<table>
<thead>
<tr>
<th>Title</th>
<th>Spatial-temporal Model with Heterogeneous Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>SS-2021-0280</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a></td>
</tr>
<tr>
<td>DOI</td>
<td>10.5705/ss.202021.0280</td>
</tr>
<tr>
<td>Complete List of Authors</td>
<td>Xingdong Feng,</td>
</tr>
<tr>
<td></td>
<td>Wenyu Li</td>
</tr>
<tr>
<td></td>
<td>Qianqian Zhu</td>
</tr>
<tr>
<td>Corresponding Author</td>
<td>Qianqian Zhu</td>
</tr>
<tr>
<td>E-mail</td>
<td><a href="mailto:zhu.qianqian@mail.shufe.edu.cn">zhu.qianqian@mail.shufe.edu.cn</a></td>
</tr>
</tbody>
</table>
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
Heterogeneous DSAR Model

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,
Heterogeneous DSAR Model

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. Specif-
ically, 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 \( \lambda_{\text{min}}(\cdot) \) and \( \lambda_{\text{max}}(\cdot) \) 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, \( 0_m \) as an \( m \times 1 \) vector of zeros and \( \iota_m \) as an \( m \times 1 \) vector of ones, and let \( J_m = \iota_m \iota_m' \). 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 $G_n = \sqrt{n}(E_n - E)$. In addition, $\ell^\infty(\mathcal{T})$ denotes the space of all uniformly bounded functions on $\mathcal{T}$. Moreover, $\rightarrow_p$ denotes convergence in probability, $\rightarrow_d$ 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:

$$y_{it} = \theta_i + \alpha y_{i,t-1} + \lambda \sum_{j=1}^{N} w_{N,ij} y_{jt} + \sum_{l=1}^{q} \gamma_l z_{lt_{it}} + \varepsilon_{it}, \quad i = 1, \ldots, N, t = 1, \ldots, T, \quad (2.1)$$

where $\{\theta_i\}$ are space-specific or individual random effects, the scalar parameters $\alpha$ and $\lambda$ characterize the dynamic effect and the spatial autoregressive effect, respectively, $\{\gamma_l\}$ are the coefficients of the time-varying exogenous variables $\{z_{lt_{it}}\}$, $\{w_{N,ij}\}$ 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 $\sigma^2$.

To flexibly depict the possible influence of space-specific covariates on an individual random effect, we assume a location-scale structure for $\theta_i$, that is,

$$\theta_i = \mu_i + \vartheta_i,$$

where $\{x_{li}\}$ are time-invariant constant regressors, $\{\psi_l\}$ and $\{\beta_l\}$ are the coefficients of $\{x_{li}\}$ and $\{|x_{li}|\}$, 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 $\sigma^2$. To see Koenker and Zhao (1994) and Zhao and Xiao (2014). Let $\mu_i = \psi_0 + \sum_{\ell=1}^{p} \psi_{l} x_{li}$ and $\vartheta_i = (\beta_0 + \sum_{\ell=1}^{p} \beta_{l} |x_{li}|) \eta_i$. It then follows that

$$\theta_i = \mu_i + \vartheta_i.$$
The location-scale structure combines the advantages of fixed and random effects. Specifically, the location \( \mu_i \) depicts the difference between individuals on the conditional mean and can be arbitrarily related to time-variant regressors \( \{ z_{it} \} \), whereas the scale effect \( \vartheta_i \) allows for conditional heteroscedasticity and reflects the influence of \( \{ x_{it} \} \) on the fluctuation. Note that if \( \psi_{\ell} = \beta_{\ell} = 0 \) for all \( \ell = 1, \ldots, p \), then \( \theta_i \) reduces to the commonly used individual random effect. Moreover, if \( \beta_0 = \beta_1 = \cdots = \beta_p = 0 \), then \( \theta_i \) 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 \( Y = (y'_1, \ldots, y'_T)' \) and \( Y_{-1} = (y'_0, \ldots, y'_{T-1})' \) with \( y_t = (y_{1t}, \ldots, y_{Nt})' \), \( Z = (Z'_1, \ldots, Z'_T)' \) with \( Z_t = (z_{1t}, \ldots, z_{Nt})' \) and \( z_{it} = (z_{1it}, \ldots, z_{qit})' \), \( X = (x_1, \ldots, x_N)' \) with \( x_i = (1, x_{1i}, \ldots, x_{pi})' \), and \( \varepsilon = (\varepsilon'_1, \ldots, \varepsilon'_T)' \) with \( \varepsilon_i = (\varepsilon_{1i}, \ldots, \varepsilon_{Nt})' \). Let \( S_{NT}(\lambda) = I_T \otimes B_N(\lambda) \) with \( B_N(\lambda) = I_N - \lambda W_N \) and \( W_N = \{ w_{N,ij}, 1 \leq i, j \leq N \} \) be the \( N \times N \) specified constant spatial weights matrix, and denote \( \tilde{Z} = (Y_{-1}, Z, (\iota_T \otimes I_N)X) \). Then, model \( (2.1) \) with \( (2.2) \) can be rewritten in the following matrix form:

\[
S_{NT}(\lambda)Y = \tilde{Z} \phi + (\iota_T \otimes I_N) \vartheta + \varepsilon, \tag{2.3}
\]

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

**Proposition 1.** Suppose that \( \eta_i \) and \( \varepsilon_{it} \) are mutually independent. If \( |\alpha B_N^{-1}(\lambda)| < 1 \) and the process \( \{ z_{it} \} \) is covariance stationary, then there exists a unique covariance stationary
Heterogeneous DSAR Model

solution to models (2.1) and (2.2). The solution takes the form

\[ y_t = \sum_{m=0}^{\infty} \alpha^m \left[ B_N^{-1}(\lambda) \right]^{m+1} (\theta + Z_{t-m}\gamma + \varepsilon_{t-m}). \]

Proposition 1 indicates that the stationarity of \{y_t\} restricts the spatial and temporal effects to some extent by imposing conditions on the weights matrix \( W_N \), spatial effect parameter \( \lambda \), and autoregressive coefficient \( \alpha \). Specifically, \(|\alpha| + |\lambda| < 1\) implies \(|\alpha B_N^{-1}(\lambda)| < 1\) if \( W_N \) is row-normalized. Moreover, the stationarity of \{y_t\} depends on that of the exogenous covariates \{z_it\}. However, assuming that both \{y_t\} and \{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 \{y_t\}.

We are interested in estimating \( \phi, \beta, \lambda, \sigma^2_\varepsilon \), and \( \sigma^2_\eta \), and providing a conditional quantile estimation of the individual random effect \( \theta_i \). Let \( Q_\eta(\tau) \) be the \( \tau \)th quantile of \( \eta_i \). Denote \( \theta = (\theta_1, \ldots, \theta_N)' \) and \( X_a = (x_{a1}, \ldots, x_{aN})' \). The conditional quantile of \( \theta \), given the space-specific covariates \( X \), has the form of

\[ Q_{\theta}(\tau|X) = X\psi + X_a\varphi(\tau), \quad (2.4) \]

where \( \varphi(\tau) = (\varphi_0(\tau), \varphi_1(\tau), \ldots, \varphi_p(\tau))' = Q_\eta(\tau)\beta \). Note that the individual effect \( \theta \) are unobservable. Thus, to estimate \( Q_{\theta}(\tau|X) \) 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 \( \theta_i \) 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 \( \lambda, \alpha, \) and \( \gamma \), and then to apply the quantile regression estimation method to model (2.4),
with \( \theta_i \) replaced by its observable approximate, to obtain estimates of \( \psi \) and \( \varphi(\tau) \). Note that \( \psi \) is independent of \( \tau \) and it is not necessary to estimate it using a conditional quantile estimation in our model setting, whereas \( \varphi(\tau) \) is \( \tau \)-dependent and can be estimated using the quantile estimation method. Moreover, \( \lambda \) and \( \phi \) can be estimated directly, and thus the observable approximate of \( \vartheta_i \) is obtained easily from model (2.3). Together with the fact that \( Q_{\theta}(\tau|X) = X \psi + Q_{\theta^*}(\tau|X) \), this motivates us to obtain observable approximates of \( \vartheta_i \), and then to apply a quantile regression, given by

\[
Q_{\theta^*}(\tau|X) = X a \phi(\tau). \tag{3.1}
\]

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

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

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

As a result, if \( S_{NT}(\hat{\lambda}) \) is invertible, then the conditional mean of the data \( Y \) can be estimated by \( \hat{E}(Y|X) = S_{NT}^{-1}(\hat{\lambda}) \hat{Z} \hat{\phi} \) based on model (2.3) and \( \hat{\zeta} \). Moreover, the conditional quantile of the random effect \( \theta \) can be estimated by \( \hat{Q}_{\theta}(\tau|X) = X \hat{\psi} + X_a \hat{\varphi}(\tau) \) based on model (2.4), \( \hat{\psi} \), and \( \hat{\varphi}(\tau) \). Statistical inference tools, such as significance tests for \( \phi \), \( \varphi(\tau) \), and \( \beta \), can
also be constructed based on \( \hat{\phi}, \hat{\varphi}(\tau), \) and \( \hat{\beta} \). 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 \( (NT)^{-1/2}\partial\ln L_{NT}(\zeta_0)/\partial\zeta \) 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 \( \vartheta \) with \( \hat{\vartheta} \) 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 \( \zeta = (\phi', \sigma_\varepsilon^2, \lambda, \beta^\tau)' \) with \( \beta^\tau = \beta\sigma_\eta/\sigma_\varepsilon \), and let \( \zeta_0 = (\phi_0', \sigma_{\varepsilon_0}^2, \lambda_0, \beta_0^\tau)' \) be the true value of \( \zeta \). Denote \( V_{NT}(\lambda, \phi) = S_{NT}(\lambda)Y - \tilde{Z}\phi \),
3.1 The Quasi-maximum Likelihood Estimation

\[ V_{NT} = (\nu_T \otimes I_N) \vartheta + \varepsilon, \quad B_{0N} = B_N(\lambda_0), \quad \text{and} \quad S_{0NT} = S_{NT}(\lambda_0). \]

Then it follows that \( V_{NT}(\lambda_0, \phi_0) = V_{NT} = S_{0NT}Y - \tilde{Z}\phi_0 \). We consider the case that initial observations are generated exogenously, that is, \( y_0 \) is exogenous such that it can be treated as a constant vector; see also Su and Yang (2015).

Assume that \( \varepsilon_i \) and \( \eta_i \) are distributed with a zero mean and variances \( \sigma^2_\varepsilon \) and \( \sigma^2_\eta \), respectively, which implies that \( \text{cov}\{V_{NT}(\lambda, \phi)\} = \sigma^2_\varepsilon \Omega_{NT}(\beta^*) \), where \( \Omega_{NT}(\beta^*) = J_T \otimes A_N(\beta^*) + I_{NT} \) with \( A_N(\beta^*) = \text{diag}\{(x'_{a1}\beta^*)^2, \ldots, (x'_{aN}\beta^*)^2\} \). The Gaussian quasi-likelihood function of \( \zeta \) is then given by

\[
\mathcal{L}_{NT}(\zeta) = \left(2\pi\sigma^2_\varepsilon\right)^{-NT/2} \left|S'_{NT}(\lambda)\Omega^{-1}_{NT}(\beta^*)S_{NT}(\lambda)\right|^{1/2} \exp \left\{-\frac{V'_{NT}(\lambda, \phi)\Omega^{-1}_{NT}(\beta^*)V_{NT}(\lambda, \phi)}{2\sigma^2_\varepsilon}\right\},
\]

and the corresponding quasi-log-likelihood function of \( \zeta \) is

\[
\ln \mathcal{L}_{NT}(\zeta) = -\frac{NT}{2} \ln (2\pi) - \frac{NT}{2} \ln \sigma^2_\varepsilon - \frac{1}{2} \ln |\Omega_{NT}(\beta^*)| + \ln |S_{NT}(\lambda)| + \frac{1}{2\sigma^2_\varepsilon} V'_{NT}(\lambda, \phi)\Omega^{-1}_{NT}(\beta^*)V_{NT}(\lambda, \phi).
\]

Maximizing (3.2) gives the Gaussian QMLE of \( \zeta \), denoted by \( \hat{\zeta} \). If \( \varepsilon_i \) and \( \eta_i \) are normally distributed, then \( \hat{\zeta} \) is the MLE.

For ease of computation, we consider the concentrated log-likelihood by concentrating out \( \phi \) and \( \sigma^2_\varepsilon \). Given \( \delta = (\lambda, \beta^*)' \), maximizing (3.2) leads to the following QMLEs of \( \phi \) and \( \sigma^2_\varepsilon \):

\[
\hat{\phi}(\delta) = \left(\tilde{Z}'\Omega^{-1}_{NT}(\beta^*)\tilde{Z}\right)^{-1} \tilde{Z}'\Omega^{-1}_{NT}(\beta^*)S_{NT}(\lambda)Y,
\]

and

\[
\hat{\sigma}^2_\varepsilon(\delta) = \frac{1}{NT} \hat{V}'_{NT}(\delta)\Omega^{-1}_{NT}(\beta^*)\hat{V}_{NT}(\delta),
\]

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

\[
\ln \ell_{NT}(\delta) = -\frac{NT}{2} \ln (2\pi) + \frac{NT}{2} \ln \hat{\sigma}_\varepsilon^2(\delta) - \frac{1}{2} \ln |\Omega_{NT}(\beta^*)| + \ln |S_{NT}(\lambda)|. \tag{3.5}
\]

The QMLE \( \hat{\delta} \) that maximizes the function \( \ell_{NT}(\delta) \) is given by

\[
\hat{\delta} = \arg \max_{\delta \in \Delta} \ln \ell_{NT}(\delta),
\]

where \( \Delta \) is the parameter space of \( \delta \), which is assumed to be compact and the true value \( \delta_0 = (\lambda_0, \beta_0^*)' \) is in the interior of \( \Delta \). Then, the QMLEs of \( \phi \) and \( \sigma_\varepsilon^2 \) are \( \hat{\phi} = \hat{\phi}(\hat{\delta}) \) and \( \hat{\sigma}_\varepsilon^2 = \hat{\sigma}_\varepsilon^2(\hat{\delta}) \), respectively. As a result, the QMLE of \( \zeta \) is given by \( \hat{\zeta} = (\hat{\phi}', \hat{\sigma}_\varepsilon^2, \hat{\lambda}, \hat{\beta}^*)' \).

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

\[
\sum_{1 \leq t_1, t_2 \leq N} \sum_{1 \leq t_3, t_4 \leq T} \text{cov}(z_{it_1t_1}, z_{it_2t_2}, z_{it_3t_3}, z_{it_4t_4}) = O(1) \tag{3.6}
\]
3.1 The Quasi-maximum Likelihood Estimation

for any $1 \leq i_1, i_2 \leq N$ and $1 \leq t_1, t_2 \leq T$; and (iv) $\frac{1}{NT} \tilde{Z} \tilde{Z}'$ is positive definite for sufficiently large $N$ and $\lambda_{\min}\left(\frac{1}{NT} \tilde{Z}' \Omega_{NT}^{-1}(\beta^*) \tilde{Z}\right)$ 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 $\theta_i$ in model (2.1), we assume $\varepsilon_{it}$ 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 $\alpha_0^{t-1}$ 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 $W_N$
Heterogeneous DSAR Model

is a row-normalized matrix, then both \( \{ B_N^{-1}(\lambda) \} \) 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 \( \{ y_t \} \) by imposing conditions on the weights matrix \( W_N \), spatial effect parameter \( \lambda \), and autoregressive coefficient \( \alpha \). However, both stationary and nonstationary cases of \( \{ y_t \} \) 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*) 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 \( x_{ti} \) (or \( z_{sit} \)) is assumed to be either constant or random variable. Here, we assume time-variant random \( z_{sit} \) but time-invariant constant \( x_{ti} \) to emphasize the randomness by time \( t \). This also allows \( x_{ti} \) to be arbitrarily related to \( z_{sit} \). Assumption 3.6(iii) guarantees that the variances associated with \( \{ z_{sit}, 1 \leq i \leq N, 1 \leq t \leq T \} \) are bounded; see Lemmas 4, 5. In particularly, (3.6) implies that given \( (i_1, t_1) \) and \( (i_2, t_2) \), \( z_{sit_1}, z_{sit_2} \) is correlated with \( z_{sit_3}, z_{sit_4} \) 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 \( \zeta = (\phi', \sigma_z^2, \delta')' \) and \( \bar{\zeta} = (\bar{\phi}'(\bar{\delta}), \bar{\sigma}_z^2(\bar{\delta}), \bar{\delta}')' \). To show the consistency of the QMLE \( \bar{\zeta} \), we first verify that \( \zeta_0 = (\phi_0', \sigma_z^2_0, \delta_0')' \) is identifiable. For the quasi-log-likelihood function at (3.2), given \( \delta \), its expectation \( E[\ln L_{NT}(\zeta)] \) is maximized at

\[
\bar{\phi}(\delta) = \left[ E \left( \tilde{Z}'\Omega_{NT}^{-1}(\beta^*)\tilde{Z} \right) \right]^{-1} E \left( \tilde{Z}'\Omega_{NT}^{-1}(\beta^*)S_{NT}(\lambda)Y \right),
\]  
(3.7)
and

\[ \hat{\sigma}^2_{\varepsilon}(\delta) = \frac{1}{NT} E \left( \tilde{V}'_{NT}(\delta) \Omega^{-1}_{NT}(\beta^*) \tilde{V}_{NT}(\delta) \right), \] (3.8)

where \( \tilde{V}_{NT}(\delta) = S_{NT}(\lambda) Y - \tilde{Z} \phi(\delta) \). Define \( \ln \ell^*_N(\delta) = \max E[\ln \mathcal{L}_{NT}(\zeta)] \). Then, by (3.7) and (3.8), we have

\[ \ln \ell^*_N(\delta) = -\frac{NT}{2} (\ln (2\pi) + 1) - \frac{NT}{2} \ln \hat{\sigma}^2_{\varepsilon}(\delta) - \frac{1}{2} \ln |\Omega_{NT}(\beta^*)| + \ln |S_{NT}(\lambda)|. \]

Denote \( \Omega_{0NT} = \Omega_{NT}(\beta^*_0) \). Because \( E \left( \tilde{Z}' \Omega^{-1}_{0NT} \tilde{V}_{NT} \right) = 0 \), by Lemma 8 in the Supplementary Material, it can be shown that \( \tilde{\phi}(\delta_0) = \phi_0 + E(\tilde{Z}' \Omega^{-1}_{0NT} \tilde{Z})^{-1} E(\tilde{Z}' \Omega^{-1}_{0NT} \tilde{V}_{NT}) = \phi_0 \). Then, it follows that \( \tilde{V}_{NT}(\delta_0) = V_{NT} \) and \( \hat{\sigma}^2_{\varepsilon}(\delta_0) = \sigma^2_{\varepsilon_0} \). The identification of \( \delta_0 \) can be based on the maximum values of \( \frac{1}{NT} \ln \ell^*_N(\delta) \), and hence that of \( \phi_0 \) and \( \sigma^2_{\varepsilon_0} \) follows. The following condition is imposed for identification.

**Assumption 3.7.** For any \( \delta \neq \delta_0 \), it holds that

\[ \lim_{N \to \infty} \frac{1}{NT} \left[ \ln |\hat{\sigma}^{-2}_{\varepsilon}(\delta)\Omega^{-1}_{NT}(\beta^*)S_{NT}(\lambda)| - \ln |\sigma^{-2}_{\varepsilon_0}\Omega^{-1}_{0NT}S_{0NT}| \right] \neq 0. \]

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

**Theorem 1.** Under Assumptions 3.4 and 3.7, \( \hat{\zeta} \to_p \zeta_0 \) as \( N \to \infty \).

Let \( \Sigma = \lim_{N \to \infty} \Sigma_N \) and \( \Sigma^* = \lim_{N \to \infty} \Sigma'_N \), where \( \Sigma_N = -E \left( \frac{1}{\sqrt{NT}} \frac{\partial^2 \ln \mathcal{L}_{NT}(\zeta_0)}{\partial \zeta \partial \zeta'} \right) \) and

\[ \Sigma'_N = E \left( \frac{1}{\sqrt{NT}} \frac{\partial \ln \mathcal{L}_{NT}(\zeta_0)}{\partial \zeta} \frac{1}{\sqrt{NT}} \frac{\partial \ln \mathcal{L}_{NT}(\zeta_0)}{\partial \zeta'} \right) + E \left( \frac{1}{NT} \frac{\partial^2 \ln \mathcal{L}_{NT}(\zeta_0)}{\partial \zeta \partial \zeta'} \right). \]
By the consistency result given in Theorem 1, the asymptotic normality of \( \hat{\zeta} \) 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 \( \Sigma \) and \( \Sigma^* \) exist and \( \Sigma \) is positive definite, then \( \sqrt{NT}(\hat{\zeta} - \zeta_0) \rightarrow_d N(0, \Sigma^{-1} + \Sigma^{-1}\Sigma^*\Sigma^{-1}) \) as \( N \rightarrow \infty \).

Theorem 2 provides the asymptotic distribution of the QMLE \( \hat{\zeta} \), where the matrices \( \Sigma_N \) and \( \Sigma_N^* \) are relegated to Section S3.1 of the Supplementary Material. If \( \varepsilon_i \) and \( \eta_i \) are normally distributed, then the QMLE reduces to the MLE, and its limiting distribution can be simplified to \( \sqrt{NT}(\hat{\zeta} - \zeta_0) \rightarrow_d N(0, \Sigma^{-1}) \) as \( N \rightarrow \infty \). 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 \( \hat{\zeta} \) in practice, consistent estimates of \( \Sigma \) and \( \Sigma^* \) are necessary. Obviously, the matrix \( \hat{\Sigma} = \lim_{N \rightarrow \infty} -E\left( \frac{1}{NT} \frac{\partial^2 \ln L_N(T)(\zeta_0)}{\partial \zeta \partial \zeta'} \right) \) can be consistently estimated by \( \hat{\Sigma} = -\frac{1}{NT} \frac{\partial^2 \ln L_N(T)(\hat{\zeta})}{\partial \zeta \partial \zeta'} \). However, a closed-form estimate of the matrix

\[
\Gamma = \Sigma^* + \Sigma = \lim_{N \rightarrow \infty} E\left( \frac{1}{\sqrt{NT}} \frac{\partial \ln L_N(T)(\zeta_0)}{\partial \zeta} \frac{1}{\sqrt{NT}} \frac{\partial \ln L_N(T)(\zeta_0)}{\partial \zeta'} \right)
\]

(3.9)

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

To estimate \( \beta_0 = (\beta_{00}, \beta_{10}, \ldots, \beta_{p0})' \) in model (2.2), the constraint \( \beta_{00} \equiv 1 \) should be imposed for the identification of \( \beta_0 \), as in Koenker and Zhao (1994). Note that \( \beta_0^* = \beta_0\sigma_{\eta_0}/\sigma_{\varepsilon_0} \). Thus, we have \( \beta_{\ell 0} = \beta_{\ell 0}/\beta_{00} = \beta_{\ell 0}^*/\beta_{00}^* \), for \( \ell = 1, \ldots, p \). Then, based on
3.2 Conditional Quantile Estimation

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

3.2.1 Weighted conditional quantile estimation

Because the scale effect $\vartheta_i$ in model (3.1) is unobservable, to study its conditional quantiles, we first obtain its observable approximate using the QMLE $\hat{\zeta}$ from the first stage. Replacing $\zeta$ by $\hat{\zeta}$ in model (2.3), one can get $T$ random approximates of $\vartheta_i$ 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 $\vartheta_i$. Therefore, the pseudo scale effect, denoted
by \( \hat{\vartheta} \), that is, the approximation of \( \vartheta \), is given by

\[
\hat{\vartheta} = (\hat{\vartheta}_1, \ldots, \hat{\vartheta}_N)' = \frac{1}{T} \sum_{t=1}^{T} \left[ B_N(\hat{\lambda}) y_t - \hat{\alpha} y_{t-1} - Z_t \hat{\gamma} - X \hat{\psi} \right].
\] (3.11)

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

\[
\varphi(\tau) = (\varphi_0(\tau), \varphi_1(\tau), \ldots, \varphi_p(\tau))' = \arg\min_{\varphi \in \Phi} \sum_{i=1}^{N} \frac{1}{x' a_i \tilde{\beta}_c} \rho_\tau(\vartheta_i - x'_i \varphi), \quad \tau \in T,
\] (3.12)

where \( \Phi \) is the parameter space of \( \varphi(\tau) \), \( T \) is a closed subinterval of \((0, 1)\), \( \rho_\tau(u) = u(\tau - I(u < 0)) \) is the check function, and \( \tilde{\beta}_c \) is a consistent estimator of \( \beta_0 \) such that \( x'_a i \tilde{\beta}_c \) 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 \( \tilde{\beta}_c \), see Remark 3.

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

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

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

Theorem 3 plays an important role in verifying the validity of the proposed hybrid estimation procedure. Specifically, Theorem 3(i) guarantees that replacing \( \vartheta \) with \( \hat{\vartheta} \) does not affect the theoretical properties of the WCQE \( \varphi(\tau) \), 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 \( \varphi(\tau) \); see equation (S5.16) of the Supplementary Material.

Denote the true value of \( \varphi(\tau) \) as \( \varphi_0(\tau) = (\varphi_{00}(\tau), \varphi_{10}(\tau), \ldots, \varphi_{p0}(\tau))' \). Assume that the parameter space \( \Phi \) is compact and \( \varphi_0(\tau) \) is its interior point; see also Assumption R2 of
3.2 Conditional Quantile Estimation

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

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

**Assumption 3.9.** For the function \( \Pi_i(\tau, \varphi, r_i) = E[h_1 x_i' a_i \beta_0 (\tau - I(\vartheta_i < x_i' \varphi - r_i)) x_i] \) with \( (\tau, \varphi, r_i) \in T \times \Phi \times \mathbb{R} \), the Jacobian matrices \( \frac{\partial}{\partial \varphi} \Pi_i(\tau, \varphi, r_i) \) and \( \frac{\partial}{\partial r_i} \Pi_i(\tau, \varphi, r_i) \) for each \( 1 \leq i \leq N \) are continuous and have full rank uniformly over \( T \times \Phi \times \mathbb{R} \).

Assumption 3.8 is needed to establish the uniform convergence of \( b_{\varphi}(\tau) \). This, together with Assumption 3.9, is used for the weak convergence of \( b_{\varphi}(\tau) \); 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, \( \sup_{\tau \in T} \| \hat{\varphi}(\tau) - \varphi_0(\tau) \| \to_p 0 \) as \( N \to \infty \).

Theorems 1, 3, and 4 ensure that the conditional quantile of the random effect \( \theta \) can be consistently estimated by

\[
\hat{Q}_{\theta}(\tau|X) = X \hat{\psi} + X_a \hat{\varphi}(\tau).
\]

(3.13)

Denote the density function of \( \eta_i \) 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

\[
D_N = \frac{1}{N} \sum_{i=1}^{N} \frac{x_{ai} x_{ai}'}{(x_{ai}' \beta_0)^2}, \quad D_0 = \lim_{N \to \infty} D_N \quad \text{and} \quad D_{1N}(\tau) = \Xi(\tau) D_N.
\]

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

\[
\sqrt{N} \left( \hat{\varphi}(\cdot) - \varphi_0(\cdot) \right) = D_{1N}^{-1}(\cdot) \mathbb{G}_N \left( \frac{x_{ai} x_{ai}' \Psi(\eta_i - Q_{\eta}(\cdot))}{x_{ai}' \beta_0} \right) + o_p(1) \sim \mathbb{G}(\cdot) \text{ in } \ell^\infty(T),
\]

19
where $\Psi_\tau(u) = \tau - I(u < 0)$ and $G(\cdot)$ 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 $\hat{b}(\cdot)$. For any fixed quantile level $\tau$, by Theorem 5, it follows that, as $N \to \infty$,

$$\sqrt{N}(\hat{b}(\tau) - \phi_0(\tau)) \to_d N(0, \tau(1 - \tau)\Xi^{-2}(\tau)D_0^{-1}). \tag{3.14}$$

To calculate the covariance matrix of $\hat{b}(\tau)$ in practice, consistent estimators of $\Xi(\tau) = f_\eta(Q_\eta(\tau))$ and $D_0$ are required. Define the error function $\hat{b}_i(\beta) = \hat{b}_i / (x_i'\beta)$. Then the residuals can be computed as $\hat{\eta}_i = \hat{b}_i(\hat{\beta}_e)$, where $\hat{\beta}_e$ is a consistent estimate of $\beta_0$, as in (3.12). The density function $f_\eta(\cdot)$ can be estimated using the kernel density estimator $\hat{f}_\eta(x) = (Nh)^{-1} \sum_{i=1}^N K((x - \hat{\eta}_i)/h)$, 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, $h = 0.9N^{-1/5} \min\{s, \hat{R}/1.34\}$, where $s$ and $\hat{R}$ 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 $\hat{f}_\eta(Q_\eta(\tau))$, where $Q_\eta(\tau)$ is the sample $\tau$th quantile of $\{\hat{\eta}_i\}_{i=1}^N$. Moreover, $D_0$ can be approximated by $D_N$, with $\beta_0$ replaced by $\hat{\beta}_e$, that is, $\hat{D}_N = \frac{1}{N} \sum_{i=1}^N \frac{x_i'x_i}{(x_i'\hat{\beta}_e)^2}$. Consequently, the asymptotic covariance matrix of $\hat{b}(\tau)$ can be estimated by plugging in $\Xi(\tau) = \hat{f}_\eta(Q_\eta(\tau))$ and $\hat{D}_N$ to replace $\Xi(\tau)$ and $D_0$, respectively.

**Remark 1.** (Discussion on the estimation of $\psi$) As the conditional mean coefficient of the random effects, $\psi$ can be estimated using either the quasi-maximum likelihood estimation in the first stage or the conditional quantile estimation for model $Q_\theta(\tau|X) = X\psi + X_a\varphi(\tau)$ in the second stage. However, comparing Theorems 3 and 5, the QMLE $\hat{\psi}$ of $\psi$ has a faster convergence rate of $\sqrt{NT}$ than that of $\sqrt{N}$ for the quantile estimator obtained using
3.2 Conditional Quantile Estimation

model $Q_\theta(\tau|X) = X\psi + X_\alpha\varphi(\tau)$. This is another reason why the proposed two-stage hybrid estimation procedure does not consider a quantile estimation for both $\psi$ and $\varphi(\tau)$ simultaneously.

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

Stage 1': Estimate $(\alpha, \gamma', \sigma^2_e, \delta)'$ by the QMLE using the transformation approach or first-differencing to eliminate individual effect $\theta_i$ (Lee and Yu, 2010; Yu et al., 2008).

Stage 2': Define the pseudo random effect $\tilde{\theta}$ similarly as in (3.11). Then, apply the WCQE to model (2.4), with $\theta$ replaced by its approximation $\tilde{\theta}$. Furthermore, $\tau$-independent $\psi$ and $\beta$ can be estimated by $\tilde{\psi}$ and $\tilde{\beta}$, 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 $\theta$ in (2.2), which makes the estimation less efficient. Moreover, $\tilde{\psi}$ obtained using a quantile regression in Stage 2' also results in an efficiency loss, as discussed in Remark 3. Thus, we prefer the two-stage hybrid estimation procedure proposed in Section 3.

Remark 3. (Choice of $\tilde{\beta}_c$ in (3.12) for the WCQE) To conduct the weighted quantile regression in (3.12), we employ $\beta$ in (3.10) as $\tilde{\beta}_c$. However, $\tilde{\beta}$ is not stable in finite-sample cases, so we construct an alternative choice for $\tilde{\beta}_c$. Note that $\varphi_0(\tau) = Q_\eta(\tau)\beta_0$. Thus, we can construct $\tilde{\beta}_c$ by combining the information of the estimator for $\varphi_0(\tau)$ over multiple quantile levels. Furthermore, $\varphi_0(\tau)$ and $Q_\eta(\tau)$ have the same sign. Thus, as an alternative, we consider $\sum_{k=1}^K |\varphi_0(\tau_k)| = \beta_0 \sum_{k=1}^K |Q_\eta(\tau_k)|$ 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 $\beta_0$ and $Q_\eta(\tau)$. As a result, we suggest the following choice for $\tilde{\beta}_c$ in (3.12):

$$\tilde{\beta}_c = \frac{\sum_{k=1}^{K} |\tilde{\varphi}(\tau_k)|}{\sum_{k=1}^{K} |\tilde{Q}_\eta(\tau_k)|},$$  \hspace{1cm} (3.15)

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

### 3.2.2 Weighted quantile average estimation

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

Recall that $\beta_0 = (\beta_{00}, \beta_{10}, \ldots, \beta_{p0})'$ is the true value of $\beta$, and $\varphi_0(\tau) = Q_\eta(\tau)\beta_0$ with $\varphi_{\ell 0}(\tau) = Q_\eta(\tau)\beta_{\ell 0}$ holds for $\ell = 0, 1, \ldots, p$. Note that for any weight vector $\pi_K = (\pi_1, \ldots, \pi_K)'$ satisfying $\sum_{k=1}^{K} \pi_k Q_\eta(\tau_k) = 1$, it holds that $\beta_0 = \sum_{k=1}^{K} \pi_k \varphi_{0}(\tau_k)$ under the constraint $\beta_{00} \equiv 1$. This motivates us to combine $\{\tilde{\varphi}(\tau_k), 1 \leq k \leq K\}$ linearly to define the WQAE of $\beta$, as follows:

$$\hat{\beta}(\pi_K) = \sum_{k=1}^{K} \pi_k \tilde{\varphi}(\tau_k) \quad \text{with} \quad \sum_{k=1}^{K} \pi_k Q_\eta(\tau_k) = 1;$$  \hspace{1cm} (3.16)

see also Zhao and Xiao (2014).
Denote $\mathcal{B}$ as the parameter space of $\beta$, and assume $\mathcal{B}$ is compact and $\beta_0$ 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 $\hat{\beta}(\pi_K)$ and optimal choice for $\pi_K$ are provided below.

**Theorem 6.** Suppose the conditions of Theorem 3 hold. Then as $N \to \infty$,

$$
\sqrt{N} \left( \hat{\beta}(\pi_K) - \beta_0 \right) \to_d N(0, \Upsilon(\pi_K)),
$$

where $\Upsilon(\pi_K) = \pi_K' KH \pi_K D_0^{-1}$. Moreover, the optimal weight is

$$
\pi_{opt,K} = \arg\min_{\pi, s.t. \pi' q = 1} \Upsilon(\pi_K) = \frac{H^{-1} q}{q' H^{-1} q},
$$

(3.17)

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

To estimate the optimal weight $\pi_{opt,K}$, consistent estimators of $q$ and $H$ are necessary. We can approximate $q$ by $\tilde{q} = (\tilde{Q}_\eta(\tau_1), \ldots, \tilde{Q}_\eta(\tau_K))'$, with $\tilde{Q}_\eta(\tau) = \tilde{\varphi}_0(\tau)$, and estimate $H$ by $\tilde{H} = \{\tilde{h}_{ij}, 1 \leq i, j \leq K\}$, with $\tilde{h}_{ij} = \Xi^{-1}(\tau_i)S(\tau_i, \tau_j)\Xi^{-1}(\tau_j)$ and $\Xi(\tau) = \tilde{f}_\eta(\tilde{Q}_\eta(\tau))$. Here, using $\tilde{Q}_\eta(\tau)$ instead of the empirical quantile $\hat{Q}_\eta(\tau)$ to estimate $q$ ensures that the estimate of $\beta_{00}$ is one. Then, a consistent estimator $\tilde{\pi}_{opt,K} = (\tilde{\pi}_{1, opt}, \ldots, \tilde{\pi}_{K, opt})'$ of $\pi_{opt,K} = (\pi_{1, opt}, \ldots, \pi_{K, opt})'$ can be obtained by plugging in $\tilde{H}$ and $\tilde{q}$. Finally, the optimal WQAE can be calculated as $\hat{\beta}(\tilde{\pi}_{opt,K}) = \sum_{k=1}^K \tilde{\pi}_{k, opt} \tilde{\varphi}(\tau_k)$. It can be shown that $\sqrt{N}(\hat{\beta}(\tilde{\pi}_{opt,K}) - \beta_0) \to_d N(0, \Upsilon(\pi_{opt,K}))$ as $N \to \infty$, which implies that $\hat{\beta}(\tilde{\pi}_{opt,K})$ and $\hat{\beta}(\pi_{opt,K})$ have the same asymptotic efficiency.

**Remark 4.** (Comparison of $\beta$, $\tilde{\beta}_c$, and $\hat{\beta}(\pi_{opt,K})$) $\beta$ and $\hat{\beta}(\pi_{opt,K})$ 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
Heterogeneous DSAR Model

\( \hat{\beta}(\pi_{opt,K}) \) outperforms \( \hat{\beta} \) 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 \( \hat{\beta}(\pi_{opt,K}) \) and \( \hat{\beta} \) are \( \sqrt{N} \) and \( \sqrt{NT} \), respectively. Thus, \( \hat{\beta} \) can be more efficient than \( \hat{\beta}(\pi_{opt,K}) \) for large samples. We also compare the finite-sample performance of \( \hat{\beta}(\pi_{opt,K}) \) and \( \hat{\beta}_c \), with the results showing that \( \hat{\beta}(\pi_{opt,K}) \) has smaller biases and empirical standard deviations than \( \hat{\beta}_c \). Therefore, we suggest improving the efficiency by using \( \hat{\beta}(\pi_{opt,K}) \) when the sample size is small or moderate, whereas \( \hat{\beta} \) 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 \( \hat{\zeta} \), the WCQE \( \hat{\varphi}(\tau) \), and the WQAE \( \hat{\beta}(\pi_{opt,K}) \).

The data are generated from the following model:

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

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 $\mathbf{\zeta}_0 = (\alpha_0, \gamma_{10}, \psi_{00}, \psi_{10}, \sigma^2_{\varepsilon 0}, \lambda_0, \beta_{00}^*, \beta_{10}^*)'$, $\mathbf{\varphi}_0(\tau) = (\varphi_{00}(\tau), \varphi_{10}(\tau))'$, and $\beta_{10}$ using the QMLE $\hat{\mathbf{\zeta}}$, WCQE $\hat{\mathbf{\varphi}}(\tau)$, and WQAE $\hat{\beta}_1(\hat{\pi}_{opt,K})$, respectively, where $(\beta_{00}^*, \beta_{10}^*) = (1, \beta_{10})\sigma_{\varepsilon 0}/\sigma_{\varepsilon 0}$ and $\mathbf{\varphi}_0(\tau) = (1, \beta_{10})Q_\eta(\tau)$. 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 $\hat{\mathbf{\varphi}}(\tau)$, owing to $\mathbf{\varphi}_0(\tau) = \beta_0 Q_\eta(\tau)$ and $\beta_0 \geq 0$. For the WQAE, we employ $K = 9$ quantile levels $\tau_k = k/10$, with $k = 1, \ldots, 9$, and the resulting $\hat{\beta}_1(\hat{\pi}_{opt,K})$ is almost nonnegative under the same sign restriction for $\hat{\mathbf{\varphi}}(\tau)$.

Table 1 reports the biases, empirical standard deviations (ESDs), and asymptotic standard deviations (ASDs) of the QMLE $\hat{\mathbf{\zeta}}$ at Stage 1, where the ASDs are calculated using a residual-based bootstrap with bootstrap sample size $B = 500$. Recall that we require $N \rightarrow \infty$ 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 $\eta_i$ are mostly smaller than those for the uniform distribution. This is expected, because the QMLE reduces to the MLE when $\eta_i$ is normally distributed.

Table 2 summarizes the biases, ESDs, and ASDs of the WCQE $\hat{\mathbf{\varphi}}(\tau)$ at Stage 2 for three
quantile levels, $\tau = 0.25, 0.5, \text{ and } 0.75$, 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 $\{\eta_t\}$, 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 $\hat{\varphi}(\tau)$ improves as the quantile level $\tau$ 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 $\hat{\beta}_1(\pi_{opt,K})$ at Stage 2 for three sample settings $(N, T) = (20, 20), (100, 50),$ and $(300, 200)$. For comparison, the results of $\tilde{\beta}_1$ in (3.10) and $\tilde{\beta}_{1c}$ in (3.15) are also reported, where the extra setting $(N, T) = (50, 20)$ is added to show the performance of $\tilde{\beta}_1$ 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 $\hat{\beta}_1(\pi_{opt,K})$ move closer to their corresponding ASDs for larger $N$; see Table S.2 in the Supplementary Material. In addition, the estimate $\tilde{\beta}_1$ is not satisfactory, because the the sample size is not large, and $\tilde{\beta}_{1c}$ is usually less efficient than $\hat{\beta}_1(\pi_{opt,K})$ 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 $\{y_{it}\}$, $\{z_{1it}\}$, $\{z_{2it}\}$, $\{x_{1it}\}$, 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

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

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 $i$th city is adjacent to the $j$th 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 4, 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 $\theta_i$, 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 $\theta_i$ in model (2.4) are

\[
\begin{align*}
\hat{Q}_{\theta_i}(0.25|\mathbf{x}_i) &= (0.171 + 0.078 x_{1i} + 0.063 x_{2i}) + (-0.066 - 0.016 |x_{1i}| - 0.046 |x_{2i}|), \\
\hat{Q}_{\theta_i}(0.75|\mathbf{x}_i) &= (0.171 + 0.078 x_{1i} + 0.063 x_{2i}) + (0.030 + 0.021 |x_{1i}| + 0.014 |x_{2i}|),
\end{align*}
\]

where the summary information of the fitted coefficients is reported in Table 3. For each equation of (5.2), the first set of parentheses corresponds to the estimated conditional location of $\theta_i$, and the second represents the conditional quantile estimate of the scale effect. Because the conditional location of $\theta_i$ is actually the conditional expectation of $\theta_i$, the fitted conditional expectation of $\theta_i$ is $\hat{E}(\theta_i|\mathbf{x}_i) = 0.171 + 0.078 x_{1i} + 0.063 x_{2i}$. 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 3, we have the following conclusions: (i) both the secondary industry share and GRP may have positive impacts on the conditional mean of $\theta_i$ 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 environment diminishes gradually; (iii) the secondary industry share has a statistically significant positive impact on the conditional location of $\theta_i$ 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 $\theta_i$ at the lower quantile level, while no statistically significant effect is observed at the higher level.
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 8 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):

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

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 \(\Sigma_N, \Sigma_N^*, \Omega^{-1}_{NT}(\beta^*)\), and the vector \(Y_{-1}\), 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


REFERENCES


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


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
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 $\hat{\zeta}$ when the innovations $\{\eta_i\}$ follow a standard normal or uniform distribution.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>Bias</th>
<th>ESD</th>
<th>ASD</th>
<th>$N(0,1)$</th>
<th>Bias</th>
<th>ESD</th>
<th>ASD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>20</td>
<td>0.0021</td>
<td>0.0234</td>
<td>0.0398</td>
<td>0.0029</td>
<td>0.0232</td>
<td>0.0341</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.0007</td>
<td>0.0146</td>
<td>0.0400</td>
<td>0.0008</td>
<td>0.0141</td>
<td>0.0332</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0006</td>
<td>0.0080</td>
<td>0.0089</td>
<td>0.0006</td>
<td>0.0079</td>
<td>0.0088</td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>20</td>
<td>0.0020</td>
<td>0.0511</td>
<td>0.0561</td>
<td>0.0013</td>
<td>0.0516</td>
<td>0.0571</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.0019</td>
<td>0.0336</td>
<td>0.0381</td>
<td>0.0026</td>
<td>0.0318</td>
<td>0.0371</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0002</td>
<td>0.0145</td>
<td>0.0147</td>
<td>0.0005</td>
<td>0.0144</td>
<td>0.0147</td>
<td></td>
</tr>
<tr>
<td>$\psi_0$</td>
<td>20</td>
<td>0.0089</td>
<td>0.3737</td>
<td>1.2303</td>
<td>0.0091</td>
<td>0.3930</td>
<td>0.8101</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.0292</td>
<td>0.3561</td>
<td>1.8826</td>
<td>0.0201</td>
<td>0.3966</td>
<td>2.1397</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0040</td>
<td>0.1537</td>
<td>0.1609</td>
<td>0.0045</td>
<td>0.1519</td>
<td>0.1596</td>
<td></td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>20</td>
<td>0.0404</td>
<td>0.6821</td>
<td>1.2488</td>
<td>0.0161</td>
<td>0.6284</td>
<td>0.9061</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.0018</td>
<td>0.6537</td>
<td>4.9535</td>
<td>0.0149</td>
<td>0.6501</td>
<td>3.0061</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0072</td>
<td>0.2782</td>
<td>0.2731</td>
<td>0.0100</td>
<td>0.2742</td>
<td>0.2735</td>
<td></td>
</tr>
<tr>
<td>$\sigma_\varepsilon^2$</td>
<td>20</td>
<td>0.0039</td>
<td>0.0714</td>
<td>0.1266</td>
<td>0.0086</td>
<td>0.0752</td>
<td>0.1138</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.0013</td>
<td>0.0464</td>
<td>0.1113</td>
<td>0.0027</td>
<td>0.0449</td>
<td>0.0913</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0001</td>
<td>0.0201</td>
<td>0.0377</td>
<td>0.0003</td>
<td>0.0193</td>
<td>0.0366</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>20</td>
<td>0.0030</td>
<td>0.0237</td>
<td>0.0435</td>
<td>0.0037</td>
<td>0.0235</td>
<td>0.0375</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.0008</td>
<td>0.0147</td>
<td>0.0428</td>
<td>0.0010</td>
<td>0.0141</td>
<td>0.0346</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0006</td>
<td>0.0080</td>
<td>0.0089</td>
<td>0.0006</td>
<td>0.0079</td>
<td>0.0088</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>20</td>
<td>0.1805</td>
<td>0.5059</td>
<td>2.7118</td>
<td>0.1295</td>
<td>0.4365</td>
<td>1.6301</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.1834</td>
<td>0.4967</td>
<td>7.5333</td>
<td>0.1041</td>
<td>0.4295</td>
<td>4.2356</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0198</td>
<td>0.2087</td>
<td>0.2581</td>
<td>0.0177</td>
<td>0.1360</td>
<td>0.1718</td>
<td></td>
</tr>
<tr>
<td>$\beta_1^*$</td>
<td>20</td>
<td>0.1256</td>
<td>0.9422</td>
<td>5.7388</td>
<td>0.0584</td>
<td>0.7995</td>
<td>3.0798</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20 50</td>
<td>0.1419</td>
<td>0.9644</td>
<td>15.7608</td>
<td>0.0312</td>
<td>0.8020</td>
<td>9.5194</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100 50</td>
<td>0.0015</td>
<td>0.4065</td>
<td>0.4525</td>
<td>0.0043</td>
<td>0.2630</td>
<td>0.2991</td>
<td></td>
</tr>
</tbody>
</table>
Heterogeneous DSAR Model

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

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$\varphi_0$</th>
<th>$\varphi_1$</th>
<th>$N(0,1)$</th>
<th>$U(-\sqrt{3}, \sqrt{3})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>20</td>
<td>0.0303 0.4880 0.9059</td>
<td>0.0913 0.7667 1.6977</td>
<td>0.1333 0.5239 0.9455</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.0219 0.3287 0.3713</td>
<td>0.0585 0.5937 0.7213</td>
<td>0.0479 0.3667 0.3934</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>200</td>
<td>0.0081 0.1935 0.1970</td>
<td>0.0225 0.3896 0.4091</td>
<td>0.0144 0.2116 0.2101</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.0164 0.2520 0.8681</td>
<td>0.0143 0.3411 1.6275</td>
<td>0.0037 0.2654 1.0016</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.0038 0.1092 0.3529</td>
<td>0.0104 0.1718 0.6852</td>
<td>0.0035 0.1334 0.4541</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>200</td>
<td>0.0045 0.0589 0.1851</td>
<td>0.0052 0.1125 0.3844</td>
<td>0.0041 0.0750 0.2427</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.0143 0.3411 1.6275</td>
<td>0.0143 0.3411 1.6275</td>
<td>0.0253 0.4375 1.8757</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.0104 0.1718 0.6852</td>
<td>0.0104 0.1718 0.6852</td>
<td>0.0028 0.2384 0.8802</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>200</td>
<td>0.0052 0.1125 0.3844</td>
<td>0.0052 0.1125 0.3844</td>
<td>0.0015 0.1558 0.5042</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.0336 0.4843 0.8838</td>
<td>0.0864 0.7897 1.6561</td>
<td>0.1093 0.5027 0.9572</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.0141 0.3197 0.3737</td>
<td>0.0300 0.5683 0.7254</td>
<td>0.0511 0.3666 0.3981</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>200</td>
<td>0.0082 0.1913 0.1960</td>
<td>0.0082 0.1913 0.1960</td>
<td>0.0063 0.2080 0.2101</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>0.0864 0.7897 1.6561</td>
<td>0.0118 0.8468 1.7927</td>
<td>0.0119 0.4215 0.4368</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Biases, ESDs, and ASDs of three estimators of $\beta_0$, that is, $\hat{\beta}$ by the QMLE, the initial estimator $\tilde{\beta}_c$, and the WQAE $\tilde{\beta}(\hat{\pi}_{opt,K})$, when the innovations $\{\eta_i\}$ follow a standard normal or uniform distribution.

<table>
<thead>
<tr>
<th>$N$</th>
<th>$T$</th>
<th>$\beta_1$</th>
<th>$\beta_{1c}$</th>
<th>$\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Bias ESD ASD</td>
<td>Bias ESD ASD</td>
<td>Bias ESD ASD</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>276805 827496 $5e+12$</td>
<td>146686 609974 $3e+12$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>20</td>
<td>3181 100581 $4e+9$</td>
<td>3281 103752 $4e+9$</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>0.1638 0.7819 0.8772</td>
<td>0.0779 0.4386 0.5052</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>200</td>
<td>0.0340 0.3032 0.3548</td>
<td>0.0167 0.1942 0.2278</td>
<td></td>
</tr>
</tbody>
</table>

$* e + 12$ 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 (%).

<table>
<thead>
<tr>
<th></th>
<th>AQI</th>
<th>TEM</th>
<th>PRE</th>
<th>WIN</th>
<th>GRP</th>
<th>Industry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>13.264</td>
<td>-283.000</td>
<td>0.000</td>
<td>4.714</td>
<td>1500100</td>
<td>18.270</td>
</tr>
<tr>
<td>Max</td>
<td>312.845</td>
<td>325.714</td>
<td>5113.000</td>
<td>78.286</td>
<td>326798700</td>
<td>72.900</td>
</tr>
</tbody>
</table>
Table 5: Summary information of fitted coefficients for models (5.1) and (5.2).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.178</td>
<td>0.027</td>
<td>6.593</td>
<td>0.000</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.648</td>
<td>0.022</td>
<td>29.455</td>
<td>0.000</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.096</td>
<td>0.015</td>
<td>-6.400</td>
<td>0.000</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.065</td>
<td>0.016</td>
<td>-10.833</td>
<td>0.000</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.078</td>
<td>0.013</td>
<td>-6.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\psi_0$</td>
<td>0.171</td>
<td>0.194</td>
<td>0.881</td>
<td>0.378</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>0.078</td>
<td>0.088</td>
<td>0.886</td>
<td>0.375</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>0.063</td>
<td>0.036</td>
<td>1.750</td>
<td>0.080</td>
</tr>
<tr>
<td>$\varphi_0(0.25)$</td>
<td>-0.066</td>
<td>0.103</td>
<td>-0.641</td>
<td>0.522</td>
</tr>
<tr>
<td>$\varphi_1(0.25)$</td>
<td>-0.016</td>
<td>0.028</td>
<td>-0.571</td>
<td>0.568</td>
</tr>
<tr>
<td>$\varphi_2(0.25)$</td>
<td>-0.046</td>
<td>0.024</td>
<td>-1.917</td>
<td>0.055</td>
</tr>
<tr>
<td>$\varphi_0(0.75)$</td>
<td>0.030</td>
<td>0.216</td>
<td>0.139</td>
<td>0.890</td>
</tr>
<tr>
<td>$\varphi_1(0.75)$</td>
<td>0.021</td>
<td>0.058</td>
<td>0.362</td>
<td>0.717</td>
</tr>
<tr>
<td>$\varphi_2(0.75)$</td>
<td>0.014</td>
<td>0.049</td>
<td>0.286</td>
<td>0.775</td>
</tr>
</tbody>
</table>