Journal of Business & Economic Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ubes20

Autoregressive Model With Spatial Dependence and Missing Data

Jing Zhou, Jin Liu, Feifei Wang & Hansheng Wang

JOURNAL OF BUSINESS & ECONOMIC STATISTICS
2022, VOL. 40, NO. 1, 28–34
https://doi.org/10.1080/07350015.2020.1766471

Autoregressive Model With Spatial Dependence and Missing Data

Jing Zhou[a,b], Jin Liu[c], Feifei Wang[a,b], and Hansheng Wang[d]

[a]Center for Applied Statistics, Renmin University of China, Beijing, China; [b]School of Statistics, Renmin University of China, Beijing, China; [c]School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tian Jin, China; [d]Guanghua School of Management, Peking University, Beijing, China

ABSTRACT
We study herein an autoregressive model with spatially correlated error terms and missing data. A logistic regression model with completely observed covariates is used to model the missingness mechanism. An autoregressive model is used to accommodate time series dependence, and a spatial error model is used to capture spatial dependence. To estimate the model, a weighted least squares estimator is developed for the temporal component, and a weighted maximum likelihood estimator is developed for the spatial component. The asymptotic properties for both estimators are investigated. The finite sample performance is assessed through extensive simulation studies. A real data example about Beijing's $PM_{2.5}$ level data is illustrated.

ARTICLE HISTORY
Received May 2019
Accepted May 2020

KEYWORDS
Missing data; Spatial error model; Spatial-temporal dependence, Weighted least squares estimator; Weighted maximum likelihood estimator

1. Introduction

Time series data with spatial dependence are common in practice. Consider, for example, the $PM_{2.5}$ air quality data released by the Beijing Municipal Environmental Monitoring Center. The $PM_{2.5}$ readings are collected at different monitoring sites and different time points in Beijing. This is a dataset of profound importance for public health. Statistical modeling in this regard is challenging because two types of dependencies are involved, namely, time series dependence and spatial dependence. In addition, due to some practical reasons, a significant portion of the data could be missing, which makes the situation even more complicated.

It is noteworthy that even with complete data, determining how to statistically model data with both time series and spatial dependencies can be difficult. To accommodate time series dependence, various time series models have been developed, including autoregressive (AR) models, moving average models (Brockwell and Davis 1991; Hamilton 1994; Fuller 1996), and others. To accommodate spatial dependence, various spatial models have been developed, including spatial autoregressive models (Ord 1975; Anselin 1980; Lee, Liu, and Lin 2010), spatial error models (SEMs, LeSage and Pace 2009), and others. However, to simultaneously consider both types of dependencies remains a challenging problem. In this regard, various dynamic panel data models have been proposed. For example, a spatial dynamic panel data (SDPD) model with fixed individual effects was proposed by Yu, de Jong, and Lee (2008). Lee and Yu (2014) extended that model and proposed a SDPD model with both individual and time effects. Lee and Yu (2010) developed spatial autoregressive panel data models with fixed individual effects and spatial autoregressive disturbances. Su and Yang (2015) proposed dynamic panel data models with spatial errors. Recently, Yang (2018) developed SDPD models with spatial errors for short panels. More discussions can be found in the above studies and references therein. In addition, Dou, Parrella, and Yao (2016) and Gao et al. (2019) developed spatiotemporal models with unknown coefficients as well as diagonal or banded matrices. It is noteworthy that all of the above models were developed for complete data.

In this work, we propose a novel autoregressive model but with spatially correlated errors. Specifically, a standard autoregressive model is used to model the temporal effect. By doing so, each time series can be fitted separately in a parallel manner. The competitive advantage could be significant if the number of time series is large. To capture spatial dependence, the time series errors from different locations are allowed to be spatially dependent. The dependency is modeled by the widely used spatial autoregressive model. In the case of complete data, the model can be fitted by a quasi-maximum likelihood method (Yu, de Jong, and Lee 2008). However, establishing an approach to conduct parameter estimation in the presence of missing data is challenging. To fill this theoretical gap, a new method needs to be developed.

To solve this problem, we first need a probabilistic model to reflect the missingness mechanism. To this end, a standard logistic regression model is used. In theory, this model can be replaced by any parametric regression model with a binary response. The covariate used in the logistic regression is assumed to be completely observed. By doing so, we implicitly adopt the *missing at random* (MAR) assumption. In addition, two other popularly used assumptions are missing completely at random (MCAR) and nonignorable missing (NM). See Rubin (1976) for a more detailed discussion. In our case, the MAR assumption implies that conditional on the observed covariates, whether the response is missing or not is independent of the response itself. Under the MAR assumption, a weighted least squares estimator (WLSE) is proposed to estimate the

CONTACT Jin Liu liujin@nankai.edu.cn School of Statistics and Data Science, LPMC & KLMDASR, Nankai University, Tian Jin, 300071, China.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

autoregressive dependence. In addition, a weighted maximum likelihood estimator (WMLE) is developed to estimate the spatial dependence. Both estimators are shown to be consistent and asymptotically normal. Extensive simulation studies are presented to demonstrate their finite sample performance. Last, a real data example based on Beijing's $PM_{2.5}$ level data is demonstrated.

The rest of the article is organized as follows. Section 2 introduces the model and methodology. The WLSE and WMLE are developed. In Section 3, the asymptotic properties of the proposed estimators are presented. Extensive numerical studies and a real data example are given in Section 4. Finally, we conclude the article with a brief discussion in Section 5. All detailed techniques are relegated to the appendix in the supplementary materials.

2. Model and Methodology

2.1. Model Setup

Let $Y_{it} \in \mathbb{R}^1$ be a continuous response collected from the ith ($1 \le i \le N$) region at time point t ($1 \le t \le T$). For a given i, $\{Y_{it} : 1 \le t \le T\}$ constitutes a time series. To model this time series dependence, the following autoregressive model is considered:

$$Y_{it} = \alpha_i Y_{i(t-1)} + \varepsilon_{it}, \tag{2.1}$$

where α_i is the autocorrelation coefficient characterizing region i's temporal dependence, and ε_{it} is the error term. In addition, to guarantee the stationarity of the time series, we assume $\max_i |\alpha_i| < \delta < 1$ for some positive constant δ. It is noted that the intercept is omitted for simplicity, which implies that the $\{Y_{it}\}$ values need to be centralized before the formal data analysis. Next, define $\mathcal{E}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$. Assume $\mathcal{E}_t$ has a spatially dependent structure, which is characterized by the following SEM (see Ord 1975; Anselin 1980; LeSage and Pace 2009; Lee, Liu, and Lin 2010):

$$\mathcal{E}_t = \rho W \mathcal{E}_t + \epsilon_t, \tag{2.2}$$

where $\rho \in \mathbb{R}^1$ is the spatial autoregressive parameter satisfying $|\rho| < 1$ (Carlin, Gelfand, and Banerjee 2014), and $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})^\top \in \mathbb{R}^N$ is the residual vector with a mean of 0 and covariance matrix $\sigma^2 I \in \mathbb{R}^{N\times N}$. Here, I stands for an $N \times N$ identity matrix. In addition, we assume that $E(\epsilon_{it}^4) < \infty$ and that ϵ_t is independent and identically distributed over time t. The weight matrix W is the row-normalized adjacency matrix. One typical adjacency matrix is defined as $A = (a_{ij}) \in \mathbb{R}^{N\times N}$, where $a_{ij} = 1$ if region i is bordered by region j, and $a_{ij} = 0$ otherwise. Then, $W = (w_{ij}) \in \mathbb{R}^{N\times N}$ is defined as $w_{ij} = a_{ij}/d_i$, where $d_i = \sum_{j=1}^N a_{ij}$ is the total number of regions that border i. By (2.2), we know that $\mathcal{E}_t = (I - \rho W)^{-1}\epsilon_t$. As a result, $\mathcal{E}_t$ follows a distribution with a mean of 0 and covariance matrix $\Sigma_t = \Sigma = (\sigma_{ij}) = \sigma^2(I - \rho W)^{-1}(I - \rho W^\top)^{-1}$.

Define $\mathcal{F}$ as the σ-field generated by $\{(Y_{it}, X_{it}) : 1 \le i \le N, 1 \le t \le T\}$. To cope with the missing data problem, define a binary indicator $Z_{it} \in \{0, 1\}$ such that

$$P(Z_{it} = 1|\mathcal{F}) = \frac{\exp(\beta^\top X_{it})}{1 + \exp(\beta^\top X_{it})} = p_{it}, \tag{2.3}$$

where $Z_{it} = 1$ indicates that Y_{it} is observed, and $Z_{it} = 0$ indicates that Y_{it} is missing. Accordingly, the Z_{it} values are conditionally independent with each other on $\mathcal{F}$. The random variable $X_{it} = (X_{it,1}, \ldots, X_{it,p})^\top \in \mathbb{R}^p$ are p-dimensional covariates with no missing value, and $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is the associated regression coefficient vector. Conditional on $\mathcal{F}$, the Z_{it} values are assumed to be mutually independent.

2.2. Least Squares Estimation of Temporal Dependence

We first consider how to estimate the AR model (2.1) in the presence of missing data. A straightforward solution would be to consider the "complete data" only. Complete data refer to those that are completely observed pairs $(Y_{it}, Y_{i(t-1)})$. Accordingly, the least squares objective function can be constructed as $Q_1(\alpha_i) = \sum_{t=2}^T Z_{it}Z_{i(t-1)}(Y_{it} - \alpha_i Y_{i(t-1)})^2$. The corresponding least squares estimator (LSE) is given by the following:

$$\hat{\alpha}_i^{\text{LSE}} = \left\{\sum_{t=2}^T Z_{it}Z_{i(t-1)}Y_{it}Y_{i(t-1)}\right\}\left\{\sum_{t=2}^T Z_{it}Z_{i(t-1)}Y_{i(t-1)}^2\right\}^{-1}.$$

One can verify that $E\{Q_1(\alpha_i)|\mathcal{F}\} = \sum_{t=2}^T p_{it}p_{i(t-1)}(Y_{it} - \alpha_i Y_{i(t-1)})^2$, which suggests that different weights (e.g., p_{it}, $p_{i(t-1)}$) are expected for different sample pairs $(Y_{it}, Y_{i(t-1)})$. In other words, each sample pair is no longer treated equally in the estimation process. This might make $\hat{\alpha}_i^{\text{LSE}}$ less efficient. To fix this problem, we propose the following weighted least squares type of objective function as $Q_2(\alpha_i) = \sum_{t=2}^T (p_{it}p_{i(t-1)})^{-1}(Z_{it}Z_{i(t-1)})(Y_{it} - \alpha_i Y_{i(t-1)})^2$. This leads to the following WLSE as follows:

$$\tilde{\alpha}_i^{\text{WLSE}} = \left\{\sum_{t=2}^T \frac{Z_{it}Z_{i(t-1)}Y_{it}Y_{i(t-1)}}{p_{it}p_{i(t-1)}}\right\}\left\{\sum_{t=2}^T \frac{Z_{it}Z_{i(t-1)}Y_{i(t-1)}^2}{p_{it}p_{i(t-1)}}\right\}^{-1}.$$

The WLSE is an infeasible estimator because its computation involves the unknown parameter β and thus p_{it} values. To address the unknown parameter, we replace β in $\tilde{\alpha}_i^{\text{WLSE}}$ by its maximum likelihood estimator $\hat{\beta}$. This leads to the following feasible WLSE as follows:

$$\hat{\alpha}_i^{\text{WLSE}} = \left\{\sum_{t=2}^T \frac{Z_{it}Z_{i(t-1)}Y_{it}Y_{i(t-1)}}{\hat{p}_{it}\hat{p}_{i(t-1)}}\right\}\left\{\sum_{t=2}^T \frac{Z_{it}Z_{i(t-1)}Y_{i(t-1)}^2}{\hat{p}_{it}\hat{p}_{i(t-1)}}\right\}^{-1}.$$

Here, $\hat{p}_{it} = \exp(\hat{\beta}^\top X_{it})/\{1 + \exp(\hat{\beta}^\top X_{it})\}$, and $\hat{\beta}$ is defined as $\hat{\beta} = \text{argmax}_\beta \ell^\star(\beta)$, where $\ell^\star(\beta) = \sum_{i=1}^N \sum_{t=2}^T \Big[Z_{it}\beta^\top X_{it} - \log\{1 + \exp(\beta^\top X_{it})\}\Big]$.

2.3. Maximum Likelihood Estimation of Spatial Dependence

We next consider how to estimate spatial dependence in the model (2.2) in the presence of missing data. In this case, complete data refer to the completely observed residual pairs $(\varepsilon_{it}, \varepsilon_{jt})$. To this end, a log maximum likelihood objective function for $\theta = (\rho, \sigma^2)^\top$ omitting some constants can be constructed as follows:

$$\ell_0(\theta) = \ell_0(\rho, \sigma^2) = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\sum_{i=1,j=1}^{N} Z_{it}Z_{i(t-1)}Z_{jt}Z_{j(t-1)}\varepsilon_{it}\varepsilon_{jt}w_{ij}(\rho),$$

where $\Omega(\rho) = (I-\rho W^\top)(I-\rho W)$, and $w_{ij}(\rho)$ is the (i,j)th element in the matrix $\Omega(\rho)$. Similarly, one can verify the following equation:

$$E\{\ell_0(\rho,\sigma^2)|\mathcal{F}\} = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\Big\{\sum_{i\neq j} p_{it}p_{i(t-1)}p_{jt}p_{j(t-1)}\varepsilon_{it}\varepsilon_{jt}w_{ij}(\rho) + \sum_{i=1}^{N} p_{it}p_{i(t-1)}\varepsilon_{it}^2 w_{ii}(\rho)\Big\}.$$

It should be noted that $\varepsilon_{it} = Y_{it} - \alpha_i Y_{i(t-1)}$. Therefore, the expected weight for $\varepsilon_{it}\varepsilon_{jt}$ is $p_{it}p_{i(t-1)}p_{jt}p_{j(t-1)}$. As a result, each residual pair is no longer treated equally in the estimation process, which leads to a less efficient estimator. To fix the problem, we propose the following weighted log-likelihood type of objective function:

$$\ell_1(\theta) = \ell_1(\rho,\sigma^2) = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}f(\rho), \tag{2.4}$$

where $f(\rho)$ is given by the following:

$$\begin{aligned} f(\rho) &= \sum_{t=2}^{T}\Big\{\sum_{i\neq j}\frac{Z_{it}Z_{i(t-1)}Z_{jt}Z_{j(t-1)}}{p_{it}p_{i(t-1)}p_{jt}p_{j(t-1)}}\varepsilon_{it}\varepsilon_{jt}w_{ij}(\rho) + \sum_{i=1}^{N}\frac{Z_{it}Z_{i(t-1)}}{p_{it}p_{i(t-1)}}\varepsilon_{it}^2 w_{ii}(\rho)\Big\} \\ &= \sum_{t=2}^{T}\mathcal{E}_t^\top \mathcal{Z}_t\mathcal{P}_t^{-1}\Big\{\Omega(\rho) - \mathrm{diag}(\Omega(\rho))\Big\}\mathcal{P}_t^{-1}\mathcal{Z}_t\mathcal{E}_t + \sum_{t=2}^{T}\mathcal{E}_t^\top\mathrm{diag}\Big\{\Omega(\rho)\Big\}\mathcal{P}_t^{-1}\mathcal{Z}_t\mathcal{E}_t \\ &= \sum_{t=2}^{T}\mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t, \end{aligned}$$

$\mathcal{P}_t = \mathrm{diag}\{p_{it}p_{i(t-1)}\} \in \mathbb{R}^{N\times N}$, $\mathcal{Z}_t = \mathrm{diag}\{Z_{it}Z_{i(t-1)}\} \in \mathbb{R}^{N\times N}$, and $A_t(\rho) = \mathcal{Z}_t\mathcal{P}_t^{-1}\{\Omega(\rho) - \mathrm{diag}(\Omega(\rho))\}\mathcal{P}_t^{-1}\mathcal{Z}_t + \mathrm{diag}(\Omega(\rho))\mathcal{P}_t^{-1}\mathcal{Z}_t$.

To optimize (2.4), we first derive the maximum likelihood estimator of σ^2 by letting its first-order derivative equal 0, which leads to $\tilde{\sigma}^2 = \{N(T-1)\}^{-1}f(\rho)$. We then replace σ^2 by $\tilde{\sigma}^2$ in (2.4) and obtain a profiled log-likelihood function as follows (omitting some constants):

$$\ell_1^*(\rho) = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log\Big\{\sum_{t=2}^{T}\mathcal{E}_t^\top A_t(\rho)\mathcal{E}_t\Big\}. \tag{2.5}$$

This leads to the WMLE as $\tilde{\rho}^{\mathrm{WMLE}} = \mathrm{argmax}_\rho \ell_1^*(\rho)$. The WMLE is an infeasible estimator since its computation allows unknown parameters α_i and β. To fix this problem, we replace α_i with $\widehat{\alpha}_i^{\mathrm{WLSE}}$ and β with $\hat{\beta}$. As a result, we can obtain the estimated $\hat{\mathcal{E}}_t$ and $\hat{A}_t(\rho)$ by replacing ε_{it} and p_{it} with $\hat{\varepsilon}_{it} = Y_{it} - \widehat{\alpha}_i^{\mathrm{WLSE}}Y_{i(t-1)}$ and $\hat{p}_{it}$, respectively. In this way, we can obtain a feasible weighted log-likelihood objective function as follows:

$$\ell_2(\theta) = \ell_2(\rho,\sigma^2) = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top\hat{A}_t(\rho)\hat{\mathcal{E}}_t, \tag{2.6}$$

where $\hat{A}_t(\rho) = \mathcal{Z}_t\hat{\mathcal{P}}_t^{-1}\{\Omega(\rho) - \mathrm{diag}(\Omega(\rho))\}\hat{\mathcal{P}}_t^{-1}\mathcal{Z}_t + \mathrm{diag}(\Omega(\rho))\hat{\mathcal{P}}_t^{-1}\mathcal{Z}_t$, and $\hat{\mathcal{E}}_t = (\hat{\varepsilon}_{1t},\ldots,\hat{\varepsilon}_{Nt})^\top$. Similar to (2.5), we can obtain a feasible profiled log-likelihood function as follows:

$$\ell_2^*(\rho) = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log\Big\{\sum_{t=2}^{T}\hat{\mathcal{E}}_t^\top\hat{A}_t(\rho)\hat{\mathcal{E}}_t\Big\}. \tag{2.7}$$

This leads to a feasible estimator $\hat{\rho}^{\mathrm{WMLE}} = \mathrm{argmax}_\rho \ell_2^*(\rho)$.

Remark. An estimator that is likely more efficient for the temporal dependence $\alpha = (\alpha_1,\ldots,\alpha_N)^\top \in \mathbb{R}^N$ can be obtained when the spatial dependence is given. Specifically, the quantity in (2.4) can be rewritten as follows:

$$\ell_3^*(\rho,\sigma^2,\alpha) = \frac{T-1}{2}\log\left|\Omega(\rho)\right| - \frac{N(T-1)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\Big\{\mathbb{Y}_t - \mathrm{diag}(\mathbb{Y}_{t-1})\alpha\Big\}^\top \times A_t(\rho)\Big\{\mathbb{Y}_t - \mathrm{diag}(\mathbb{Y}_{t-1})\alpha\Big\}, \tag{2.8}$$

which is a function of ρ, σ^2, and α. Here, $\mathbb{Y}_t = (Y_{1t},\ldots,Y_{Nt})^\top \in \mathbb{R}^N$. By temporarily fixing ρ and σ^2 and optimizing ℓ_3^* with respect to α, this leads to another estimated α as $\bar{\alpha}$ as follows:

$$\bar{\alpha} = \Big\{\sum_{t=2}^{T}\mathrm{diag}(\mathbb{Y}_{t-1})A_t(\rho)\mathrm{diag}(\mathbb{Y}_{t-1})\Big\}^{-1} \times \Big\{\sum_{t=2}^{T}\mathrm{diag}(\mathbb{Y}_{t-1})A_t(\rho)\mathbb{Y}_t\Big\}.$$

The unreported simulation studies show that $\bar{\alpha}$ has basically the same efficiency as the proposed $\hat{\alpha}^{\mathrm{WLSE}}$. The reason might be because the quantity in $A_t(\rho)$ of (2.8) can be split into two parts according to whether $i = j$ or not. For the part with $i \neq j$, it requires that Z_{it}, $Z_{i(t-1)}$, Z_{jt}, and $Z_{j(t-1)}$ are all nonmissing. This is a particularly difficult condition to satisfy in the context of missing data. However, the part with $i = j$ only requires that Z_{it} and $Z_{i(t-1)}$ are nonmissing, which happens to be the proposed estimator (e.g., $\hat{\alpha}^{\mathrm{WLSE}}$). For simplicity, we focus on the asymptotic properties for the proposed $\hat{\alpha}^{\mathrm{WLSE}}$ estimator.

3. Theoretical Results

3.1. Technical Conditions

Let $\dot{\Omega}(\rho)$ and $\ddot{\Omega}(\rho)$ be the first- and second-order of $\Omega(\rho)$, that is, $\dot{\Omega}(\rho) = d\Omega(\rho)/d\rho = -(W^\top + W) + 2\rho W^\top W$, $\ddot{\Omega}(\rho) = d^2\Omega(\rho)/d^2\rho = 2W^\top W$. Denote $\dot{A}_t(\rho) = dA_t(\rho)/d\rho$, and $\ddot{A}_t(\rho) = d^2 A_t(\rho)/d\rho^2$. To investigate the asymptotic properties of the proposed estimators of α_i (i.e., infeasible and feasible WLSEs), and ρ (i.e., infeasible and feasible WMLEs), we consider the following technical conditions.

(C1) (Missing rate) Assume the expectation $\sigma_{1i}^2 = E(Y_{i(t-1)}^2 \varepsilon_{it}^2 p_{it}^{-1} p_{i(t-1)}^{-1})$ exists. There also exists a positive matrix $\Delta = (\Delta_{11}, \Delta_{12}; \Delta_{21}, \Delta_{22})$ such that

$$\Delta_{11} = \lim_{N\to\infty} (4N)^{-1} E\Big(2\mathrm{tr}\Big[\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}^2\Big] + \mathrm{tr}^2\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}\Big) - (4N)^{-1}\mathrm{tr}^2\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\}$$

$$\Delta_{22} = \lim_{N\to\infty} (4N\sigma^4)^{-1} E\Big(2\mathrm{tr}\Big[\Big\{A_t(\rho)\Omega(\rho)^{-1}\Big\}^2\Big] + \mathrm{tr}^2\{A_t(\rho)\Omega(\rho)^{-1}\}\Big) - N(4\sigma^4)^{-1}$$

$$\Delta_{12} = \Delta_{21} = \lim_{N\to\infty} (4\sigma^2)^{-1}\mathrm{tr}\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\} - (4N\sigma^2)^{-1} E\Big[2\mathrm{tr}\Big\{\dot{A}_t(\rho)A_t(\rho)\Omega(\rho)^{-2}\Big\} + \mathrm{tr}\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}\mathrm{tr}\Big\{A_t(\rho)\Omega(\rho)^{-1}\Big\}\Big]$$

(C2) (Weight matrix) There exists a positive matrix $\Lambda = (\Lambda_{11}, \Lambda_{12}; \Lambda_{21}, \Lambda_{22})$ such that $\Lambda_{11} = \lim_{N\to\infty}(2N)^{-1} \mathrm{tr}\Big[\Big\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\Big\}^2\Big]$, $\Lambda_{22} = (2\sigma^4)^{-1}$, and $\Lambda_{12} = \Lambda_{21} = \lim_{N\to\infty} -(2N\sigma^2)^{-1}\mathrm{tr}\{\Omega^{-1}(\rho)\dot{\Omega}(\rho)\}$.

(C3) (Diverging speed) For some finite positive constant $c > 0$, assume $N^{1+c}/T = O(1)$. As $N \to \infty$, T also goes to infinity, but at a faster speed.

These conditions are commonly used in the literature. Specifically, Condition (C1) provides the restrictions of the missing rate. It is worth noting that the missing rate should not be too large. Furthermore, when there are no missing observations, these conditions hold because they will reduce to the traditional conditions for the time series and spatial autoregressive model. Condition (C2) puts the constraint on the weight matrix. A similar condition was also used by Sun and Wang (2019). It should be noted that Δ_{22} is an extension of Λ_{22}. When there is no missingness, $A_t(\rho)$ is equal to $\Omega(\rho)$, and then Δ_{22} becomes Λ_{22}. Condition (C3) is a special case of $T/N \to \infty$. It assumes that T should be diverging faster than N. Otherwise, the uniform convergence rate of $\hat{\alpha}_i$ would be too slow. Accordingly, unnecessary bias can be produced for the global parameter $\hat{\rho}$.

3.2. Asymptotic Results

With the help of the above conditions, we then have the following theorems. We first present the asymptotic theories for the infeasible and feasible WLSEs of α_i.

Theorem 1. Assume conditions (C1)–(C3). For any $i \geq 1$, as $T \to \infty$, we have

$$\sqrt{T}(\widetilde{\alpha}_i^{\mathrm{WLSE}} - \alpha_i) \to_d N\Big(0, \frac{\sigma_{1i}^2}{\sigma_{Y_i}^4}\Big),$$

Theorem 2. Assume conditions (C1)–(C3). For any $i \geq 1$, as $T \to \infty$, we have

$$\sqrt{T}(\widehat{\alpha}_i^{\mathrm{WLSE}} - \alpha_i) \to_d N\Big(0, \frac{\sigma_{1i}^2}{\sigma_{Y_i}^4}\Big),$$

where $\sigma_{Y_i}^2 = \sigma_{ii}^2/(1-\alpha_i^2)$, and σ_{ii}^2 is the ith diagonal element of Σ. The proof of Theorems 1 and 2 are given in Appendices B and C in the supplementary materials, respectively. According to these two theorems, we can draw the conclusion that both the infeasible estimator $\widetilde{\alpha}_i^{\mathrm{WLSE}}$ and the feasible estimator $\widehat{\alpha}_i^{\mathrm{WLSE}}$ are consistent and asymptotically normal. As one can see, the theoretical results given in Theorems 1 and 2 hold as long as T goes to infinity, having nothing to do with N. This result is not surprising because Theorems 1 and 2 are established based on model (2.1) and (2.3) from the given individual i only.

Next, we investigate the asymptotic properties for the infeasible and feasible estimators of the spatial dependence of ρ. To perform this, we further define $\theta = (\rho, \sigma^2)^\top \in \mathbb{R}^2$. Theorem 3 presents the property of the infeasible WMLE of θ.

Theorem 3. Assume conditions (C1)–(C3) hold. As $\min\{N, T\} \to \infty$, we have

$$\sqrt{NT}(\widetilde{\theta}^{\mathrm{WMLE}} - \theta) \to_d N\Big(0, \Lambda^{-1}\Delta\Lambda^{-1}\Big),$$

where Δ and Λ are defined as in conditions (C1) and (C2), respectively. The proof of this theorem is given in Appendix D in the supplementary materials; it suggests that the infeasible $\widetilde{\theta}^{\mathrm{WMLE}}$ is consistent and asymptotically normal when some conditions hold. A two-step estimation procedure is used to estimate the covariance. We assume $\mathcal{E}_t$ and (α, β) are given and use the sample versions of Λ and Δ. For example, Δ_{11} can be estimated by $\hat{\Delta}_{11} = \{4N(T-1)\}^{-1}\sum_{t=2}^{T}\Big(2\mathrm{tr}\Big[\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}^2\Big] + \mathrm{tr}^2\Big\{\dot{A}_t(\rho)\Omega(\rho)^{-1}\Big\}\Big) - (4N)^{-1}\mathrm{tr}^2\Big\{\dot{\Omega}(\rho)\Omega(\rho)^{-1}\Big\}$. In the first step, we estimate heterogeneous dynamic (temporal) effects α_i. Then, the spatial effect ρ is estimated in the second step, which requires the order of $\sup_i(\hat{\alpha}_i - \alpha)$. Thus, we assume $N^{1+c}/T = O(1)$. Similar to $\widetilde{\theta}^{\mathrm{WMLE}}$, the asymptotic property for the feasible $\widehat{\theta}^{\mathrm{WMLE}}$ is given in Theorem 4.

Theorem 4. Assume conditions (C1)–(C3) hold. As $\min\{N, T\} \to \infty$, we have

$$\sqrt{NT}(\widehat{\theta}^{\mathrm{WMLE}} - \theta) \to_d N\Big(0, \Lambda^{-1}\Delta\Lambda^{-1}\Big).$$

The proof of this theorem is shown in Appendix E in the supplementary materials. For the estimation of the covariance matrix, we use the estimated $\hat{\beta}$ and $\hat{\alpha}_i^{\mathrm{WLSE}}$, and $\hat{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i^{\mathrm{WLSE}} Y_{i(t-1)}$.

4. Numerical Studies

4.1. Simulation Models

To assess the finite sample performance of the proposed methods, we present several simulation studies. For a given network size N, the network adjacency matrix $A = (a_{ij})$ is simulated as follows. We first generate N independent and identically distributed random variables according to an exponential distribution with a mean of 10. Denote these variables by U_i with $1 \le i \le N$. For each node i, we randomly select a sample size of $[U_i]$ from $\mathcal{S}_F = \{1, 2, \ldots, N\}$ without replacement, where $[U_i]$ stands for the smallest integer no less than U_i. Denote these selected samples by $\mathcal{S}_i$. Define $a_{ij} = 1$ if $j \in \mathcal{S}_i$ and $a_{ij} = 0$ otherwise. Last, let $a_{ii} = 0$ for every $1 \le i \le N$. This leads to the adjacency matrix A. Subsequently, $W = (w_{ij})$ can be obtained by normalizing each row of A so that $\sum_{j=1}^{N} w_{ij} = 1$ for any i. According to the model setup, once W is simulated, it is fixed across all time points.

Next, for a given time point t, the spatial error term $\mathcal{E}_t$ is generated according to $\mathcal{E}_t = (I - \rho W)^{-1}\epsilon_t$, where ϵ_t is simulated from a normal distribution with a mean of 0 and covariance $\sigma^2 I$. We set $\sigma^2 = 4$, and consider $\rho = 0.1$. Then, with the simulated spatial error term, the Y_{it} sequence can be generated as follows. We first set $Y_{i0} = 0$ for $i = 1, 2, \ldots, N$. Then, we generate Y_{it} sequentially according to model (2.1) for $t = 1, \ldots, T_0 + T$, where α_i is stimulated from a uniform distribution between 0.2 and 0.7, and T_0 is a prespecified integer. For example, in this work, we assume $T_0 = 1000$. We then redefine $Y_{it} = Y_{i,t-T_0}$, where $t = T_0 + 1, \ldots, T + T_0$. This leads to the final sequence of $\{Y_{it} : 1 \le i \le N, 1 \le t \le T\}$.

We next consider two missing mechanism cases, that is, MCAR and MAR. According to model (2.3), for each node i, Y_{it} is set to be nonmissing with the probability of p_{it}. For illustration, we consider here one covariate $X_{it} \in \mathbb{R}^1$. Then, the nonmissing probability is set to be $\{\exp(\beta_0 + \beta_1 X_{it})\}/\{1 + \exp(\beta_0 + \beta_1 X_{it})\}$, where β_0 is a tuning parameter controlling the missing rate, and β_1 is the corresponding coefficient. We consider two different values of β_0, that is, $\beta_0 = 1$ and 0.5, which results in a missing rate of approximately 25% and 35% on average, respectively. In the MCAR case, the covariate X_{it} is simulated from a standard normal distribution. In the MAR case, we set $X_{it} = Y_{it} Y_{i(t-1)} + e_{it}$, where e_{it} is generated from a standard normal distribution. Finally, for each experiment, β_1 is set to be 0.1. For a reliable evaluation, each experiment is randomly replicated $M = 1000$ times.

Table 1. Simulation results for the MCAR case with a missing rate of 25% ($\beta_0 = 1$) and 35% ($\beta_0 = 0.5$).

		$\rho = 0.1, \beta_0 = 1$			$\rho = 0.1, \beta_0 = 0.5$		
N	T	$\hat{\alpha}_i^{\mathrm{OLS}}$	$\tilde{\alpha}_i^{\mathrm{WLS}}$	$\hat{\alpha}_i^{\mathrm{WLS}}$	$\hat{\alpha}_i^{\mathrm{OLS}}$	$\tilde{\alpha}_i^{\mathrm{WLS}}$	$\hat{\alpha}_i^{\mathrm{WLS}}$
100	200	8.55	8.55(93.9)	8.55(93.9)	10.10	10.11(94.0)	10.11(94.3)
	500	5.36	5.36(94.7)	5.36(94.7)	6.30	6.31(94.6)	6.31(94.7)
	1000	3.76	3.77(94.9)	3.77(94.9)	4.43	4.43(94.8)	4.43(94.9)
	2000	2.66	2.66(94.9)	2.66(94.9)	3.12	3.13(95.1)	3.13(95.2)
200	200	8.62	8.62(94.0)	8.62(94.0)	10.15	10.16(94.0)	10.16(94.4)
	500	5.39	5.40(94.7)	5.40(94.7)	6.35	6.36(94.6)	6.36(94.8)
	1000	3.80	3.81(94.9)	3.81(94.9)	4.47	4.48(94.9)	4.48(95.0)
	2000	2.69	2.69(95.1)	2.69(95.0)	3.16	3.16(95.0)	3.16(95.0)
400	200	8.60	8.60(94.1)	8.60(94.1)	10.16	10.18(93.9)	10.18(94.4)
	500	5.41	5.41(94.7)	5.41(94.7)	6.36	6.37(94.7)	6.37(94.8)
	1000	3.81	3.81(94.9)	3.81(94.9)	4.48	4.48(94.9)	4.48(95.0)
	2000	2.69	2.70(95.0)	2.70(95.0)	3.17	3.17(95.0)	3.17(95.0)

NOTE: The RMSE values ($\times 10^{-2}$) are reported for every estimator. The CP (in %) for the infeasible WLSE ($\tilde{\alpha}_i^{\mathrm{WLS}}$) and feasible WLSE ($\hat{\alpha}_i^{\mathrm{WLS}}$) are given in parentheses.

Table 2. Simulation results for the MAR case with a missing rate of 25% ($\beta_0 = 1$) and 35% ($\beta_0 = 0.5$).

		$\rho = 0.1, \beta_0 = 1$			$\rho = 0.1, \beta_0 = 0.5$		
N	T	$\hat{\alpha}_i^{\mathrm{OLS}}$	$\tilde{\alpha}_i^{\mathrm{WLS}}$	$\hat{\alpha}_i^{\mathrm{WLS}}$	$\hat{\alpha}_i^{\mathrm{OLS}}$	$\tilde{\alpha}_i^{\mathrm{WLS}}$	$\hat{\alpha}_i^{\mathrm{WLS}}$
100	200	9.47	8.16(94.5)	8.16(94.2)	11.59	9.37(94.7)	9.37(94.6)
	500	7.88	5.16(94.9)	5.16(94.7)	10.15	5.95(94.8)	5.95(94.8)
	1000	7.27	3.64(94.9)	3.64(94.9)	9.63	4.21(95.0)	4.20(95.1)
	2000	6.97	2.58(95.0)	2.57(95.0)	9.37	2.98(95.1)	2.98(95.1)
200	200	9.64	8.26(94.6)	8.26(94.3)	11.86	9.54(94.6)	9.54(94.5)
	500	8.01	5.23(94.9)	5.23(94.8)	10.35	6.05(94.9)	6.06(94.9)
	1000	7.43	3.70(95.0)	3.7(95.0)	9.84	4.28(95.0)	4.28(95.0)
	2000	7.13	2.62(95.0)	2.62(95.0)	9.59	3.04(95.0)	3.04(95.0)
400	200	9.64	8.25(94.6)	8.25(94.3)	11.85	9.51(94.7)	9.51(94.6)
	500	8.04	5.24(94.9)	5.24(94.8)	10.39	6.06(94.9)	6.06(94.9)
	1000	7.46	3.70(95.0)	3.70(94.9)	9.87	4.29(95.1)	4.29(95.0)
	2000	7.14	2.62(95.0)	2.62(95.0)	9.61	3.04(95.0)	3.04(95.0)

NOTE: The RMSE values ($\times 10^{-2}$) are reported for every estimator. The CP (in %) for the infeasible WLSE ($\tilde{\alpha}_i^{\mathrm{WLS}}$) and feasible WLSE ($\hat{\alpha}_i^{\mathrm{WLS}}$) are given in parentheses.

4.2. Performance of Least Squares Estimation

According to Theorems 1 and 2, we know that both the infeasible and feasible WLSEs are $\sqrt{T}$-consistent and asymptotically normal. To investigate the asymptotic performance of these estimators, we consider here various time spans (i.e., $T = 200, 500, 1000, 2000$). In addition, different network sizes are considered (i.e., $N = 100, 200, 400$). Let $\hat{\theta}^{(m)} = (\hat{\theta}_k^{(m)})^\top = (\hat{\alpha}_i^{\mathrm{LSE}(m)}, \tilde{\alpha}_i^{\mathrm{WLSE}(m)}, \hat{\alpha}_i^{\mathrm{WLSE}(m)})^\top$ be the estimators obtained in the mth replication. We employ the following two measures to gauge the performance of the proposed method. First, for a given parameter θ_k with $1 \le k \le 3$, the root-mean-square error is evaluated by $\mathrm{RMSE}_k = \{M^{-1}\sum_{m=1}^{M}(\hat{\theta}_k^{(m)} - \theta_k)^2\}^{1/2}$. Second, for $1 \le k \le 3$, a 95% confidence interval is constructed for θ_k as $\mathrm{CI}_k^{(m)} = (\hat{\theta}_k^{(m)} - z_{0.975}\widehat{\mathrm{SE}}_k^{(m)}, \hat{\theta}_k^{(m)} + z_{0.975}\widehat{\mathrm{SE}}_k^{(m)})$, where $\widehat{\mathrm{SE}}_k^{m}$ is the computed standard error according to the asymptotic covariance in Theorem 1 and Theorem 2 by plugging in the resulting estimators, respectively, and z_α is the αth quantile of a standard normal distribution. Consequently, the empirical coverage probability (ECP) is computed as $\mathrm{ECP}_k = M^{-1}\sum_{m=1}^{M} I(\theta_k \in \mathrm{CI}_k^{(m)})$, where $I(\cdot)$ is the indicator function.

For a given sample size T, we compute the average values of RMSE_k and ECP_k across different i. Simulation results are presented in Tables 1 and 2. Table 1 displays the results for the MCAR case. We can see (i.e., $N = 100$) that the average RMSE value decreases as the sample size T increases for all three estimators. Moreover, the empirical coverage probability for $\tilde{\alpha}_i^{\mathrm{WLSE}}$ and $\hat{\alpha}_i^{\mathrm{WLSE}}$ remains stable around the nominal level of 95%. Another interesting finding is that with a lower missing rate, the average RMSE is smaller. The pattern is nearly the same across different settings of N. Table 2 presents the results for the MAR case. From the table, we can see that the traditional OLS estimator no longer works, whereas both WLSEs work well. The

Table 3. Simulation results for the MCAR case with a missing rate of 25% ($\beta_0 = 1$) and 35% ($\beta_0 = 0.5$).

		$\rho = 0.1, \beta_0 = 1$		$\rho = 0.1, \beta_0 = 0.5$	
N	T	$\tilde{\rho}^{\text{WML}}$	$\hat{\rho}^{\text{WML}}$	$\tilde{\rho}^{\text{WML}}$	$\hat{\rho}^{\text{WML}}$
100	100	3.97(94.2)	4.00(94.1)	5.31(95.1)	5.34(94.2)
	200	2.71(94.7)	2.71(95.0)	3.63(95.2)	3.64(95.0)
	500	1.83(94.0)	1.83(93.8)	2.50(93.7)	2.51(94.1)
200	100	2.72(94.3)	2.69(94.6)	3.79(94.0)	3.79(93.8)
	200	1.86(95.0)	1.89(95.3)	2.55(95.3)	2.54(95.0)
	500	1.27(94.8)	1.24(93.9)	1.72(94.6)	1.70(94.8)
500	100	1.75(94.3)	1.70(94.5)	2.41(93.9)	2.41(94.4)
	200	1.17(95.2)	1.20(94.9)	1.64(94.6)	1.63(94.9)
	500	0.79(94.3)	0.79(94.5)	1.09(95.1)	1.09(94.9)

The RMSE values ($\times 10^{-2}$) are reported for every estimator. The CP (in %) for the infeasible WMLE ($\tilde{\rho}^{\text{WML}}$) and feasible WMLE ($\hat{\rho}^{\text{WML}}$) are given in parentheses.

Table 4. Simulation results for the MAR case with a missing rate of 25% ($\beta_0 = 1$) and 35% ($\beta_0 = 0.5$).

		$\rho = 0.1, \beta_0 = 1$		$\rho = 0.1, \beta_0 = 0.5$	
N	T	$\tilde{\rho}^{\text{WML}}$	$\hat{\rho}^{\text{WML}}$	$\tilde{\rho}^{\text{WML}}$	$\hat{\rho}^{\text{WML}}$
100	100	3.91(92.7)	3.92(92.9)	5.26(93.3)	5.24(93.3)
	200	2.67(93.6)	2.64(93.7)	3.61(93.4)	3.58(93.8)
	500	1.77(92.3)	1.77(92.3)	2.42(93.4)	2.41(92.1)
200	100	2.62(93.7)	2.61(94.0)	3.64(93.6)	3.61(94.1)
	200	1.83(94.4)	1.82(95.0)	2.47(93.6)	2.44(94.6)
	500	1.21(94.7)	1.20(94.5)	1.66(94.6)	1.65(94.9)
500	100	1.68(93.7)	1.67(93.7)	2.29(94.1)	2.27(93.9)
	200	1.12(95.2)	1.12(94.5)	1.60(94.4)	1.59(94.4)
	500	0.81(92.1)	0.81(93.0)	1.07(94.1)	1.07(94.2)

NOTE: The RMSE values ($\times 10^{-2}$) are reported for every estimator. The CP (in %) for the infeasible WMLE ($\tilde{\rho}^{\text{WML}}$) and feasible WMLE ($\hat{\rho}^{\text{WML}}$) are given in parentheses.

RMSE value for $\tilde{\alpha}_i^{\text{WLSE}}$ and $\hat{\alpha}_i^{\text{WLSE}}$ decreases as the sample size T increases. Furthermore, the empirical coverage probability remains around the nominal level of 95%. All of the results corroborate the theoretical results in Theorems 1 and 2.

4.3. Performance of Maximum Likelihood Estimation

Theorem 3 and 4 show that both the infeasible and feasible WMLEs for parameter ρ are $\sqrt{NT}$-consistent and asymptotically normal. To verify the asymptotic performance of these estimators, we consider different network sizes (N = 100, 200, and 500) and different time spans (T = 100, 200, and 500). Similar to the LSEs of α_i, we also use the RMSE and ECP to evaluate the performance of the proposed estimators of ρ. Detailed simulation results are summarized in Tables 3 and 4, from which we can draw the following conclusions. First, in the MCAR case presented in Table 3, we find that both the infeasible and feasible WMLEs are consistent because their RMSE values decrease toward 0 as $N \to \infty$ and $T \to \infty$. Second, the empirical coverage probabilities are fairly close to their nominal level 95%, which suggests that the estimated standard errors (i.e., $\widehat{\text{SE}}$) are well approximated. Third, quantitatively similar results are obtained for the MAR case shown in Table 4, which implies the two estimators also work well in the MAR missingness mechanism. All of these findings confirm that the proposed estimators $\tilde{\rho}^{\text{WMLE}}$ and $\hat{\rho}^{\text{WMLE}}$ are indeed consistent and asymptotically normal.

4.4. A Real Data Example

As our last example, we present here a real data study based on Beijing $PM_{2.5}$ levels. The original data were generously donated by Professor Songxi Chen from Peking University. The dataset can be found at the UC Irvine Machine Learning Repository (*http://archive.ics.uci.edu/ml/index.php*). It contains hourly $PM_{2.5}$ readings collected from 11 different state-owned monitoring sets. For illustration purposes, we use the data from one particular month, which leads to a total of $T = 744$ hourly recordings for each of the $N = 11$ monitoring sites.

For this example, the response Y_{it} is the standardized $PM_{2.5}$ reading collected from the ith location (i.e., one state-owned monitoring site) and the tth time point (i.e., one particular hour in a day). Due to some unknown practical reasons, the collected $PM_{2.5}$ data are often missing for some sites. For a given time point t, we define this time point to be "complete" if all Y_{it} values are observed for every $1 \leq i \leq N$. We then find that the percentage of the complete time points is as low as 89%. For the remaining 11% of the time points, a standard spatial autoregression model cannot be estimated due to incompleteness.

To capture the spatial dependence, two different types of adjacency matrices are constructed. The first is a binary adjacency matrix. In this case, we define $a_{ij} = 1$ if the geographical distance of the two monitoring sites (i.e., $i \neq j$) is less than 10 kilometers. Otherwise, we define $a_{ij} = 0$. The second adjacency matrix is a value-based adjacency matrix. In this case, we define a_{ij} to be the straight-line distance between two sites. In this case a_{ij}'s are positive continuous variables and thus are no longer binary. Before the formal analysis, both adjacency matrices are row normalized.

To study the missingness mechanism, the following three covariates are considered. They are temperature (°C), pressure (hPa), and wind speed (m/s). The logistic regression results show that the coefficients for the above three covariates are −0.06, −0.05, and 0.29, respectively. All of them are statistically significant at the 1% level. This result suggests that all three meteorological indicators have significant influence on the missingness of $PM_{2.5}$. With the help of this logistic regression model, we are able to compute the spatial autocorrelation estimate. We find that the estimated spatial correlation coefficients are given by 0.441 and 0.509 for the two different adjacency matrices. The estimated standard errors are 0.010 and 0.013, respectively. This result suggests that the estimated spatial correlation coefficients are statistically significant at the 1% level for both cases. Thus, the spatial dynamics of the $PM_{2.5}$ readings from different monitoring sites can be statistically measured.

5. Conclusion

In this article, we develop a novel estimation method to analyze missing response issues in a dynamic SEM. The proposed method makes use of the information from temporal dependence, spatial dependence and exogenous regression covariates. To capture the temporal dependence, we apply an AR(1) process with varied autocorrelation coefficients. The spatial dependence is characterized implementing the widely used SEM. A logistic regression model with exogenous covariates is used to reflect the missing mechanism. The WLSE for temporal dependence and the WMLE for spatial dependence are proposed for parameter estimation in the presence of missing data. Compared with traditional estimators, the newly proposed WLSE and WMLE are more efficient. The associated consistency and asymptotic normality of these two estimators are also established. Finally, the performance of the WLSE and WMLE are demonstrated by both simulation studies and a real data example.

To conclude this article, we discuss here several interesting topics for future study. First, the proposed AR(1) effect can be generalized to an AR(p) effect to characterize a more sufficient temporal dynamic dependence. The associated theoretical results can be inferred from the current theorems. However, the practical efficiency would be reduced because of the substantially decreased sample size. Therefore, determining how to improve the estimation efficiency would be an interesting and important research problem for a separate study in the future. Second, the cross-sectional relationships between different individuals could be considered in the future. Third, the missingness mechanism can be easily extended to a new mechanism with heterogeneous β coefficients for better flexibility, which requires further research. Finally, the imputation techniques based on our proposed estimators can also be developed.

Supplementary Materials

This article's appendices are accompanied by supplementary material. The appendices consist of five sections. The first section provides the technical lemmas that facilitate the proofs of Theorems 1–4. The remaining four sections are the technical proofs of theoretical results (Theorems 1–4) in the main text, respectively.

Funding

This research is supported by the National Natural Science Foundation of China (Nos. 71702185, 71873137, 11971504, 71532001, 11525101, 71332006). The Beijing Municipal Social Science Foundation (No. 19GLC052). The Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 18XNLG02). Ministry of Education Focus on Humanities and Social Science Research Base (Major Research Plan 17JJD910001). China's National Key Research Special Program (No. 2016YFC0207704). The fund for building world-class universities (disciplines) of Renmin University of China and the Center for Applied Statistics of Renmin University of China.

References

Anselin, L. (1980), *Estimation Methods for Spatial Autoregressive Structures*, Regional Science Dissertation and Monograph Series (Vol. 8), Ithaca, NY: Cornell University. [28,29]

Brockwell, P., and Davis, R. (1991), *Time Series: Theory and Methods*, Springer Series in Statistics (2nd ed.), Berlin: Springer. [28]

Carlin, B. P., Gelfand, A. E., and Banerjee, S. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman and Hall/CRC. [29]

Dou, B., Parrella, M. L., and Yao, Q. (2016), "Generalized Yule-Walker Estimation for Spatio-Temporal Models With Unknown Diagonal Coefficients," *Journal of Econometrics*, 194, 369–382. [28]

Fuller, W. (1996), *Introduction to Statistical Time Series*, New York: Wiley. [28]

Gao, Z., Ma, Y., Wang, H., and Yao, Q. (2019), "Banded Spatio-Temporal Autoregressions," *Journal of Econometrics*, 208, 211–230. [28]

Hamilton, J. (1994), *Time Series Analysis*, Princeton, NJ: Princeton University Press. [28]

Lee, L., Liu, X., and Lin, X. (2010), "Specification and Estimation of Social Interaction Models With Network Structures," *The Econometrics Journal*, 13, 145–176. [28,29]

Lee, L., and Yu, J. (2010), "Estimation of Spatial Autoregressive Panel Data Models With Fixed Effects," *Journal of Econometrics*, 154, 165–185. [28]

——— (2014), "Efficient GMM Estimation of Spatial Dynamic Panel Data Models With Fixed Effects," *Journal of Econometrics*, 180, 174–197. [28]

LeSage, J., and Pace, R. (2009), *Introduction to Spatial Econometrics*, Boca Raton, FL: Chapman & Hall/CRC. [28,29]

Ord, K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126. [28,29]

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592. [28]

Su, L., and Yang, Z. (2015), "QML Estimation of Dynamic Panel Data Models With Spatial Errors," *Journal of Econometrics*, 185, 230–258. [28]

Sun, Z., and Wang, H. (2019), "Network Imputation for a Spatial Autoregression Model With Incomplete Data," *Statistica Sinica* (forthcoming). [31]

Yang, Z. (2018), "Unified M-Estimation of Fixed-Effects Spatial Dynamic Models With Short Panels," *Journal of Econometrics*, 205, 423–447. [28]

Yu, J., de Jong, R., and Lee, L. (2008), "Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data With Fixed Effects When Both n and T Are Large," *Journal of Econometrics*, 146, 118–134. [28]