Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Regional Science and Urban Economics 37 (2007) 363–374
regional SCIENCE & urban ECONOMICS
www.elsevier.com/locate/regec

The relative efficiencies of various predictors in spatial econometric models containing spatial lags

Harry H. Kelejian *, Ingmar R. Prucha

University of Maryland, College Park, MD 20742, United States

Received in revised form 28 September 2006; accepted 6 November 2006
Available online 2 January 2007

Abstract

The purpose of this paper is to describe prediction efficiencies of various suboptimal predictors relative to the efficient (kriging) minimum mean square error predictor in spatial models containing spatial lags in both the dependent variable and the error term. Suboptimal predictors have been suggested in the literature. One reason is that they are suggested on an intuitive level; another is that they are computationally less tedious. We describe these relative efficiencies theoretically, as well as empirically. Among other things our results suggest that one of the intuitively suggestive suboptimal predictors is especially inefficient.

JEL classification: C1; C21
Keywords: Spatial models with spatial lags; Optimal and suboptimal prediction efficiencies; BLUP; Kriging

1. Introduction

Linear spatial models have wide applications in economics, geography, and regional science, among other areas of research.[1] As with many research efforts, prediction is one of the applications of this modeling. Although the determination of an efficient predictor is fairly straight forward, [2]

* Corresponding author. Tel.: +1 301 405 3492; fax: +1 301 405 3542.
E-mail address: Kelejian@econ.umd.edu (H.H. Kelejian).

[1] Classic references on spatial models are Cliff and Ord (1973, 1981), Anselin (1988), and Cressie (1993). For a variety of recent studies which relate to spatial techniques see e.g., Cohen and Morrison Paul (2004), Rey and Boarnet (2004), Yuzefovich (2003), Kapoor (2003), Pinske, Slade, and Brett (2003), Bell and Bockstael (2000), Kelejian and Robinson (2000), Buettner (1999), LeSage (1999), Bollinger and Ihlanfeldt (1997), and Audretsch and Feldmann (1996), Bernat (1996), and Besley and Case (1995).

[2] An early study relating to best linear unbiased prediction (BLUP) in a GLS-type model is Goldberger (1962); see also Cressie (1993, Chapter 3) who describes optimal prediction in a spatial framework.

0166-0462/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2006.11.005

suboptimal predictors have been considered in the literature. One reason for this is that suboptimal predictors are often suggested on an intuitive level; another is that they are typically computationally simpler than efficient predictors.[3]

Essentially, the purpose of this paper is to give results which illustrate the extent of inefficiencies of various predictors in a spatial model. Specifically, we consider prediction issues in the context of a linear spatial model which contains exogenous variables, a spatially lagged dependent variable, and a spatially lagged error term.[4] In the context of this model, we consider three nested information sets which a researcher would have access to and might be considered for purposes of prediction. Corresponding to these information sets we consider three predictors defined as conditional means based on these information sets. We refer to the predictor based on the largest information set as the full information predictor, and to the other two predictors as limited information predictors. For future reference we note that the smallest of these information sets only contains the exogenous variables and the weighting matrix. As expected, predictors corresponding to the larger information sets are more complex than those corresponding to smaller sets, and so there are trade-offs between simplicity and prediction efficiency. We also consider a "user-friendly and intuitive" predictor which is biased, namely, the right hand side of the regression model. The bias arises because of the correlation between the spatially lagged dependent variable and the error term.[5] Finally, we consider an intuitive but biased predictor, as well as the full information predictor in the context of a spatial error model as a special case of our general model.

For each of our considered predictors we give an estimate of its predictive efficiency relative to the full information predictor. All of our results specialize to models in which one, or both of these spatial lags are absent. In addition, qualitative extensions to space-time models will become evident.

As a preview, it turns out that in our general model the worst predictor, by far, is the conditional mean which is based only on the exogenous variables and the weighting matrix. For example, in the numerical experiments we considered, its predictive efficiency relative to that of the full information predictor is, on average, only between 4% and 12.2%. Although the biased predictor is a considerable improvement, it is still significantly worse than the full information predictor, as well as the other conditional mean predictor considered which recognizes spatial lags in both the dependent variable and in the error term. Interestingly, in the context of a spatial error model, the intuitive but biased predictor performs reasonably well in that its prediction efficiency relative to the full information predictor is, on average, between, roughly, 91.7% and 97.7%. Again, in this model the predictor determined as the conditional mean on the exogenous variables and weighting matrix is substantially worse than the other considered predictors.

We also find that the prediction inefficiencies involved for all of our considered predictors relative to the corresponding full information predictor generally increase as the sparseness of the

[3] Specific cases will be indicated below; at this point we note that Bannerjee, Carlin, and Gelfand (2003) have criticized the way researchers often use spatial models in an *ad hoc* way to form predictions.

[4] Anselin (1988, pp. 87–88) gave results which have been interpreted as suggesting that such models are not identified if the weighting matrix relating to the spatial lag of the dependent variable is the same as that relating to the error term. This may be one reason that such models are typically not considered in practice, see e.g., Dubin (2003, 2004). This is unfortunate because such models are rich in patterns of spatial correlations and are, under reasonable conditions, clearly identified — see, e.g., Kelejian and Prucha (1998, 1999) and Lee (2003).

[5] Among others, such a predictor was considered by Dubin (2004); Kelejian and Yuzefovich (2004) considered the conditional mean predictor based only on the exogenous variables and weighting matrix.

weighting matrix increases. The suggestion, therefore, is to avoid simpler but inefficient predictors when the weighting matrix is relatively sparse. Finally, the inefficiencies involved in the simpler predictors increase dramatically at certain extreme values of the autoregressive parameters.

In Section 2 we specify the model. Section 3 describes the predictors considered, and their mean squared errors of prediction. The experimental design and corresponding numerical results relating to the relative prediction efficiencies are given in Section 4. Concluding comments are given in Section 5.

2. Model specification

Consider the model

$$\begin{aligned} y &= \lambda Wy + X\beta + u, \\ u &= \rho Wu + \varepsilon, \end{aligned} \tag{1}$$

where y is the $n\times 1$ vector of values of the dependent variable, W is an $n\times n$ nonstochastic weighting matrix, X is an $n\times k$ nonstochastic matrix of observations on k exogenous variables (i.e., our analysis is conditional on the exogenous variables), u is an $n\times 1$ vector of disturbances, ε is an $n\times 1$ vector of innovations, λ and ρ are scalar autoregressive parameters, and β is a $k\times 1$ vector of parameters. The above model has been referred to as a SARAR(1,1) model in the literature — see, e.g., Kelejian and Prucha (1998). It is a variant of the spatial model introduced by Cliff and Ord (1973, 1981).[6]

Our discussion of prediction based on Eq. (1) considers the case in which the dependent vector is observed except for its i-th ($1\leq i\leq n$) element, and all elements of W and X are observed. We refer to this case as the full information case.

We make the following assumptions.

Assumption 1. The diagonal elements of W are all zero.

Assumption 2. $|\lambda|<1$, $|\rho|<1$, and $(I-\alpha W)$ is non-singular for all $|\alpha|<1$.

Assumption 3. $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$.

Given W and X are matrices of known constants, Assumptions 2 and 3 imply

$$\begin{aligned} u &= (I-\rho W)^{-1}\varepsilon, \\ y &= (I-\lambda W)^{-1}X\beta + (I-\lambda W)^{-1}(I-\rho W)^{-1}\varepsilon, \end{aligned} \tag{2}$$

[6] We note that the model in Eq. (1) can also be thought of in a panel data framework. As one example, suppose there are data on R cross sectional units for each of T time periods. In this case $n=RT$. In order to avoid notational confusion with material below, let Y denote n × 1 dependent in vector in Eq. (1). Then one would have

$$Y' = (Y_1', \ldots, Y_T')$$

where Y_t is the $R\times 1$ vector of observations on the dependent variable at time $t=1$,…,T. The matrix X and the vector u would be defined in a similar fashion. Note, in this case, that time dynamics involving the dependent variable would correspond to the case in which the $RT\times RT$ weighting matrix is a lower block diagonal matrix. Clearly this discussion can be extended to the case in which the number of cross sectional units are not the same in all of the time periods.

and so

$$u \sim N\left(0, \sigma_\varepsilon^2 \sum\nolimits^u\right),$$
$$y \sim N\left(\mu_y, \sigma_\varepsilon^2 \sum\nolimits^y\right), \tag{3}$$

with

$$\mu_y = (I-\lambda W)^{-1} X\beta,$$
$$\sum\nolimits^u = (I-\rho W)^{-1}(I-\rho W')^{-1},$$
$$\sum\nolimits^y = (I-\lambda W)^{-1} \sum\nolimits^u (I-\lambda W')^{-1}.$$

Let S_{-i} be the $n-1 \times n$ selector matrix which is identical to the $n \times n$ identity matrix I except that the i-th row of I is deleted. Let y_{-i} be the available $n-1$ observations on the dependent variable. Then

$$y_{-i} = S_{-i} y,$$

and, given Eq. (3)

$$y_{-i} \sim N(S_{-i}\mu_y, \sigma_\varepsilon^2 S_{-i} \sum\nolimits^y S'_{-i}). \tag{4}$$

For future reference note that the distribution in Eq. (4) involves all of the model parameters.

Our exact theoretical results below relate to the *limits* of relative prediction efficiencies of the considered predictors because those results are conditional on the parameters λ, ρ, and β. Of course, in practice the model parameters will not be known and therefore must be estimated prior to prediction. In such cases relative prediction efficiencies would clearly depend upon just how well the model parameters are estimated; this, in turn, would depend upon a number of factors, such as the particular exogenous variables considered, the sample size, etc.[7] Given the model in Eq. (1), Assumptions 1–3, data on X, W, and y_{-i}, and further reasonable conditions these parameters can, e.g., be consistently estimated by the maximum likelihood procedure based on Eq. (4), or by a variant of the Kelejian and Prucha (1998) S2SLS procedure which requires less data.[8]

3. Predictors and their mean squared errors

3.1. Predictors

Given the structure of the model in Eq. (1), Assumptions 1–3, data on X, W, and y_{-i}, and λ, β, ρ the objective is to predict the i-th element of y say y_i. In the following we discuss the efficient predictor corresponding to this so-called full information set. We also discuss less efficient but simpler predictors that are based on smaller information sets. The considered predictors and their motivations are described below. At this point, for the convenience of the reader, we note that the

[7] In a model such as Eq. (1) an analysis of prediction efficiencies in this case would clearly be numerical, e.g. via Monte Carlo methods.

[8] Further details relating to such an estimation can be obtained by writing to the authors.

minimum mean squared error predictor based on a given information set is the conditional mean corresponding to that information set.[9]

Given Eq. (1), y_i is determined as

$$\begin{aligned} y_i &= \lambda w_{i.}\, y + x_{i.}\beta + u_i, \\ u_i &= \rho w_{i.} u + \varepsilon_i, \end{aligned} \tag{5}$$

where $w_{i.}$ and $x_{i.}$ are, respectively, the i-th rows of W and X, u_i and ε_i are the i-th elements of u and ε, and $w_{i.}y$ and $w_{i.}u$ denote the i-th elements of the spatial lags Wy and Wu. Note that $w_{i.}\, y$ does not include y_i in light of Assumption 1.

We consider three information sets, namely

$$\begin{aligned} \Lambda_1 &= \{X, W\}, \\ \Lambda_2 &= \{X, W, w_{i.}y\}, \\ \Lambda_3 &= \{X, W, y_{-1}\}. \end{aligned} \tag{6}$$

Clearly Λ_1 and Λ_2 are subsets of the full information set Λ_3, and Λ_1 is a subset of Λ_2. The information set Λ_3 includes all available $n-1$ observations on the dependent vector. In contrast Λ_2 only contains information on a linear combination of them, while no information on the dependent vector is contained in Λ_1. As will become evident below, the use of Λ_1 corresponds to predictions motivated by the reduced form in Eq. (2), and was considered in Kelejian and Yuzefovich (2004); the use of Λ_2 corresponds to predictions motivated by Eq. (5).

We consider five predictors of y_i, which will be denoted as $y_i^{(p)}$, $p=1,\ldots,5$. The first three are the conditional means corresponding to the information sets Λ_p, $p=1,2,3$, in Eq. (6). By construction these predictors are unbiased (conditional on the corresponding information set) and given by:

$$\begin{aligned} y_i^{(1)} &= E(y_i|\Lambda_1) \\ &= (I-\lambda W)_{i.}^{-1} X\beta \end{aligned} \tag{7}$$

$$\begin{aligned} y_i^{(2)} &= E(y_i|\Lambda_2) \\ &= \lambda w_{i.}y + x_{i.}\beta + \frac{\mathrm{cov}(u_i, w_{i.}y)}{\mathrm{var}(w_{i.}y)}[w_{i.}y - E(w_{i.}y)], \end{aligned} \tag{8}$$

$$\begin{aligned} y_i^{(3)} &= E(y_i|\Lambda_3) \\ &= \lambda w_{i.}y + x_{i.}\beta + \mathrm{cov}(u_i, y_{-i})[\mathrm{VC}(y_{-1})]^{-1}[y_{-i} - E(y_{-i})], \end{aligned} \tag{9}$$

[9] For the convenience of the reader, we note the following. Using evident notation, let $(Z_1, Z_2) \sim N(\mu, V)$ where

$$\mu' = (\mu_1', \mu_2');\ V = \{V_{ij}\}, i,j = 1,2.$$

Then the minimum mean squared error predictor of Z_1 given Z_2, and the corresponding predictor variance–covariance matrix are

$$\begin{aligned} E(Z_1|Z_2 = z_2) &= \mu_1 + V_{12}V_{22}^{-1}(z_2 - \mu_2) \\ VC(Z_1|Z_2 = z_2) &= V_{11} - V_{12}V_{22}^{-1}V_{21}, \end{aligned}$$

see, e.g. Greene (2003, p. 872).

where

$$
\begin{aligned}
E(w_{i.}y) &= w_{i.}(I-\lambda W)^{-1}X\beta, \\
\mathrm{var}(w_{i.}y) &= \sigma_\varepsilon^2 w_{i.} \sum\nolimits^{y} w'_{i.}, \\
\mathrm{cov}(u_i, w_{i.}y) &= \sigma_\varepsilon^2 \sigma^u_{i.}(I-\lambda W')^{-1} w'_{i.}, \\
E(y_{-i}) &= S_{-i}(I-\lambda W)^{-1}X\beta, \\
\mathrm{VC}(y_{-i}) &= \sigma_\varepsilon^2 S_{-i} \sum\nolimits^{y} S'_{-i}, \\
\mathrm{cov}(u_i, y_{-i}) &= \sigma_\varepsilon^2 \sigma^u_{i.}(I-\lambda W')^{-1} S'_{-i}.
\end{aligned}
\tag{10}
$$

In the above expressions $(I-\lambda W)^{-1}_{i.}$ and $\sigma^u_{i.}$ denote the *i*-th rows respectively, of $(I-\lambda W)^{-1}$ and $\sum^{\mathrm{u}}$.

The fourth predictor we consider is given by

$$
y_i^{(4)} = \lambda w_{i.}y + x_{i.}\beta. \tag{11}
$$

The reason for including this predictor in our analysis is that, via Eq. (5), it has intuitive appeal and has been considered in the literature, as was discussed in the introduction. It may be viewed as a restricted version of the predictor $y_i^{(2)}$, which implicitly assumes that $cov(u_i, w_{i.}y)=0$. Of course, as seen from Eq. (10), in general $cov(u_i,w_{i.}y)\neq 0$ unless both ρ and λ are zero. Thus, conditional on the information set Λ_2 the predictor is biased, and the bias is given by

$$
\mathrm{bias}_i = y_i^{(4)} - y_i^{(2)} = -\frac{\mathrm{cov}(u_i, w_{i.}y)}{\mathrm{var}(w_{i.}y)}[w_{i.}y - E(w_{i.}y)]. \tag{12}
$$

In passing we note, from Eqs. (7) and (11), that if $\lambda=0$, $y_i^{(4)}=y_i^{(1)}$.

Our fifth, and biased predictor, $y_i^{(5)}$ relates to a special case of model Eq. (1), namely the spatial error model. In this model $\lambda=0$ and so, via Eq. (5),

$$
\begin{aligned}
y_i &= x_{i.}\beta + u_i \\
u_i &= \rho w_{i.}u + \varepsilon_i
\end{aligned}
\tag{13}
$$

For this model, we take $y_i^{(5)}$ to be

$$
\begin{aligned}
y_i^{(5)} &= x_i\beta + \rho w_{i.}u \\
&\equiv x_i\beta + \rho w_{i.}[y - X\beta]
\end{aligned}
\tag{14}
$$

Given our scenario $y_i^{(5)}$ is a feasible predictor in that it does not involve y_i because the *i-th* element of $w_{i.}$ is zero. Clearly, $y_i^{(5)}$ would be suggested by arguments which are similar to those suggesting $y_i^{(4)}$ in Eq. (11). At this point note that the bias arises because the covariance between $\rho w_{i.}u$ and ε_i is not zero.

For future reference we also note that if $\rho=0$ in the spatial error model, or if $\lambda=\rho=0$ in our general model Eq. (5), then all five of our predictors are the same: $y_i^{(1)}=y_i^{(5)}$, for $i=1,2,3,4$.

3.2. Mean squared errors

Let $e_i^{(j)}$ be the error in predicting y_i when using the predictor $y_{\mathrm{i}}^{(j)}$:

$$
e_i^{(j)} = y_i - y_i^{(j)}, j = 1, \ldots, 5 \tag{15}
$$

Then, the forecast variances corresponding to the first three minimum mean squared error predictors in Eqs. (7)–(9) are given, respectively, by:

$$\begin{aligned}\mathrm{var}(e_i^{(1)}|\Lambda_1) &= \mathrm{var}((I-\lambda W)_{i.}^{-1}u|\Lambda_1)\\ &= \sigma_\varepsilon^2(I-\lambda W)_{i.}^{-1}\sum\nolimits^u(I-\lambda W)_{i.}^{-1\prime},\end{aligned} \tag{16}$$

$$\begin{aligned}\mathrm{var}(e_i^{(2)}|\Lambda_2) &= \mathrm{var}(u_i|\Lambda_2)\\ &= \sigma_\varepsilon^2\sigma_{ii}^u-[\mathrm{cov}(u_i,w_{i.}y)]^2[\mathrm{var}(w_{i.}y)]^{-1},\end{aligned} \tag{17}$$

$$\begin{aligned}\mathrm{var}(e_i^{(3)}|\Lambda_3) &= \mathrm{var}(u_i|\Lambda_3)\\ &= \sigma_\varepsilon^2\sigma_{ii}^u-\mathrm{cov}(u_i,y_{-i})[\mathrm{VC}(y_{-i})]^{-1}\mathrm{cov}(u_i,y_{-i})',\end{aligned} \tag{18}$$

where σ_{ii}^u is the i-th diagonal element of $\sum^{\mathrm{u}}$. As indicated, conditional on the corresponding information set, the predictors $y_i^{(1)}$, $y_i^{(2)}$, and $y_i^{(3)}$ are unbiased and so the forecast variances in Eqs. (16)–(18) are also mean squared errors. Because of this we will, at times, use the notation

$$\mathrm{var}(e_i^{(p)}|\Lambda_p) = \mathrm{MSE}(y_i^{(p)}), \quad p = 1,2,3. \tag{19}$$

Now consider the biased predictor $y_i^{(4)}$. In light of Eq. (12) have $y_i^{(4)}=y_i^{(2)}+\mathrm{bias}_i$ and thus the mean squared error of $y_i^{(4)}$ conditional on Λ_2 is given by

$$\begin{aligned}\mathrm{MSE}(y_i^{(4)}|\Lambda_2) &= \mathrm{bias}_i^2 + \mathrm{MSE}(y_i^{(2)})\\ &\geq \mathrm{MSE}(y_i^{(2)})\end{aligned} \tag{20}$$

For future reference note from Eq. (12) that $(y_i^{(4)}|\Lambda_2)$ depends upon y_{-i} via the *bias* term. Since the first three mean squared error $\mathrm{MSE}(y_i^{(p)})$, $p=1,2,3$ do not depend upon y_{-i} they can be thought of as being averaged over the realizations of y_i. In order to obtain a comparable measure of predictive efficiency relating to $y_i^{(4)}$, we therefore average $\mathrm{MSE}(y_i^{(4)}|\Lambda_2)$ over y_{-i} by taking the expected value of $\mathrm{MSE}(y_i^{(4)}|\Lambda_2)$ conditional only upon Λ_1 namely, via Eq. (20)

$$\begin{aligned}E\left[\mathrm{MSE}(y_i^{(4)}|\Lambda_2)\right]|\Lambda_1] &= \mathrm{MSE}(y_i^{(4)})\\ &= \left(\frac{\mathrm{cov}(u_i,w_{i.}y)}{\mathrm{var}(w_{i.}y)}\right)^2\mathrm{var}(w_{i.}y) + \mathrm{MSE}(y_i^{(2)})\\ &= \sigma_\varepsilon^2\sigma_{ii}^u.\end{aligned} \tag{21}$$

The last line in Eq. (21) should be evident because, given λ and β, the prediction error is, via Eq. (5), $y_i-\lambda w_{i.}y-x_{i.}\beta=u_i$.

Finally, consider the spatial error model and the intuitive but biased predictor $y_i^{(5)}$ in Eq. (14). Given the values of the model parameters β and ρ it should be clear that $y_i-y_i^{(5)}=\varepsilon_i$ and so an argument similar to that explaining the last line of Eq. (21) implies that the mean square error of $y_i^{(5)}$ average over the realizations of y_{-i} is just

$$\mathrm{MSE}(y_i^{(5)}) = \sigma_\varepsilon^2 \tag{22}$$

Our empirical results below relate to $\mathrm{MSE}(y_i^{(p)})$, $p=1,\ldots,5$. On a theoretical level, since $\Lambda_1 \subset \Lambda_2 \subset \Lambda_3$ and given the results in Eqs. (18), (21) and (22), it follows that[10]

$$\begin{array}{lcl} \mathrm{MSE}(y_i^{(1)}) & \geq & \mathrm{MSE}(y_i^{(2)}) \quad \geq \quad \mathrm{MSE}(y_i^{(3)}), \\ \mathrm{MSE}(y_i^{(4)}) & \geq & \mathrm{MSE}(y_i^{(2)}), \\ \mathrm{MSE}(y_i^{(5)}) & \geq & \mathrm{MSE}(y_i^{(3)}), \\ \text{If } \lambda & = & 0 : \mathrm{MSE}(y_i^{(4)}) \quad = \quad \mathrm{MSE}(y_i^{(1)}). \\ \text{If } \lambda & = & \rho = 0 : \mathrm{MSE}(y_i^{(j)}) \quad = \quad \mathrm{MSE}(y_i^{(1)}), j \quad = \quad 2, \ldots, 5. \end{array} \tag{23}$$

Note if $\lambda \neq 0$, the relative magnitudes of $\mathrm{MSE}(y_i^{(4)})$ and $\mathrm{MSE}(y_i^{(1)})$ are not certain. The reason for this is that although $(y_i^{(4)})$ is biased, it is based on a larger information set than that of $(y_i^{(1)})$. Therefore, the reduction in the variance due to that larger information set could compensate for the bias. For the case of the spatial error model ($\lambda=0$) the magnitude of $\mathrm{MSE}(y_i^{(5)})$ is greater than $\mathrm{MSE}(y_i^{(3)})$ because (y_i^{3}) is the full information conditional mean predictor.[11] On the other hand, in general, the magnitude of $\mathrm{MSE}(y_i^{(5)})$ relative to those of $\mathrm{MSE}(y_i^{(1)})$, $\mathrm{MSE}(y_i^{(2)})$, and $\mathrm{MSE}(y_i^{(4)})$ is not certain. Of course if $\lambda=\rho=0$ all five predictors are equivalent and so their mean square errors are the same.

4. Comparisons of prediction efficiencies

4.1. Design of the experiments

In this section we give illustrative numerical results regarding the mean squared errors described in (16)–(18), (21) and (22) in order to gain insights as to the quantitative importance of available information and its proper use for prediction purposes in models such as Eq. (1). These results are based on two weighting matrices which differ in their degree of sparseness. The first matrix is such that each unit is directly related to the two units which are immediately after it and immediately before it in the ordering. Specifically, nonzero elements in the i-th row of this matrix for $i=3,\ldots, n-2$ are $w_{i,i+1}$ and $w_{i,i+2}$ (two ahead) and $w_{i,i-1}$ and $w_{i,i-2}$ (two behind). This matrix is defined in a circular world so that, e.g., in the first row the nonzero elements are $w_{1,2}$ and $w_{1,3}$ (two ahead) and $w_{1,n}$ and $w_{1,n-1}$ (two behind). Rows 2, $n-1$ and n are defined in a corresponding circular fashion. The matrix is row normalized, and all of its nonzero elements are equal to 1/4. Kelejian and Prucha (1999) describe a matrix of this sort as "two ahead and two behind". The second weighting matrix is identical in structure to the first except that it is "eight ahead and eight behind" so that its nonzero elements are all 1/16. These matrices were considered because they easily capture aspects of sparseness. Of course, as is typical of numerical prediction results in spatial models, our numerical results below depend upon the use of these weighting matrices.[12]

In addition to the weighting matrix, the model parameters that enter into the mean squared errors of the predictors are λ, ρ and σ_ε^2. For each weighting matrix we give results relating to 25

[10] The result in the first line of Eq. (23) follows as a straight forward application of the result in Mood, Graybill, and Boes (1974, p.159).

[11] Also note that Eqs. (9), (18) and (22) imply that

$$\mathrm{MSE}(y_i^{(3)}) = \mathrm{var}(\varepsilon_i | y_{-i}) \leq \mathrm{var}(\varepsilon_i) = \mathrm{MSE}(y_i^{(5)}).$$

[12] Clearly, it might be of interest to consider a larger numerical study of prediction efficiencies in which various other characteristics of weighting matrices are considered.

combinations of λ and ρ, specifically, for all combinations of $\lambda,\rho = -.9, -.4, 0, .4, .9$. For relative comparisons of mean squared errors of the predictors, σ_ε^2 cancels and so σ_ε^2 is set arbitrarily equal to one. Finally, all of our results are given for the case in which $n=100$.

4.2. Results

Numerical results relating to the prediction efficiencies based on the weighting matrix “two ahead and two behind” are given in Table 1; results relating to the weighting matrix “eight ahead and eight behind” are given in Table 2. Specifically, we report sample averages over $i=1,\ldots,n$ for $\mathrm{MSE}(y_i^{(p)})$ for $p=1,\ldots,5$. The results for the spatial error model correspond to the case in which $\lambda=0$.

Consider first the results in Table 1 and note that those results are consistent with the theoretical notions in Eq. (23). Note also that, by far, the largest mean squared errors, and therefore, the lowest efficiencies correspond to $y_i^{(1)}$, i.e., the predictor based on $\Lambda_1=\{X,W\}$. Note also that the MSEs of this predictor are “extreme” when $\lambda=\rho=.9$. The only other predictor which has more moderated but still “extreme” MSEs is the biased predictor $y_i^{(4)}$. The “extreme” MSE

Table 1
MSEs based on the weighting matrix two ahead and two behind

λ	ρ	MSEs				
		$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(3)}$	$y_i^{(4)}$	$y_i^{(5)}$
−.9	−.9	4.35395	0.60998	0.38750	1.68204	*N*/*R*
−.9	−.4	2.31239	0.68523	0.58532	1.11021	*N*/*R*
−.9	0.0	1.68204	0.85038	0.83160	1.00000	*N*/*R*
−.9	0.4	1.40991	1.17256	1.15344	1.17308	*N*/*R*
−.9	0.9	3.16929	1.67722	1.45504	7.98954	*N*/*R*
−.4	−.9	2.31239	0.68523	0.58532	1.68204	*N*/*R*
−.4	−.4	1.37494	0.81520	0.78889	1.11021	*N*/*R*
−.4	0.0	1.11021	0.96250	0.96154	1.00000	*N*/*R*
−.4	0.4	1.09287	1.08780	1.08272	1.17308	*N*/*R*
−.4	0.9	4.60197	1.07524	1.07012	7.98954	*N*/*R*
0.0	−.9	1.68204	0.85038	0.83160	1.68204	1.0
0.0	−.4	1.11021	0.96250	0.96154	1.11021	1.0
0.0	0.0	1.00000	1.00000	1.00000	1.00000	1.0
0.0	0.4	1.17308	0.96273	0.96154	1.17308	1.0
0.0	0.9	7.98954	0.88041	0.83160	7.98954	1.0
0.4	−.9	1.40991	1.17256	1.15344	1.68204	*N*/*R*
0.4	−.4	1.09287	1.08780	1.08272	1.11021	*N*/*R*
0.4	0.0	1.17308	0.96273	0.96154	1.00000	*N*/*R*
0.4	0.4	1.77454	0.84597	0.81994	1.17308	*N*/*R*
0.4	0.9	19.7150	0.82555	0.65233	7.98954	*N*/*R*
0.9	−.9	3.16929	1.67722	1.45504	1.68204	*N*/*R*
0.9	−.4	4.60197	1.07524	1.07012	1.11021	*N*/*R*
0.9	0.0	7.98954	0.88041	0.83160	1.00000	*N*/*R*
0.9	0.4	19.71500	0.82555	0.65233	1.17308	*N*/*R*
0.9	0.9	469.81046	1.40530	0.49167	7.98954	*N*/*R*
Col. Ave. $p=1,\ldots,5$		22.673	1.001	0.906	2.591	1.0
$\frac{\text{Column Average}_3}{\text{Column Average}_p} * 100$		4.0	90.5	100	35.0	91.7

Table 2
MSEs based on the weighting matrix eight ahead and eight behind

λ	ρ	MSEs				
		$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(3)}$	$y_i^{(4)}$	$y_i^{(5)}$
−.9	−.9	1.29624	0.79928	0.69037	1.09708	*N/R*
−.9	−.4	1.16707	0.87272	0.83753	1.02295	*N/R*
−.9	0.0	1.09708	0.95545	0.95181	1.00000	*N/R*
−.9	0.4	1.06332	1.04432	1.04077	1.04906	*N/R*
−.9	0.9	1.64018	1.09616	1.08237	3.18349	*N/R*
−.4	−.9	1.16707	0.87272	0.83753	1.09708	*N/R*
−.4	−.4	1.06935	0.94011	0.93329	1.02295	*N/R*
−.4	0.0	1.02295	0.99027	0.99010	1.00000	*N/R*
−.4	0.4	1.02357	1.02014	1.01939	1.04906	*N/R*
−.4	0.9	2.12209	1.01097	1.00974	3.18349	*N/R*
0.0	−.9	1.09708	0.95545	0.95181	1.09708	1.0
0.0	−.4	1.02295	0.99027	0.99010	1.02295	1.0
0.0	0.0	1.00000	1.00000	1.00000	1.00000	1.0
0.0	0.4	1.04906	0.99033	0.99010	1.04906	1.0
0.0	0.9	3.18349	0.96447	0.95181	3.18349	1.0
0.4	−.9	1.06332	1.04432	1.04077	1.09708	*N/R*
0.4	−.4	1.02357	1.02014	1.01939	1.02295	*N/R*
0.4	0.0	1.04906	0.99033	0.99010	1.00000	*N/R*
0.4	0.4	1.22877	0.95749	0.95194	1.04906	*N/R*
0.4	0.9	6.86717	0.94325	0.89509	3.18349	*N/R*
0.9	−.9	1.64018	1.09616	1.08237	1.09708	*N/R*
0.9	−.4	2.12209	1.01097	1.00974	1.02295	*N/R*
0.9	0.0	3.18349	0.96447	0.95181	1.00000	*N/R*
0.9	0.4	6.86717	0.94325	0.89509	1.04906	*N/R*
0.9	0.9	150.65631	1.12765	0.82693	3.18349	*N/R*
Col. Ave., $p=1,\ldots 5$		7.829	0.984	0.958	1.471	1.0
$\frac{\text{Column Average}_3}{\text{Column Average}_p} * 100$		12.2	98.4	100	65.1	97.7

values of this predictor always correspond to cases in which $\rho=.9$. There are no extreme MSE values associated with the use of either $y_i^{(2)}$ $y_i^{(3)}$ or $y_i^{(5)}$. Clearly the moral of the story is that the full information predictor, $y_i^{(3)}$, should be the one used for predictions when the available information set is $\Lambda_3=\{X,W,y_{-i}\}$. Correspondingly, if Λ_3 is available the worst predictor, by far, would be the one based on the reduced form, namely $y_i^{(1)}$ in Eq. (7). One case in which an information set such as Λ_3 may be available relates to the price prediction of a housing unit in a sampled neighborhood. In such a case the available data may relate to the values of the regressors, the weighting matrix, and the prices of neighboring units.

Now consider the results in Table 1 which relate to the spatial error model, $\lambda=0$. Using evident notation, the average of the mean square errors for the five predictors corresponding to the five cases in Table 1 in which $\lambda=0$ are $\overline{\text{MSE}}(y_i^{(1)})=\overline{\text{MSE}}(y_i^{(4)})=2.591$; $\overline{\text{MSE}}(y_i^{(2)})=.931$; $\overline{\text{MSE}}(y_i^{(3)})=.917$; and $\overline{\text{MSE}}(y_i^{(5)})=1.0$. Clearly, as expected $y_i^{(3)}$ is the best predictor, which is followed closely by $y_i^{(2)}$. Both of these predictor are unbiased conditional mean predictors which utilize information relating to y_{-i}. The intuitively suggestive predictor, namely $y_i^{(5)}$ while substantially better than $y_i^{(1)}$ and $y_i^{(4)}$, is still roughly 8% less efficient than the efficient predictor, $y_i^{(3)}$.

A glance at Table 2 suggests results which are again consistent with our theoretical notions in Eq. (23), as well as with the numerical "extreme" values given in Table 1 in somewhat more moderated form. For example, the extreme values of the MSEs relating to $y_i^{(1)}$ and $y_i^{(4)}$ are roughly a third of their corresponding values in Table 1. We also note that the relative efficiencies of the other four predictors are all significantly higher in Table 2 than their corresponding values in Table 1. Thus, in our framework, as the extent of sparseness in the weighting matrix decreases, the relative predictive efficiencies of the inefficient predictors increases.

Results relating to the spatial error model are also qualitatively similar to those in Table 1. For instance, again using evident notation, the average of the mean square errors for the five predictors corresponding to the five cases in Table 2 in which $\lambda=0$ are $\bar{\text{M}}\text{SE}(y_i^{(1)})=\bar{\text{M}}\text{SE}(y_i^{(4)})=$ 1.471; $\bar{\text{M}}\text{SE}(y_i^{(2)})=.980$; $\bar{\text{M}}\text{SE}(y_i^{(3)})=.977$; $\bar{\text{M}}\text{SE}(y_i^{(5)}).=1.0$. Clearly, $y_i^{(3)}$ remains the best predictor but in this case it is quite closely followed by $y_i^{(2)}$ as well as by the intuitive but biased predictor $y_i^{(5)}$. Again, the predictors which do not utilize any information relating to y_{-1} namely $y_i^{(1)}$ and $y_i^{(4)}$, have a mean square error which is roughly 50% higher than those that do utilize such information, even if in a biased manner!

5. Summary and conclusions

In this paper we considered the prediction of the dependent variable in a spatial model containing spatial lags in both the dependent variable and disturbance term. A special case of this model is the spatial error model. Five predictors were considered, one of which is the full information predictor; the other four predictors are simpler and are either based on limited information, or an assumption that a certain covariance is zero when in fact it, in general, is not. Our empirical results are consistent with theoretical notions in that predictors based on "properly used" larger information sets are indeed more efficient than those based on less information. Our results also suggest that a predictor suggested in the literature is considerably less efficient than other (biased) predictors that one might consider in such models. That predictor is the mean of the dependent variable conditional only on the exogenous variables and the weighting matrix. Finally, our results suggest that the relative inefficiencies of predictors increase as the sparseness of the weighting matrix increases.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Boston.

Audretsch, D.B., Feldmann, M.P., 1996. R&D spillovers and the geography of innovation and production. American Economic Review 86, 630–640.

Bannerjee, S., Carlin, B., Gelfand, A., 2003. Hierarchical Modeling and Analysis for Spatial Data, Monographs on Statistics and Applied Probability. Chapman and Hall, Boston.

Bell, K.P., Bockstael, N.E., 2000. Applying the generalized-moments estimation approach to spatial problems involving microlevel data. Review of Economics and Statistics 82, 72–82.

Bernat Jr., G., 1996. Does manufacturing matter? A spatial econometric view of Kaldor's Laws. Journal of Regional Science 36, 463–477.

Besley, T., Case, A., 1995. Incumbent behavior: vote-seeking, tax-setting, and yardstick competition. American Economic Review 85, 25–45.

Bollinger, C., Ihlanfeldt, K., 1997. The impact of rapid rail transit on economic development: The case of Atlanta's Marta. Journal of Urban Economics 42, 179–204.

Buettner, T., 1999. The effect of unemployment, aggregate wages, and spatial contiguity on local wages: an investigation with German district level data. Papers in Regional Science 78, 47–67.

Cliff, A., Ord, J., 1973. Spatial Autocorrelation. Pion, London.

Cliff, A., Ord, J., 1981. Spatial Processes, Models and Applications. Pion, London.

Cohen, J.P., Morrison, P.C.J., 2004. Public infrastructure investment, interstate spatial spillovers, and manufacturing cost. Review of Economic and Statistics 86, 551–560.

Cressie, N.A.C., 1993. Statistics of Spatial Data. Wiley, New York.
Dubin, R., 2003. Robustness of spatial autocorrelation specifications: some Monte Carlo evidence. Journal of Regional Science 43, 221–248.
Dubin, R., 2004. Spatial lags and spatial errors revisited: some Monte Carlo evidence. Manuscript, Department of Economics, Case Western Reserve University.
Goldberger, A., 1962. Best linear unbiased prediction in the generalized regression model. Journal of the American Statistical Association 57, 369–375.
Greene, W., 2003. Econometric Analysis, 5-th ed. Prentice Hall, Upper Saddle River.
Kapoor, M. 2003. Panel data models with spatial correlation: Estimation theory and an empirical investigation of the U.S. wholesale gasoline industry, University of Maryland, Ph.D. Thesis.
Kelejian, H.H., 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 and Economics 17, 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.
Kelejian, H., Robinson, D., 2000. Returns to investment in navigation infrastructure: an equalibrium approach. Annuals of Regional Science 34, 83–108.
Kelejian, H.H., Yuzefovich, E., 2004. Properties of tests for spatial error components: a further analysis. In: Getis, A., Mur, J., Zoller, H.G. (Eds.), Spatial Econometrics and Spatial Statistics. Palgrave, Macmillan, Basingstoke, Hampshire, England, pp. 35–149.
Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, 307–335.
LeSage, J.P., 1999. A Spatial econometric analysis of China's economic growth. Journal of Geographic Information Sciences 5, 143–153.
Mood, A., Graybill, F., Boes, D., 1974. Introduction to the Theory of Statistics. McGraw-Hill, Boston.
Pinske, J., Slade, M.E., Brett, C., 2003. Spatial price competition: a semiparametric approach. Econometrica 70, 1111–1153.
Rey, S.J., Boarnet, M.G., 2004. A taxonomy of spatial econometric models for simultaneous equation systems. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Springer, New York, pp. 99–119.
Yuzefovich, Y., 2003. Two essays on spatial econometrics, University of Maryland, Ph.D. Thesis.