and hence that of $\boldsymbol{\phi}_0$ and $\sigma_{\varepsilon 0}^2$ follows. The following condition is imposed for identification.

Assumption 3.7. *For any $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0$, it holds that*

$$\lim_{N\to\infty} \frac{1}{NT}\left[\ln\left|\widetilde{\sigma}_{\varepsilon}^{-2}(\boldsymbol{\delta})\Omega_{NT}^{-1}(\boldsymbol{\beta}^{\star})S_{NT}(\lambda)\right| - \ln\left|\sigma_{\varepsilon 0}^{-2}\Omega_{0NT}^{-1}S_{0NT}\right|\right] \neq 0.$$

Assumption 3.7 is equivalent to $\lim_{N\to\infty} \frac{1}{NT}[\ln \ell^{\star}_{NT}(\boldsymbol{\delta}) - \ln \ell^{\star}_{NT}(\boldsymbol{\delta_0})] \neq 0$ for any $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0$, which ensures the identification of $\boldsymbol{\delta}_0$; see also Assumption 9 of Lee (2004) and Assumption R(iv) of Su and Yang (2015). Finally, the consistency of $\widehat{\boldsymbol{\delta}}$ holds by the identification and uniform convergence of $\frac{1}{NT}[\ln \ell_{NT}(\boldsymbol{\delta}) - \ln \ell^{\star}_{NT}(\boldsymbol{\delta})]$ to zero on Δ; thus the consistency of $\widehat{\boldsymbol{\phi}}$ and $\widehat{\sigma}_{\varepsilon}^2$ follows.

Theorem 1. *Under Assumptions 3.1, and 3.4–3.7, then $\widehat{\boldsymbol{\zeta}} \to_p \boldsymbol{\zeta}_0$ as $N \to \infty$.*

Let $\Sigma = \lim_{N\to\infty}\Sigma_N$ and $\Sigma^{\star} = \lim_{N\to\infty}\Sigma^{\star}_N$, where $\Sigma_N = -E\left(\frac{1}{NT}\frac{\partial^2 \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'}\right)$ and

$$\Sigma^{\star}_N = E\left(\frac{1}{\sqrt{NT}}\frac{\partial \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial\boldsymbol{\zeta}}\frac{1}{\sqrt{NT}}\frac{\partial \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial\boldsymbol{\zeta}'}\right) + E\left(\frac{1}{NT}\frac{\partial^2 \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}'}\right).$$

By the consistency result given in Theorem 1, the asymptotic normality of $\widehat{\boldsymbol{\zeta}}$ can be obtained using the Taylor expansion and the central limit theorem for linear-quadratic forms in Kelejian and Prucha (2001).

Theorem 2. *Under Assumptions 3.2, and 3.4–3.7, if* Σ *and* $\Sigma^{\star}$ *exist and* Σ *is positive definite, then* $\sqrt{NT}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) \to_d N(\mathbf{0}, \Sigma^{-1} + \Sigma^{-1}\Sigma^{\star}\Sigma^{-1})$ *as* $N \to \infty$.

Theorem 2 provides the asymptotic distribution of the QMLE $\widehat{\boldsymbol{\zeta}}$, where the matrices Σ_N and $\Sigma_N^{\star}$ are relegated to Section S3.1 of the Supplementary Material. If ε_{it} and η_i are normally distributed, then the QMLE reduces to the MLE, and its limiting distribution can be simplified to $\sqrt{NT}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) \to_d N(\mathbf{0}, \Sigma^{-1})$ as $N \to \infty$. Note that Theorems 1 and 2 allow T to be fixed or tend to infinity, which makes the QMLE applicable to both short and long panel data.

To calculate the covariance matrix of $\widehat{\boldsymbol{\zeta}}$ in practice, consistent estimates of Σ and $\Sigma^{\star}$ are necessary. Obviously, the matrix $\Sigma = \lim_{N\to\infty} -E\left(\frac{1}{NT}\frac{\partial^2 \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\zeta}'}\right)$ can be consistently estimated by $\widehat{\Sigma} = -\frac{1}{NT}\frac{\partial^2 \ln \mathcal{L}_{NT}(\widehat{\boldsymbol{\zeta}})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\zeta}'}$. However, a closed-form estimate of the matrix

$$\Gamma = \Sigma^{\star} + \Sigma = \lim_{N\to\infty} E\left(\frac{1}{\sqrt{NT}}\frac{\partial \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial \boldsymbol{\zeta}}\frac{1}{\sqrt{NT}}\frac{\partial \ln \mathcal{L}_{NT}(\boldsymbol{\zeta}_0)}{\partial \boldsymbol{\zeta}'}\right) \tag{3.9}$$

is not readily available; see also Su and Yang (2015). Therefore, we use the residual-based bootstrap and estimate Γ by its bootstrap estimate $\widehat{\Gamma}_B$; see the Supplementary Material for the detailed procedure and its validity. Consequently, the asymptotic covariance matrix of $\widehat{\boldsymbol{\zeta}}$ can be estimated by $\widehat{\Sigma}^{-1}\widehat{\Gamma}_B\widehat{\Sigma}^{-1}$.

To estimate $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{10}, \ldots, \beta_{p0})'$ in model (2.2), the constraint $\beta_{00} \equiv 1$ should be imposed for the identification of $\boldsymbol{\beta}_0$, as in Koenker and Zhao (1994). Note that $\boldsymbol{\beta}_0^{\star} = \boldsymbol{\beta}_0\sigma_{\eta 0}/\sigma_{\varepsilon 0}$. Thus, we have $\beta_{\ell 0} = \beta_{\ell 0}/\beta_{00} = \beta_{\ell 0}^{\star}/\beta_{00}^{\star}$, for $\ell = 1, \ldots, p$. Then, based on the QMLE $\widehat{\boldsymbol{\zeta}} = (\widehat{\boldsymbol{\phi}}', \widehat{\sigma}_\varepsilon^2, \widehat{\lambda}, \widehat{\boldsymbol{\beta}}^{\star\prime})'$ with $\widehat{\boldsymbol{\beta}}^\star = (\widehat{\beta}_0^\star, \widehat{\beta}_1^\star, \ldots, \widehat{\beta}_p^\star)'$, we can construct a consistent estimator for $\boldsymbol{\beta}_0$, denoted by $\check{\boldsymbol{\beta}} = (1, \check{\beta}_1, \ldots, \check{\beta}_p)'$, where

$$\check{\beta}_\ell = \widehat{\beta}_\ell^\star / \widehat{\beta}_0^\star, \quad \ell = 1, \ldots, p. \tag{3.10}$$

Under the conditions of Theorem 2, we can use the delta method (Doob, 1935) to show that $\sqrt{NT}(\check{\beta}_\ell - \beta_{\ell 0}) \to_d N(0, \nabla h_\ell(\boldsymbol{\zeta}_0)' \Sigma_\zeta \nabla h_\ell(\boldsymbol{\zeta}_0))$ as $N \to \infty$, where $\Sigma_\zeta = \Sigma^{-1} + \Sigma^{-1}\Sigma^\star\Sigma^{-1}$, and $\nabla h_\ell(\boldsymbol{\zeta}) = \partial h_\ell(\boldsymbol{\zeta})/\partial \boldsymbol{\zeta} = (\mathbf{0}'_{4+p+q}, -\beta_\ell^\star/\beta_0^{\star 2}, \mathbf{0}'_{\ell-1}, 1/\beta_0^\star, \mathbf{0}'_{p-1})'$ with $h_\ell(\boldsymbol{\zeta}) = \beta_\ell^\star/\beta_0^\star$. Furthermore, we can calculate the asymptotic variance of $\check{\beta}_\ell$ in practice by plugging in $\nabla h_\ell(\widehat{\boldsymbol{\zeta}})$ and $\widehat{\Sigma}_\zeta = \widehat{\Sigma}^{-1}\widehat{\Gamma}_B\widehat{\Sigma}^{-1}$. However, $\check{\boldsymbol{\beta}}$ performs very poorly when the sample size is small or even moderate; see Table 3. As a result, we propose another estimator for $\boldsymbol{\beta}_0$ in Section 3.2.2, that is much more efficient for small and moderate samples.

3.2 Conditional Quantile Estimation

In Section 3.2.1, we first investigate the conditional quantile estimation for ϑ_i. Then, we construct an efficient estimator for the conditional scale coefficient $\boldsymbol{\beta}$ of the random effects by optimally combining information over multiple quantile levels τ in Section 3.2.2.

3.2.1 Weighted conditional quantile estimation

Because the scale effect ϑ_i in model (3.1) is unobservable, to study its conditional quantiles, we first obtain its observable approximate using the QMLE $\widehat{\boldsymbol{\zeta}}$ from the first stage. Replacing $\boldsymbol{\zeta}$ by $\widehat{\boldsymbol{\zeta}}$ in model (2.3), one can get T random approximates of ϑ_i for each i. To fully use the information in these approximations, a natural way is to take the average of T random approximations as the final approximation for ϑ_i. Therefore, the pseudo scale effect, denoted by $\widehat{\boldsymbol{\vartheta}}$, that is, the approximation of $\boldsymbol{\vartheta}$, is given by

$$\widehat{\boldsymbol{\vartheta}} = (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_N)' = \frac{1}{T}\sum_{t=1}^{T}\left[B_N(\widehat{\lambda})\boldsymbol{y}_t - \widehat{\alpha}\boldsymbol{y}_{t-1} - Z_t\widehat{\boldsymbol{\gamma}} - X\widehat{\boldsymbol{\psi}}\right]. \tag{3.11}$$

Thus, the pseudo random effect follows by $\widehat{\boldsymbol{\theta}} = X\widehat{\boldsymbol{\psi}} + \widehat{\boldsymbol{\vartheta}}$. Next, the conditional quantile estimation can be applied to model (3.1), with the latent variable $\boldsymbol{\vartheta}$ replaced by its observable pseudo scale effect $\widehat{\boldsymbol{\vartheta}}$. Owing to the presence of conditional heteroscedasticity in ϑ_i, we consider the WCQE given by

$$\widehat{\boldsymbol{\varphi}}(\tau) = (\widehat{\varphi}_0(\tau), \widehat{\varphi}_1(\tau), \ldots, \widehat{\varphi}_p(\tau))' = \underset{\boldsymbol{\varphi}\in\Phi}{\operatorname{argmin}} \sum_{i=1}^{N} \frac{1}{\boldsymbol{x}_{ai}'\widetilde{\boldsymbol{\beta}}_c}\rho_\tau(\widehat{\vartheta}_i - \boldsymbol{x}_{ai}'\boldsymbol{\varphi}), \quad \tau \in \mathcal{T}, \tag{3.12}$$

where Φ is the parameter space of $\boldsymbol{\varphi}(\tau)$, $\mathcal{T}$ is a closed subinterval of $(0,1)$, $\rho_\tau(u) = u\left(\tau - I(u<0)\right)$ is the check function, and $\widetilde{\boldsymbol{\beta}}_c$ is a consistent estimator of $\boldsymbol{\beta}_0$ such that $\boldsymbol{x}_{ai}'\widetilde{\boldsymbol{\beta}}_c$ is bounded away from zero for all $1 \leq i \leq N$; see also Koenker and Zhao (1994) and Zhao and Xiao (2014). For the choice of $\widetilde{\boldsymbol{\beta}}_c$, see Remark 3.

Theorem 3. *Under Assumptions 3.3–3.7, as $N \to \infty$, the following results hold:*

(i) $\widehat{\vartheta}_i - \vartheta_i = o_p(1)$, *for* $1 \leq i \leq N$;

(ii) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\vartheta}_i - \vartheta_i) = o_p(1)$.

Theorem 3 plays an important role in verifying the validity of the proposed hybrid estimation procedure. Specifically, Theorem 3(i) guarantees that replacing $\boldsymbol{\vartheta}$ with $\widehat{\boldsymbol{\vartheta}}$ does not affect the theoretical properties of the WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$, and its proof requires $T \to \infty$, which is implied by $N \to \infty$ and Assumption 3.3. Theorem 3(ii) simplifies the derivation of the Bahadur representation for $\widehat{\boldsymbol{\varphi}}(\tau)$; see equation (S5.16) of the Supplementary Material.

Denote the true value of $\boldsymbol{\varphi}(\tau)$ as $\boldsymbol{\varphi}_0(\tau) = (\varphi_{00}(\tau), \varphi_{10}(\tau), \ldots, \varphi_{p0}(\tau))'$. Assume that the parameter space Φ is compact and $\boldsymbol{\varphi}_0(\tau)$ is its interior point; see also Assumption R2 of

Chernozhukov and Hansen (2006) and Assumption 4.1 (b) of Canay (2011). To establish the theoretical properties of $\widehat{\boldsymbol{\varphi}}(\tau)$, we need the following basic assumptions.

Assumption 3.8. *For $1 \le i \le N$, $\vartheta_i \in \mathbb{R}$ has a bounded conditional density a.s., that is, $\sup_{\vartheta \in \mathbb{R}} f_{\vartheta_i|\boldsymbol{x}_i}(\vartheta) < K$ for some $K > 0$ a.s., where $\mathbb{R} = (-\infty, \infty)$ and $f_{\vartheta_i|\boldsymbol{x}_i}(\vartheta)$ is bounded away from zero at the point $\boldsymbol{x}_i' \varphi_0(\tau)$.*

Assumption 3.9. *For the function $\boldsymbol{\Pi}_i(\tau, \boldsymbol{\varphi}, r_i) = E\left[\frac{1}{\boldsymbol{x}_{ai}'\boldsymbol{\beta}}\left(\tau - I(\vartheta_i < \boldsymbol{x}_{ai}'\boldsymbol{\varphi} - r_i)\right)\boldsymbol{x}_{ai}\right]$ with $(\tau, \boldsymbol{\varphi}, r_i) \in \mathcal{T} \times \Phi \times \mathbb{R}$, the Jacobian matrices $\frac{\partial}{\partial \boldsymbol{\varphi}'}\boldsymbol{\Pi}_i(\tau, \boldsymbol{\varphi}, r_i)$ and $\frac{\partial}{\partial r_i}\boldsymbol{\Pi}_i(\tau, \boldsymbol{\varphi}, r_i)$ for each $1 \le i \le N$ are continuous and have full rank uniformly over $\mathcal{T} \times \Phi \times \mathbb{R}$.*

Assumption 3.8 is needed to establish the uniform convergence of $\widehat{\boldsymbol{\varphi}}(\tau)$. This, together with Assumption 3.9, is used for the weak convergence of $\widehat{\boldsymbol{\varphi}}(\tau)$; see Assumption R3 of Chernozhukov and Hansen (2006) and Assumption 4.1(c) of Canay (2011).

Theorem 4. *Suppose the conditions of Theorem 3 hold. Then, under Assumption 3.8, $\sup_{\tau \in \mathcal{T}} \|\widehat{\boldsymbol{\varphi}}(\tau) - \boldsymbol{\varphi}_0(\tau)\| \to_p 0$ as $N \to \infty$.*

Theorems 1, 3, and 4 ensure that the conditional quantile of the random effect $\boldsymbol{\theta}$ can be consistently estimated by

$$\widehat{Q}_{\boldsymbol{\theta}}(\tau|X) = X\widehat{\boldsymbol{\psi}} + X_a\widehat{\boldsymbol{\varphi}}(\tau). \tag{3.13}$$

Denote the density function of η_i by $f_\eta(\cdot)$. Let $S(\tau, \tau') = \min\{\tau, \tau'\} - \tau\tau'$ and $\Xi(\tau) = f_\eta(Q_\eta(\tau))$. Define the $(p+1) \times (p+1)$ matrices

$$D_N = \frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{x}_{ai}\boldsymbol{x}_{ai}'}{(\boldsymbol{x}_{ai}'\boldsymbol{\beta}_0)^2}, D_0 = \lim_{N\to\infty} D_N \text{ and } D_{1N}(\tau) = \Xi(\tau)D_N.$$

Theorem 5. *Suppose the conditions of Theorem 4 hold. If D_0 exists and is positive definite, then under Assumption 3.9, as $N \to \infty$,*

$$\sqrt{N}\left(\widehat{\boldsymbol{\varphi}}(\cdot) - \boldsymbol{\varphi}_0(\cdot)\right) = D_{1N}^{-1}(\cdot)\mathbb{G}_N\left(\frac{\boldsymbol{x}_{ai}}{\boldsymbol{x}_{ai}'\boldsymbol{\beta}_0}\Psi_{\boldsymbol{\cdot}}\left(\eta_i - Q_\eta(\cdot)\right)\right) + o_p(1) \rightsquigarrow \mathbb{G}(\cdot) \text{ in } \ell^\infty(\mathcal{T}),$$

where $\Psi_\tau(u) = \tau - I(u < 0)$ *and* $\mathbb{G}(\cdot)$ *is a zero mean Gaussian process with covariance kernel* $\Xi^{-1}(\tau)S(\tau,\tau')\Xi^{-1}(\tau')D_0^{-1}$.

Theorem 5 states the weak convergence of the quantile regression process $\widehat{\boldsymbol{\varphi}}(\cdot)$. For any fixed quantile level τ, by Theorem 5, it follows that, as $N \to \infty$,

$$\sqrt{N}(\widehat{\boldsymbol{\varphi}}(\tau) - \boldsymbol{\varphi}_0(\tau)) \to_d N(\mathbf{0}, \tau(1-\tau)\Xi^{-2}(\tau)D_0^{-1}). \tag{3.14}$$

To calculate the covariance matrix of $\widehat{\boldsymbol{\varphi}}(\tau)$ in practice, consistent estimators of $\Xi(\tau) = f_\eta(Q_\eta(\tau))$ and D_0 are required. Define the error function $\widehat{\eta}_i(\boldsymbol{\beta}) = \widehat{\vartheta}_i/(\boldsymbol{x}'_{ai}\boldsymbol{\beta})$. Then the residuals can be computed as $\widetilde{\eta}_i = \widehat{\eta}_i(\widetilde{\boldsymbol{\beta}}_c)$, where $\widetilde{\boldsymbol{\beta}}_c$ is a consistent estimate of $\boldsymbol{\beta}_0$, as in (3.12). The density function $f_\eta(\cdot)$ can be estimated using the kernel density estimator $\widetilde{f}_\eta(x) = (Nh)^{-1}\sum_{i=1}^N K((x-\widetilde{\eta}_i)/h)$, where $K(\cdot)$ is the kernel function and h is the bandwidth. In practice, we suggest using the Gaussian kernel for $K(\cdot)$ and its rule-of-thumb bandwidth, $h = 0.9N^{-1/5}\min\{s, \widehat{R}/1.34\}$, where s and $\widehat{R}$ are the sample standard deviation and interquartile of the residuals, respectively; see also Zhao and Xiao (2014). As a result, $\Xi(\tau)$ can be estimated using $\widetilde{f}_\eta(\bar{Q}_\eta(\tau))$, where $\bar{Q}_\eta(\tau)$ is the sample τth quantile of $\{\widetilde{\eta}_i\}_{i=1}^N$. Moreover, D_0 can be approximated by D_N, with $\boldsymbol{\beta}_0$ replaced by $\widetilde{\boldsymbol{\beta}}_c$, that is, $\widetilde{D}_N = \frac{1}{N}\sum_{i=1}^N \frac{\boldsymbol{x}_{ai}\boldsymbol{x}'_{ai}}{(\boldsymbol{x}'_{ai}\widetilde{\boldsymbol{\beta}}_c)^2}$. Consequently, the asymptotic covariance matrix of $\widehat{\boldsymbol{\varphi}}(\tau)$ can be estimated by plugging in $\bar{\Xi}(\tau) = \widetilde{f}_\eta(\bar{Q}_\eta(\tau))$ and $\widetilde{D}_N$ to replace $\Xi(\tau)$ and D_0, respectively.

Remark 1. *(Discussion on the estimation of* $\boldsymbol{\psi}$*) As the conditional mean coefficient of the random effects,* $\boldsymbol{\psi}$ *can be estimated using either the quasi-maximum likelihood estimation in the first stage or the conditional quantile estimation for model* $Q_{\boldsymbol{\theta}}(\tau|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\psi} + \boldsymbol{X}_a\boldsymbol{\varphi}(\tau)$ *in the second stage. However, comparing Theorems 2 and 5, the QMLE* $\widehat{\boldsymbol{\psi}}$ *of* $\boldsymbol{\psi}$ *has a faster convergence rate of* $\sqrt{NT}$ *than that of* $\sqrt{N}$ *for the quantile estimator obtained using* *model* $Q_{\boldsymbol{\theta}}(\tau|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\psi} + \boldsymbol{X}_a\boldsymbol{\varphi}(\tau)$. *This is another reason why the proposed two-stage hybrid estimation procedure does not consider a quantile estimation for both* $\boldsymbol{\psi}$ *and* $\boldsymbol{\varphi}(\tau)$ *simultaneously.*

Remark 2. *(Discussion on an alternative estimation procedure) To estimate the DSAR model* (2.1) *with the random effect* θ_i *specified by* (2.2), *we consider the following alternative estimation procedure:*

Stage $1'$*: Estimate* $(\alpha, \boldsymbol{\gamma}', \sigma_\varepsilon^2, \boldsymbol{\delta})'$ *by the QMLE using the transformation approach or first-differencing to eliminate individual effect* θ_i *(Lee and Yu, 2010; Yu et al., 2008).*

Stage $2'$*: Define the pseudo random effect* $\widetilde{\boldsymbol{\theta}}$ *similarly as in* (3.11). *Then, apply the WCQE to model* (2.4), *with* $\boldsymbol{\theta}$ *replaced by its approximation* $\widetilde{\boldsymbol{\theta}}$. *Furthermore,* τ*-independent* $\boldsymbol{\psi}$ *and* $\boldsymbol{\beta}$ *can be estimated by* $\widetilde{\boldsymbol{\psi}}$ *and* $\widetilde{\boldsymbol{\beta}}$, *respectively, by combining the information from the quantile estimators at multiple quantile levels.*

Note that the transformation approach or first-differencing in Stage $1'$ *ignores the structure of* $\boldsymbol{\theta}$ *in* (2.2), *which makes the estimation less efficient. Moreover,* $\widetilde{\boldsymbol{\psi}}$ *obtained using a quantile regression in Stage* $2'$ *also results in an efficiency loss, as discussed in Remark 1. Thus, we prefer the two-stage hybrid estimation procedure proposed in Section 3.*

Remark 3. *(Choice of* $\widetilde{\boldsymbol{\beta}}_c$ *in* (3.12) *for the WCQE) To conduct the weighted quantile regression in* (3.12), *we employ* $\check{\boldsymbol{\beta}}$ *in* (3.10) *as* $\widetilde{\boldsymbol{\beta}}_c$. *However,* $\check{\boldsymbol{\beta}}$ *is not stable in finite-sample cases, so we construct an alternative choice for* $\widetilde{\boldsymbol{\beta}}_c$. *Note that* $\boldsymbol{\varphi}_0(\tau) = Q_\eta(\tau)\boldsymbol{\beta}_0$. *Thus, we can construct* $\widetilde{\boldsymbol{\beta}}_c$ *by combining the information of the estimator for* $\boldsymbol{\varphi}_0(\tau)$ *over multiple quantile levels. Furthermore,* $\boldsymbol{\varphi}_0(\tau)$ *and* $Q_\eta(\tau)$ *have the same sign. Thus, as an alternative, we consider* $\sum_{k=1}^K |\boldsymbol{\varphi}_0(\tau_k)| = \boldsymbol{\beta}_0 \sum_{k=1}^K |Q_\eta(\tau_k)|$ *to avoid the offset among positive and negative values for both terms at different quantile levels. Following Koenker and Zhao (1994), we* *adopt the constraint* $\beta_{00} \equiv 1$ *to ensure the identification of* $\boldsymbol{\beta}_0$ *and* $Q_\eta(\tau)$*. As a result, we suggest the following choice for* $\widetilde{\boldsymbol{\beta}}_c$ *in* (3.12)*:*

$$\widetilde{\boldsymbol{\beta}}_c = \frac{\sum_{k=1}^K |\widetilde{\boldsymbol{\varphi}}(\tau_k)|}{\sum_{k=1}^K |\widetilde{Q}_\eta(\tau_k)|}, \tag{3.15}$$

where $\widetilde{\boldsymbol{\varphi}}(\tau_k)$ *with* $\tau_k = k/(K+1)$*, for* $1 \leq k \leq K$*, is the unweighted conditional quantile estimator defined by* $\widetilde{\boldsymbol{\varphi}}(\tau_k) = (\widetilde{\varphi}_0(\tau_k), \ldots, \widetilde{\varphi}_p(\tau_k))' = \operatorname{argmin}_{\boldsymbol{\varphi}\in\Phi} \sum_{i=1}^N \rho_{\tau_k}(\widehat{\vartheta}_i - \boldsymbol{x}'_{ai}\boldsymbol{\varphi})$*, and* $\widetilde{Q}_\eta(\tau_k) = \widetilde{\varphi}_0(\tau_k)$*; see also Zhu et al. (2018). Clearly,* (3.15) *guarantees the positivity of* $\widetilde{\boldsymbol{\beta}}_c$*.*

3.2.2 Weighted quantile average estimation

For the conditional scale coefficient $\boldsymbol{\beta}$ of the random effects, we have proposed two consistent estimators $\breve{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\beta}}_c$, in (3.10) and (3.15), respectively. However, $\breve{\boldsymbol{\beta}}$ is not suggested, because it is not stable in small or even moderate samples. For $\widetilde{\boldsymbol{\beta}}_c$, it is actually an equally weighted quantile average estimator by assigning equal weights $[\sum_{k=1}^K |\widetilde{Q}_\eta(\tau_k)|]^{-1}$ to the WCQE $\widehat{\boldsymbol{\varphi}}(\cdot)$ at each quantile level, which may lead to a less efficient estimator. We next introduce an efficient estimator of $\boldsymbol{\beta}$ by optimally combining the information of $\widehat{\boldsymbol{\varphi}}(\cdot)$ across K quantile levels: $\tau_k = k/(K+1)$, for $1 \leq k \leq K$, where K is a fixed integer.

Recall that $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{10}, \ldots, \beta_{p0})'$ is the true value of $\boldsymbol{\beta}$, and $\boldsymbol{\varphi}_0(\tau) = Q_\eta(\tau)\boldsymbol{\beta}_0$ with $\varphi_{\ell 0}(\tau) = Q_\eta(\tau)\beta_{\ell 0}$ holds for $\ell = 0, 1, \ldots, p$. Note that for any weight vector $\boldsymbol{\pi}_K = (\pi_1, \ldots, \pi_K)'$ satisfying $\sum_{k=1}^K \pi_k Q_\eta(\tau_k) = 1$, it holds that $\boldsymbol{\beta}_0 = \sum_{k=1}^K \pi_k \boldsymbol{\varphi}_0(\tau_k)$ under the constraint $\beta_{00} \equiv 1$. This motivates us to combine $\{\widehat{\boldsymbol{\varphi}}(\tau_k), 1 \leq k \leq K\}$ linearly to define the WQAE of $\boldsymbol{\beta}$, as follows:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\pi}_K) = \sum_{k=1}^K \pi_k \widehat{\boldsymbol{\varphi}}(\tau_k) \quad \text{with} \quad \sum_{k=1}^K \pi_k Q_\eta(\tau_k) = 1; \tag{3.16}$$

see also Zhao and Xiao (2014).

Denote $\mathscr{B}$ as the parameter space of $\boldsymbol{\beta}$, and assume $\mathscr{B}$ is compact and $\boldsymbol{\beta}_0$ is an interior point. Define the $K \times K$ matrix $H = \{h_{ij}, 1 \leq i, j \leq K\}$, with $h_{ij} = \Xi^{-1}(\tau_i) S(\tau_i, \tau_j) \Xi^{-1}(\tau_j)$. The asymptotics of $\widehat{\boldsymbol{\beta}}(\boldsymbol{\pi}_K)$ and optimal choice for $\boldsymbol{\pi}_K$ are provided below.

Theorem 6. *Suppose the conditions of Theorem 5 hold. Then as $N \to \infty$,*

$$\sqrt{N}\left(\widehat{\boldsymbol{\beta}}(\boldsymbol{\pi}_K) - \boldsymbol{\beta}_0\right) \to_d N\left(0, \Upsilon(\boldsymbol{\pi}_K)\right),$$

where $\Upsilon(\boldsymbol{\pi}_K) = \boldsymbol{\pi}_K' H \boldsymbol{\pi}_K D_0^{-1}$. Moreover, the optimal weight is

$$\boldsymbol{\pi}_{opt,K} = \underset{\boldsymbol{\pi},\, s.t. \boldsymbol{\pi}'\boldsymbol{q}=1}{\operatorname{argmin}} \Upsilon(\boldsymbol{\pi}_K) = \frac{H^{-1}\boldsymbol{q}}{\boldsymbol{q}' H^{-1} \boldsymbol{q}}, \tag{3.17}$$

where $\boldsymbol{q} = (Q_\eta(\tau_1), \ldots, Q_\eta(\tau_K))'$. Thus, the asymptotic covariance of the optimal WQAE $\widehat{\boldsymbol{\beta}}(\boldsymbol{\pi}_{opt,K})$ is $\Upsilon(\boldsymbol{\pi}_{opt,K}) = (\boldsymbol{q}' H^{-1} \boldsymbol{q})^{-1} D_0^{-1}$.

To estimate the optimal weight $\boldsymbol{\pi}_{opt,K}$, consistent estimators of $\boldsymbol{q}$ and H are necessary. We can approximate $\boldsymbol{q}$ by $\widehat{\boldsymbol{q}} = (\widehat{Q}_\eta(\tau_1), \ldots, \widehat{Q}_\eta(\tau_K))'$, with $\widehat{Q}_\eta(\tau) = \widehat{\varphi}_0(\tau)$, and estimate H by $\widehat{H} = \{\widehat{h}_{ij}, 1 \leq i, j \leq K\}$, with $\widehat{h}_{ij} = \widehat{\Xi}^{-1}(\tau_i) S(\tau_i, \tau_j) \widehat{\Xi}^{-1}(\tau_j)$ and $\widehat{\Xi}(\tau) = \widetilde{f}_\eta(\widehat{Q}_\eta(\tau))$. Here, using $\widehat{Q}_\eta(\tau)$ instead of the empirical quantile $\bar{Q}_\eta(\tau)$ to estimate $\boldsymbol{q}$ ensures that the estimate of β_{00} is one. Then, a consistent estimator $\widehat{\boldsymbol{\pi}}_{opt,K} = (\widehat{\pi}_{1,opt}, \ldots, \widehat{\pi}_{K,opt})'$ of $\boldsymbol{\pi}_{opt,K} = (\pi_{1,opt}, \ldots, \pi_{K,opt})'$ can be obtained by plugging in $\widehat{H}$ and $\widehat{\boldsymbol{q}}$. Finally, the optimal WQAE can be calculated as $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K}) = \sum_{k=1}^K \widehat{\pi}_{k,opt} \widehat{\boldsymbol{\varphi}}(\tau_k)$. It can be shown that $\sqrt{N}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K}) - \boldsymbol{\beta}_0) \to_d N\left(0, \Upsilon(\boldsymbol{\pi}_{opt,K})\right)$ as $N \to \infty$, which implies that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ and $\widehat{\boldsymbol{\beta}}(\boldsymbol{\pi}_{opt,K})$ have the same asymptotic efficiency.

Remark 4. *(Comparison of $\check{\boldsymbol{\beta}}$, $\widetilde{\boldsymbol{\beta}}_c$, and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$) $\check{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ are estimated using different methods, and thus a theoretical comparison is infeasible. Instead, we compare their finite-sample performance by means of a simulation. The results in Section 4 indicate that*

$\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ outperforms $\check{\boldsymbol{\beta}}$ in terms of the bias and empirical and asymptotic standard deviations when the sample size is small or even moderate. In contrast, note that the convergence rates of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ and $\check{\boldsymbol{\beta}}$ are $\sqrt{N}$ and $\sqrt{NT}$, respectively. Thus, $\check{\boldsymbol{\beta}}$ can be more efficient than $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ for large samples. We also compare the finite-sample performance of $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ and $\widetilde{\boldsymbol{\beta}}_c$, with the results showing that $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ has smaller biases and empirical standard deviations than $\widetilde{\boldsymbol{\beta}}_c$. Therefore, we suggest improving the efficiency by using $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$ when the sample size is small or moderate, whereas $\check{\boldsymbol{\beta}}$ is preferred in large samples owing to its faster convergence rate.

4. Simulation Study

This section conducts a simulation experiment to evaluate the finite-sample performance of the proposed two-stage hybrid estimation procedure in Section 3, including the Gaussian QMLE $\widehat{\boldsymbol{\zeta}}$, the WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$, and the WQAE $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$.

The data are generated from the following model:

$$\begin{cases} y_{it} = \theta_i + 0.5 y_{i,t-1} + 0.5 \sum_{j=1}^{N} w_{N,ij} y_{jt} + z_{it} + \varepsilon_{it}, \\ \theta_i = 1 + x_i + (1 + |x_i|)\, \eta_i, \end{cases}$$

where the innovations $\{\varepsilon_{it}\}$ are independently standard normal, $\{\eta_i\}$ are independently standard normal or follow the uniform distribution $U(-\sqrt{3}, \sqrt{3})$, and $\{x_i\}$ and $\{z_{it}\}$ are independently generated randomly from the uniform distribution $U(-1,1)$ and the standard normal distribution, respectively. The spatial weights matrix is generated according to rook contiguity and is row normalized (Yu et al., 2008; Su and Yang, 2015). Note that the QMLE at Stage 1 has a $\sqrt{NT}$-convergence rate under Assumption 3.2, which allows two cases: (i) both N and T are large; and (ii) N is large, but T is fixed. In contrast, the WCQE and WQAE at Stage 2 both have a $\sqrt{N}$-convergence rate under Assumption 3.3, which allows only case (i). Hence, we consider three sample settings, $(N,T) = (20,20), (20,50)$, and $(100,50)$, for the QMLE, and $(N,T) = (20,20), (100,50)$, and $(300,200)$ for the WCQE and WQAE, with 1000 replications generated for each sample setting.

We aim to estimate the parameters $\boldsymbol{\zeta}_0 = (\alpha_0, \gamma_{10}, \psi_{00}, \psi_{10}, \sigma^2_{\varepsilon 0}, \lambda_0, \beta^\star_{00}, \beta^\star_{10})'$, $\boldsymbol{\varphi}_0(\tau) = (\varphi_{00}(\tau), \varphi_{10}(\tau))'$, and β_{10} using the QMLE $\widehat{\boldsymbol{\zeta}}$, WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$, and WQAE $\widehat{\beta}_1(\widehat{\boldsymbol{\pi}}_{opt,K})$, respectively, where $(\beta^\star_{00}, \beta^\star_{10}) = (1, \beta_{10})\sigma_{\eta 0}/\sigma_{\varepsilon 0}$ and $\boldsymbol{\varphi}_0(\tau) = (1, \beta_{10})Q_\eta(\tau)$. The quasi-Newton method (Doob, 1935) is employed to solve the QMLE. For the WCQE, we use the R package "nloptr" (Johnson, 2021), with the same sign restriction imposed on each element of $\widehat{\boldsymbol{\varphi}}(\tau)$, owing to $\boldsymbol{\varphi}_0(\tau) = \boldsymbol{\beta}_0 Q_\eta(\tau)$ and $\boldsymbol{\beta}_0 \geq 0$. For the WQAE, we employ $K = 9$ quantile levels $\tau_k = k/10$, with $k = 1, \ldots, 9$, and the resulting $\widehat{\beta}_1(\widehat{\boldsymbol{\pi}}_{opt,K})$ is almost nonnegative under the same sign restriction for $\widehat{\boldsymbol{\varphi}}(\tau)$.

Table 1 reports the biases, empirical standard deviations (ESDs), and asymptotic standard deviations (ASDs) of the QMLE $\widehat{\boldsymbol{\zeta}}$ at Stage 1, where the ASDs are calculated using a residual-based bootstrap with bootstrap sample size $B = 500$. Recall that we require $N \to \infty$ theoretically. Thus, we focus only on the comparison of the (N,T) settings with N increasing. Comparing the results of the settings $(N,T) = (20,20)$ and $(100,50)$, in which T increases at a slower rate than N, most of the biases, ESDs, and ASDs become smaller as N increases. We have the same finding when comparing the results of $(N,T) = (20,50)$ with that of $(100,50)$, where T is fixed as N increases. Moreover, the biases for the normal distribution of η_i are mostly smaller than those for the uniform distribution. This is expected, because the QMLE reduces to the MLE when η_i is normally distributed.

Table 2 summarizes the biases, ESDs, and ASDs of the WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$ at Stage 2 for three quantile levels, $\tau = 0.25, 0.5$, and 0.75, and three sample settings, $(N,T) = (20,20), (100,50)$, and $(300,200)$. We find the following: (i) for both normally and uniformly distributed innovations $\{\eta_i\}$, as N increases, most of the biases, ESDs, and ASDs become smaller, and the ESDs get closer to their corresponding ASDs; (ii) the performance of $\widehat{\boldsymbol{\varphi}}(\tau)$ improves as the quantile level τ gets closer to the center. In addition, the ESDs are obviously smaller than the corresponding ASDs at $\tau = 0.5$, because $Q_\eta(0.5) \equiv 0$ holds for symmetric distributions centered at zero, and thus $\varphi_{00}(0.5) = \varphi_{10}(0.5) \equiv 0$, which makes ESDs abnormally small under the same sign constraint. However the ESDs are closer to their corresponding ASDs for larger N; see Table S.1 in the Supplementary Material.

Table 3 provides the biases, ESDs, and ASDs of the WQAE $\widehat{\beta}_1(\widehat{\boldsymbol{\pi}}_{opt,K})$ at Stage 2 for three sample settings $(N,T) = (20,20), (100,50)$, and $(300,200)$. For comparison, the results of $\check{\beta}_1$ in (3.10) and $\widetilde{\beta}_{1c}$ in (3.15) are also reported, where the extra setting $(N,T) = (50,20)$ is added to show the performance of $\check{\beta}_1$ in cases with moderate sample sizes. It can be seen that, as N increases, the biases, ESDs, and ASDs of all estimators become smaller, and the ESDs get closer to the corresponding ASDs. Moreover, the ESDs of $\widehat{\beta}_1(\widehat{\boldsymbol{\pi}}_{opt,K})$ move closer to their corresponding ASDs for larger N; see Table S.2 in the Supplementary Material. In addition, the estimate $\check{\beta}_1$ is not satisfactory, because the the sample size is not large, and $\widetilde{\beta}_{1c}$ is usually less efficient than $\widehat{\beta}_1(\widehat{\boldsymbol{\pi}}_{opt,K})$ in terms of ESDs and ASDs.

In summary, the finite-sample performance of the two-stage hybrid estimation procedure is reasonable, which validates the accuracy of the asymptotics in Section 3.

5. Empirical Analysis of Air Quality

This section demonstrates the usefulness of the proposed model and its estimation procedure by analyzing the pure influence of socio-economic factors on air quality in China, after eliminating the spatial-temporal effects.

We choose the air quality index (AQI) as the indicator for air quality, downloaded from China National Environmental Monitoring Centre (*http://www.cnemc.cn/*), spanning the period January 1, 2018, to December 16, 2018, among 143 cities in China. Because the AQI is closely related to meteorological factors (Yang et al., 2007; Lee et al., 2012), we select temperature (TEM), precipitation (PRE), and wind speed (WIN) to represent exogenous spatial-temporal effects in our model; the data are downloaded from National Meteorological Information Center (*http://data.cma.cn/en*). Moreover, the AQI can be affected by socio-economic factors related to each city (Chen and Xu, 2017; Yu and Liu, 2020). Hence, we use the gross regional product (GRP) and secondary industry share in GRP (Industry) as space-specific factors; the data are obtained from China City Statistical Yearbook. Many references indicate that population or population density is also an important socio-economic factor that can influence air quality. However, Figure 1 shows that there is a significant positive correlation between GRP and population (or population density) in China. Hence we only include GRP to avoid multicollinearity. To balance the size and information of the data in the time dimension, we transform the original average hourly data of the AQI and the average daily data of the TEM and WIN into average weekly data, and transform the original cumulative daily data of PRE into cumulative weekly data by summation. Finally, we obtain a spatial-temporal data set with $N = 143$ and $T = 49$.

In Figure 2, the heat maps of the AQI of some local areas of China from the first to the ninth weeks in 2018 imply that there are spatial and temporal effects in the air quality. This motivates us to consider both effects using our model. In this study, we aim to understand the influence of local socio-economic factors on air quality, controlling for spatial-temporal effects, so that governments can evaluate the effectiveness of local policies in terms of improving air quality.

Note that the scales of AQI, TEM, PRE, WIN, GRP, and Industry differ from each other, as clearly demonstrated in Table 4, so we divide them by the estimated standard deviations. Moreover, we centralize TEM, PRE, and WIN, owing to the zero mean requirement of Assumption 3.6(ii). In accordance with the model setting in (2.1), for $i = 1, \ldots, 143$ and $t = 0, \ldots, 49$, we denote the processed variables AQI, TEM, PRE, WIN, GRP, and Industry as $\{y_{it}\}$, $\{z_{1it}\}$, $\{z_{2it}\}$, $\{z_{3it}\}$, $\{x_{1i}\}$, and $\{x_{2i}\}$, respectively. We consider the proposed two-stage hybrid estimation procedure in Section 3. The fitted model of (2.1) is then given by

$$y_{it} = \widehat{\theta}_i + 0.178\, y_{i,t-1} + 0.648 \sum_{j=1}^{N} w_{N,ij} y_{jt} - 0.096\, z_{1it} - 0.065\, z_{2it} - 0.078\, z_{3it} + \widehat{\varepsilon}_{it}, \quad (5.1)$$

where the spatial weights matrix $W_N = \{w_{N,ij} = a_{ij}/d_i\}$ is a row-normalized binary adjacency matrix with $d_i = \sum_{j=1}^{N} a_{ij}$, $a_{ii} = 0$, and $a_{ij} = 1$ if the ith city is adjacent to the jth city geographically, otherwise $a_{ij} = 0$ for $i \neq j$. The summary information of the fitted coefficients in model (5.1) is provided in Table 5, where the standard errors are calculated using a residual-based bootstrap with bootstrap sample size $B = 500$. Note that the coefficients are significantly nonzero at the 10% significance level. The fitted model (5.1) indicates that the first lag of the AQI and its neighbors have positive impacts on the AQI, whereas the temperature, precipitation, and wind speed have negative impacts on the AQI. These findings are consistent with common sense.

We are particularly interested in the influence of x_{1i} and x_{2i} on the low and high quantile levels of the individual random effect θ_i, that is, the influence of socio-economic factors on air quality in good or poor status after spatial-temporal effects have been removed. For this purpose, we choose two quantile levels, $\tau = 0.25$ and 0.75, for illustration. The estimated conditional quantile functions of θ_i in model (2.4) are

$$\begin{cases} \widehat{Q}_{\theta_i}(0.25|\boldsymbol{x}_i) = (0.171 + 0.078\,x_{1i} + 0.063\,x_{2i}) + (-\,0.066 - 0.016\,|x_{1i}| - 0.046\,|x_{2i}|)\,, \\ \widehat{Q}_{\theta_i}(0.75|\boldsymbol{x}_i) = (0.171 + 0.078\,x_{1i} + 0.063\,x_{2i}) + (0.030 + 0.021\,|x_{1i}| + 0.014\,|x_{2i}|)\,, \end{cases} \tag{5.2}$$

where the summary information of the fitted coefficients is reported in Table 5. For each equation of (5.2), the first set of parentheses corresponds to the estimated conditional location of θ_i, and the second represents the conditional quantile estimate of the scale effect. Because the conditional location of θ_i is actually the conditional expectation of θ_i, the fitted conditional expectation of θ_i is $\widehat{E}(\theta_i|\boldsymbol{x}_i) = 0.171 + 0.078x_{1i} + 0.063x_{2i}$. Note that $x_{1i}, x_{2i} > 0$ for all i in our situation. Then, $|x_{\ell i}|$ is equal to $x_{\ell i}$, for $\ell = 1, 2$. As a result, based on model (5.2) and Table 5, we have the following conclusions: (i) both the secondary industry share and GRP may have positive impacts on the conditional mean of θ_i and its low and high quantiles; (ii) with the spatial-temporal effects eliminated, the nonsignificant positive impact on the AQI of GRP implies that China may locate near the apex of the environmental Kuznets curve (Stern et al., 1996); that is, the adverse impact of China's economic development on the enviroment diminishes gradually; (iii) the secondary industry share has a statistically significant positive impact on the conditional location of θ_i at the 10% significance level, which indicates that the secondary industry may cause a deterioration of the air quality in China. Moreover, the variable x_{2i} is significantly related to the conditional scale of θ_i at the lower quantile level, while no statistically significant effect is observed at the higher level

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

at the 10% significance level, which provides evidence of possible heterogeneity in the AQI, owing to the secondary industry share.

6. Conclusion

We have introduced a dynamic spatial autoregressive model with heterogeneous random effects, which is useful in cases with complicated correlation structures. Although we focus on space-specific factors, this idea can be used similarly in other random-effect models to account for possible heterogeneity based on a quantile regression.

There are several potential research topics in the framework of the proposed model. First, note that the asymptotic covariance of the QMLE in Theorem 2 is infeasible to compute, and the limiting distributions of the WCQE and WQAE in (3.14) and Theorem 6 are not convenient to estimate, owing to the unknown density function $f_\eta(\cdot)$ of those innovations with estimations that need some nonparametric methods with tuning parameters. The bootstrap method we used to approximate the covariance matrix of the QMLE introduces a perturbation at the individual level, but ignores the time level, which may make it less efficient. As a result, we can consider a new method that perturbs the information of individual and time levels simultaneously for the proposed model. Hopefully, the bootstrapping can simplify the estimation of the asymptotic distributions and lead to more accurate inference results when the sample size is moderate or small. Second, missing values in individuals or time points are common for spatial panel data in practice. For example, in our real application, there are missing observations for the AQI of certain cities (e.g., Yantai), and we adopt the likewise deletion method for simplicity, which may result in information loss. Hence, it is beneficial to consider an estimation procedure in the presence of missing values at the response. Third, we can consider a more general quantile regression model than the location-scale model given in (2.2):

$$Q_{\theta_i}(\tau|x_{1i},\cdots,x_{pi}) = \tilde{\psi}_0(\tau) + \sum_{\ell=1}^{p} \tilde{\psi}_\ell(\tau)x_{\ell i},$$

where $\{\tilde{\psi}_\ell(\cdot)\}$ are quantile-dependent coefficients. We leave these topics for future research.

Supplementary Material

The online Supplementary Material contains a bootstrapping procedure for the QMLE, additional simulation results, detailed forms of the matrices Σ_N, $\Sigma_N^\star$, $\Omega_{NT}^{-1}(\boldsymbol{\beta}^\star)$, and the vector $\boldsymbol{Y}_{-1}$, and technical details for Theorems 1–6 and Propositions 1–2, and Lemmas 1–10.

Acknowledgments

We are grateful to the co-editor, associate editor, and two anonymous referees for their valuable comments and suggestions. This research was partly supported by the National Natural Science Foundation of China grant (11971292, 11690012 and 12001355), Shanghai Pujiang Program 2019PJC051, Chenguang Program 19CG44, and Shanghai Research Center for Data Science and Decision Technology. The authors contributed equally to this paper and their names are listed in alphabetical order.

References

Baltagi, B. H., P. Egger, and M. Pfaffermayr (2013). A generalized spatial panel data model with random effects. *Econometric Reviews 32*, 650–685.

Baltagi, B. H., B. C. Jung, and S. H. Song (2010). Testing for heteroskedasticity and serial correlation in a random effects panel data model. *Journal of Econometrics 154*, 122–124.

Bowatte, G., C. Lodge, A. J. Lowe, B. Erbas, J. Perret, M. J. Abramson, M. Matheson, and S. C. Dharmage (2015). The influence of childhood traffic-related air pollution exposure on asthma, allergy and sensitization: a systematic review and a meta-analysis of birth cohort studies. *Allergy 70*, 245–256.

Busch, P.-O. and H. Jörgens (2005). International patterns of environmental policy change and convergence. *European Environment 15*, 80–101.

Cakmak, S., R. Dales, J. Leech, and L. Liu (2011). The influence of air pollution on cardiovascular and pulmonary function and exercise capacity: Canadian health measures survey (CHMS). *Environmental Research 111*, 1309–1312.

Canay, I. A. (2011). A simple approach to quantile regression for panel data. *The Econometrics Journal 14*, 368–386.

Chen, J., B. Wang, S. Huang, and M. Song (2020). The influence of increased population density in China on air pollution. *Science of the Total Environment 735*, 139456.

Chen, N. and L. Xu (2017). Relationship between air quality and economic development in the provincial capital cities of China. *Environmental Science and Pollution Research 24*, 2928–2935.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics 132*, 491–525.

Cramer, J. C. (1998). Population growth and air quality in California. *Demography 35*, 45–56.

De Bruyn, S. M., J. C. van den Bergh, and J. B. Opschoor (1998). Economic growth and emissions: reconsidering the empirical basis of environmental Kuznets curves. *Ecological Economics 25*, 161–175.

Doob, J. L. (1935). The limiting distributions of certain statistics. *The Annals of Mathematical Statistics 6*, 160–169.

Giri, D., M. Krishna, and P. Adhikary (2008). The influence of meteorological conditions on PM10 concentrations in Kathmandu Valley. *International Journal of Environmental Research 2*, 49–60.

Gulia, S., S. S. Nagendra, M. Khare, and I. Khanna (2015). Urban air quality management-A review. *Atmospheric Pollution Research 6*, 286–304.

Hansen, B. (2021). *Econometrics*. University of Wisconsin. Available at https://www.ssc.wisc.edu/~bhansen/econometrics/Econometrics.pdf.

Johnson, S. G. (2021). The nlopt nonlinear-optimization package. http://github.com/stevengj/nlopt.

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

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

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

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics 104*, 219–257.

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

Koenker, R. and G. J. Bassett (1978). Regression quantiles. *Econometrica 46*, 33–50.

Koenker, R. and Q. Zhao (1994). L-estimatton for linear heteroscedastic models. *Journaltitle of Nonparametric Statistics 3*, 223–235.

Lee, C. C., T. J. Ballinger, and N. A. Domino (2012). Utilizing map pattern classification and surface weather typing to relate climate to the Air Quality Index in Cleveland, Ohio. *Atmospheric Environment 63*, 50–59.

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

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

Lee, L.-F. and J. Yu (2014). Efficient gmm estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics 180*, 174–197.

Li, L., Y. Lei, S. Wu, C. He, J. Chen, and D. Yan (2018). Impacts of city size change and industrial structure change on CO_2 emissions in Chinese cities. *Journal of Cleaner Production 195*, 831–838.

Li, L. and Z. Yang (2021). Spatial dynamic panel data models with correlated random effects. *Journal of Econometrics 221*, 424–454.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica 46*, 69–85.

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

Shi, W. and L.-F. Lee (2017). Spatial dynamic panel data models with interactive fixed effects. *Journal of Econometrics 197*, 323–347.

Stern, D. I., M. S. Common, and E. B. Barbier (1996). Economic growth and environmental degradation: the environmental Kuznets curve and sustainable development. *World Development 24*, 1151–1160.

Streck, C. (2004). New partnerships in global environmental policy: The clean development mechanism. *The Journal of Environment & Development 13*, 295–322.

Su, L. and Z. Yang (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics 185*, 230–258.

Wang, S. and J. Hao (2012). Air quality management in China: Issues, challenges, and options. *Journal of Environmental Sciences 24*, 2–13.

Yang, L., D. Wang, S. Cheng, Z. Wang, Y. Zhou, X. Zhou, and W. Wang (2007). Influence of meteorological conditions and particulate matter on visual range impairment in Jinan, China. *Science of the Total Environment 383*, 164–173.

Yu, J., R. De Jong, and L.-F. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with

doi:10.5705/ss.202021.0280

fixed effects when both n and T are large. *Journal of Econometrics 146*, 118–134.

Yu, Y. and H. Liu (2020). Economic growth, industrial structure and nitrogen oxide emissions reduction and prediction in China. *Atmospheric Pollution Research 11*, 1042–1050.

Zhao, Z. and Z. Xiao (2014). Efficient regressions via optimally combining quantile information. *Econometric Theory 30*, 1272–1314.

Zhou, J., J. Liu, F. Wang, and H. Wang (2022). Autoregressive model with spatial dependence and missing data. *Journal of Business & Economic Statistics 40*, 28–34.

Zhu, Q., Y. Zheng, and G. Li (2018). Linear double autoregression. *Journal of Econometrics 207*, 162–174.

Xingdong Feng, Shanghai University of Finance and Economics, School of Statistics and Management

E-mail: feng.xingdong@mail.shufe.edu.cn

Wenyu Li, Shanghai University of Finance and Economics, School of Statistics and Management

E-mail: liwenyu2018@163.sufe.edu.cn

Qianqian Zhu, Shanghai University of Finance and Economics, School of Statistics and Management

E-mail: zhu.qianqian@mail.shufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

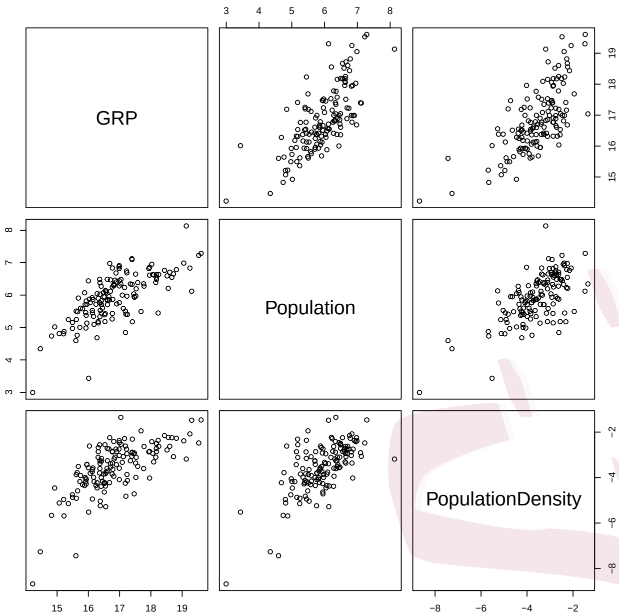

Figure 1: Scatterplots of the log-transformed variables "GRP," "population," and "population density."

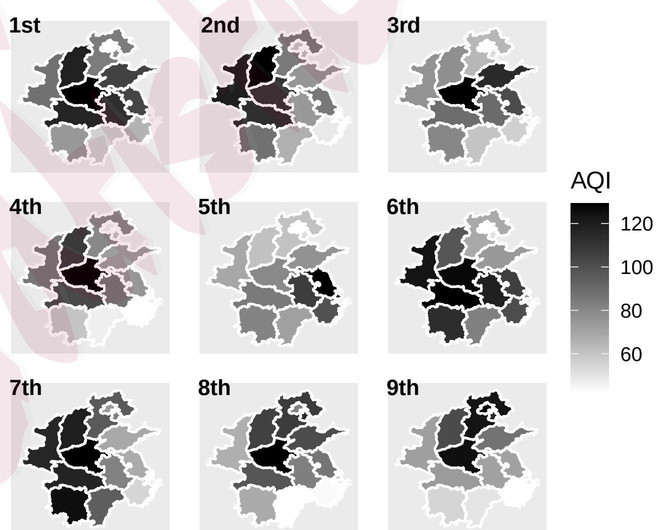

Figure 2: Heat maps of the AQI of part of central and eastern China from the first to the ninth weeks in 2018.

Table 1: Biases, ESDs, and ASDs of the QMLE $\widehat{\boldsymbol{\zeta}}$ when the innovations $\{\eta_i\}$ follow a standard normal or uniform distribution.

			$N(0,1)$			$U(-\sqrt{3},\sqrt{3})$		
	N	T	Bias	ESD	ASD	Bias	ESD	ASD
α	20	20	0.0021	0.0234	0.0398	0.0029	0.0232	0.0341
	20	50	0.0007	0.0146	0.0400	0.0008	0.0141	0.0332
	100	50	0.0006	0.0080	0.0089	0.0006	0.0079	0.0088
γ_1	20	20	0.0020	0.0511	0.0561	0.0013	0.0516	0.0571
	20	50	0.0019	0.0336	0.0381	0.0026	0.0318	0.0371
	100	50	0.0002	0.0145	0.0147	0.0005	0.0144	0.0147
ψ_0	20	20	0.0089	0.3737	1.2303	0.0091	0.3930	0.8101
	20	50	0.0292	0.3561	1.8826	0.0201	0.3966	2.1397
	100	50	0.0040	0.1537	0.1609	0.0045	0.1519	0.1596
ψ_1	20	20	0.0404	0.6824	1.2488	0.0161	0.6284	0.9061
	20	50	0.0018	0.6537	4.9535	0.0149	0.6501	3.0061
	100	50	0.0072	0.2782	0.2731	0.0100	0.2742	0.2735
σ_ε^2	20	20	0.0039	0.0714	0.1266	0.0086	0.0752	0.1138
	20	50	0.0013	0.0464	0.1113	0.0027	0.0449	0.0913
	100	50	0.0001	0.0201	0.0377	0.0003	0.0193	0.0366
λ	20	20	0.0030	0.0237	0.0435	0.0037	0.0235	0.0375
	20	50	0.0008	0.0147	0.0428	0.0010	0.0141	0.0346
	100	50	0.0006	0.0080	0.0089	0.0006	0.0079	0.0088
$\beta_0^\star$	20	20	0.1805	0.5059	2.7118	0.1295	0.4365	1.6301
	20	50	0.1834	0.4967	7.5333	0.1041	0.4295	4.2356
	100	50	0.0198	0.2087	0.2581	0.0177	0.1360	0.1718
$\beta_1^\star$	20	20	0.1256	0.9422	5.7388	0.0584	0.7995	3.0798
	20	50	0.1419	0.9644	15.7608	0.0312	0.8020	9.5194
	100	50	0.0015	0.4065	0.4525	0.0043	0.2630	0.2991

Table 2: Biases, ESDs, and ASDs of the WCQE $\widehat{\boldsymbol{\varphi}}(\tau)$ at $\tau = 0.25, 0.5, 0.75$ when the innovations $\{\eta_i\}$ follow a standard normal or uniform distribution.

				$N(0,1)$			$U(-\sqrt{3},\sqrt{3})$		
		N	T	Bias	ESD	ASD	Bias	ESD	ASD
$\tau = 0.25$	φ_0	20	20	0.0303	0.4880	0.9059	0.1333	0.5239	0.9455
		100	50	0.0219	0.3287	0.3713	0.0479	0.3667	0.3934
		300	200	0.0081	0.1935	0.1970	0.0144	0.2116	0.2101
	φ_1	20	20	0.0913	0.7667	1.6977	0.0247	0.8428	1.7688
		100	50	0.0585	0.5937	0.7213	0.0496	0.6657	0.7625
		300	200	0.0225	0.3896	0.4091	0.0056	0.4270	0.4368
$\tau = 0.5$	φ_0	20	20	0.0164	0.2520	0.8681	0.0037	0.2654	1.0016
		100	50	0.0038	0.1092	0.3529	0.0035	0.1334	0.4541
		300	200	0.0045	0.0589	0.1851	0.0041	0.0750	0.2427
	φ_1	20	20	0.0143	0.3411	1.6275	0.0253	0.4375	1.8757
		100	50	0.0104	0.1718	0.6852	0.0028	0.2384	0.8802
		300	200	0.0052	0.1125	0.3844	0.0015	0.1558	0.5042
$\tau = 0.75$	φ_0	20	20	0.0336	0.4843	0.8838	0.1093	0.5027	0.9572
		100	50	0.0141	0.3197	0.3737	0.0511	0.3666	0.3981
		300	200	0.0082	0.1913	0.1960	0.0063	0.2080	0.2101
	φ_1	20	20	0.0864	0.7897	1.6561	0.0118	0.8468	1.7927
		100	50	0.0300	0.5683	0.7254	0.0517	0.6734	0.7715
		300	200	0.0126	0.3703	0.4069	0.0119	0.4215	0.4368

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0280

Table 3: Biases, ESDs, and ASDs of three estimators of $\boldsymbol{\beta}_0$, that is, $\check{\boldsymbol{\beta}}$ by the QMLE, the initial estimator $\widetilde{\boldsymbol{\beta}}_c$, and the WQAE $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{\pi}}_{opt,K})$, when the innovations $\{\eta_i\}$ follow a standard normal or uniform distribution.

			$N(0,1)$			$U(-\sqrt{3},\sqrt{3})$		
	N	T	Bias	ESD	ASD	Bias	ESD	ASD
$\check{\beta}_1$	20	20	276805	827496	5e+12	146686	609974	3e+12
	50	20	3181	100581	4e+9	3281	103752	4e+9
	100	50	0.1638	0.7819	0.8772	0.0779	0.4386	0.5052
	300	200	0.0340	0.3032	0.3548	0.0167	0.1942	0.2278
$\widetilde{\beta}_{1c}$	20	20	0.5346	1.2984	-	0.4840	1.2580	-
	100	50	0.2714	0.7978	-	0.2020	0.5663	-
	300	200	0.0668	0.3621	-	0.0637	0.3044	-
$\widehat{\beta}_1$	20	20	0.5116	1.2505	1.1530	0.3418	1.1946	1.0260
	100	50	0.1165	0.8976	0.4603	0.0669	0.5035	0.3477
	300	200	0.0348	0.3578	0.2455	0.0606	0.2372	0.1730

$*\ e+12$ denotes 10^{12} .

Table 4: Summary statistics for AQI, TEM (0.1°C), PRE (0.1mm), WIN (0.1m/s), GRP (10000 yuan), and Industry (%).

	AQI	TEM	PRE	WIN	GRP	Industry
Min	13.264	-283.000	0.000	4.714	1500100	18.270
Max	312.845	325.714	5113.000	78.286	326798700	72.900

doi:10.5705/ss.202021.0280

Table 5: Summary information of fitted coefficients for models (5.1) and (5.2).

	Estimate	Std. Error	z statistic	p-value
α	0.178	0.027	6.593	0.000
λ	0.648	0.022	29.455	0.000
γ_1	-0.096	0.015	-6.400	0.000
γ_2	-0.065	0.006	-10.833	0.000
γ_3	-0.078	0.013	-6.000	0.000
ψ_0	0.171	0.194	0.881	0.378
ψ_1	0.078	0.088	0.886	0.375
ψ_2	0.063	0.036	1.750	0.080
$\varphi_0(0.25)$	-0.066	0.103	-0.641	0.522
$\varphi_1(0.25)$	-0.016	0.028	-0.571	0.568
$\varphi_2(0.25)$	-0.046	0.024	-1.917	0.055
$\varphi_0(0.75)$	0.030	0.216	0.139	0.890
$\varphi_1(0.75)$	0.021	0.058	0.362	0.717
$\varphi_2(0.75)$	0.014	0.049	0.286	0.775