

Article

Information Theory Estimators for the First-Order Spatial Autoregressive Model

Evgeniy V. Perevodchikov ¹, Thomas L. Marsh ^{2,*} and Ron C. Mittelhammer ²

¹ The Institute for Innovation, Tomsk State University of Control Systems and Radioelectronics, Tomsk 634050, Russia; E-Mail: evp@2i.tusur.ru

² School of Economic Sciences, Washington State University, PO Box 646210, Pullman, WA 99164, USA; E-Mail: mittelha@wsu.edu

* Author to whom correspondence should be addressed; E-Mail: tl_marshall@wsu.edu; Tel.: +1-509-335-8597; Fax: +1-509-335-1173.

Received: 25 May 2012; in revised form: 29 June 2012 / Accepted: 30 June 2012 /

Published: 4 July 2012

Abstract: Information theoretic estimators for the first-order spatial autoregressive model are introduced, small sample properties are investigated, and the estimator is applied empirically. Monte Carlo experiments are used to compare finite sample performance of more traditional spatial estimators to three different information theoretic estimators, including maximum empirical likelihood, maximum empirical exponential likelihood, and maximum log Euclidean likelihood. Information theoretic estimators are found to be robust to selected specifications of spatial autocorrelation and may dominate traditional estimators in the finite sample situations analyzed, except for the quasi-maximum likelihood estimator which competes reasonably well. The information theoretic estimators are illustrated via an application to hedonic housing pricing.

Keywords: information theoretic estimators; first order spatial autocorrelation; maximum empirical likelihood; maximum empirical exponential likelihood; maximum log-Euclidean likelihood; finite sample properties

PACS Codes: 02.50.Ng; 02.50.Tt

1. Introduction

Information Theoretic (IT) estimators are alternatives to traditional estimators [1–3] and have been applied in a variety of modeling contexts [4–8]. The IT estimators are semi-nonparametric in nature and lead to a wide and flexible class of estimators that are minimally power divergent from any reference distribution and both consistent with a nonparametric specification of the model and a parametric specification of empirical moment conditions. The estimators are generally asymptotically efficient and can have superior sampling properties in terms of mean squared error relative to traditional estimators [9]. Application of IT estimators to spatial regression models is currently limited [10,11]. In order for such estimators to be more widely adopted, it is important that analysts be aware of the ways in which they can be implemented empirically, and that they also have an understanding of the estimators' finite sample properties for work with smaller-to-medium size samples prevalent in applied work.

Spatial estimators are particularly relevant when variables, related through their location, can influence economic behavior and equilibrium outcomes. For instance, spatial patterns are found in regional adoption of agricultural technology [12] and in a household's behavior when a household gains utility in consuming bundles similar to those consumed by its neighbors [13]. Spatial spillover effects arise from technical innovations [14] and geographic proximity of a firm's competitors might directly determine its marketing strategy [15]. Recent empirical studies in agriculture focused on spatial aspects of technology adoption, structure of production, market efficiency, arbitrage, and integration [16–23].

The objectives of this paper are to introduce a generalized information theoretic estimator of the first-order spatial autoregressive model, compare finite sample properties of selected IT estimators with traditional ones, and provide an illustration of the estimator's implementation in the context of an empirical application. The IT estimators analyzed include the maximum empirical likelihood, maximum empirical exponential likelihood, and maximum log Euclidean likelihood estimators. Finite sample properties of these estimators are investigated in the context of an extensive Monte Carlo analysis conducted over a range of finite sample sizes, a range of spatial autoregressive coefficients, selected forms of heteroscedasticity, different distributional assumptions on disturbance terms, and alternative forms of spatial weight matrices typically used in applied econometric work. The estimators are compared to each other as well as to traditional estimators on the basis of root mean square error, and response functions are used to summarize findings of all Monte Carlo experiments. Then each of the IT estimators are applied to the Harrison and Rubinfeld [24] example of hedonic housing prices and demand for clean air [25]. We draw implications and provide concluding remarks in the final section of the paper.

2. First-Order Spatial Autoregressive Model and Traditional Estimators

Consider a first-order spatial autoregressive model:

$$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad |\rho| < 1 \quad (1)$$

where $\mathbf{Y}$ is a $n \times 1$ dependent variable vector, $\mathbf{X}$ is $n \times k$ matrix of exogenous variables with full column rank, $\mathbf{W}$ is a $n \times n$ spatial proximity matrix of constants, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, ρ is a

scalar spatial autoregressive parameter, and ε is an $n \times 1$ vector of unobserved residuals with $E(\varepsilon) = \mathbf{0}$ and $E(\varepsilon\varepsilon') = \sigma^2\mathbf{\Omega}$ [26–28]. The spatial proximity matrix is row normalized [27] or a “row stochastic” [29] matrix, so that row sums are unity and all diagonal elements are zero, *i.e.*, the spatial weight matrix is nonsingular and the reduced form of (1) is well defined [30]. The weights are fixed in repeated sampling and are constructed based on the spatial configuration of points making up a spatial sample. LeSage and Pace [31] provide proper limits on the parameter space for the spatial autoregressive parameter and justification for using an interval from -1 to 1 in most applied situations. This specification of disturbances allows for general patterns of autocorrelation and heteroscedasticity.

A relatively wide variety of alternative estimation procedures for the model, with and without spatially correlated disturbances, have been proposed in the literature. Bayesian estimators are not considered in the current study, but can be found in [32]. The ordinary least squares (OLS) estimator of (1) is generally biased and inconsistent due to simultaneity bias [33–35]. Under appropriate regularity assumptions, the maximum likelihood (ML) estimator of (1) (with multivariate normal distribution of disturbances) is consistent and asymptotically efficient [27,36]. The quasi maximum likelihood (QML) estimator is consistent and asymptotically normal [36]. The maximum likelihood approaches were not practical for large datasets until innovations were made (e.g., Barry and Pace [37]). Generalized method of moments (GMM) estimation has been proposed for its computational simplicity, less restrictive distributional assumptions, and good asymptotic properties [38–42]. The generalized spatial two-stage least squares (GS2SLS) estimator is consistent and asymptotically normal [38], and the best spatial two-stage least squares (BS2SLS) estimator is also an asymptotically optimal instrumental variable estimator [43]. However the two-stage least squares estimators are inefficient relative to ML estimator and may be inconsistent [39,43]. The best generalized method of moments (BGMM) estimator incorporates additional moment conditions and attains the same limiting distribution as the ML estimator (with normal disturbances) [41]. A computationally simple sequential GMM estimator based on optimization of a concentrated objective function with respect to a single spatial effect parameter may be as efficient as BGMM estimator [42]. All of the above estimators of the first-order spatial autoregressive model and their asymptotic properties were derived under the assumption that the disturbance term is homoscedastic.

Regarding the computational implementation of the preceding estimators, assuming a normally distributed disturbance term, $\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{\Omega})$, as is often done in applications, the ML estimator of the parameter vector in (1) is given by:

$$\begin{bmatrix} \hat{\beta}_{ML} \\ \hat{\rho}_{ML} \\ \hat{\sigma}_{ML}^2 \end{bmatrix} = \arg \max_{\beta, \rho, \sigma^2} \left[-\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln(|\mathbf{\Omega}|) - \frac{1}{2\sigma^2} \varepsilon' \mathbf{\Omega}^{-1} \varepsilon + \ln(|(\mathbf{I} - \rho \mathbf{W})|) \right] \quad (2)$$

where $\mathbf{I}$ is a conformable identity matrix, $\varepsilon = (\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\beta$, and $|\cdot|$ is a determinant operator [27,43]. A difficulty in practice is that there is generally insufficient information with which to specify the parametric form of the likelihood function, so that (2) then represents a quasi-ML approach to estimation. Moreover, the structure of the covariance matrix of the disturbance term is generally unknown. One might contemplate estimation based on less restrictive assumptions about the existence of zero valued moment conditions.

The model in (1) can be estimated by the computationally less complicated two-stage least squares (2SLS) method. Instrumental variables are generated from exogenous regressors and a spatial weight matrix [38,41]. The 2SLS estimator is specified by:

$$\begin{bmatrix} \hat{\beta}_{2SLS} \\ \hat{\rho}_{2SLS} \end{bmatrix} = \left[\begin{pmatrix} (\mathbf{WY})' \\ \mathbf{X}' \end{pmatrix} \mathbf{Z}_{(p)} (\mathbf{WY}, \mathbf{X}) \right]^{-1} \left[\begin{pmatrix} (\mathbf{WY})' \\ \mathbf{X}' \end{pmatrix} \mathbf{Z}_{(p)} \mathbf{Y} \right] \quad (3)$$

where $\mathbf{Z}$ is an $n \times m$ instrumental variable (IV) matrix with full column rank and $m \geq k$, and $\mathbf{Z}_{(p)} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the orthogonal projector of the column space of $\mathbf{Z}$.

The consistent and asymptotically normal generalized spatial 2SLS estimator (GS2SLS) suggested by Kelejian and Prucha [38] defines the IV matrix as $\mathbf{Z} = [\mathbf{X}, \mathbf{WX}, \mathbf{WWX}]$. The asymptotically optimal best spatial 2SLS (BS2SLS) estimator proposed by Lee [43] defines the IV matrix as $\mathbf{Z} = [\mathbf{X}, \mathbf{W}(\mathbf{I} - \hat{\rho}\mathbf{W})^{-1}\mathbf{X}^*\hat{\beta}]$, where $\mathbf{X}^*$ has no intercept column, and $\hat{\beta}$ and $\hat{\rho}$ are the estimates from the first stage.

The consistent generalized method of moments (GMM) estimator of unknown parameters in (1) can be derived from the empirical moments:

$$n^{-1}\mathbf{Z}'[(\mathbf{I} - \rho\mathbf{W})\mathbf{Y} - \mathbf{X}\beta] = \mathbf{0} \quad (4)$$

where $\mathbf{Z}$ is an $(n \times m)$ instrumental variables matrix with full column rank and $m \geq k$. The parameters can be estimated as:

$$\begin{bmatrix} \hat{\beta}_{GMM} \\ \hat{\rho}_{GMM} \end{bmatrix} = \arg \min_{\beta, \rho} \left[n^{-1}\mathbf{Z}'[(\mathbf{I} - \rho\mathbf{W})\mathbf{Y} - \mathbf{X}\beta] \right]' \hat{\mathbf{W}}_n \left[n^{-1}\mathbf{Z}'[(\mathbf{I} - \rho\mathbf{W})\mathbf{Y} - \mathbf{X}\beta] \right] \quad (5)$$

where traditionally $\hat{\mathbf{W}}_n$ is an estimate of the asymptotically optimal weight matrix equal to the inverse of the estimated covariance matrix of the moment conditions [44].

3. Information Theoretic Estimators

Previous information theoretic econometric literature relating to the spatial autoregressive model of type (1) has focused on generalized maximum entropy. Marsh and Mittelhammer [10] formulate generalized maximum entropy estimators for the general linear model and the censored regression model when there is first order spatial autoregression in the dependent variable. Monte Carlo experiments were provided to compare the performance of spatial entropy estimators relative to classical spatial estimators. Fernandez-Vazquez, Mayor-Fernandez, and Rodriguez-Valez [11] compare some traditional spatial estimators with generalized maximum entropy and generalized cross entropy estimators by means of Monte Carlo simulations. Bernardini Papalia [45] applied generalized cross entropy for modeling economic aggregates and estimating their sub-group (sub-area) decomposition when no individual or sub-group data are available.

To extend the literature relating to IT estimators for the spatial autoregressive model, we first modify the moment conditions in (4). The empirical moments for the IT estimator of the parameters in (1) are specified as:

$$(\mathbf{p} \odot \mathbf{Z})' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}] = \mathbf{0} \quad (6)$$

where $\mathbf{p}$ is an $(n \times 1)$ vector of unknown empirical probability weights supported on a sample outcome $(\mathbf{Y}, \mathbf{X})$, and $\odot$ denotes the extended Hadamard (elementwise) product operator. For example, the concept of empirical probability weights is developed in the framework of empirical likelihood function [46]. The value of the empirical likelihood function is the maximum empirical probability $\prod_{i=1}^n p_i$ that can be assigned to a random sample outcome of $(\mathbf{Y}, \mathbf{X})$ among all distributions of probability $\mathbf{p}$ supported on the (X_i, Y_i) 's that satisfy the empirical moment conditions, adding up restriction $\sum_{i=1}^n p_i = 1$, and nonnegativity constraint $p_i \geq 0, \forall i = 1, \dots, n$ [47].

Similarly in (6) the empirical weights p_i are treated as unknown parameters of a multinomial distribution for n different types of data outcomes presented by the sample such that $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0, \forall i = 1, \dots, n$. In contrast, the empirical moment conditions of the GMM approach in (4) restrict $p_i = \frac{1}{n}$ for $i = 1, \dots, n$.

A generalized IT estimator of the unknown parameters in (1) is specified as the values of the parameters that solve the following optimization problem subject to empirical moments, normalization, and nonnegativity constraints given by:

$$\arg \max_{\mathbf{p}, \boldsymbol{\beta}, \rho} \left[\phi(\mathbf{p}) \text{ s.t. } (\mathbf{p} \odot \mathbf{Z})' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}] = \mathbf{0}, \sum_{i=1}^n p_i = 1, p_i \geq 0, \forall i \right] \quad (7)$$

The objective function $\phi(\mathbf{p})$ in (7) is the negative of the Cressie–Read power divergence statistic [48] given by:

$$CR(\mathbf{p}, \mathbf{q}, \gamma) = \frac{1}{\gamma(\gamma + 1)} \sum_{i=1}^n p_i \left[\left(\frac{p_i}{q_i} \right)^\gamma - 1 \right] \quad (8)$$

where γ is a given real-valued constant, and $\mathbf{p} = [p_1, \dots, p_n]'$ and $\mathbf{q} = [q_1, \dots, q_n]'$ are $(n \times 1)$ vectors of estimated and empirical probability densities supported on a sample outcome $(\mathbf{Y}, \mathbf{X})$. The empirical probability density is taken to be fixed and the estimated probability weights p_i are calculated to solve the moment equations in (6) while being as minimally divergent from the empirical probability distribution as possible. The specification in (7) is a general representation of the objective function which defines a class of IT estimators for the model in (1). This estimation approach circumvents the need for estimating a weight matrix in the GMM procedure.

The Lagrange form of the extremum problem in (7) is given by:

$$L(\mathbf{p}, \boldsymbol{\beta}, \rho, \boldsymbol{\lambda}, \eta) = \phi(\mathbf{p}) - \boldsymbol{\lambda}' (\mathbf{p} \odot \mathbf{Z})' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}] - \eta \left(\sum_{i=1}^n p_i - 1 \right) \quad (9)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]'$ is $(m \times 1)$ vector and η is scalar of Lagrange multipliers. First order conditions are:

$$\frac{\partial L}{\partial p_i} = \frac{\partial \phi(\mathbf{p})}{\partial p_i} - \boldsymbol{\lambda}' \mathbf{Z}[i, \cdot]' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}[i, \cdot]\boldsymbol{\beta}] - \eta = 0, \forall i = 1, \dots, n \quad (10)$$

$$\frac{\partial L}{\partial \beta} = \mathbf{X}'(\mathbf{p} \odot \mathbf{Z})\lambda = \mathbf{0} \quad (11)$$

$$\frac{\partial L}{\partial \rho} = \lambda'(\mathbf{p} \odot \mathbf{Z})' \mathbf{W} \mathbf{Y} = 0 \quad (12)$$

$$\frac{\partial L}{\partial \lambda} = (\mathbf{p} \odot \mathbf{Z})' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\beta] = \mathbf{0} \quad (13)$$

$$\frac{\partial L}{\partial \eta} = \sum_{i=1}^n p_i - 1 = 0 \quad (14)$$

and $p_i \geq 0, \forall i$. The first set of equations in (10) links the empirical probability weights p_i 's to the unknown parameters β , ρ , and λ through the empirical moment conditions. The set of equations in (11)–(13) modifies traditional orthogonality conditions of GMM estimation. Equation (14) is a standard normalization condition for the empirically estimated probability weights.

In order to derive the IT estimator, let $\mathbf{f}(\mathbf{p}) = (f_1(\mathbf{p}), \dots, f_n(\mathbf{p}))'$ where $f_i(\mathbf{p}) = \frac{\partial \phi(\mathbf{p})}{\partial p_i} \forall i$. Then the general solution for $\mathbf{p}$ in terms of parameters and Lagrange multipliers is:

$$\mathbf{p}(\beta, \rho, \lambda, \eta) = \mathbf{f}^{-1}(\lambda' \mathbf{Z}' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\beta] + \eta) \quad (15)$$

where $\mathbf{f}^{-1}(\cdot)$ is a well-defined vector-valued inverse function.

Substituting (15) into (6) leads to a well-defined solution for λ under general regularity conditions (Qin and Lawless [49]) which is an implicit function of parameters β and ρ denoted by:

$$\lambda(\beta, \rho) = \arg_{\lambda} \left[\sum_{i=1}^n p_i(\beta, \rho, \lambda, \eta^*) \mathbf{Z}[i, \cdot]' (Y_i - \rho \mathbf{W}[i, \cdot] \mathbf{Y} - \mathbf{X}[i, \cdot] \beta) = 0 \right] \quad (16)$$

where η^* is the optimal value obtained from the first order conditions.

Substituting $\mathbf{p}(\beta, \rho, \lambda(\beta, \rho))$, $\lambda(\beta, \rho)$, and η^* in (7) produces the concentrated objective function:

$$\phi(\beta, \rho, \lambda(\beta, \rho)) = \arg \max_{\mathbf{p}} \left\{ \phi(\mathbf{p}) - \lambda(\beta, \rho)' (\mathbf{p} \odot \mathbf{Z})' [(\mathbf{I} - \rho \mathbf{W})\mathbf{Y} - \mathbf{X}\beta] - \eta^* \left(\sum_{i=1}^n p_i - 1 \right) \right\} \quad (17)$$

which assigns the most favorable empirical weights to each value of the parameter vector from within a family of multinomial distributions supported on the sample $(\mathbf{Y}, \mathbf{X})$ and satisfying the empirical moment equations in (6).

The IT estimators behave like conventional likelihood estimators in the sense that they are obtained by maximizing the objective function in (17) over the domain of parameter space given by $\beta \in R^k$ and $-1 \leq \rho \leq 1$ as:

$$\begin{bmatrix} \hat{\beta}_{IT} \\ \hat{\rho}_{IT} \end{bmatrix} = \arg \max_{\beta, \rho} \{ \phi(\beta, \rho, \lambda(\beta, \rho)) \} \quad (18)$$

Let the empirical probability density in (8) be fixed to a discrete uniform distribution $\mathbf{q} = n^{-1} \mathbf{1}$ where $\mathbf{1}$ represents a $(n \times 1)$ vector of ones. Then the objective functions in (7) encompasses: (a) the

traditional empirical log-likelihood function $\phi(\mathbf{p}) = \sum_{i=1}^n \ln(p_i)$ since $\lim_{\gamma \rightarrow -1} CR(\mathbf{p}, n^{-1}\mathbf{1}, \gamma) = -\sum_{i=1}^n \ln(p_i)$,

(b) the empirical exponential likelihood (or negative entropy) function $\phi(\mathbf{p}) = -\sum_{i=1}^n p_i \ln(p_i)$ since

$\lim_{\gamma \rightarrow 0} CR(\mathbf{p}, n^{-1}\mathbf{1}, \gamma) = \sum_{i=1}^n p_i \ln(p_i)$, and (c) log Euclidean likelihood function $\phi(\mathbf{p}) = n^{-1} \sum_{i=1}^n (n^2 p_i^2 - 1)$

for $\gamma = 1$. The corresponding IT estimators are maximum empirical likelihood (MEL), maximum exponential empirical likelihood (MEEL), and maximum log Euclidean empirical likelihood (MLEL).

The resulting optimal weights p_i for the MEL estimator implied by (10) are given by:

$$p_i(\boldsymbol{\beta}, \rho, \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)) = n^{-1} [1 + \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)' \mathbf{Z}[i, \cdot]' (Y_i - \rho \mathbf{W}[i, \cdot] \mathbf{Y} - \mathbf{X}[i, \cdot] \boldsymbol{\beta})]^{-1} \quad (19)$$

where $\eta^* = 1$. The optimal p_i for the MEEL estimator can be expressed as:

$$p_i(\boldsymbol{\beta}, \rho, \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)) = \frac{\exp[1 + \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)' \mathbf{Z}[i, \cdot]' (Y_i - \rho \mathbf{W}[i, \cdot] \mathbf{Y} - \mathbf{X}[i, \cdot] \boldsymbol{\beta})]}{\sum_{i=1}^n \exp[1 + \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)' \mathbf{Z}[i, \cdot]' (Y_i - \rho \mathbf{W}[i, \cdot] \mathbf{Y} - \mathbf{X}[i, \cdot] \boldsymbol{\beta})]} \quad (20)$$

The optimal p_i for the MLEL estimator can be expressed as:

$$p_i(\boldsymbol{\beta}, \rho, \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)) = (2n)^{-1} [\eta^* + \boldsymbol{\lambda}(\boldsymbol{\beta}, \rho)' \mathbf{Z}[i, \cdot]' (Y_i - \rho \mathbf{W}[i, \cdot] \mathbf{Y} - \mathbf{X}[i, \cdot] \boldsymbol{\beta})] \quad (21)$$

where $\eta^* = 2 - \frac{1}{n} \sum_{i=1}^n [\boldsymbol{\lambda}' \mathbf{Z}[i, \cdot]' (Y_i - \rho \mathbf{W}[i, \cdot] \mathbf{Y} - \mathbf{X}[i, \cdot] \boldsymbol{\beta})]$. The IT estimators $\hat{\boldsymbol{\beta}}_{IT}$ and $\hat{\rho}_{IT}$ for $IT \in \{MEL, MEEL, MLEL\}$ are obtained by maximizing the corresponding concentrated objective function in (17) with respect to the $(k+1)$ unknown parameters with the appropriate $p_i(\boldsymbol{\lambda}(\boldsymbol{\beta}, \rho), \boldsymbol{\beta}, \rho)$ substituted for p_i .

4. Monte Carlo Experiments

In order to compare the finite sampling properties of the estimators, extensive Monte Carlo sampling experiments were conducted. For a range of small sample sizes, spatial autocorrelation parameters, and commonly used weight matrices, finite sample performances of the IT estimators are compared to the OLS, 2SLS, GMM, and QML estimators noted in Section 2.

4.1. Monte Carlo Design

The Monte Carlo experiments are constructed in the following manner: three different sample sizes were analyzed, namely, $n = 25, 60$, and 90 . For each sample size, three different specifications of spatial weight matrices were used. Weight matrices differ in their degree of sparseness, they have equal weights, and an average number of neighbors per unit, J , is chosen to be $J = 2, 6$ and 10 [50]. Eleven values for the spatial autoregressive coefficient ρ in (1) are chosen, namely, $-0.9, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8$ and 0.9 . Three values of σ^2 are $1, 2.5$, and 5 . The remaining elements of the parameter vector are specified as $\boldsymbol{\beta} = (\beta_1, \beta_2) = (1, 1)$. Two regressors, $\mathbf{X} = (X_1, X_2)$, are

specified as $X_1 \sim \text{Gamma}(2, 2)$ and $X_2 \sim \text{Uniform}(0, 1)$. The n observations on independent variables are normalized so that their sample means and variances are, respectively, zero and one. The same regressors are used in all experiments and 1000 repetitions were conducted for each Monte Carlo experiment. In summary, there are then three values of n , eleven values of ρ , three values of σ^2 , and three values of J . All combinations of n , ρ , σ^2 , and J result in $3 \times 11 \times 3 \times 3 = 297$ Monte Carlo experiments.

In addition, three different parametric functional forms for the distribution of the disturbance term ε in (1) are analyzed, together with varying specifications relating to heteroskedasticity. First, normally distributed homoscedastic disturbance terms, $\varepsilon_i \sim N(0, \sigma^2)$ for $\forall i = 1, \dots, n$, are considered and referred to as $H = 1$. Three forms of heteroscedasticity are also specified for the normal distribution such that $\sigma_i^2 = [0.5X_2]\sigma^2$, $\sigma_i^2 = [0.5X_2^2]\sigma^2$, and $\sigma_i^2 = \exp(0.5X_2)\sigma^2$, referred to as $H = 2$, $H = 3$, $H = 4$, correspondingly. In addition, the log-normal distribution ($H = 5$) is chosen because of its asymmetric nature such that $\varepsilon_i = \exp(\xi_i) - \exp(0.5\sigma^2)$, where $\xi_i \sim N(0, \sigma^2)$. Finally, a mixture of normal is considered ($H = 6$), where a normally distributed variable is contaminated by another with a larger variance, producing thicker tails than a normal distribution, *i.e.*, $\varepsilon_i = \varsigma_i \xi_i + (1 - \varsigma_i) \zeta_i$, the ς_i are *iid* Bernoulli variables with $\Pr(\varsigma_i = 1) = 0.95$, $\xi_i \sim N(0, \sigma^2)$ and $\zeta_i \sim N(0, 100)$. There are 297 experiments for the 6 distributional assumptions considered, which result in a total of $297 \times 6 = 1782$ Monte Carlo experiments performed.

To keep the Monte Carlo study manageable in terms of reporting results, bias and spread of the estimator distributions are measured by root mean square error (RMSE). Since sample moments for the calculation of the standard RMSE might not exist, the RMSE's measure proposed by [50] is used, which guarantees the existence of necessary components. According to Kelejian and Prucha [48], $RMSE = [bias^2 + (\frac{IQ}{1.35})^2]^{0.5}$ where the *bias* is the difference between the median and the true parameter values, and $IQ = c_1 - c_2$ is the inter-quantile range where c_1 and c_2 are 0.75 and 0.25 quantiles, correspondingly.

In addition, response functions are used to summarize the relationship between RMSEs of estimators and model parameters over the set of all considered parameter values. In particular, let ρ_i , β_{1i} , β_{2i} , J_i and σ_i^2 be the values, respectively, of ρ , β_1 , β_2 , J and σ^2 in the i^{th} experiment, $i = 1, \dots, 297$, for each distributional assumption, and let n_i correspond to the sample size. Then the functional form of the response function [51] is:

$$RMSE_i = \frac{\sigma_i}{\sqrt{n_i}} \exp[a_1 + a_2(\frac{1}{J_i}) + a_3(\frac{\rho_i}{J_i}) + a_4\rho_i + a_5\rho_i^2 + a_6(\frac{J_i}{n_i}) + a_7(\frac{J_i}{n_i})^2 + a_8(\frac{1}{n_i}) + a_9\sigma_i^2], \quad i = 1, \dots, 297 \quad (22)$$

where $RMSE_i$ is the RMSE of an estimator of a given parameter in the i^{th} experiment and the parameters $a_1, \dots, a_9$ are different for each estimator of each parameter. For each case considered, the parameters of (22) are estimated using the entire set of 297 Monte Carlo experiments for each distributional assumption by least squares after taking natural logs on both sides.

4.2. Results

There are three types of results presented. First, tables of root mean square errors of the estimators for a subset of experiments are reported. Second, average RMSEs for the entire Monte Carlo study are provided. Finally, response functions for the RMSEs of estimators are estimated, and graphs of these response functions are presented.

In order to conserve space, Tables 1–3 report RMSEs of the estimators for a selected subset of experimental parameter values. These tables report RMSEs of the estimators of the parameters ρ , β_1 and β_2 , correspondingly, for 33 Monte Carlo experiments representing a combination of 11 parameter values for ρ and 3 parameter values for σ^2 , with the average number of neighbors $J = 6$, sample size $n = 60$, and normally distributed disturbance term $H = 1$.

According to Table 1, the RMSEs of the QML estimator of ρ are the lowest and the OLS estimator are the largest, since the OLS estimator is inconsistent and the QML estimator with a normally distributed homoscedastic disturbance term is consistent and efficient.

Table 1. Root mean square errors of estimators of ρ , $n = 60$, $J = 6$, $H = 1$.

ρ	σ^2	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
0.9	1	0.055	0.048	0.048	0.036	0.051	0.074	0.051
0.9	2.5	0.077	0.061	0.063	0.040	0.065	0.093	0.069
0.9	5	0.110	0.115	0.111	0.046	0.113	0.098	0.127
0.8	1	0.077	0.069	0.070	0.055	0.071	0.120	0.074
0.8	2.5	0.155	0.128	0.138	0.073	0.143	0.121	0.153
0.8	5	0.189	0.178	0.183	0.079	0.165	0.146	0.180
0.6	1	0.173	0.148	0.159	0.097	0.149	0.116	0.155
0.6	2.5	0.237	0.227	0.233	0.113	0.236	0.133	0.243
0.6	5	0.258	0.236	0.260	0.124	0.258	0.146	0.267
0.4	1	0.159	0.166	0.170	0.116	0.160	0.128	0.166
0.4	2.5	0.229	0.241	0.245	0.139	0.247	0.157	0.251
0.4	5	0.271	0.295	0.288	0.150	0.267	0.181	0.274
0.2	1	0.196	0.194	0.204	0.137	0.200	0.190	0.207
0.2	2.5	0.260	0.295	0.294	0.153	0.276	0.222	0.289
0.2	5	0.315	0.400	0.389	0.216	0.354	0.233	0.369
0	1	0.204	0.209	0.214	0.153	0.210	0.164	0.219
0	2.5	0.305	0.315	0.331	0.201	0.302	0.220	0.310
0	5	0.389	0.563	0.593	0.220	0.506	0.255	0.532
−0.2	1	0.231	0.213	0.225	0.166	0.205	0.184	0.207
−0.2	2.5	0.404	0.390	0.414	0.219	0.389	0.238	0.397
−0.2	5	0.424	0.465	0.496	0.216	0.393	0.274	0.410
−0.4	1	0.360	0.282	0.321	0.200	0.284	0.230	0.288
−0.4	2.5	0.476	0.421	0.438	0.214	0.405	0.256	0.415
−0.4	5	0.555	0.556	0.612	0.232	0.533	0.328	0.533
−0.6	1	0.360	0.256	0.282	0.193	0.242	0.221	0.246
−0.6	2.5	0.581	0.434	0.497	0.240	0.397	0.315	0.408
−0.6	5	0.669	0.598	0.627	0.239	0.549	0.350	0.546
−0.8	1	0.532	0.368	0.382	0.208	0.302	0.266	0.307
−0.8	2.5	0.591	0.442	0.463	0.247	0.364	0.307	0.368
−0.8	5	0.789	0.609	0.569	0.240	0.504	0.337	0.498
−0.9	1	0.328	0.224	0.186	0.175	0.178	0.144	0.189
−0.9	2.5	0.656	0.411	0.387	0.217	0.323	0.244	0.329
−0.9	5	0.757	0.576	0.392	0.235	0.396	0.225	0.403
Column Average		0.345	0.307	0.312	0.163	0.280	0.203	0.287

Table 2. Root mean square errors of estimators of β_1 , $n = 60$, $J = 6$, $H = 1$.

ρ	σ^2	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
0.9	1	0.141	0.142	0.141	0.139	0.142	0.154	0.146
0.9	2.5	0.217	0.217	0.219	0.222	0.221	0.247	0.224
0.9	5	0.294	0.309	0.313	0.303	0.315	0.334	0.312
0.8	1	0.132	0.131	0.132	0.128	0.137	0.151	0.138
0.8	2.5	0.225	0.228	0.229	0.222	0.221	0.254	0.228
0.8	5	0.301	0.312	0.315	0.293	0.307	0.321	0.304
0.6	1	0.132	0.133	0.135	0.130	0.136	0.175	0.137
0.6	2.5	0.206	0.209	0.209	0.203	0.202	0.249	0.208
0.6	5	0.305	0.299	0.300	0.293	0.305	0.346	0.310
0.4	1	0.129	0.130	0.129	0.129	0.135	0.192	0.134
0.4	2.5	0.204	0.206	0.206	0.202	0.204	0.272	0.209
0.4	5	0.303	0.305	0.316	0.304	0.310	0.351	0.306
0.2	1	0.130	0.134	0.135	0.128	0.135	0.190	0.139
0.2	2.5	0.217	0.219	0.220	0.219	0.223	0.262	0.221
0.2	5	0.295	0.305	0.314	0.302	0.314	0.356	0.319
0	1	0.133	0.136	0.137	0.129	0.144	0.171	0.146
0	2.5	0.211	0.209	0.213	0.211	0.210	0.252	0.210
0	5	0.296	0.309	0.320	0.294	0.315	0.339	0.315
-0.2	1	0.126	0.129	0.127	0.126	0.132	0.165	0.142
-0.2	2.5	0.199	0.202	0.203	0.201	0.199	0.236	0.204
-0.2	5	0.297	0.305	0.302	0.307	0.320	0.341	0.318
-0.4	1	0.139	0.142	0.145	0.137	0.142	0.197	0.144
-0.4	2.5	0.205	0.212	0.222	0.208	0.207	0.259	0.215
-0.4	5	0.295	0.292	0.303	0.282	0.299	0.334	0.299
-0.6	1	0.131	0.130	0.130	0.129	0.131	0.185	0.137
-0.6	2.5	0.240	0.224	0.234	0.220	0.223	0.281	0.223
-0.6	5	0.303	0.317	0.314	0.296	0.317	0.333	0.323
-0.8	1	0.142	0.134	0.139	0.136	0.140	0.180	0.140
-0.8	2.5	0.216	0.219	0.236	0.210	0.219	0.276	0.219
-0.8	5	0.293	0.298	0.312	0.290	0.301	0.317	0.309
-0.9	1	0.142	0.129	0.132	0.126	0.134	0.175	0.137
-0.9	2.5	0.237	0.214	0.225	0.209	0.211	0.274	0.218
-0.9	5	0.285	0.294	0.307	0.284	0.300	0.337	0.295
Column Average		0.216	0.217	0.222	0.213	0.220	0.258	0.222

This is not surprising since in this case, the QML estimator is actually a correctly specified classical ML estimator. The IT estimators of ρ have the second lowest RMSEs, outperforming the OLS, 2SLS, and GMM estimators. The MEL estimator outperforms the other IT estimators but is, on average, slightly less efficient than the QML estimator. Tables 2 and 3 report the RMSEs of the QML and IT estimators of β_1 and β_2 which are on average roughly the same. A loss of efficiency of the IT estimators relative to the QML estimators is small and mostly arises from estimation of the spatial autocorrelation coefficient ρ [11]. The same pattern holds for other distributional assumptions, namely $H = 2, \dots, 5$, and only the averages of the RMSEs are reported.

In order to summarize the results of the total 1782 Monte Carlo experiments, average RMSEs of the estimators, namely, for ρ , β_1 and β_2 , are presented in Tables 4–9. A table's entry is the averages of RMSEs of estimators over 33 Monte Carlo experiments for a given sample size n , average number of neighbors J , and distributional assumptions H .

Table 3. Root mean square errors of estimators of β_2 , $n = 60$, $J = 6$, $H = 1$.

ρ	σ^2	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
0.9	1	0.131	0.130	0.131	0.126	0.139	0.149	0.141
0.9	2.5	0.233	0.236	0.243	0.230	0.246	0.249	0.243
0.9	5	0.297	0.307	0.308	0.301	0.314	0.330	0.310
0.8	1	0.140	0.137	0.143	0.136	0.141	0.156	0.146
0.8	2.5	0.225	0.225	0.239	0.221	0.237	0.244	0.237
0.8	5	0.274	0.278	0.277	0.275	0.292	0.306	0.293
0.6	1	0.127	0.125	0.124	0.125	0.128	0.173	0.132
0.6	2.5	0.199	0.198	0.199	0.204	0.205	0.249	0.214
0.6	5	0.301	0.311	0.308	0.302	0.304	0.345	0.307
0.4	1	0.137	0.132	0.131	0.133	0.137	0.186	0.140
0.4	2.5	0.211	0.215	0.219	0.210	0.226	0.265	0.226
0.4	5	0.299	0.303	0.312	0.312	0.311	0.340	0.316
0.2	1	0.135	0.136	0.137	0.134	0.140	0.182	0.144
0.2	2.5	0.219	0.224	0.224	0.221	0.220	0.269	0.226
0.2	5	0.308	0.309	0.314	0.306	0.319	0.334	0.321
0	1	0.131	0.132	0.131	0.133	0.140	0.179	0.141
0	2.5	0.228	0.230	0.227	0.226	0.233	0.259	0.236
0	5	0.290	0.302	0.304	0.292	0.304	0.333	0.309
−0.2	1	0.128	0.128	0.131	0.126	0.141	0.182	0.147
−0.2	2.5	0.205	0.210	0.208	0.206	0.221	0.251	0.221
−0.2	5	0.282	0.302	0.303	0.281	0.296	0.310	0.306
−0.4	1	0.134	0.137	0.135	0.133	0.141	0.178	0.143
−0.4	2.5	0.216	0.217	0.221	0.214	0.222	0.250	0.230
−0.4	5	0.302	0.300	0.312	0.294	0.312	0.331	0.319
−0.6	1	0.133	0.131	0.131	0.126	0.140	0.179	0.142
−0.6	2.5	0.236	0.214	0.212	0.202	0.214	0.257	0.213
−0.6	5	0.316	0.321	0.336	0.298	0.314	0.361	0.318
−0.8	1	0.156	0.149	0.155	0.142	0.147	0.190	0.148
−0.8	2.5	0.227	0.212	0.234	0.212	0.222	0.269	0.218
−0.8	5	0.311	0.324	0.320	0.308	0.324	0.343	0.315
−0.9	1	0.137	0.137	0.139	0.136	0.140	0.192	0.145
−0.9	2.5	0.216	0.199	0.205	0.200	0.213	0.265	0.212
−0.9	5	0.305	0.327	0.342	0.296	0.330	0.360	0.344
Column Average		0.218	0.219	0.223	0.214	0.225	0.257	0.227

The results of the Monte Carlo experiments for other distributional assumptions are consistent with the results for a subset of parameter space for the case of $H = 1$ reported in Tables 1–3. In fact, Tables 4 and 5 indicate that the RMSEs of the QML estimator of ρ are the lowest, regardless of sample size, average number of neighbors, and distributional assumptions considered. The IT estimators of ρ have the second lowest RMSEs across all Monte Carlo experiments, being less efficient than the QML estimator, but often outperforming OLS, 2SLS, and GMM estimators.

Tables 6–9 indicate that the RMSEs of QML and the IT estimators of β_1 and β_2 are roughly the same with the exception of the MEL estimator which is, on average, less efficient than the QML estimator for β_1 and β_2 . Indeed, many cases exist where IT estimators (e.g., MLEL) outperform QML. In any case, the loss of efficiency of the IT estimators relative to the QML estimator is, on average, small and mostly arises from estimation of the spatial autocorrelation coefficient ρ .

Table 4. Average root mean square errors of estimators of ρ for $H = 1-3$.

H	J	n	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
1	2	25	0.212	0.221	0.247	0.115	0.210	0.191	0.223
1	2	60	0.170	0.140	0.148	0.077	0.139	0.112	0.146
1	2	90	0.156	0.113	0.118	0.060	0.113	0.097	0.117
1	6	25	0.493	0.522	0.540	0.272	0.456	0.402	0.465
1	6	60	0.345	0.307	0.312	0.163	0.280	0.203	0.287
1	6	90	0.301	0.251	0.256	0.132	0.230	0.159	0.236
1	10	25	0.693	0.721	0.698	0.372	0.598	0.524	0.610
1	10	60	0.463	0.440	0.466	0.226	0.413	0.284	0.418
1	10	90	0.400	0.351	0.352	0.180	0.319	0.208	0.325
2	2	25	0.114	0.106	0.114	0.077	0.106	0.113	0.109
2	2	60	0.082	0.066	0.069	0.050	0.065	0.070	0.065
2	2	90	0.080	0.057	0.058	0.038	0.055	0.061	0.056
2	6	25	0.260	0.262	0.276	0.189	0.244	0.240	0.246
2	6	60	0.197	0.170	0.177	0.120	0.154	0.145	0.154
2	6	90	0.162	0.125	0.133	0.095	0.117	0.116	0.117
2	10	25	0.395	0.422	0.435	0.273	0.370	0.360	0.370
2	10	60	0.272	0.244	0.257	0.189	0.225	0.214	0.224
2	10	90	0.225	0.184	0.196	0.133	0.169	0.156	0.169
3	2	25	0.148	0.134	0.140	0.081	0.126	0.144	0.127
3	2	60	0.143	0.117	0.117	0.061	0.106	0.111	0.105
3	2	90	0.132	0.099	0.104	0.046	0.084	0.093	0.083
3	6	25	0.365	0.375	0.386	0.217	0.336	0.341	0.334
3	6	60	0.285	0.240	0.257	0.134	0.207	0.217	0.203
3	6	90	0.269	0.236	0.246	0.115	0.203	0.178	0.197
3	10	25	0.444	0.473	0.510	0.277	0.418	0.425	0.420
3	10	60	0.367	0.336	0.360	0.183	0.300	0.288	0.294
3	10	90	0.325	0.325	0.343	0.146	0.260	0.241	0.253

Table 5. Average root mean square errors of estimators of ρ for $H = 4-6$.

H	J	n	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
4	2	25	0.219	0.232	0.250	0.120	0.211	0.204	0.222
4	2	60	0.188	0.167	0.177	0.076	0.157	0.131	0.164
4	2	90	0.185	0.144	0.152	0.064	0.138	0.112	0.142
4	6	25	0.506	0.560	0.579	0.266	0.488	0.448	0.496
4	6	60	0.380	0.380	0.387	0.162	0.333	0.239	0.338
4	6	90	0.338	0.307	0.306	0.132	0.271	0.189	0.274
4	10	25	0.701	0.795	0.805	0.365	0.647	0.589	0.657
4	10	60	0.494	0.515	0.510	0.225	0.446	0.303	0.452
4	10	90	0.452	0.450	0.460	0.186	0.410	0.252	0.412
5	2	25	0.175	0.169	0.182	0.100	0.149	0.151	0.155
5	2	60	0.148	0.119	0.125	0.066	0.102	0.105	0.103
5	2	90	0.142	0.101	0.105	0.050	0.087	0.090	0.088
5	6	25	0.439	0.452	0.474	0.244	0.373	0.368	0.374
5	6	60	0.304	0.275	0.282	0.143	0.228	0.210	0.226
5	6	90	0.283	0.235	0.244	0.120	0.196	0.168	0.195
5	10	25	0.659	0.685	0.716	0.349	0.544	0.532	0.549
5	10	60	0.420	0.408	0.442	0.206	0.362	0.325	0.360
5	10	90	0.357	0.307	0.336	0.163	0.270	0.240	0.266
6	2	25	0.160	0.160	0.169	0.083	0.143	0.148	0.148
6	2	60	0.147	0.120	0.126	0.057	0.096	0.103	0.096
6	2	90	0.142	0.100	0.105	0.046	0.080	0.086	0.080
6	6	25	0.343	0.375	0.386	0.193	0.312	0.311	0.315
6	6	60	0.286	0.269	0.278	0.131	0.213	0.206	0.209
6	6	90	0.261	0.222	0.227	0.108	0.171	0.159	0.168
6	10	25	0.486	0.597	0.580	0.271	0.473	0.451	0.478
6	10	60	0.383	0.401	0.431	0.183	0.330	0.332	0.322
6	10	90	0.343	0.324	0.344	0.152	0.257	0.251	0.250
Column Average			0.304	0.295	0.305	0.153	0.255	0.230	0.257

Table 6. Average root mean square errors of estimators of β_1 for $H = 1-3$.

H	J	n	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
1	2	25	0.349	0.356	0.374	0.337	0.365	0.384	0.372
1	2	60	0.226	0.223	0.231	0.215	0.227	0.254	0.231
1	2	90	0.186	0.179	0.183	0.172	0.181	0.215	0.182
1	6	25	0.345	0.352	0.364	0.338	0.359	0.383	0.366
1	6	60	0.216	0.217	0.222	0.212	0.220	0.258	0.222
1	6	90	0.177	0.177	0.179	0.173	0.178	0.227	0.180
1	10	25	0.343	0.355	0.368	0.334	0.360	0.384	0.368
1	10	60	0.219	0.221	0.227	0.212	0.224	0.266	0.227
1	10	90	0.172	0.172	0.173	0.169	0.174	0.226	0.175
2	2	25	0.327	0.329	0.331	0.329	0.339	0.356	0.353
2	2	60	0.229	0.230	0.231	0.227	0.235	0.249	0.241
2	2	90	0.186	0.184	0.185	0.184	0.181	0.201	0.185
2	6	25	0.335	0.334	0.335	0.331	0.338	0.354	0.349
2	6	60	0.224	0.224	0.225	0.223	0.221	0.244	0.225
2	6	90	0.184	0.183	0.184	0.183	0.181	0.200	0.182
2	10	25	0.317	0.316	0.320	0.314	0.321	0.342	0.330
2	10	60	0.222	0.223	0.225	0.222	0.221	0.243	0.225
2	10	90	0.187	0.186	0.187	0.186	0.184	0.205	0.186
3	2	25	0.758	0.773	0.763	0.781	0.802	0.820	0.833
3	2	60	0.589	0.600	0.599	0.599	0.602	0.616	0.600
3	2	90	0.546	0.561	0.564	0.559	0.554	0.578	0.540
3	6	25	0.883	0.884	0.881	0.888	0.901	0.929	0.916
3	6	60	0.768	0.778	0.781	0.778	0.784	0.794	0.781
3	6	90	0.631	0.635	0.635	0.640	0.623	0.649	0.614
3	10	25	0.782	0.784	0.786	0.781	0.795	0.820	0.801
3	10	60	0.732	0.733	0.731	0.732	0.735	0.742	0.734
3	10	90	0.655	0.664	0.663	0.660	0.649	0.671	0.637

Table 7. Average root mean square errors of estimators of β_1 for $H = 4-6$.

H	J	n	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
4	2	25	0.799	0.828	0.828	0.839	0.844	0.871	0.854
4	2	60	0.751	0.795	0.800	0.804	0.799	0.835	0.796
4	2	90	0.505	0.523	0.521	0.524	0.518	0.543	0.513
4	6	25	0.844	0.857	0.866	0.851	0.870	0.906	0.878
4	6	60	0.616	0.628	0.634	0.627	0.626	0.657	0.624
4	6	90	0.655	0.666	0.665	0.669	0.656	0.696	0.648
4	10	25	0.828	0.850	0.873	0.840	0.870	0.899	0.882
4	10	60	0.706	0.715	0.711	0.716	0.707	0.751	0.703
4	10	90	0.579	0.585	0.586	0.580	0.579	0.611	0.574
5	2	25	0.270	0.271	0.282	0.257	0.258	0.294	0.257
5	2	60	0.190	0.185	0.190	0.178	0.170	0.221	0.165
5	2	90	0.168	0.158	0.162	0.152	0.145	0.194	0.140
5	6	25	0.276	0.274	0.285	0.265	0.262	0.301	0.259
5	6	60	0.182	0.180	0.183	0.177	0.165	0.219	0.161
5	6	90	0.155	0.154	0.155	0.149	0.139	0.194	0.134
5	10	25	0.270	0.268	0.284	0.257	0.259	0.298	0.256
5	10	60	0.187	0.185	0.190	0.181	0.171	0.229	0.167
5	10	90	0.153	0.152	0.153	0.148	0.138	0.196	0.133
6	2	25	0.234	0.236	0.244	0.229	0.231	0.263	0.230
6	2	60	0.181	0.178	0.183	0.170	0.155	0.212	0.147
6	2	90	0.161	0.152	0.156	0.146	0.129	0.193	0.122
6	6	25	0.241	0.240	0.249	0.234	0.232	0.267	0.230
6	6	60	0.178	0.176	0.179	0.172	0.152	0.217	0.145
6	6	90	0.148	0.148	0.150	0.144	0.125	0.196	0.117
6	10	25	0.238	0.243	0.254	0.228	0.233	0.269	0.232
6	10	60	0.173	0.173	0.177	0.168	0.151	0.215	0.143
6	10	90	0.149	0.148	0.151	0.145	0.125	0.194	0.118
Column Average			0.382	0.386	0.390	0.382	0.383	0.418	0.383

Table 8. Average root mean square errors of estimators of β_2 for $H = 1-3$.

H	J	n	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
1	2	25	0.351	0.361	0.376	0.342	0.371	0.386	0.379
1	2	60	0.223	0.221	0.230	0.213	0.228	0.250	0.231
1	2	90	0.185	0.178	0.183	0.172	0.182	0.213	0.184
1	6	25	0.343	0.351	0.367	0.332	0.360	0.385	0.369
1	6	60	0.218	0.219	0.223	0.214	0.225	0.257	0.227
1	6	90	0.178	0.175	0.178	0.172	0.177	0.221	0.181
1	10	25	0.343	0.353	0.364	0.336	0.362	0.387	0.370
1	10	60	0.216	0.217	0.221	0.212	0.220	0.258	0.223
1	10	90	0.175	0.175	0.177	0.172	0.177	0.225	0.179
2	2	25	0.152	0.151	0.155	0.147	0.151	0.183	0.152
2	2	60	0.108	0.102	0.106	0.100	0.099	0.151	0.098
2	2	90	0.092	0.085	0.087	0.083	0.082	0.144	0.081
2	6	25	0.160	0.160	0.161	0.156	0.155	0.191	0.154
2	6	60	0.109	0.107	0.108	0.106	0.103	0.158	0.101
2	6	90	0.086	0.085	0.085	0.084	0.081	0.150	0.080
2	10	25	0.162	0.162	0.165	0.158	0.157	0.193	0.157
2	10	60	0.100	0.098	0.100	0.097	0.096	0.153	0.095
2	10	90	0.085	0.084	0.085	0.084	0.082	0.154	0.081
3	2	25	0.254	0.254	0.255	0.250	0.239	0.282	0.231
3	2	60	0.188	0.187	0.190	0.181	0.169	0.227	0.161
3	2	90	0.182	0.171	0.176	0.166	0.151	0.221	0.140
3	6	25	0.238	0.235	0.243	0.227	0.220	0.263	0.217
3	6	60	0.205	0.204	0.203	0.200	0.183	0.241	0.172
3	6	90	0.174	0.179	0.179	0.179	0.148	0.238	0.134
3	10	25	0.239	0.239	0.250	0.231	0.230	0.272	0.227
3	10	60	0.209	0.208	0.208	0.204	0.185	0.257	0.176
3	10	90	0.185	0.186	0.185	0.185	0.164	0.242	0.153

Table 9. Average root mean square errors of estimators of β_2 for $H = 4-6$.

H	J	n	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
3	6	90	0.174	0.179	0.179	0.179	0.148	0.238	0.134
4	2	25	0.436	0.452	0.473	0.437	0.461	0.486	0.464
4	2	60	0.316	0.321	0.329	0.305	0.318	0.354	0.312
4	2	90	0.257	0.255	0.262	0.247	0.248	0.297	0.243
4	6	25	0.426	0.440	0.459	0.421	0.441	0.474	0.441
4	6	60	0.302	0.306	0.313	0.293	0.297	0.350	0.292
4	6	90	0.255	0.257	0.262	0.254	0.248	0.315	0.242
4	10	25	0.445	0.458	0.480	0.444	0.458	0.495	0.461
4	10	60	0.303	0.310	0.314	0.299	0.304	0.358	0.299
4	10	90	0.243	0.243	0.246	0.240	0.236	0.302	0.233
5	2	25	0.278	0.280	0.289	0.271	0.267	0.304	0.261
5	2	60	0.198	0.192	0.199	0.186	0.173	0.238	0.167
5	2	90	0.172	0.163	0.168	0.158	0.149	0.213	0.143
5	6	25	0.279	0.285	0.291	0.273	0.267	0.310	0.261
5	6	60	0.189	0.189	0.192	0.184	0.171	0.243	0.166
5	6	90	0.159	0.157	0.158	0.153	0.141	0.217	0.136
5	10	25	0.275	0.276	0.287	0.266	0.259	0.304	0.253
5	10	60	0.192	0.192	0.193	0.189	0.172	0.249	0.166
5	10	90	0.159	0.158	0.161	0.156	0.144	0.224	0.137
6	2	25	0.252	0.253	0.262	0.246	0.245	0.276	0.242
6	2	60	0.194	0.191	0.198	0.186	0.165	0.230	0.154
6	2	90	0.172	0.164	0.168	0.159	0.138	0.208	0.128
6	6	25	0.249	0.250	0.256	0.243	0.240	0.275	0.237
6	6	60	0.188	0.187	0.189	0.185	0.160	0.231	0.148
6	6	90	0.160	0.159	0.159	0.156	0.132	0.209	0.124
6	10	25	0.248	0.253	0.260	0.241	0.241	0.276	0.237
6	10	60	0.185	0.184	0.188	0.181	0.157	0.230	0.147
6	10	90	0.160	0.158	0.161	0.156	0.132	0.210	0.123
Column Average			0.220	0.220	0.225	0.214	0.210	0.263	0.207

The response functions (22) are used to depict a relationship between the RMSEs of estimators over the set of all considered parameter values. Figures 1–3 describe 18 sets of response functions for the RMSEs of the QML and the IT estimators of ρ for average number of neighbors equal $J = 2$ and three sample sizes $n = 25, 60, 90$. Each figure describes six response functions of the RMSEs for six distributional assumptions.

Figure 1. Response function RMSEs of QML, MEEL, MLEL and MEL of ρ ($n = 25$ and $J = 2$).

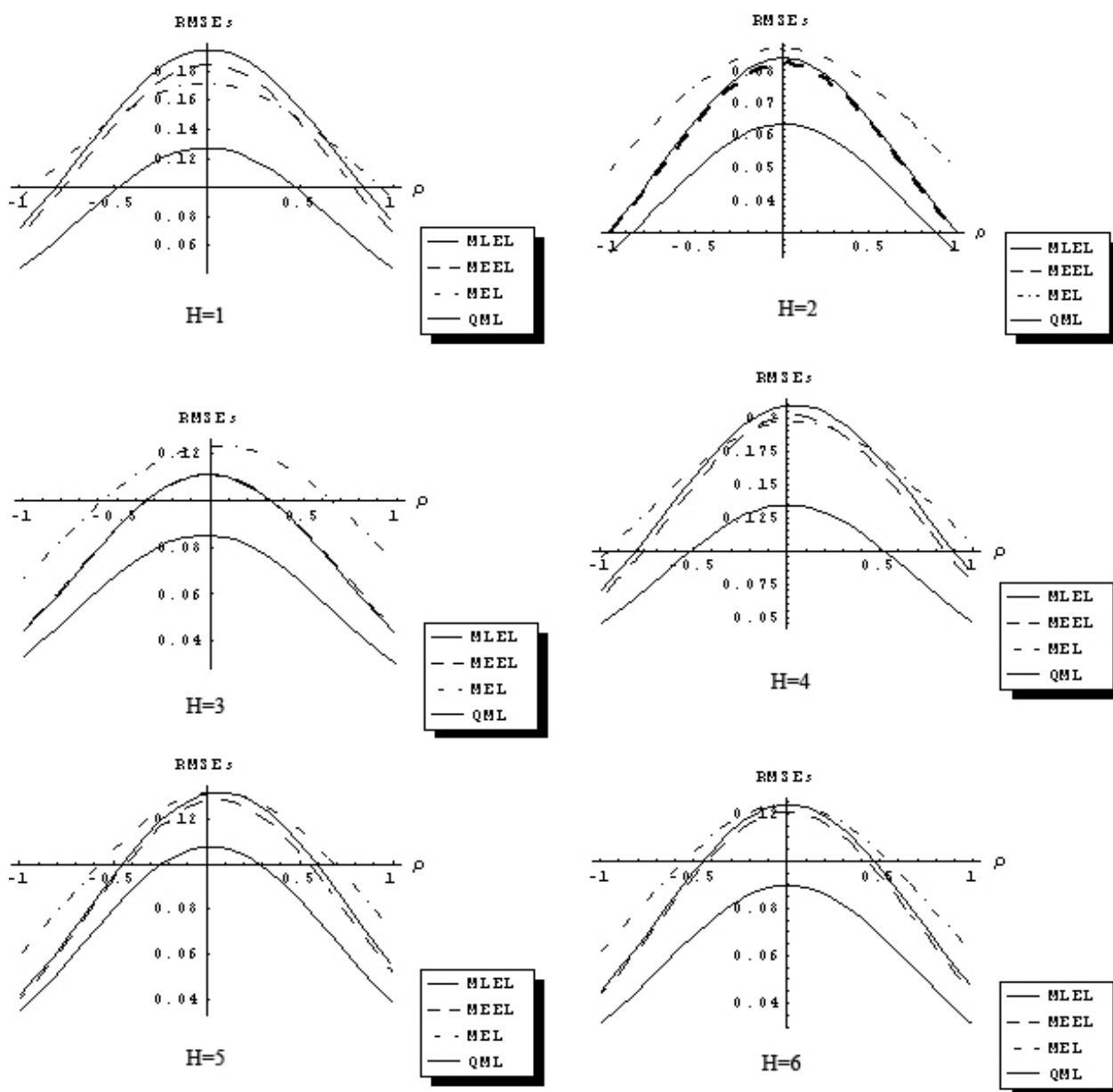
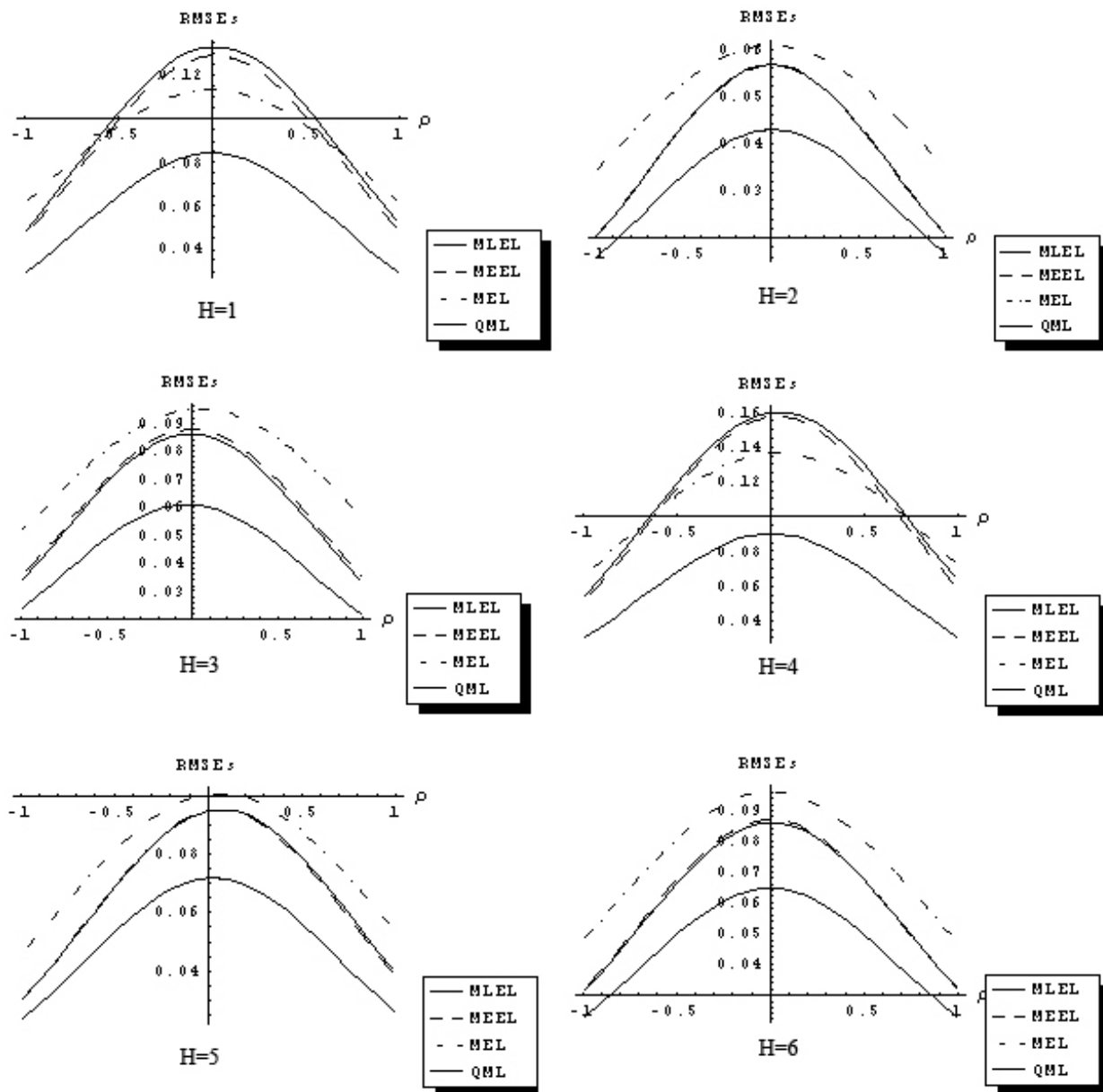
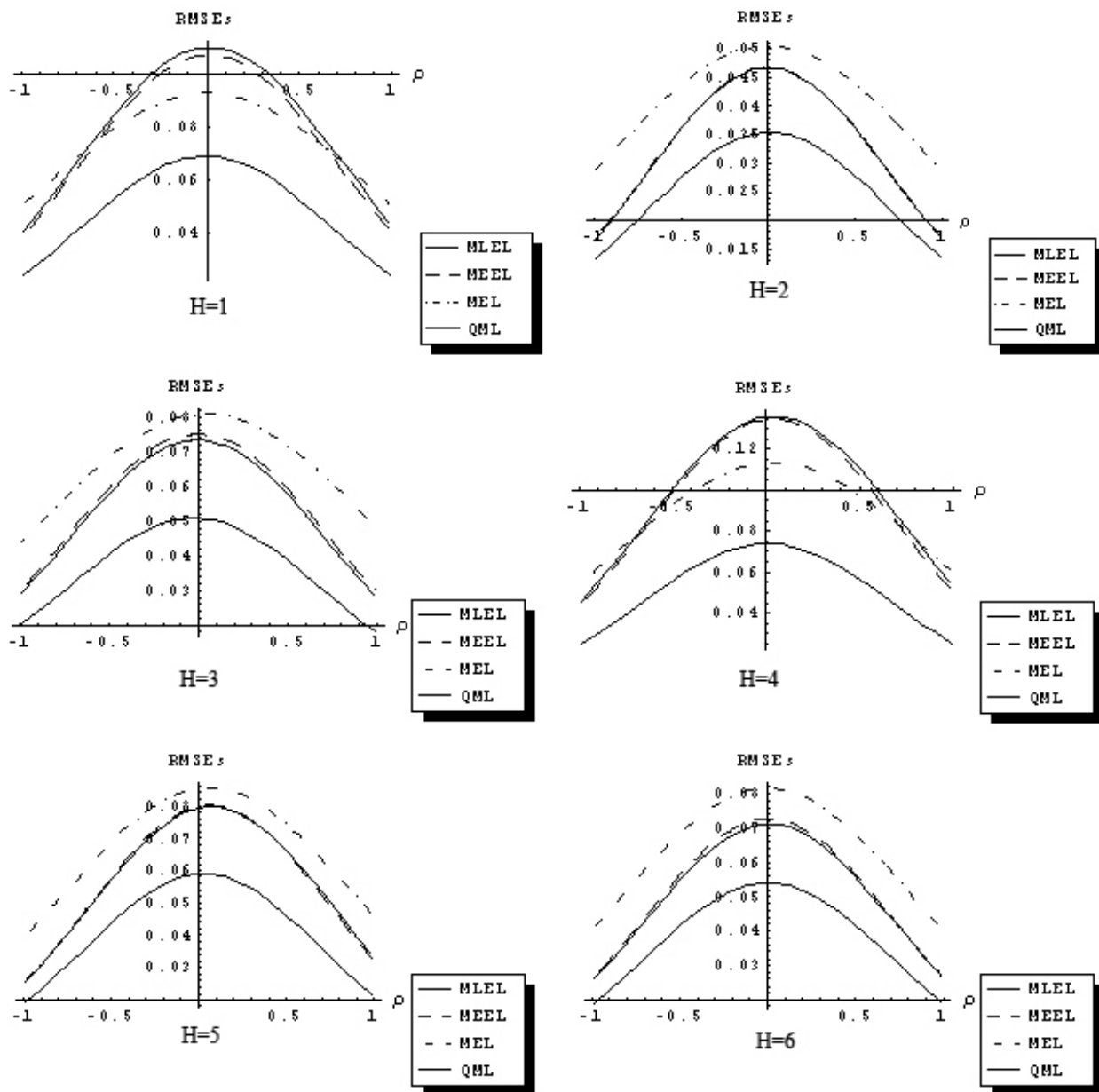


Figure 2. Response function RMSEs of QML, MEEL, MLEL and MEL of ρ ($n = 60$ and $J = 2$).

The RMSEs are related to the spatial autocorrelation parameter, ρ , in a concave fashion. The difference in the RMSEs between QML and IT estimators increases as ρ approaches zero, and it decreases as ρ approaches ± 1 . The QML estimator outperforms the IT estimators for all distributional assumptions considered and for $J = 2, 6, 10$. The response functions for the three IT estimators considered are related to the ρ in similar fashion and have comparable magnitude. As sample size increases, differences between the response functions of QML and IT estimators declines.

The results of the Monte Carlo study suggest that in certain sampling situations information theoretic estimators of the first-order spatial autoregressive model have superior sampling properties and outperform traditional OLS, 2SLS and GMM estimators with the exception of quasi-maximum likelihood estimator.

Figure 3. Response function RMSEs of QML, MEEL, MLEL and MEL of ρ ($n = 90$ and $J = 2$).

5. Illustrative Application: Hedonic Housing Value Model

Harrison and Rubinfeld [24] estimated demand for clean air or the dollar benefits of air quality improvements. The approach estimates willingness to pay for better air quality based on the presumption that consumers pay more for an identical housing unit located in an area with good air quality than in an area with poor air quality. The hedonic housing value function translates housing attributes into a price. The household's annual willingness to pay for a small improvement in air quality is the increased cost of purchasing a house with identical attributes except for marginal improvement in air quality found by calculating the derivatives of the hedonic housing price equation with respect to the air pollution attribute.

The housing data for census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970 contains 506 observations (one observation per census tract) on 14 independent variables. The

dependent variable is the median value of owner-occupied homes in a census tract (MV). Independent variables include two structural attribute variables, eight neighborhood variables, and two accessibility variables such as average number of rooms (RM), proportion of structures built before 1940 (AGE), black population proportion (B), lower status population proportion (LSTAT), crime rate (CRIM), proportion of area zoned with large lots (ZN), proportion of non retail business area (INDUS), property tax rate (TAX), pupil-teacher ratio (PTRATIO), location contiguous to the Charles River (CHAS), weighted distances to the employment centers (DIS), and an index of accessibility (RAD) [25]. In addition, the pollution variable (NOX), which is the concentration of nitrogen oxides, is used to proxy air quality.

Gilley and Pace [52] specify the best functional form of the hedonic housing equation given by:

$$\begin{aligned} \log(MV) = & \beta_0 + \rho W \log(MV) + \beta_1 CRIM + \beta_2 ZN + \beta_3 INDUS + \beta_4 CHAS \\ & \beta_5 NOX^2 + \beta_6 RM^2 + \beta_7 AGE + \beta_8 \log(DIS) + \beta_9 \log(RAD) \\ & \beta_{10} TAX + \beta_{11} PTRATIO + \beta_{12} B + \beta_{13} LSTAT + \varepsilon \end{aligned} \quad (23)$$

where W is a distance matrix, and ρ is a scalar spatial autoregressive parameter.

The OLS, 2SLS, GMM, QML, MEEL, MEL, and MLEL estimates are reported in Table 10. All coefficients have the expected signs [24]. OLS exhibits biasness consistent with the outcomes of the Monte Carlo experiments. The 2SLS, MEEL, and MEL estimates are very comparable across parameters. The information theoretic and the QML estimators provide similar estimates of the model parameters with an exception being the spatial autoregressive parameter ρ .

Table 10. Estimated parameters of the hedonic housing model.

Parameters	OLS	2SLS	GMM	QML	MEEL	MEL	MLEL
β_0	0.441	0.276	0.033	0.393	0.276	0.276	0.432
β_2	1.034	1.237	1.850	1.137	1.237	1.237	1.137
β_3	−0.004	−0.004	−0.004	−0.004	−0.004	−0.004	−0.004
β_4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β_5	0.001	0.001	0.000	0.001	0.001	0.001	0.001
β_6	0.009	0.022	0.002	0.012	0.022	0.022	0.012
β_7	−0.131	−0.141	0.396	−0.146	−0.143	−0.141	−0.146
β_8	0.003	0.004	0.000	0.003	0.004	0.004	0.003
β_9	0.000	0.000	0.001	0.000	0.000	0.000	0.000
β_{10}	−0.157	−0.142	0.108	−0.161	−0.141	−0.142	−0.161
β_{11}	0.083	0.083	0.150	0.083	0.084	0.083	0.083
β_{12}	0.000	0.000	0.000	0.000	0.000	0.000	0.000
β_{13}	−0.003	−0.004	−0.008	−0.004	−0.004	−0.004	−0.004
β_{14}	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ρ	−0.286	−0.296	−0.623	−0.295	−0.294	−0.296	−0.295

6. Concluding Remarks

A general information theoretic estimator of the first order spatial autoregressive model was proposed, and special cases of Maximum Empirical Likelihood, Maximum Exponential Empirical Likelihood, and Maximum Log Euclidean Likelihood estimators were explored.

In extensive Monte Carlo experiments, small sample performances of the information theoretic estimators were evaluated and found to be robust to a range of sample sizes, spatial autocorrelation coefficients, distributional assumptions of disturbance term, and spatial weight matrices. Compared to

traditional estimators, it was found that in certain sampling situations information theoretic estimators of the first-order spatial autoregressive model have superior sampling properties and outperform traditional OLS, 2SLS and GMM estimators. However, while generally comparable in RMSE, the IT estimators do not consistently outperform the quasi-maximum likelihood estimator. Findings suggest that one can confidently use information theoretic estimators for estimating the first order spatial autocorrelation model for small sample sizes, and they represent defensible alternatives when maximum likelihood estimators cannot be easily computed. The estimation of the housing value model provided a practical illustration of the implementation and results forthcoming from OLS, 2SLS, QML, and information theoretic estimators. Future work could include comparison of IT estimators with Lin and Lee [53] and Kelejian and Prucha [54] who examined robust GMM estimators incorporating spatial autoregressive structures.

References

1. Owen, A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **1988**, *75*, 237–249.
2. Owen, A. Empirical likelihood for linear models. *Ann. Stat.* **1991**, *19*, 1725–1747.
3. Kitamura; Y.; Stutzer, M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica* **1997**, *65*, 861–874.
4. Fraser, I. An application of maximum entropy estimation: The demand for mean in the United Kingdom. *Appl. Econ.* **2000**, *32*, 45–59.
5. Golan; A.; Judge, G.; Zen, E. Estimating a demand system with nonnegativity constraints: Mexican meat demand. *Rev. Econ. Stat.* **2001**, *83*, 541–550.
6. Preckel, P.V. Least squares and entropy: A penalty function perspective. *Am. J. Agr. Econ.* **2001**, *83*, 366–377.
7. Miller, D.J. Entropy-based methods of modeling stochastic production efficiency. *Am. J. Agr. Econ.* **2002**, *84*, 1264–1270.
8. Zohrabian; A.; Traxler, G.; Caudill, S.; Smale, M. Valuing pre-commercial genetic resources: A maximum entropy approach. *Am. J. Agr. Econ.* **2003**, *85*, 429–436.
9. Imbens; G.W.; Spady, R.H.; Johnson, P. Information theoretic approaches to inference in moment condition models. *Econometrica* **1998**, *66*, 333–357.
10. Marsh, T.L.; Mittelhammer, R. Spatial and spatiotemporal econometrics. In *Advances in Econometrics*; Elsevier: Oxford, UK, 2004; Volume 18, Chapter 7, pp. 203–238.
11. Fernandez-Vazquez, E.; Mayor-Fernandez, M.; Rodriguez-Valez, J. Estimating spatial autoregressive models by GME-GCE techniques. *Int. Reg. Sci. Rev.* **2009**, *32*, 148–172.
12. Case, A.C. Neighborhood influence and technological change. *Reg. Sci. Urban Econ.* **1992**, *22*, 491–508.
13. Case, A.C. Spatial patterns in household demand. *Econometrica* **1991**, *59*, 953–965.
14. Anselin, L.; Vagra, A.; Acs, Z. Local geographic spillovers between university research and high technology innovations. *J. Urban Econ.* **1997**, *42*, 422–448.
15. Haining, R.P. Testing a spatial interacting-markets hypothesis. *Rev. Econ. Stat.* **1984**, *66*, 576–583.

16. Wu, J. Environmental amenities and the spatial pattern of urban sprawl. *Am. J. Agr. Econ.* **2001**, *83*, 691–997.
17. Goodwin, B.K.; Piggott, N.E. Spatial market integration in the presence of threshold effects. *Am. J. Agr. Econ.* **2001**, *83*, 302–317.
18. Roe, B.; Irwin, E.; Sharp, J.S. Pigs in space: Modeling the spatial structure of hog production in traditional and nontraditional production regions. *Am. J. Agr. Econ.* **2002**, *84*, 259–278.
19. Thompson, S.R.; Donggyu, S.; Bohl, M.T. Spatial market efficiency and policy regime change: Seemingly unrelated error correction model estimation. *Am. J. Agr. Econ.* **2002**, *84*, 1042–1053.
20. Sephton, P.S. Spatial market arbitrage and threshold cointegration. *Am. J. Agr. Econ.* **2003**, *85*, 1041–1046.
21. Saak, A.E. Spatial and temporal marketing considerations under marketing loan programs. *Am. J. Agr. Econ.* **2003**, *85*, 872–887.
22. Sarmiento, C.; Wilson, W.W. Spatial modeling in technology adoption decisions: The case of shuttle train elevators. *Am. J. Agr. Econ.* **2005**, *87*, 1034–1045.
23. Negassa, A.; Myers, R.J. Estimating policy effects on spatial market efficiency: An extension to the parity bounds model. *Am. J. Agr. Econ.* **2007**, *89*, 338–352.
24. Harrison, D.; Rubinfeld, D. Hedonic housing prices and the demand for clean air. *J. Environ. Econ. Manag.* **1978**, *5*, 81–102.
25. Gilley, O.W.; Pace, R. On the Harrison and Rubinfeld data. *J. Environ. Econ. Manag.* **1996**, *31*, 403–405.
26. Cliff, A.; Ord, K. *Spatial Processes, Models, and Applications*; Pion: London, UK, 1981.
27. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer: Boston, MA, USA, 1988.
28. Cressie, N. *Statistics for Spatial Data*; Wiley: New York, NY, USA, 1993.
29. LeSage, J.; Pace, R.K. Spatial and spatiotemporal econometrics. In *Advances in Econometrics*; Elsevier: Oxford, UK, 2004; Volume 18, Chapter Introduction, pp. 1–32.
30. Horn, R.; Johnson, C. *Matrix Analysis*; Cambridge University Press: New York, NY, USA, 1985.
31. LeSage, J.P.; Pace, R.K. *Introduction to Spatial Econometrics*; Taylor-Francis/CRC Press: Boca Raton, FL, USA, 2009.
32. LeSage, J. Bayesian estimation of spatial autoregressive models. *Int. Reg. Sci. Rev.* **1997**, *20*, 113–129.
33. Whittle, P. On stationary processes in the plane. *Biometrika* **1954**, *41*, 434–449, Parts 3 and 4.
34. Ord, K. Estimation methods for models of spatial interaction. *J. Am. Stat. Assoc.* **1975**, *70*, 120–126.
35. Lee, L. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Economet. Theor.* **2002**, *18*, 252–277.
36. Lee, L. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **2004**, *72*, 1899–1925.
37. Barry, R.P.; Pace, R.K. A Monte Carlo Estimator of the log determinant of large sparse matrices. *Lin. Algebra Appl.* **1999**, *289*, 41–54.
38. Kelejian, H.H.; Prucha, I. A Generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. R. Estate Finance Econ.* **1998**, *17*, 99–121.

39. Kelejian, H.H.; Prucha, I. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Reg. Sci. Urban Econ.* **2002**, *32*, 691–707.
40. Lee, L. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Department of Economics, Ohio State University, Columbus, OH, USA. Unpublished work, 2001.
41. Lee, L. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J. Econometrics* **2007**, *137*, 489–514.
42. Lee, L. The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models. *J. Econometrics* **2007**, *140*, 155–189.
43. Lee, L. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Economet. Rev.* **2003**, *22*, 307–335.
44. Hansen, L. Large sample properties of generalized method of moments estimators. *Econometrica* **1982**, *50*, 1029–1054.
45. Bernardini Papalia, R. Incorporating spatial structures in ecological inference: An information theoretic approach. *Entropy* **2010**, *12*, 2171–2185.
46. Thomas, D.; Grunkemeier, G. Confidence interval estimation of survival probabilities for censored data. *J. Am. Stat. Assoc.* **1975**, *70*, 865–871.
47. Owen, A. Empirical likelihood ratio confidence regions. *Ann. Stat.* **1990**, *18*, 90–120.
48. Cressie, N.; Read, T. Multinomial goodness-of-fit tests. *J. Roy. Stat. Soc. B Stat. Meth.* **1984**, *46*, 440–464.
49. Qin, J.; Lawless, J. Empirical likelihood and general estimating equations. *Ann. Stat.* **1994**, *22*, 300–325.
50. Kelejian, H.H.; Prucha, I. A generalized moments estimator for the autoregressive parameter in a spatial model. *Int. Econ. Rev.* **1999**, *40*, 509–533.
51. Das, D.; Kelejian, H.H.; Prucha, I. Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Paper Reg. Sci.* **2003**, *82*, 1–26.
52. Gilley, O.W. Pace, R.K. The Harrison and Rubinfeld data revisited. *J. Environ. Econ. Manag.* **1996**, *31*, 403–405.
53. Lin, X.; Lee, L.F. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *J. Econometrics* **2010**, *157*, 34–52.
54. Kelejian, H.H.; Prucha, I.R. Specication and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *J. Econometrics* **2010**, *157*, 53–67.