The Canadian Journal of Statistics

Vol. , 2011, Pages 1–31

La revue canadienne de statistique

A Combined Moment Equation Approach for Spatial Autoregressive Models

Key words and phrases: Spatial autoregression models; Maximum likelihood estimation; Generalized method of moment; Spatial two-stage least squares.

MSC 2010: Primary 62???; secondary 62???

Abstract: Existing methods for fitting spatial autoregressive models have various strengths and weaknesses. For example, the maximum likelihood estimation (MLE) approach yields efficient estimates, but is computationally burdensome. Computationally efficient methods, such as generalized method of moment (GMM) and spatial two-stage least squares (2SLS), typically require exogenous covariates to be significant, a restrictive assumption that may fail in practice. We propose a new estimating equation approach, termed combined moment equation (COME), which combines the first moment and the covariance conditions of residual terms. The proposed estimator is less computationally demanding than MLE, and does not need the restrictive exogenous conditions as required by GMM and 2SLS. We show that proposed estimator is consistent and establish the asymptotical distribution. Extensive simulations demonstrate that the proposed method outperforms the competitors in bias, efficiency and computation. We apply the proposed method to analyze an air pollution study, and obtain some interesting results about the spatial distribution of PM2.5 concentrations in Beijing. *The Canadian Journal of Statistics* : 1–31; 2011

1. INTRODUCTION

Spatial data are collected often from studies in meteorology, environmental science, ecology, epidemiology and economics; see Cliff and Ord (1973), Anselin (1988a), Cressie (1993), and Anselin and Bera (1998). Among many spatial models, spatial autoregressive (SAR) models (Cliff and Ord, 1973) have emerged as a powerful tool when modeling spatially correlated data based on neighborhood relationships. Though SAR models have been known for decades in the econometrical and statistical literature, their application and research have been limited because of intensive computation (Cressie, 1993; Rangel and Bini, 2006), and elusive statistical properties. For example, the asymptotic properties of the maximum likelihood estimator were not established until recently by Lee (2004).

A key assumption of SAR modelling is that the value of the dependent variable in a location is related to its values measured in the neighborhood. SAR models are typically estimated by *maximum likelihood estimation (MLE) or quasi-likelihood method (QML)* (Ord, 1975; Smirnov and Anselin, 2001; Kazar and Celik, 2012; LeSage and Pace, 2009; Lee, 2004), *generalized method of moment (GMM) estimation* (Kelejian and Prucha, 1999; Lee, 2007a, 2001; Bell and Bockstael, 2000; Lee, 2007b; Liu et al., 2010), *spatial two-stage least squares (2SLS) method* (Land and Deane, 1992; Lee, 2003; Anselin, 1990; Kelejian and Prucha, 1997, 1998; Anselin, 1988b), and *least squares estimation (LSE)* (Huang

* *Author to whom correspondence may be addressed.*

et al., 2019; Zhu et al., 2020). However, MLE or QML requires the computation of the determinant of an $n \times n$ matrix with an unknown autoregression coefficient, which is computationally demanding when the sample size is large. GMM or 2SLS reduces computational burden and also relaxes the normality assumption on the error terms which is required by MLE. However, GMM may lose the efficiency if the instrumental variables(IVs) are not optimal. Lee (2007a) further propose an efficient GMM method by using optimal IVs, However, the optimal IVs involve the inverse of an $n \times n$ matrix and fourth order moments, as a result, the efficient GMM is computationally intensive when the sample size is large and may be unstable for small sample size. In fact, the simulation results from Lee (2007a) show that the two GMMs perform similar with the efficient GMM being slightly worse for small sample size. On the other hand, 2SLS builds an estimator with a matrix of instrumental variables (IVs), which require that the exogenous covariates in the model to be statistically significant to ensure the consistency of the resulting estimators. In reality, significant exogenous covariates may not always be available; even if they do exist, testing their significance is challenging Lee (2007a). Recently, under the assumption that $\mathbf{Y}$ follows the multivariate normal distribution, Huang et al. (2019) and Zhu et al. (2020) propose a least squares estimation (LSE) based on the conditional expectation of response Y_i given $Y_{(-i)} = (Y_1, \cdots, Y_{i-1}, Y_{i+1}, Y_n)$. The assumption of normal distribution is crucial for LSE, which may not hold in practice. In addition, LSE is less efficient due to the partially using of correlation information among $\mathbf{Y}$. These observations are also confirmed by the results of our simulation studies in Tables 1-3, which shows that LSE had a much bigger SD and RMSE than the proposed COME and MLE even the normal assumption is satisfied, and get worse when the measurement error is generated from the uniform distribution.

We propose a new estimating equation approach, which is derived based on the combination of the first moment and the covariance conditions of residual terms. Our approach, termed combined moment equation (COME), presents several advantages. First, based on the observation that autoregression coefficient ρ fully express the spatial autocorrelation, we construct the equation for ρ by exactly removing the correlation among n individuals. Hence, deviating from the IV based methods such as GMM and 2SLS, our method does not require IV variables or significant exogenous covariates. Second, COME is based on moment equation, consequently, allows the distribution of error terms to be unspecified. Third, COME is computationally efficient without the need to compute the determinant or inverse of an $n \times n$ matrix. Finally, we have established consistency and asymptotic distributions for the proposed estimator.

The rest of the paper is organized as follows. In Section 2, we present the COME method. In Section 3, we establish the large sample properties of the COME estimator, and in Section 4, we conduct simulations to investigate the performance of the proposed method. We apply the proposed method to analyze a PM2.5 air pollution dataset in Section 5 and conclude the paper in Section 6. Technical proofs are relegated to the Appendix.

2. PROPOSED COMBINED MOMENT EQUATIONS (COME)

For $i = 1, \ldots, n$, a SAR model stipulates that

$$Y_i = \rho W_i \mathbf{Y} + \mathbf{X}_i' \boldsymbol{\beta} + \epsilon_i, \tag{1}$$

where Y_i is the response and $\mathbf{X}_i$ is the p-dimensional covariate vector of individual i, $W_i = (w_{i1}, \ldots, w_{in})$ with $w_{ij} = 1$ if $j \in N(i)$ and 0 otherwise, and $\mathbf{Y} = (Y_1, \cdots, Y_n)'$. Here, $N(i)$ is the index set for the neighborhoods of individual i, $i \notin N(i)$, and the ϵ_is are identically and independently distributed random errors with mean zero. The vector $\boldsymbol{\beta}$ and the autoregressive spatial parameter ρ are the coefficients to be estimated.

Given $\rho = \rho_0$, since $\mathbf{X}_i$ and ϵ_i are independent, we can estimate $\boldsymbol{\beta}$ by the traditional least square errors, that minimizing

$$Q_1(\rho, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i^2(\rho, \boldsymbol{\beta}), \tag{2}$$

with respect to $\boldsymbol{\beta}$, which is based on the first moment of the error term, where $\epsilon_i(\rho, \boldsymbol{\beta}) = Y_i - \rho W_i \mathbf{Y} - \mathbf{X}_i' \boldsymbol{\beta}$.

Because ϵ_i is correlated with Y_j, when $j \in N(i)$, through its connection with Y_i, $W_i \mathbf{Y}$ is correlated with ϵ_i and is termed an endogeneous term (Lee, 2004). As a result, the estimator for ρ based on minimizing $Q_1(\rho, \boldsymbol{\beta})$ is biased. A key observation that leads to our proposed estimation for ρ is that ρ measures the spatial autocorrelation. That is, if $\rho = \rho_0$, then $Y_i - \rho W_i \mathbf{Y} - \mathbf{X}_j' \boldsymbol{\beta}_0, i = 1, \cdots, n$ are uncorrelated; otherwise, they are correlated. To see this, with $i \neq j$, we have

$$E(Y_i - \rho W_i \mathbf{Y} - \mathbf{X}_i' \boldsymbol{\beta}_0)(Y_j - \rho W_j \mathbf{Y} - \mathbf{X}_j' \boldsymbol{\beta}_0)$$
$$= (\rho - \rho_0) \left\{ (\rho - \rho_0) E(W_i \mathbf{Y} W_j \mathbf{Y}) - E(W_i \mathbf{Y} \epsilon_j + W_j \mathbf{Y} \epsilon_i) \right\}.$$

Therefore, $E(Y_i - \rho W_i \mathbf{Y} - \mathbf{X}_i' \boldsymbol{\beta})(Y_j - \rho W_j \mathbf{Y} - \mathbf{X}_j' \boldsymbol{\beta}) = 0$ if $\rho = \rho_0$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, where ρ_0 and $\boldsymbol{\beta}_0$ are the true values of ρ and $\boldsymbol{\beta}$, respectively. Replacing the theoretical quantity by its empirical version, we can estimate ρ and $\boldsymbol{\beta}$ by minimizing

$$Q_2(\rho, \boldsymbol{\beta}) = \left\{ \frac{1}{n(n-1)} \sum_{i \neq j} \epsilon_i(\rho, \boldsymbol{\beta}) \epsilon_j(\rho, \boldsymbol{\beta}) \right\}^2.$$

This leads to the estimating equations

$$\frac{\partial Q_2(\rho, \boldsymbol{\beta})}{\partial \rho} = 0, \text{ and } \frac{\partial Q_2(\rho, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0. \tag{3}$$

As a result, we estimate $\boldsymbol{\gamma}$ by combining the estimating equations as follows

$$\frac{\partial Q_2(\rho, \boldsymbol{\beta})}{\partial \rho} = 0, \text{ and } \frac{\partial Q_2(\rho, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} + \lambda \frac{\partial Q_1(\rho, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0, \tag{4}$$

where $\lambda \geq 0$ is a weight, striking a balance between the equations based on the covariances and the first moment of the error terms. Particularly, when $\lambda = 0$, the estimator for $\boldsymbol{\beta}$ is based on only the second moment condition $Q_2(\rho, \boldsymbol{\beta})$, and is more determined by the first moment condition $Q_1(\rho, \boldsymbol{\beta})$ when λ gets larger. By (4), we estimate $\boldsymbol{\beta}$ using extra information from the second moment condition, hence is more efficient than the traditional LSE only based on the first moment condition $Q_1(\rho, \boldsymbol{\beta})$. Equations (4) are termed combined moment equations (COMEs) of the first and second moments of the error terms.

Directly solving (4) is computationally difficult. By recognizing the solutions to (4) solve two optimization problems, we propose an efficient iterative algorithm as follows.

Step 0. Choose initial estimates of $\boldsymbol{\beta}$ and ρ, denoted by $\boldsymbol{\beta}^{[0]}$ and $\rho^{[0]}$. For example, the 2SLS estimates can serve as the initial estimates.

Step 1. Given an estimate of ρ at the previous step, i.e. $\rho^{[k-1]}$, estimate $\boldsymbol{\beta}$ by minimizing

$$\left\{\frac{1}{n(n-1)}\sum_{i\neq j}^{n}(Y_i-\rho^{[k-1]}W_i\mathbf{Y}-\mathbf{X}_i'\boldsymbol{\beta})(Y_j-\rho^{[k-1]}W_j\mathbf{Y}-\mathbf{X}_j'\boldsymbol{\beta})\right\}^2 + \frac{\lambda}{n}\sum_{i=1}(Y_i-\rho^{[k-1]}W_i\mathbf{Y}-\mathbf{X}_i'\boldsymbol{\beta})^2.$$

Step 2. Given an updated estimate of $\boldsymbol{\beta}$, i.e. $\boldsymbol{\beta}^{[k]}$, re-estimate ρ by minimizing

$$\left\{\sum_{i\neq j}^{n}(Y_i-\rho W_i\mathbf{Y}-\mathbf{X}_i'\boldsymbol{\beta}^{[k]})(Y_j-\rho W_j\mathbf{Y}-\mathbf{X}_j'\boldsymbol{\beta}^{[k]})\right\}^2.$$

Steps 1 and 2 are iterated until convergence.

Remark 1 *We recommend to use the R* optim *for Steps 1 and 2. Specifically, we adopt the* BFGS *method for Step 1 and the* Brent *method for Step 2. In addition, the simulation studies in Section 4 show that the COME estimator is not sensitive* *to the choice of* λ*. For simplicity, we choose* $\lambda = 1$ *in all of the simulation studies and real data analysis, the resulting estimator suggests* $\lambda = 1$ *works well. In practice, we can use the following BIC (Schwarz et al., 1978; Zhang and Yu, 2018) to select a rough* λ*,*

$$BIC(\lambda) = n * \log(RSS/n) + p * \log(n), \tag{5}$$

where $RSS = \sum_{i=1}(Y_i - \widehat{\rho} W_i \mathbf{Y} - \mathbf{X}_i' \widehat{\boldsymbol{\beta}})^2$*. A brief simulation study shows that the BIC criterion works well.*

3. ASYMPTOTIC PROPERTIES

We establish consistency and the asymptotic distributions for our estimator and defer the proofs to the Appendix. Denote by $\mathbf{X} = (\mathbf{X}_1, \cdots, \mathbf{X}_n)'$, $S_n(\rho) = I_n - \rho\mathbf{W}$ for any value of ρ, $S_n = S_n(\rho_0)$, $b_n(\boldsymbol{\gamma}) = (I_n - \rho\mathbf{W})S_n^{-1}\mathbf{X}\boldsymbol{\beta}_0 - \mathbf{X}\boldsymbol{\beta}$, $G_n = \mathbf{W}S_n^{-1}$, $q_n = (S_n^{-1}\mathbf{X}\boldsymbol{\beta}_0)'\mathbf{W}'C_n$, $C_n = \frac{1}{(n-1)}1_n 1_n' - \frac{1}{(n-1)} I_n$ where 1_n is a $n \times 1$ all-ones vector and I_n is a $n \times n$ identity matrix, $b_{1,n}(\boldsymbol{\gamma}) = b_n(\boldsymbol{\gamma})' C_n b_n(\boldsymbol{\gamma})$ and $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_p)$ with $\mathbf{x}_j$ being column j of $\mathbf{X}$. The following regularity assumptions are set.

Assumption 1. The ϵ_is are i.i.d. with mean zero and $var(\epsilon_i) = \sigma^2 < \infty$. Its moment $E(|\epsilon_i|^{4+\eta})$ for some $\eta > 0$ exits.

Assumption 2. The matrices $\mathbf{W}$ and S_n^{-1} are uniformly bounded in both row and columns sums in absolute value.

Assumption 3. The matrix S_n is nonsingular.

Assumption 4. The $\mathbf{X}_i$'s are bounded. The elements of $C_n\mathbf{X}\mathbf{X}'C_n$, $q_n'q_n$ and $C_n\mathbf{x}_jq_n$ for $j=1,\cdots,p$ are at most $O_p(n)$, $O_p(1)$ and $O_p(1)$, and the corresponding matrices in row and columns sums $C_n\mathbf{X}\mathbf{X}'C_n$ are uniformly $O_p(h_nn)$, $O_p(h_n)$ and $O_p(h_n)$ respectively. Furthermore, we assume that the limits for $\lim_{n\to\infty}\mathbf{X}'\mathbf{X}/n$, $\lim_{n\to\infty}1_n'\mathbf{X}/n$, $\lim_{n\to\infty}tr(q_n'q_n)/n$, $\lim_{n\to\infty}\mathbf{X}'\mathbf{W}S_n^{-1}\mathbf{X}\boldsymbol{\beta}_0/n$ and $\lim_{n\to\infty}\frac{1}{n}(tr(C_n\mathbf{x}_1q_n),\cdots,tr(C_n\mathbf{x}_pq_n))'$ exist, denoted as $\mathbf{A}$, $\boldsymbol{a}$, c_1, $\boldsymbol{c}_0$, $\boldsymbol{b}$ respectively. Let $\Psi=\begin{pmatrix}c_1, & \boldsymbol{b}'\\ \lambda\boldsymbol{c}_0, & \lambda\boldsymbol{A}\end{pmatrix}$ be non-singular matrix.

Assumption 5. Let the sum of vector $b_n(\gamma)$ are $O(\sqrt{n})$ uniformly in a compact parameter space Γ. The true parameter $\boldsymbol{\gamma}_0$ is in the interior of Γ.

Assumption 6. $\lim_{n\to\infty}\inf_{\boldsymbol{\gamma}\in\overline{N}(\boldsymbol{\gamma}_0)}|EU(\boldsymbol{\gamma})|>0$, where $\overline{N}(\boldsymbol{\gamma}_0)$ is the complement of an open neighborhood of $\boldsymbol{\gamma}_0$.

Assumptions 1-3 are specified in Lee (2004), concerning model (1). Particularly, Assumptions 1 demonstrate the features of the disturbance and the weights matrix. Assumption 2 ensures the existence of mean and variance of $\mathbf{Y}$ while Assumption 3 is a condition that limits the spatial correlation to a manageable degree. Assumption 5 requires that the regressors are bounded, and that multi-collinearity among the regressors of $\mathbf{X}$ is ruled out. In fact, combining Assumptions 2-3, we have $\frac{1}{n}\mathbf{X}'\mathbf{W}S_n^{-1}\mathbf{X}\boldsymbol{\beta}_0=O_p(1)$, $tr(q_n'q_n)/n=O_p(1)$, $\mathbf{X}'\mathbf{X}_n/n=O_p(1)$ and $\frac{1}{n}(tr(C_n\mathbf{x}_1q_n),\cdots,tr(C_n\mathbf{x}_pq_n))'=O_p(1)$ which renders

Assumption 4 is reasonable. That Ψ^{-1} exists ensures the existence of asymptotic variance of the proposed estimator for $\boldsymbol{\gamma}$. Assumption 5 is a mild regularity condition assumed in Lee (2007b) and Amemiya (1985). Assumption 6 ensures the existence of a locally unique solution of $EU(\boldsymbol{\gamma}) = 0$.

Theorem 1. *Under Assumptions 1-6,* $\boldsymbol{\gamma}$ *is identifiable and* $\widehat{\boldsymbol{\gamma}}$ *is a consistent estimator of* $\boldsymbol{\gamma}_0$.

Theorem 2. *Under Assumptions 1-6,*

$$\sqrt{n}\begin{pmatrix} \widehat{\rho} - \rho_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \rightsquigarrow \Psi^{-1} g(\boldsymbol{Z}), \tag{6}$$

where $\boldsymbol{Z}$ *is a multivariate normal variable with mean* $\mathbf{0}$ *and covariance* Σ *and* $g(\cdot) : R^{2p+3} \to R^{p+1}$ *is a continuous mapping, which are defined in the Appendix.*

As the asymptotic variances of $\widehat{\rho}$ and $\widehat{\boldsymbol{\beta}}$ in Theorem 2 involve unknown parameters and are infeasible to compute, we adopted a bootstrap procedure (Jin and Lee, 2012) to estimate the variances.

(1) Obtain the estimates for $\boldsymbol{\gamma}$.

(2) Compute $\widehat{\boldsymbol{\epsilon}} = (\widehat{\epsilon}_1, \cdots, \widehat{\epsilon}_n)'$ by $\widehat{\epsilon}_i = Y_i - (\widehat{\rho} W_i \boldsymbol{Y} + \mathbf{X}_i' \widehat{\boldsymbol{\beta}})$. For each $\mathbf{X}_i$, draw the independent residuals $\widehat{\epsilon}_i^*$ from the empirical distribution of the centered residuals $(I - 1_n 1_n'/n)\widehat{\boldsymbol{\epsilon}}$. Let

$$Y_i^* = \widehat{\rho} W_i \boldsymbol{Y} + \mathbf{X}_i' \widehat{\boldsymbol{\beta}} + \widehat{\epsilon}_i^*,$$

and calculate the estimator $\boldsymbol{\gamma}^*$ based on $\{\mathbf{X}_i, Y_i^*\}_{i=1}^n$.

(3) Repeat Step (2) for a total of B times, obtaining B estimated $\boldsymbol{\gamma}^*$, say, $\boldsymbol{\gamma}_b^*, b = 1, \cdots, B$, and compute the sample covariance.

4. NUMERICAL STUDIES

We compared the finite sample performance of COME with the existing methods, which are 2SLS (Kelejian and Prucha, 1997); GMM (Lee, 2007a); MLE (Ord, 1975) and LSE (Huang et al., 2019). We use $(\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X})$ as IV to obtain 2SLS. To perform GMM estimator, we use IV's $(\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X})$ and $(\mathbf{W}, \mathbf{W}^2 - \{tr(\mathbf{W}^2)/n\}I_n)$ for linear moments and quadratic moments, respectively. Based on Trefethen and Bau (1997), the computational complexity is $O(n^3)$ for MLE, while $O(n^2)$ for 2SLS, GMM, LSE and COME methods. Hence 2SLS, GMM, LSE and COME have computational advantage than MLE. We used the R **spdep** package (Bivand and Anselin, 2011) to implement 2SLS and MLE.

Data were generated from the model,

$$\mathbf{Y} = (I_n - \rho\mathbf{W})^{-1}(1_n\beta_1 + \mathbf{X}_1\beta_2 + \mathbf{X}_2\beta_3 + \boldsymbol{\epsilon}), \tag{7}$$

where $\mathbf{X}_k = (X_{k1}, \ldots, X_{kn})'$, for $k = 1, 2$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ were independent vectors. Also, $\mathbf{X}_1, \mathbf{X}_2$ and $\boldsymbol{\epsilon}$ had i.i.d components. We considered three cases.

Case (1): $\rho = 0.9$, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = -1$ and $X_{1i}, X_{2i}, \epsilon_i \sim N(0, 0.5^2)$;

Case (2): $\rho = 0.9$, $\beta_1 = 1$, $\beta_2 = 0.1$, $\beta_3 = -0.1$ and $X_{1i}, X_{2i}, \epsilon_i \sim N(0, 0.5^2)$;

Case (3): $\rho = -0.9$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = -0.3$ and $X_{1i}, X_{2i}, \epsilon_i \sim N(0, 0.5^2)$.

Case (4): $\rho = -0.9$, $\beta_1 = 1$, $\beta_2 = 0.3$, $\beta_3 = 0.3$ and $X_{1i}, X_{2i} \sim N(0, 0.5^2)$, $\epsilon_i \sim U(-1, 1)$.

In Cases (1)-(3), the effects of exogenous covariates were designed to be strong, weak and moderate, respectively. Case (4) is used to investigate the performance of various methods when the measurement error ϵ_i is generated from non-Gaussian distribution. For each case, we varied the sample size by taking $n = 49, 98, 490, 980$ and 2450. When $n = 49$, we took $\mathbf{W} = \mathbf{W}_0$, where $\mathbf{W}_0$ was specified as in Anselin (1988b); when $n = 98, 490, 980$ and 2450, we took $\mathbf{W}$ to be $I_2 \otimes \mathbf{W}_0$, $I_{10} \otimes \mathbf{W}_0$, $I_{20} \otimes \mathbf{W}_0$ and $I_{50} \otimes \mathbf{W}_0$, respectively, where $\otimes$ is the Kronecker product operator. We evaluate the performance of the competing estimators using the criteria of absolute bias (BIAS), empirical standard deviation (SD), the root of the mean square error (RMSE) and the average CPU time per run in seconds ($T(s)$). For each parameter configuration, we generated a total of 1,000 independent datasets, and Tables 1-3 report the results based on these 1000 replications for Cases (1)-(4). We set $\lambda = 1$ for Cases (1)-(4). We summarize the findings below.

First, the coefficients of the exogenous covariates, β_2 and β_3, estimated by the methods except of LSE, were comparable, which confirmed that the asymptotic variances for β_2 and β_3 were the same. LSE estimates for β_2 and β_3 had slightly larger SD and RMSE than others. Second, the estimates of COME for ρ and β_1, on average, were close to the truth in all of the cases considered. In contrast, the MLEs of ρ and β_1 were much biased when the sample size was small ($n = 49, 98$); see Cases (1) and (2) in Tables 1 and 2. The 2SLS estimate of ρ and β_1 were biased when the effect of exogenous covariates was small, as shown in Case (2). Even in Cases (1) and (3) where the exogenous covariates had moderate to large effects, COME had a much smaller SD and RMSE than 2SLS, especially when the sample size was small ($n = 49, 98$); the gap became larger for Case (2). The GMM and LSE had larger bias and SD for ρ and β_1 in all of the cases considered, and the LSE get worse when the measurement error is far away from the normal distribution, see Case (4) in Table 3.

Finally, when comparing the computing time in Figure 1(a), we found that MLE took much more CPU time than COME, GMM and 2SLS, especially with large sample sizes. For example, the $T(s)$ of MLE increased drastically from $.96$ seconds to around 100 minutes, while that of COME increased from 0.36 seconds to about 8 minutes.

To investigate the performance of the bootstrap described in Section 3, we calculated the standard errors(SE) based on the bootstrap with 200 bootstrap samples. The standard deviations, denoted by SD, based on 200 simulations can be regarded as the true SE. The average and the SD of 200 estimated SE, denoted by SE.ave and SE.std, summarize the performance of the bootstrap. As shown in Table 5, we conclude that the performance of the standard error, obtained from bootstrap, is quite satisfactory.

Finally, we also investigate the effect of varying λ on the resulting estimates. The bias, SD and RMSE against λ for Case (1) are shown in Table 4, which suggest that the resulting estimates are insensitive to λ. In practice, we can choose a rough λ by BIC criterion. We plot the average of BIC versus RMSE when $\lambda = 0.001, 0.05, 0.5, 1$ for Case (1) in Figure 1(b), which shows that BIC increases as RMSE increases, suggesting that formula (5) provides a reasonable estimator of λ.

TABLE 1: Comparisons of competing methods in bias, empirical standard error (SD), root of mean squared error (RMSE) and CPU time ($T(s)$) for Case (1) based on 1000 replications.

		ρ =0.9	$\beta_2 = 1$	$\beta_3 = -1$	$\beta_1 = 1$	T(s)
n		bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	
49	COME	.019(.090)[.092]	.010(.310)[.311]	.011(.304)[.304]	.173(1.249)[1.261]	.014
	LSE	.061(.153)[.165]	.008(.366)[.366]	.009(.357)[.357]	.566(1.862)[1.946]	.029
	MLE	.048(.076)[.090]	.013(.307)[.307]	.017(.297)[.297]	.486(1.131)[1.231]	.148
	GMM	.014(.154)[.155]	.006(.317)[.317]	.005(.309)[.309]	.121(1.782)[1.786]	.011
	2SLS	.049(.150)[.158]	.025(.306)[.307]	.021(.294)[.295]	.486(1.711)[1.778]	.004
98	COME	.003(.048)[.048]	.002(.206)[.206]	.003(.211)[.211]	.014(.758)[.758]	.024
	LSE	.039(.114)[.120]	.004(.247)[.247]	.002 (.245)[.245]	.395(1.32)[1.38]	.047
	MLE	.019(.038)[.043]	.007(.204)[.204]	.004(.207)[.207]	.190(.690)[.718]	.187
	GMM	.008(.087)[.087]	.005(.211)[.211]	.006(.219)[.219]	.087(1.05)[1.05]	.026
	2SLS	.014(.111)[.112]	.016(.208)[.209]	.019(.211)[.212]	.151(1.25)[1.26]	.009
490	COME	.002(.016)[.016]	.002(.091)[.091]	.001(.091)[.091]	.011(.306)[.307]	.356
	LSE	.007(.045)[.046]	.003(.109)[.109]	.002(.107)[.107]	.069(.548)[.552]	.541
	MLE	.004(.014)[.015]	.003(.090)[.090]	.000(.091)[.091]	.036(.299)[.301]	.964
	GMM	.001(.024)[.024]	.001(.093)[.093]	.001(.095)[.095]	.006(.387)[.387]	.581
	2SLS	.002(.034)[.034]	.002(.093)[.093]	.006(.093)[.093]	.002(.416)[.417]	.157
980	COME	.001(.012)[.012]	.000(.064)[.064]	.000(.066)[.066]	.010(.211)[.211]	1.358
	LSE	.003(.030)[.030]	.000(.077)[.077]	.002(.080)[.080]	.032(.372)[.374]	16.90
	MLE	.002(.010)[.010]	.000(.064)[.064]	.000(.066)[.066]	.019(.210)[.210]	17.210
	GMM	.001(.012)[.012]	.000(.066)[.066]	.001(.069)[.069]	.009(.234)[.234]	2.109
	2SLS	.003(.022)[.022]	.002(.065)[.065]	.003(.067)[.067]	.002(.290)[.290]	.787
2450	COME	.000(.008)[.008]	.001(.043)[.043]	.000(.041)[.041]	.006(.140) .140]	14.944
	LSE	.001(.019)[.019]	.001(.048)[.048]	.001(.047)[.047]	.015(.232)[.233]	97.833
	MLE	.001(.006)[.006]	.001(.042)[.042]	.000(.041)[.041]	.004(.135)[.136]	99.577
	GMM	.000(.007)[.007]	.002(.043)[.043]	.001(.042)[.042]	.001(.145)[.145]	43.951
	2SLS	.000(.014)[.014]	.000(.043)[.043]	.001(.042)[.042]	.001(.188)[.188]	10.362

TABLE 2: Comparisons of competing methods in bias, empirical standard error (SD), root of mean squared error (RMSE) and CPU time ($T(s)$) for Cases (2) and (3) based on 1000 replications.

		Case (2)				
		$\rho = 0.9$	$\beta_2 = 0.1$	$\beta_3 = -0.1$	$\beta_1 = 1$	T(s)
n		bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	
98	COME	.001(.065)[.065]	.001(.204)[.204]	.005(.208)[.208]	0.021(0.878)[0.878]	0.018
	LSE	.039(.112)[.119]	.004(.247)[.247]	.002(.244)[.244]	0.392(1.307)[1.364]	0.037
	MLE	.023(.045)[.050]	.001(.203)[.203]	.003(.207)[.207]	0.224(0.735)[0.768]	0.146
	GMM	.167(.210)[.268]	.004(.226)[.226]	.011(.231)[.231]	1.684(2.252)[2.812]	0.042
	2SLS	.114(.181)[.214]	.004(.203)[.203]	.006(.210)[.210]	1.149(1.923)[2.241]	0.007
490	COME	.001(.035)[.035]	.001(.090)[.090]	.002(.091)[.091]	0.001(0.428[0.428]	0.143
	LSE	.007(.045)[.046]	.003(.109)[.109]	.002(.107)[.107]	0.069(0.548)[0.552]	0.399
	MLE	.005(.016)[.016]	.001(.090)[.090]	.002(.091)[.091]	0.040(0.309)[0.312]	0.711
	GMM	.138(.160)[.211]	.010(.100)[.100]	.012(.100)[.101]	1.388(1.646)[2.153]	0.663
	2SLS	.090(.154)[.180]	.004(.091)[.091]	.007(.092)[.092]	0.900(1.538)[1.781]	0.114
980	COME	.001(.015)[.015]	.000(.063)[.063]	.000(.066)[.066]	0.013(0.232)[0.233]	1.465
	LSE	.003(.031)[.031]	.000(.077)[.077]	.002(.080)[.080]	0.036(0.379)[0.381]	4.095
	MLE	.002(.011)[.011]	.001(.063)[.063]	.000(.066)[.066]	0.024(0.218)[0.220]	4.169
	GMM	.030(.091)[.096]	.003(.067)[.067]	.005(.069)[.070]	0.313(0.945)[0.995]	2.025
	2SLS	.082(.135)[.157]	.004(.063)[.063]	.005(.066)[.066]	0.819(1.356)[1.584]	0.593
		Case (3)				
		$\rho = -0.9$	$\beta_2 = 0.3$	$\beta_3 = -0.3$	$\beta_1 = 1$	T(s)
n		bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	
49	COME	.025(.223)[.224]	.001(.300)[.300]	.001(.291)[.291]	0.011(0.831)[0.831]	0.002
	LSE	.304(.540)[.619]	.029(.502)[.503]	.033(.471)[.472]	0.166(1.396)[1.405]	0.005
	MLE	.001(.181)[.181]	.006(.299)[.299]	.001(.287)[.287]	0.014(0.816)[0.816]	0.162
	GMM	.033 (.222)[.224]	.025(.391)[.392]	.023(.360)[.360]	0.130(1.288)[1.295]	0.003
	2SLS	.060(.774)[.776]	.022(.317)[.318]	.008(.305)[.305]	0.058(0.925)[0.927]	0.001
98	COME	.008(.167)[.167]	.001(.202)[.202]	.006(.204)[.205]	0.008(0.577)[0.577]	0.004
	LSE	.100(.215)[.237]	.002(.290)[.290]	.000(.284)[.284]	0.048(0.809) [0.810]	0.014
	MLE	.000(.134)[.134]	.002(.201)[.201]	.005(.205)[.205]	0.006(0.577)[0.577]	0.167
	GMM	.016(.154)[.154]	.002(.204)[.204]	.004(.214)[.214]	0.002(0.609)[0.609]	0.009
	2SLS	.027(.743)[.743]	.021(.217)[.218]	.018(.216)[.217]	0.020(0.741)[0.741]	0.002
490	COME	.003(.095)[.095]	.001(.090) [.090]	.002(0.091[0.091]	0.005(0.262)[0.263]	0.106
	LSE	.025(.074)[.078]	.006(.116)[.116]	.001(.120)[.120]	0.004(0.340)[0.340]	0.283
	MLE	.000(.057)[.057]	.001(.090)[.090]	.002(.090)[.090]	0.006(0.262)[0.262]	0.574
	GMM	.005(.065)[.065]	.000(.093)[.093]	.002(.093)[.093]	0.001(0.280)[0.280]	0.460
	2SLS	.021(.397)[.398]	.006(.093)[.093]	.009(.095)[.095]	0.018(0.336)[0.336]	0.147

TABLE 3: Comparisons of competing methods in bias, empirical standard error (SD), root of mean squared error (RMSE) and CPU time ($T(s)$) for Cases (4) based on 500 replications.

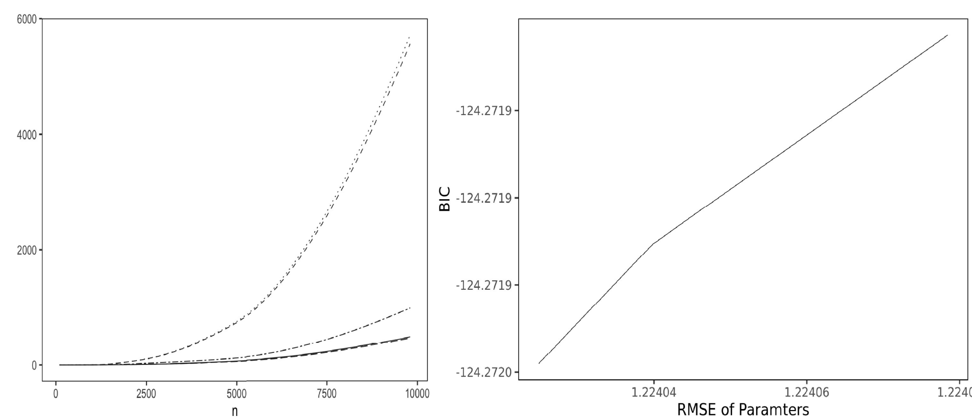
		$\rho = -0.9$	$\beta_2 = 0.3$	$\beta_3 = 0.3$	$\beta_1 = 1$	T(s)
n		bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	
49	COME	.005(.224)[.224]	.019(.366)[.367]	.013(.341)[.341]	.056(1.065)[1.067]	0.006
	LSE	.291(.574)[.644]	.031(.563)[.564]	.011(.528)[.528]	.298 (1.545)[1.574]	0.004
	MLE	.003(.191)[.191]	.020(.363)[.364]	.016(.336)[.337]	.072(1.027)[1.029]	0.149
	GMM	.033(.222)[.224]	.025(.391)[.392]	.023(.360)[.360]	.130(1.288)[1.295]	0.005
	2SLS	.014(.770) [.770]	.025(.387)[.387]	.040(.357)[.359]	.111(1.574)[1.578]	0.001
98	COME	.007(.175)[.175]	.009(.246)[.246]	.007(.232)[.232]	.012(0.721)[0.721]	0.008
	LSE	.117(.277)[.301]	.009(.331)[.331]	.011(.321)[.322]	.143(1.007) [1.017]	0.016
	MLE	.006(.137)[.137]	.004(.241)[.241]	.009(.230)[.230]	.019(0.693)[0.693]	0.130
	GMM	.014(.155)[.156]	.005(.269)[.269]	.011(.258)[.258]	.031(0.817)[0.818]	0.012
	2SLS	.052(.744)[.746]	.010(.263)[.263]	.024(.244)[.245]	.131(1.237) [1.243]	0.002
490	COME	.003(.074)[.074]	.002(.104)[.104]	.001(.097)[.097]	.001(0.292)[0.292]	0.142
	LSE	.030(.083)[.088]	.000(.136)[.136]	.001(.134)[.134]	.036(0.390)[0.392]	0.390
	MLE	.004(.059)[.059]	.002(.104)[.104]	.001(.097)[.097]	.001(0.291)[0.291]	0.556
	GMM	.008(.068)[.069]	.002(.108)[.108]	.001(.101)[.101]	.006(0.319)[0.319]	0.232
	2SLS	.013(.415)[.416]	.006(.108)[.108]	.009(.101)[.102]	.013(0.658)[0.658]	0.047

TABLE 4: The results of bias, empirical standard error (SD), root of mean squared error (RMSE) and CPU time $(T(s))$ of COME for Case (1) under different λ.

		ρ =0.9	$\beta_2 = 1$	$\beta_3 = -1$	$\beta_1 = 1$	T(s)
n	λ	bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	bias(SD)[RMSE]	
98	.001	.002(.048)[.048]	.002(.206)[.206]	.004(.211)[.211]	.014(.744)[.744]	.027
	.005	.003(.048)[.049]	.003(.206)[.206]	.004(.211)[.211]	.015(.753)[.753]	.028
	0.50	.003(.048)[.048]	.002(.206)[.206]	.003(.211)[.211]	.014(.758)[.758]	.025
	1.00	.003(.048)[.048]	.002(.206)[.206]	.003(.211)[.211]	.014(.758)[.758]	.024

TABLE 5: True and estimated standard errors and CPU time $(T(s))$ of COME for Case(1).

	ρ =0.9	$\beta_2 = 1$	$\beta_3 = -1$	$\beta_1 = 1$	T(s)
n	SD(SE.ave)[SE.std]	SD(SE.ave)[SE.std]	SD(SE.ave)[SE.std]	SD(SE.ave)[SE.std]	
98	.089(.074)[.045]	.202(.212)[.059]	.206(.214)[.062]	1.030(.962)[.389]	0.835
490	.033(.034)[.009]	.092(.093)[.006]	.093(.093)[.007]	0.408(.428)[.078]	15.515



(b)

FIGURE 1: (a) Computing time $(T(s))$ of the LSE (dashed), MLE (dotted), GMM (twodash), 2SLS (longdash) and COME (solid) for Case (1) as n increases from 49 to 9800, (b) BIC versus RMSE using $\lambda = 0.001, 0.05, 0.5, 1$ for Case (1).

5. A STUDY OF PM2.5 DISTRIBUTIONS IN BEIJING

Small particulate matters (PM) are detrimental to human beings' health. $\mathrm{PM}_{2.5}$, with a mean aerodynamic diameter less than 2.5 μm, poses more serious risks on individuals, as small matters can be inhaled deeply into lungs and even bloodstreams. As the sources of PM2.5 are still debatable, it is of interest to understand the distribution of $\mathrm{PM}_{2.5}$ in a specific region and how it is impacted by human activities, such as driving. We analyze an atmospheric particulate matter data collected from the internet. The data consist of hourly measurements of $\mathrm{PM}_{2.5}$ from 36 air-quality monitoring sites displayed in Figure 2 from Zhang et al. (2017) in Beijing. By exploring the spatial correlation patterns of $\mathrm{PM}_{2.5}$ concentration, we investigate the difference in the spatial autocorrelation of $\mathrm{PM}_{2.5}$ between different time points, such as between morning rush hours and evening rush hours or between weekdays and weekends, to understand better the possible causes, occurrence and development of $\mathrm{PM}_{2.5}$. The data document the concentration of $\mathrm{PM}_{2.5}$ measured in $n = 36$ monitoring sites from 7am to 9 pm on January 12, 2017 and January 14, 2017, which are a weekday (Thursday) and a weekend day (Saturday), respectively. As wind speed ($Speed$), temperature ($Temp$), atmospheric pressure ($Press$) may influence $\mathrm{PM}_{2.5}$ concentrations, we control for them when fitting the following SAR model at a given t:

$$Y_{t,i} = \rho_t W_i \boldsymbol{Y}_t + \beta_{1t} + Temp_{t,i}\beta_{2t} + Speed_{t,i}\beta_{3t} + Press_{t,i}\beta_{4t} + \epsilon_{t,i}, \tag{8}$$

where $Y_{t,i}, Temp_{t,i}$, $Speed_{t,i}$ and $Press_{t,i}$ represent the $PM_{2.5}$ concentration, temperature, wind speed and atmospheric pressure, respectively, at location i of hour t, $i = 1, \cdots, n$, $\mathbf{W} = (w_{ij})$ is an $n \times n$ normalized spatial weight matrix which element $w_{ij} \neq 0$ if location i is adjacent to j, $w_{ij} = 0$ otherwise, $w_{ii} = 0$, and $\boldsymbol{Y}_t = (Y_{t,1}, \ldots, Y_{t,n})'$. The adjacent relationship is generated by clustering areas into several groups. The measurement errors $\epsilon_{t,i}$s, $i = 1, \cdots, n$ are i.i.d. for any given t, $\boldsymbol{\epsilon}_t = (\epsilon_{t,1}, \ldots, \epsilon_{t,n})'$. We apply COME, MLE, 2SLS and GMM to estimate ρ_t and $\boldsymbol{\beta}_t = (\beta_{1t}, \beta_{2t}, \beta_{3t}, \beta_{4t})'$, where ρ_t represents the degree of spatial correlation of $PM_{2.5}$ at hour t. The standard errors of the estimates are calculated using the resampling method described in Section 3 with 500 bootstrap samples, where 500 is determined by monitoring the stability of the estimated standard errors.

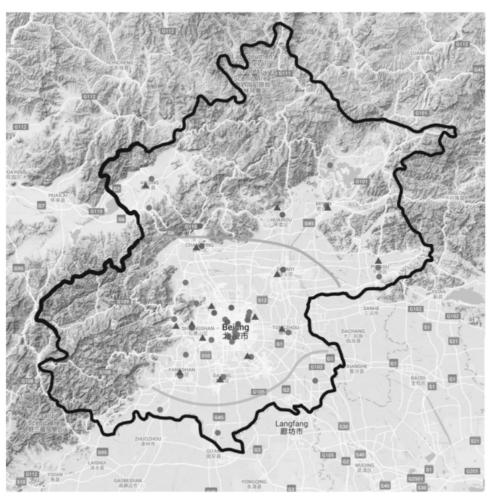


FIGURE 2: Locations of the 36 air-quality monitoring sites in Beijing, marked with red or purple dots.

The estimates (est), along with the bootstrap-based standard errors (SD) and p-values (PV), of ρ_t and $\boldsymbol{\beta}_t$ are displayed in Tables 3 and 4. The total CPU time consumed for each method is recorded as T_{time}. COME and MLE provide similar estimates for $\boldsymbol{\beta}_t$ in all of cases. The results, however, differ from those of GMM and 2SLS, especially when the covariates are not significant. This might be supported by our simulation studies, which found that GMM and 2SLS produce biased results when the covariates are not significant. Though COME and MLE provide similar estimates of ρ_t, COME provide more significant estimates than MLE, which could be attributed to the relatively small number of locations in our data, leading to unstable MLE estimates. Figure 4 shows that the trends of ρ_t over time on a weekday and a weekend day manifested different patterns. On a Thursday, ρ_t was fairly stable and significantly above 0 until 8pm, and a slight peak occurred at 10am. In contrast, on a Saturday, ρ_t for the most time points is around 0 but an obvious peak, which was significantly above 0, happened at 7pm. Coincidentally, as Yao et al. (2015) noted when studying the PM2.5 concentrations in Beijing, the traffic has been constantly heavy in Beijing on a business day while the roads are the most jammed around 10am, whereas the traffic is relatively heavier on Saturdays only after 3pm because of the increased human activities, due to, for example, dinner appointments. Our modeling results, coupled with the observations on Beijing's human activities, suggested that PM2.5 be dispersed by vehicles, whose exhaustion might be a source of PM2.5 in Beijing. Finally, in terms of computing time, COME is comparable with GMM and 2SLS and is much faster than MLE.

TABLE 6: Estimates of ρ_t and $\boldsymbol{\beta}_t$ for the $PM_{2.5}$ model on a workday

	COME	**MLE**	**GMM**	**2SLS**
	est(SD, PV)	est(SD, PV)	est(SD, PV)	est(SD, PV)
		$t = 7:00am \quad Thursday$		
ρ	0.419(.272, 0.062)	0.115(.236, 0.313)	1.066(.260, 0**)	4.264(.718, 0**)
Intercept	0.096(.077, 0.105)	0.113(.038, 0.001**)	0.138(.521, 0.396)	0.837(.189, 0**)
Temp	-0.298(.205, 0.074)	-0.355(.193, 0.033*)	-0.185(.260, 0.238)	-0.097(.158, 0.270)
Speed	0.243(.244, 0.159)	0.255(.305, 0.202)	0.226(.419, 0.295)	-0.089(.156, 0.288)
Press	0.124(.193, 0.261)	0.134(.213, 0.264)	0.147(.213, 0.245)	0.060(.158, 0.352)
T_{times}	6.2	81.7	4.1	1.7
		$t = 8:00am \quad Thursday$		
ρ	0.646(.239, 0.003**)	0.206(.265, 0.219)	1.271(.282, 0**)	3.455(.415, 0**)
Intercept	0.137(.112, 0.110)	0.161(.069, 0.010*)	0.184(.176, 0.147)	0.595(.107, 0**)
Temp	-0.218(.251, 0.192)	-0.246(.206, 0.116)	-0.083(.175, 0.317)	-0.011(.129, 0.466)
Speed	0.147(.293, 0.308)	0.129(.332, 0.348)	0.053(.121, 0.331)	-0.126(.120, 0.146)
Press	0.052(.195, 0.395)	0.056(.241, 0.407)	0.079(.384, 0.418)	0.021(.130, 0.437)
T_{time}	4.7	81.4	3.7	1.6
		$t = 9:00am \quad Thursday$		
ρ	0.752(.228, 0**)	0.298(.252, 0.119)	0.964(.183, 0**)	2.049(.148, 0**)
Intercept	0.121(.110, 0.136)	0.150(.075, 0.023*)	0.188(.720, 0.397)	0.235(.048, 0**)
Temp	-0.151(.272, 0.289)	-0.176(.214, 0.205)	0.046(.376, 0.451)	-0.016(.161, 0.459)
Speed	0.099(.285, 0.364)	0.086(.336, 0.399)	-0.126(.792, 0.437)	-0.064(.133, 0.313)
Press	-0.067(.187, 0.360)	-0.097(.228, 0.336)	-0.145(.237, 0.269)	0.044(.154, 0.388)
T_{time}	5.5	79	5.3	1.4
		$t = 5:00pm \quad Thursday$		
ρ	0.611(.172, 0**)	0.266(.143, 0.032*)	0.287(.213, 0.089)	0.597(.187, 0.001**)
Intercept	-0.030(.021, 0.07)	-0.021(.010, 0.014*)	-0.016(.010, 0.056)	-0.031(.015, 0.018*)
Temp	-0.442(.123, 0**)	-0.680(.147, 0**)	-0.722(.199, 0**)	-0.458(.166, 0.003**)
Speed	0.453(.134, 0**)	0.754(.168, 0**)	0.792(.225, 0**)	0.472(.184, 0.005**)
Press	-0.001(.088, 0.493)	-0.035(.094, 0.354)	-0.029(.094, 0.379)	-0.003(.092, 0.487)
T_{time}	13.1	63.1	2.5	1.7
		$t = 6:00pm \quad Thursday$		
ρ	0.482(.157, 0.001**)	0.371(.152, 0.007**)	0.476(.206, 0.01**)	0.511(.208, 0.007**)
Intercept	-0.011(.013, 0.197)	-0.011(.013, 0.183)	-0.013(.013, 0.163)	-0.012(.012, 0.162)
Temp	-0.429(.093, 0**)	-0.491(.120, 0**)	-0.429(.138, 0.001**)	-0.417(.139, 0.001**)
Speed	0.455(.094, 0**)	0.535(.140, 0**)	0.468(.166, 0.002**)	0.443(.158, 0.002**)
Press	0.205(.084, 0.007**)	0.215(.089, 0.008**)	0.201(.087, 0.01**)	0.203(.085, 0.008**)
T_{time}	13.1	62.6	2.5	1.7
		$t = 7:00pm \quad Thursday$		
ρ	0.709(.23, 0.001**)	0.325(.175, 0.031*)	0.443(.228, 0.026*)	0.707(.205, 0**)
Intercept	-0.008(.037, 0.417)	-0.020(.018, 0.123)	-0.019(.022, 0.192)	-0.010(.025, 0.339)
Temp	-0.286(.151, 0.029*)	-0.478(.164, 0.002**)	-0.437(.179, 0.007**)	-0.293(.163, 0.036*)
Speed	0.251(.224, 0.131)	0.480(.255, 0.030*)	0.464(.274, 0.045)	0.263(.238, 0.135)
Press	0.085(.208, 0.341)	0.098(.246, 0.345)	0.158(.246, 0.261)	0.089(.211, 0.337)
T_{time}	14.2	62.3	2.7	1.5

TABLE 7: Estimates of ρ_t and $\boldsymbol{\beta}_t$ for the $PM_{2.5}$ model on a weekend day

	COME	**MLE**	**GMM**	**2SLS**
	est(SD, PV)	est(SD, PV)	est(SD, PV)	est(SD, PV)
		$t = 7:00am$ $Saturday$		
ρ	-0.481(.217, 0.013*)	-0.253(.196, 0.099)	-0.579(.576, 0.158)	-0.972(.290, 0**)
Intercept	-0.012(.020, 0.270)	-0.016(.015, 0.141)	-0.027(.043, 0.267)	-0.115(.030, 0**)
Temp	0.234(.271, 0.194)	0.250(.299, 0.201)	0.338(.392, 0.194)	0.407(.297, 0.085)
Speed	-0.340(.290, 0.121)	-0.393(.383, 0.153)	-0.466(.447, 0.148)	-0.622(.313, 0.023*)
Press	-0.107(.170, 0.256)	-0.114(.185, 0.269)	-0.189(.216, 0.191)	-0.093(.182, 0.305)
T_{time}	6.0	86.6	6.6	1.7
		$t = 8:00am$ $Saturday$		
ρ	-0.388(.224, 0.041*)	-0.218(.194, 0.131)	-0.509(.598, 0.197)	1.010(.033, 0**)
Intercept	-0.006(.017, 0.358)	-0.007(.015, 0.313)	0.004(.035, 0.460)	0.093(.291, 0.374)
Temp	0.130(.281, 0.321)	0.154(.319, 0.314)	0.312(.472, 0.254)	-0.080(.336, 0.406)
Speed	-0.313(.290, 0.14)	-0.390(.395, 0.162)	-0.569(.529, 0.141)	0.117(.375, 0.377)
Press	-0.206(.163, 0.104)	-0.226(.189, 0.115)	-0.347(.274, 0.103)	-0.143(.189, 0.224)
T_{time}	5.3	84.6	5.5	1.3
		$t = 9:00am$ $Saturday$		
ρ	-0.468(.184, 0.006**)	-0.214(.201, 0.142)	-0.672(.864, 0.218)	2.320(.395, 0**)
Intercept	-0.005(.013, 0.362)	-0.005(.013, 0.349)	-0.008(.048, 0.433)	0.158(.094, 0.046*)
Temp	-0.282(.374, 0.226)	-0.282(.380, 0.229)	-0.136(.579, 0.407)	-0.572(.561, 0.154)
Speed	-0.152(.337, 0.326)	-0.211(.404, 0.301)	-0.469(.580, 0.209)	0.942(.536, 0.040*)
Press	0.106(.194, 0.292)	0.111(.203, 0.292)	0.100(.270, 0.355)	0.030(.275, 0.456)
T_{time}	5.8	91.0	4.1	1.6
		$t = 5:00pm$ $Saturday$		
ρ	0.515(.334, 0.062)	0.191(.180, 0.145)	0.192(.248, 0.219)	0.493(.334, 0.007**)
Intercept	0.034(.036, 0.174)	0.041(.023, 0.034*)	0.028(.020, 0.085)	0.036(.032, 0.127)
Temp	-0.442(.239, 0.033*)	-0.624(.249, 0.006**)	-0.754 (.272, 0.003**)	-0.456(.255, 0.037*)
Speed	0.273(.223, 0.111)	0.369(.277, 0.092)	0.473(.277, 0.044*)	0.280(.239, 0.121)
Press	-0.143(.154, 0.176)	-0.205(.189, 0.140)	-0.157(.182, 0.194)	-0.150(.161, 0.176)
T_{time}	14.0	62.8	2.6	1.8
		$t = 6:00pm$ $Saturday$		
ρ	0.578(.459, 0.104)	0.016(.195, 0.468)	-0.048(.368, 0.448)	0.565(.466, 0.112)
Intercept	0.032(.052, 0.266)	0.013(.018, 0.237)	0.010(.033, 0.375)	0.033(.050, 0.255)
Temp	-0.371(.278, 0.092)	-0.681(.278, 0.007**)	-1.041(.386, 0.003**)	-0.382(.298, 0.100)
Speed	0.294(.258, 0.127)	0.534(.279, 0.028*)	0.826(.353, 0.010**)	0.301(.275, 0.137)
Press	-0.008(.151, 0.48)	-0.065(.155, 0.338)	-0.035(.157, 0.412)	-0.008(.153, 0.480)
T_{time}	14.0	62.6	2.4	1.7
		$t = 7:00pm$ $Saturday$		
ρ	0.972(.185, 0**)	0.107(.200, 0.297)	0.038(.266, 0.444)	0.926(.170, 0**)
Intercept	0.019(.084, 0.411)	0.006(.019, 0.368)	-0.003(.024, 0.456)	0.025(.068, 0.358)
Temp	-0.116(.189, 0.269)	-0.635(.212, 0.001**)	-0.819(.266, 0.001**)	-0.159(.194, 0.206)
Speed	0.057(.229, 0.401)	0.447(.242, 0.032*)	0.584(.274, 0.016*)	0.094(.235, 0.344)
Press	0.049(.152, 0.374)	0.068(.172, 0.346)	0.025(.178, 0.444)	0.043(.159, 0.394)
T_{time}	13.1	62.6	2.7	1.7

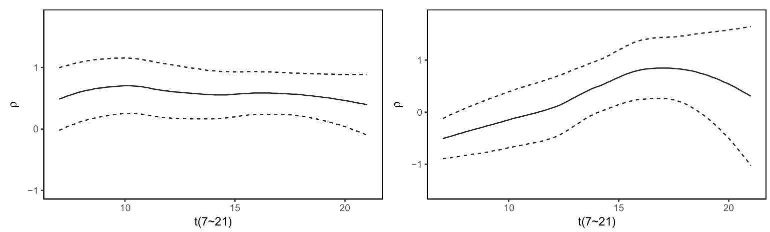


FIGURE 3: The COME estimates of ρ_t (solid) and the point-wise confidence intervals (dotted) for Thursday (left) and Saturday (right) as t varies from 7am to 9pm.

6. CONCLUSION

We propose a new method, namely, COME, that combines the covariance and the first moment conditions to draw inference based on SAR models. Under the regular conditions, we show that the estimator for the regression coefficients $\boldsymbol{\beta}$ and ρ are $\sqrt{n}-$consistent, and establish the asymptotically distribution. The simulation studies indicated a good performance of the proposal, and found that COME was more efficient than GMM and 2SLS in the cases considered. It was comparable to GMM and 2SLS in of computational efficiency, was more computational efficient than MLE.

SAR models require the specification of the spatial weights matrix (Bhattacharjee and Jensen-Butler, 2013). However, the estimation results may be highly sensitive to the specification of the spatial weights (Ahrens and Bhattacharjee, 2015). A data-driven estimator for the spatial weights used in SAR models has attracted attention. Moreover, how to extend SAR models to accommodate temporally correlated data warrant more investigation. One need to carefully design inference procedures spatio-temporal data.

This paper has focused on low dimensional covariates. With more predictors collected in air pollution studies, it will be of interest to extend the current theoretical results to spatial data with high-dimensional covariates. We will pursue this elsewhere.

BIBLIOGRAPHY

Ahrens, A. and Bhattacharjee, A. (2015). Two-step lasso estimation of the spatial weights matrix. *Econometrics*, 3(1):128–155.

Amemiya, T. (1985). *Advanced econometrics*. Harvard University Press, Cambridge.

Anselin, L. (1988a). The maximum likelihood approach to spatial process models. *A Companion to Theoretical Econometrics*.

Anselin, L. (1988b). *Spatial Econometrics: methods and models.* The Netherlands: Kluwer Academic Publishers.

Anselin, L. (1990). Some robust approaches to testing and estimation in spatial econometrics. *Regional Science and Urban Economics*, 20(2):141–163.

Anselin, L. and Bera, A. (1998). *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*. New York: Marcel Dekker.

Bell, K. P. and Bockstael, N. E. (2000). Applying the generalized-moments estimation approach to spatial problems involving micro-level data. *Review of Economics and Statistics*, 82(1):72–82.

Bhattacharjee, A. and Jensen-Butler, C. (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics*, 43(4):617–634.

Bivand, R. and Anselin, e. (2011). Spatial dependence: weighting schemes, statistics and models.

Cliff, A. D. and Ord, J. K. (1973). Spatial autocorrelation. *International Encyclopedia of Human Geography*, 14(5):308–316.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley Classics Library.

Huang, D., Lan, W., Zhang, H. H., and Wang, H. (2019). Least squares estimation of spatial autoregressive models for large-scale social networks. *Electronic Journal of Statistics*, 13(1):1135–1165.

Jin, F. and Lee, L. F. (2012). On the bootstrap for spatial econometric models. *Preprint*.

Kazar, B. M. and Celik, M. (2012). *Spatial autoregression (SAR) model: parameter estimation techniques*. Springer Science & Business Media.

Kelejian, H. H. and Prucha, I. R. (1997). Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. *International regional science review*, 20(1-2):103–111.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance & Economics*, 17(1):99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

Kelejian, H. H. and Prucha, I. R. (2001). On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, 104(2):219–257.

Land, K. C. and Deane, G. (1992). On the large-sample estimation of regression models with spatial-or network-effects terms: A two-stage least squares approach. *Sociological methodology*, 22:221–248.

Lee, L. F. (2001). Generalized method of moments estimation of spatial autoregressive processes. *Ohio State University, Columbus, OH*.

Lee, L. F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, 22(4):307–335.

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

Lee, L. F. (2007a). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137(2):489–514.

Lee, L. F. (2007b). The method of elimination and substitution in the gmm estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 140(1):155–189.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Liu, X., Lee, L. F., and R, B. C. (2010). An efficient gmm estimator of spatial autoregressive models. *Journal of Econometrics*, 159(2):303–319.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Publications of the American Statistical Association*, 70(349):120–126.

Rangel, T. F. L. V. B. and Bini, L. M. (2006). Towards an integrated computational tool for spatial analysis in macroecology and biogeography. *Global Ecology & Biogeography*, 15(4):321–327.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Smirnov, O. and Anselin, L. (2001). *Fast maximum likelihood estimation of very large spatial autoregression models: a characteristic polynomial approach*. Elsevier Science Publishers B. V.

Trefethen, L. N. and Bau, D. (1997). *Numerical linear algebra*, volume 50. Siam.

Yao, L., Lu, N., Yue, X., Du, J., and Yang, C. (2015). Comparison of hourly pm2. 5 observations between urban and suburban areas in beijing, china. *International journal of environmental research and public health*, 12(10):12264–12276.

Zhang, S., Guo, B., Dong, A., He, J., Xu, Z., and Chen, S. X. (2017). Cautionary tales on air-quality improvement in beijing. *Scientia Forestalis*, 473(2205).

Zhang, X. and Yu, J. (2018). Spatial weights matrix selection and model averaging for spatial autoregressive models. *Journal of Econometrics*, 203(1):1–18.

Zhu, X., Huang, D., Pan, R., and Wang, H. (2020). Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics*, 215(2):591–606.

Appendix A: notation

Let $\boldsymbol{\gamma} = (\rho, \boldsymbol{\beta}')'$, $\boldsymbol{\gamma}_0 = (\rho_0, \boldsymbol{\beta}_0')'$ the true value of $\boldsymbol{\gamma}$, $S_n(\rho) = I_n - \rho\mathbf{W}$ for any value of ρ, $S_n = S_n(\rho_0) = I_n - \rho_0\mathbf{W}$.

The following notation is used in Theorem 1. Denote $k_n(\boldsymbol{\gamma}) = \mathbf{W}'C_n b_n(\boldsymbol{\gamma})$, $P_n(\rho) = (I_n - \rho\mathbf{W})S_n^{-1}$, $K_n(\rho) = \mathbf{W}'C_n P_n(\rho)$, $J_n(\rho) = P_n(\rho)'C_n P_n(\rho)$ and $h_n(\boldsymbol{\gamma}) = P_n(\rho)'C_n b_n(\boldsymbol{\gamma})$. Additionally, the following notation is used in Theorem 2. Let $q_n = (S_n^{-1}\mathbf{X}\boldsymbol{\beta}_0)'\mathbf{W}'C_n$ and

$$
\begin{aligned}
&\Sigma \\
&= \frac{1}{n}\begin{pmatrix}
\mathrm{Var}(1_n'\boldsymbol{\epsilon}), \mathrm{Cov}(1_n'\boldsymbol{\epsilon}, q_n\boldsymbol{\epsilon}), \mathrm{Cov}(1_n'\boldsymbol{\epsilon}, \mathbf{X}'C_n\boldsymbol{\epsilon}), \mathrm{Cov}(1_n'\boldsymbol{\epsilon}, \mathbf{X}'\boldsymbol{\epsilon}) \\
\mathrm{Cov}(1_n'\boldsymbol{\epsilon}, q_n\boldsymbol{\epsilon}), \mathrm{Var}(q_n\boldsymbol{\epsilon}), \mathrm{Cov}(q_n\boldsymbol{\epsilon}, \mathbf{X}'C_n\boldsymbol{\epsilon}), \mathrm{Cov}(q_n\boldsymbol{\epsilon}, \mathbf{X}'\boldsymbol{\epsilon}) \\
\mathrm{Cov}(1_n'\boldsymbol{\epsilon}, \mathbf{X}'C_n\boldsymbol{\epsilon}), \mathrm{Cov}(q_n\boldsymbol{\epsilon}, \mathbf{X}'C_n\boldsymbol{\epsilon}), \mathrm{Var}(\mathbf{X}'C_n\boldsymbol{\epsilon}), \mathrm{Cov}(\mathbf{X}'C_n\boldsymbol{\epsilon}, \mathbf{X}'\boldsymbol{\epsilon}) \\
\mathrm{Cov}(1_n'\boldsymbol{\epsilon}, \mathbf{X}'\boldsymbol{\epsilon}), \mathrm{Cov}(q_n\boldsymbol{\epsilon}, \mathbf{X}'\boldsymbol{\epsilon}), \mathrm{Cov}(\mathbf{X}'\boldsymbol{\epsilon}, \mathbf{X}'C_n\boldsymbol{\epsilon}), \mathrm{Var}(\mathbf{X}'\boldsymbol{\epsilon}).
\end{pmatrix}
\end{aligned}
$$

Appendix B: Lemmas

Appendix B: Proofs of Theorems 1 and 2

We rewrite (4) as $U(\boldsymbol{\gamma}) = \begin{pmatrix} U_1(\boldsymbol{\gamma}) \\ U_2(\boldsymbol{\gamma}) \end{pmatrix}$, where

$$
\begin{aligned}
U_1(\boldsymbol{\gamma}) &= \frac{-4M_n(\boldsymbol{\gamma})}{n}(\mathbf{WY})'C_n(\mathbf{Y} - \rho\boldsymbol{W}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \\
U_2(\boldsymbol{\gamma}) &= \frac{-4M_n(\boldsymbol{\gamma})}{n^2}\mathbf{X}'C_n(\mathbf{Y} - \rho\boldsymbol{W}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) - \frac{2\lambda}{n}\mathbf{X}'(\mathbf{Y} - \rho\boldsymbol{W}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}).
\end{aligned} \tag{9}
$$

To prove Theorems 1 and 2, we introduce two needed Lemmas.

Lemma C.1 (van der Vaart, 1998, p.61) *Let $U(\boldsymbol{\gamma})$ be a random vector-value functions and $EU(\boldsymbol{\gamma})$ be a deterministic vector valued function of $\boldsymbol{\gamma}$. Suppose* *that there is a* $\boldsymbol{\gamma}_0 \in \Gamma$ *such that*

$$\inf_{d(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)\geq\epsilon} \|EU(\boldsymbol{\gamma})\| > 0 = EU(\boldsymbol{\gamma}_0), \tag{A.1}$$

and for every $\epsilon > 0$

$$\sup_{\boldsymbol{\gamma}\in\Gamma} \|U(\boldsymbol{\gamma}) - EU(\boldsymbol{\gamma})\| \xrightarrow{P} 0. \tag{A.2}$$

Then any sequence of estimators $\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\gamma}}_n$ *such that* $U(\widehat{\boldsymbol{\gamma}}) = o_p(1)$ *converge in probability to* $\boldsymbol{\gamma}_0$.

Proof of Theorem 1. With Lemma C.1, we only need to verify the conditions $(A.1)$ and $(A.2)$, which lead to $\|\widehat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0\| = o_p(1)$.

Proof of $(A.1)$. We first verify $(A.1)$. Applying **Lemma C.1**, there exists a convergence sequence $\boldsymbol{\gamma}_n$ such that $\boldsymbol{\gamma}_n \to \boldsymbol{\gamma}^*$ in probability, and it follows that $\boldsymbol{\gamma}^* \in \Gamma$ where Γ is a compact set. As

$$EU(\boldsymbol{\gamma}^*) = EU(\boldsymbol{\gamma}^*) - EU(\widehat{\boldsymbol{\gamma}}) + EU(\widehat{\boldsymbol{\gamma}}) - U(\widehat{\boldsymbol{\gamma}}),$$

it follows from $(A.2)$ that $EU(\boldsymbol{\gamma}^*) = 0$. Since $EU(\boldsymbol{\gamma}) = 0$ has a unique solution at $\boldsymbol{\gamma}_0$, we conclude that $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}^*$ which ensures the uniform consistency of $\boldsymbol{\gamma}$. This completes the proof of Theorem 1.

Proof of $(A.2)$. For convenience, we only give the proof of $\sup_{\boldsymbol{\gamma}\in\Gamma} |U_1(\boldsymbol{\gamma}) - EU_1(\boldsymbol{\gamma})| \xrightarrow{P} 0$. Similar arguments lead to the conclusion about $\sup_{\boldsymbol{\gamma}\in\Gamma} |U_2(\boldsymbol{\gamma}) - EU_2(\boldsymbol{\gamma})| \xrightarrow{P} 0$. After replacing $\boldsymbol{Y} = S_n^{-1}\mathbf{X}\boldsymbol{\beta}_0 + S_n^{-1}\boldsymbol{\epsilon}$, $\mathbf{Y} - \rho\mathbf{W}\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} =$

$b_n(\boldsymbol{\gamma}) + P_n(\rho)\boldsymbol{\epsilon}$ in (9), we consider the following decomposition

$$\sup_{\boldsymbol{\gamma}\in\Gamma} |U_1(\boldsymbol{\gamma}) - EU_1(\boldsymbol{\gamma})| \le I_1 + I_2 + I_3 + I_4 + I_5,$$

where

$$\begin{aligned}
I_1 &= \frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|b_{1,n}(\boldsymbol{\gamma})q_n b_n(\boldsymbol{\gamma}) - E\{b_{1,n}(\boldsymbol{\gamma})q_n b_n(\boldsymbol{\gamma})\}\Big|,\\
I_2 &= \frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|b_{1,n}(\boldsymbol{\gamma})k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon} + b_{1,n}(\boldsymbol{\gamma})q_n P_n(\rho)\boldsymbol{\epsilon} + 2h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}q_n b_n(\boldsymbol{\gamma})\\
&\qquad - E\{b_{1,n}(\boldsymbol{\gamma})k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon} + b_{1,n}(\boldsymbol{\gamma})q_n P_n(\rho)\boldsymbol{\epsilon} + 2h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}q_n b_n(\boldsymbol{\gamma})\}\Big|,\\
I_3 &= \frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|b_{1,n}(\boldsymbol{\gamma})\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon} + 2\boldsymbol{\epsilon}'h_n(\boldsymbol{\gamma})(S_n^{-1}\mathbf{X}\beta_0)'K_n(\rho)\boldsymbol{\epsilon}\\
&+ \boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}q_n b_n(\boldsymbol{\gamma}) + 2\boldsymbol{\epsilon}'h_n(\boldsymbol{\gamma})k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon} - E\{b_{1,n}(\boldsymbol{\gamma})\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\\
&+ 2\boldsymbol{\epsilon}'h_n(\boldsymbol{\gamma})(S_n^{-1}\mathbf{X}\beta_0)'K_n(\rho)\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}q_n b_n(\boldsymbol{\gamma}) + 2\boldsymbol{\epsilon}'h_n(\boldsymbol{\gamma})k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon}\}\Big|,\\
I_4 &= \frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|2h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{q_n P_n(\rho)\boldsymbol{\epsilon} + k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon}\right\}\\
&+ \boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\} - 2Eh_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\\
&- E\Big[\boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\}\Big] - E\boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{q_n P_n(\rho)\boldsymbol{\epsilon} + k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon}\right\}\Big|.
\end{aligned}$$

For I_1, we can directly get that $I_1 = 0$. Applying Assumptions 2-5 we can derive $I_2 = O_p(\frac{1}{\sqrt{n}})$, and $I_3 = o_p(1)$.

Note that

$$
\begin{aligned}
I_4 \leq & \frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|2h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon} - 2Eh_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\Big| \\
& +\frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|\boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{q_nP_n(\rho)\boldsymbol{\epsilon} + k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon}\right\} - E\boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{q_nP_n(\rho)\boldsymbol{\epsilon} + k_n(\boldsymbol{\gamma})'S_n^{-1}\boldsymbol{\epsilon}\right\}\Big| \\
& +\frac{4}{n}\sup_{\boldsymbol{\gamma}\in\Gamma}\Big|\boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\} - E\big[\boldsymbol{\epsilon}'J_n(\rho)\boldsymbol{\epsilon}\left\{\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\}\big]\Big| \\
& =: I_{41} + I_{42} + I_{43}.
\end{aligned}
$$

To show the linear-quadratic form in Kelejian and Prucha (2001) asymptotically follows a normal distribution, we can show

$$
\begin{aligned}
&\left\{h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\}^2 - E\left\{h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\}^2 \\
&= O_p(1),
\end{aligned} \tag{10}
$$

and obtain

$$
\begin{aligned}
&\left\{h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\}^2 - E\left\{h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon} + \boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}\right\}^2 \\
&= O_p(1) + 2h_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon} - 2Eh_n(\boldsymbol{\gamma})'\boldsymbol{\epsilon}\boldsymbol{\epsilon}'(S_n^{-1})'K_n(\rho)\boldsymbol{\epsilon}.
\end{aligned} \tag{11}
$$

Combining (10) with (11), we have $I_{41} = o_p(1)$. Similarly, we can obtain $I_{42} = I_{43} = o_p(1)$. Consequently we complete the proof of $(A.2)$.

Proof of Theorem 2. Using the Taylor expansion, we have

$$
U(\widehat{\rho}, \widehat{\boldsymbol{\beta}}) = U(\rho_0, \boldsymbol{\beta}_0) + \begin{pmatrix} \dot{U}_1{}'(\rho_0, \boldsymbol{\beta}_0) \\ \dot{U}_2{}'(\rho_0, \boldsymbol{\beta}_0) \end{pmatrix} \begin{pmatrix} \widehat{\rho} - \rho_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} + o_p(1), \tag{12}
$$

where $\dot{U}_1(\rho_0, \boldsymbol{\beta}_0) = \frac{\partial U_1(\rho_0, \boldsymbol{\beta}_0)}{\partial \gamma}|_{\gamma=\gamma_0} = \begin{pmatrix} \dot{U}_{1\rho}(\boldsymbol{\gamma}_0) \\ \dot{U}_{1\boldsymbol{\beta}}(\boldsymbol{\gamma}_0) \end{pmatrix}$, $\dot{U}_2(\rho_0, \boldsymbol{\beta}_0) = \frac{\partial U_2(\rho_0, \boldsymbol{\beta}_0)}{\partial \gamma}|_{\gamma=\gamma_0} = \begin{pmatrix} \dot{U}_{2\rho}(\boldsymbol{\gamma}_0) \\ \dot{U}_{2\boldsymbol{\beta}}(\boldsymbol{\gamma}_0) \end{pmatrix}$. Using Lemmas 1-3 and Theorem 1 in Kelejian and Prucha (2001) can show that

$$\begin{pmatrix} \dot{U}_1{}'(\rho_0, \boldsymbol{\beta}_0) \\ \dot{U}_2{}'(\rho_0, \boldsymbol{\beta}_0) \end{pmatrix} = \Psi + o_p(1), \tag{13}$$

where Ψ are defined in Assumption 4.

On the other hand, the estimating equations of $U(\boldsymbol{\gamma}_0)$ can be

$$U(\boldsymbol{\gamma}_0) = \begin{pmatrix} -4\boldsymbol{\epsilon}' C_n \boldsymbol{\epsilon}\{\frac{1}{n} q_n \boldsymbol{\epsilon} + \frac{1}{n}\boldsymbol{\epsilon}' G_n' C_n \boldsymbol{\epsilon}\} \\ -4\boldsymbol{\epsilon}' C_n \boldsymbol{\epsilon} \frac{1}{n^2} \mathbf{X}' C_n \boldsymbol{\epsilon} - \frac{2\lambda}{n} \mathbf{X}' \boldsymbol{\epsilon} \end{pmatrix} \widehat{=} \begin{pmatrix} U_1(\boldsymbol{\gamma}_0) \\ U_2(\boldsymbol{\gamma}_0) \end{pmatrix},$$

where $G_n = \mathbf{W} S_n^{-1}$. Since $\frac{1}{n}\boldsymbol{\epsilon}'\boldsymbol{\epsilon} \xrightarrow{P} \sigma^2$ and $\frac{1}{n}\boldsymbol{\epsilon}' G_n' C_n \boldsymbol{\epsilon} \xrightarrow{P} \frac{1}{n}\frac{\mathrm{tr}(G_n' C_n)}{n}$, $\sqrt{n} U(\boldsymbol{\gamma}_0)$ can be rewritten as

$$\sqrt{n} U(\boldsymbol{\gamma}_0) = \begin{pmatrix} -(\frac{2}{\sqrt{n}} 1_n' \boldsymbol{\epsilon})^2 \cdot \{\frac{1}{\sqrt{n}} q_n \boldsymbol{\epsilon} + \frac{\mathrm{tr}(G_n' C_n)}{n}\} - 4\sigma^2 \{\frac{1}{\sqrt{n}} q_n \boldsymbol{\epsilon} + \frac{\mathrm{tr}(G_n' C_n)}{n}\} \\ -(\frac{2}{\sqrt{n}} 1_n' \boldsymbol{\epsilon})^2 \frac{1}{\sqrt{n}} \mathbf{X}' C_n \boldsymbol{\epsilon} - 4\sigma^2 \frac{1}{\sqrt{n}} \mathbf{X}' C_n \boldsymbol{\epsilon} - \frac{2\lambda}{\sqrt{n}} \mathbf{X}' \boldsymbol{\epsilon} \end{pmatrix}.$$

Then, multivariate CLT can yield

$$\begin{pmatrix} \frac{1}{\sqrt{n}} 1_n' \boldsymbol{\epsilon} \\ \frac{1}{\sqrt{n}} q_n \boldsymbol{\epsilon} \\ \frac{1}{\sqrt{n}} \mathbf{X}' C_n \boldsymbol{\epsilon} \\ \frac{1}{\sqrt{n}} \mathbf{X}' \boldsymbol{\epsilon} \end{pmatrix} \rightsquigarrow \boldsymbol{N}(0, \Sigma),$$

where Σ is defined in Appendix A. Then we use **Z** to represent a multivariate normal variable with mean **0** and covariance Σ. Let $g: R^{p+2} \to R^{p+1}$ be a continuous map and $g(u,v,s,w) = \begin{pmatrix} -4u^2\{v + \frac{\mathrm{tr}(G_n' C_n)}{n}\} - 4\sigma^2\{v + \frac{\mathrm{tr}(G_n' C_n)}{n}\} \\ -4u^2 s - 4\sigma^2 s - 2\lambda w \end{pmatrix}$ correspond with $u, v \in R$ and $w, s \in R^p$. Furthermore, by Assumption 4 and the Law of Large Numbers, we note that Ψ^{-1} exists. Therefore, by the Continuous Mapping Theorem, we have

$$\sqrt{n} \begin{pmatrix} \widehat{\rho} - \rho_0 \\ \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \rightsquigarrow \Psi^{-1} g(\boldsymbol{Z}), \tag{14}$$

after combining (12) with (13).