REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

MONALISA SEN[a] – ANIL K. BERA[b]

The Improbable Nature of the Implied Correlation Matrix from Spatial Regression Models*

Abstract

Spatial lag dependence in a regression model is similar to the inclusion of a serially autoregressive term for the dependent variable in a time-series context. However, unlike in the time-series model, the implied covariance structure matrix from the spatial autoregressive model can have a very counterintuitive and improbable structure. A single value of spatial autocorrelation parameter can imply a large band of values of pair-wise correlations among different observations of the dependent variable, when the weight matrix for the spatial model is specified exogenously. This is illustrated using cigarette sales data (1963–1992) of 46 US states. It can be seen that that two "close" neighbours can have very low implied correlations compared to distant neighbours when the weighting scheme is the first-order contiguity matrix. However, if the weight matrix can capture the underlying dependence structure of the observations, then this unintuitive behaviour of implied correlation is corrected to a large extent. From this, the possibility of constructing the weight matrix (or the overall spatial dependence in the data) that is consistent with the underlying correlation structure of the dependent variable is explored. The suggested procedures produced very positive results indicating further research.

Keywords: Spatial Dependence, Variance-Covariance matrix, Implied Correlation Structure, Weight Matrix.

Introduction

The key idea of modelling spatial data is that a set of locations can characterise the dependence among the observations. One of the many general ways to do this is to define a neighbourhood structure based on the shape of a lattice. Among others, this measures the distance between centroids of the regions. Once this spatial dependence structure is determined or assumed based on distance (social/economic/physical) or adjacency, models resembling time-series autoregressive structures are considered. The two very popular models that take into account such spatial dependence structure are simultaneously

a) Universite catholique de Louvain, CORE, Voie du Roman Pays, 34, L1.03.01, B-1348 Louvain-la-Neuve, Belgium. E-mail: monalisa.sen@uclouvain.be

b) University of Illinois at Urbana Champaign, Department of Economics, 225 E David Hall, 1407 W. Gregory Hall, Urbana, Illinois 61801. E-mail: abera@illinois.edu

* We are grateful to an anonymous referee for many pertinent comments that helped us to improve the paper. An earlier version of this paper was presented at the IV[th] World Conference of the Spatial Econometric Association. We are thankful to Tim Conley, John Paul Elhorst and Roger Bivand for the comments and suggestions. Any remaining errors are our responsibility.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

autoregressive (SAR) and conditionally autoregressive (CAR) models. The SAR and CAR models were originally developed by Whittle (1954) and Besag (1974), respectively, mainly on the doubly infinite regular lattice. On a regular lattice, these models resemble the well-understood stationary time-series model defined on the integers. On an irregular lattice, however, which is most common in economic applications, the effect that the exogenously defined arbitrary neighbourhood structure and spatial correlation parameter have on implied covariance structure is not well understood. Wall (2004) was probably the first to carry out a systematic analysis of the impractical nature of the correlation structure implied by the SAR and CAR models, and this issue has spurred some further inquiries, see for instance Martellosio (2009).

This paper highlights the problem of implied structure of the SAR model in case of an irregular lattice and suggests a possible solution. Although the proposal is for the SAR model, it can be easily extended to the CAR model. Section 2 provides a summary of the existing literature. Section 3 sets up the notations and states the SAR model. Section 4 presents a spatial regression example on cigarette sales data on 46 US states where the spatial model is compared with ordinary least square case and highlights the unintuitive and impractical behaviour of the implied correlation structure when the usual neighbourhood matrix is used. The findings reconfirm the results of Wall (2004). Section 5 first gives the basic idea behind the authors W matrix construction and then estimates W using the Levenberg-Marquardt nonlinear optimization procedure. Section 6 demonstrates how the developed W matrix helps to correct the implied correlation structure and gives a more intuitive result using the same dataset as in Section 4, Section 7 concludes the paper.

Summary of relevant previous work

Although the implied correlation structures of the spatial models have such a peculiar pattern, it is quite surprising that this issue has received relatively little attention in the literature, given that these models are so widely used in a variety of applications. Haining (1990) and Besag and Kooperberg (1995) mentioned resulting heteroscedasticity from the SAR model with homoscedastic error term. They also pointed out about the unequal covariance between regions that are at same distance apart. The very first systematic treatment of this problem was probably done by Wall (2004). She provided a detailed description of the implied structure of SAR and CAR models, and in particular, considered the dependence and covariance structures on an irregular lattice. Using the US state level summary data of SAT verbal score for the year 1999, she investigated the relationship between the correlation ρ and the implied pair-wise correlations among the scores of various states when W was based on first-order neighbours. The implied spatial correlations between the different states using the SAR and CAR models did not seem to follow an intuitive or practical scheme. For example, Wall (2004) found that for the SAR model Missouri and Tennessee are constrained to be the least spatially correlated states, than Tennessee and Arkansas, although all of them are first-order neighbours. Martellosio (2009) shed some further light on how the correlation structure of the SAR model depends on W and ρ and explained this inconsistency using graph theory. He showed that implied correlation between two spatial units depends on a particular type of walks (in a graph theoretic sense) connecting the units. When $|\rho|$ is small, the correlation is largely

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

determined by short walks; however, for large values of $|\rho|$, longer walks have more importance. Since ρ can be estimated only after W has been chosen, one cannot control the correlation properties by specifying W. Defining W based on graph, his work explains the inconsistency of ranking of implied correlations between pair of locations as ρ changes.

The SAR model

Let $\{y(A_i): A_i \in (A_1 \ldots A_n)\}$ be a Gaussian random process where $(A_1 \ldots A_n)$ are n different locations. The value of the variable y in location A_i depends on the values in its neighbouring locations A_j. One way to model this dependence is by the simultaneous autoregressive (SAR) model:

$$y=\rho Wy+X\beta+\varepsilon \qquad (1)$$

where y is a n×1 vector observation on the dependent variable, ρ is the spatial autoregressive parameter, $W \equiv \left((w_{ij})\right)$ is n×n spatial weight matrix representing degree of potential interactions between neighbouring locations (geographic/economic/social), X is n×k matrix of observations on the explanatory (exogenous) variables, β is k×1 vector of regression coefficients and ε is a n×1 vector of error term with $\varepsilon \sim (0, \sigma^2 I_n)$.

Spatial effects are incorporated using the row-standardised weight matrix W. One common way to do this is to define $W = (w_{ij})$ is

$$w_{ij} = \begin{cases} 1 \text{ if } A_i \text{ shares a common edge or border with region } A_j \ (i \neq j) \\ 0 \text{ otherwise} \end{cases}$$

The other ways to define the neighbourhood structure W is to express weights as functions of the distance between two points or as functions of the length of borders. For ease of interpretation, the weight matrix is often standardised so that the elements of each row sum to one. Ensuring that all the weights are between 0 and 1, facilitates the interpretation of operations with the weight matrix as an averaging of neighbourhood values. It also ensures that the spatial parameters of different models are comparable. This is not intuitively obvious, but relates to constraints imposed in a maximum likelihood estimation framework, specifically the spatial autocorrelation parameter must be in the interval $[1/\omega_{min}, 1/\omega_{max}]$, where ω_{min} and ω_{max} are, respectively, the smallest and largest eigen values of W [Cliff and Ord (1980)]. For a row-standardised matrix, the largest eigen value is always +1, and this facilitates the interpretation of ρ as "correlation" coefficient.

It is easy to see that the implied covariance matrix of y for model (1) is given by

$$Var(y) = \sigma^2(I-\rho W)^{-1}(I-\rho W)^{t-1} \qquad (2)$$

Using (2), the pair-wise correlations $corr(y_i, y_j) = \rho_{ij}$, $i, j=1,2,\ldots, n, i \neq j$ can be calculated. However, these ρ_{ij} values can apparently have "no connection" with the values of w_{ij} and ρ. To demonstrate this, we use the widely applied cigarette sales data on 46 US States. We show that a single value of ρ can imply a large band of values of ρ_{ij} with the same w_{ij}values. The findings confirm the results of Wall (2004). We then construct a W matrix that is "consistent" with the underlying correlation structure of y. Finally, we further investigate the behaviour of implied correlations for the same model using the constructed W matrix, and show that the use of the weight matrix eliminates all the unintuitive behaviour of implied correlations.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

An example

In order to analyse the spatial interaction and implied correlation structure of the SAR model, the 1963–1992 cigarette sales data on 46 US States, that has been widely used for panel data analysis by Baltagi and Levin (1992) and Baltagi, Griffin and Xiong (2000), and later by Elhorst (2005) for spatial panel analysis, is considered. The underlying model is:

$$\log(C) = \alpha + \rho W \log(C) + \beta_1 \log(P) + \beta_2 \log(Y) + \beta_3 \log(Pn) + \epsilon, \qquad (3)$$

where C is real per capita sales of cigarettes to persons of smoking age (14 years and older), measured in packs of cigarettes per capita; P is the average retail price of a pack of cigarettes measured in real terms; Y is the real per capita disposable income, and Pn denotes the minimum real price of cigarettes in any neighbouring state. This last variable is a proxy for the casual smuggling effect across state borders, and acts as a substitute price attracting consumers from high-tax states to cross over to low-tax states. As in Elhorst (2005), we follow the conventional form of row-standardised first-order neighbourhood weight matrix, and in Table 1 present the estimation results based on 1992 cross-section data of the 46 states.

Table 1

Estimation Results of Model (3) (Standard errors are in parentheses)

Parameters	OLS	SAR(W=row-standardised first-order contiguity)
β_1	–1.24(0.31)	–1.15 (0.29)
β_2	0.17(0.32)	0.27(0.30)
β_3	1.03(0.19)	0.74(.15)
P	N/A	0.28(0.14)
σ^2	0.05	0.04
Log Likelihood		25.78
R^2	0.15	0.18

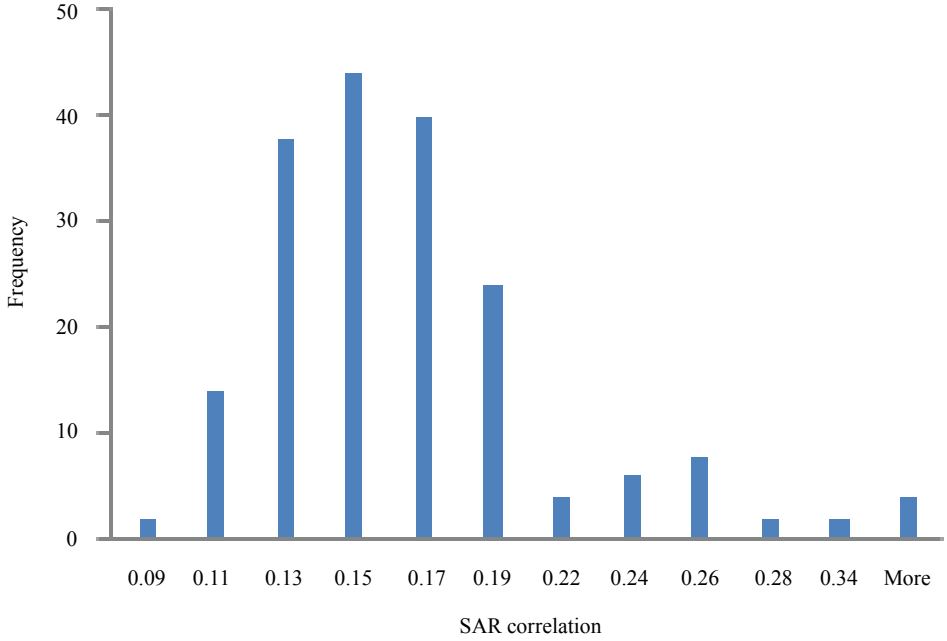
To illustrate the behaviour of the implied correlation structure from the estimated SAR model, in Figure 1, the histogram covers of all the implied *first-order* neighbour correlations and demonstrates a wide variation. The smallest correlation is 0.09 that occurs between Missouri and Tennessee and the largest correlation, equal to 0.37, occurs between New Hampshire and Maine. Wall (2004) also noted smallest and largest implied correlations *exactly* for these states, although she used *different* data (1999 US statewide average SAT verbal scores) and model. The common feature between Wall's and this example is the W matrix, more specifically, Maine has only one neighbour, i.e., New Hampshire, and Tennessee and Missouri have 7 and 8 neighbours, respectively. Also, the qualitative nature of the histograms of Wall (in her Figure 3 with ρ =0.60) and this paper are very similar. Therefore, it can be stated that implied correlation is simply a function of the first-order neighbours each state has.

To elaborate further on the implied correlations of Missouri and Tennessee with their 8 and 7 neighbours, respectively, from Table 2 it should be noted that Missouri is more correlated with Kansas than with Tennessee, and Tennessee is more correlated with its neighbour Alabama than with Missouri. Such peculiarity arises mainly due to the nature of covariance matrix (2) that involves inversion of the sparse matrix $(I - \rho W)$.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

Figure 1

Histogram of implied correlations



The relative ranking of the states using implied spatial correlation almost coincides with that of Wall (2004). These two datasets have no connection economically, and ranking of implied spatial correlation is determined by the prior fixed weight matrix.

Table 2

Implied correlation between first-order neighbours of Missouri and Tennessee

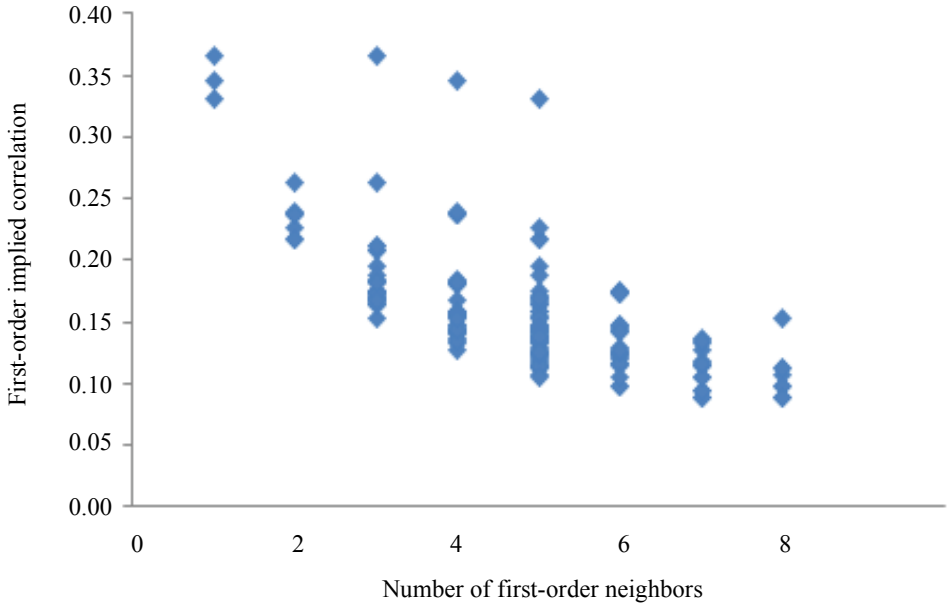
Missouri		Tennessee	
1st order neighbours	Implied correlation	1st order neighbours	Implied correlation
Arkansas	0.0965	Alabama	0.1354
Illinois	0.1062	Arkansas	0.1036
Iowa	0.0977	Georgia	0.1256
Kansas	0.1516	Kentucky	0.0931
Kentucky	0.0879	Mississippi	0.1325
Nebraska	0.1108	Missouri	0.0873
Oklahoma	0.1110	Virginia	0.1044
Tennessee	0.0873		

Figure 2 demonstrates that the relationship between the implied correlation and number of neighbours is not that simple. If number of neighbours is less, then implied correlation is strong. There is a band in which the implied correlations vary for a given number of neighbours, with less heterogeneity for an extreme number of neighbours.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

Figure 2

Implied correlations of SAR model

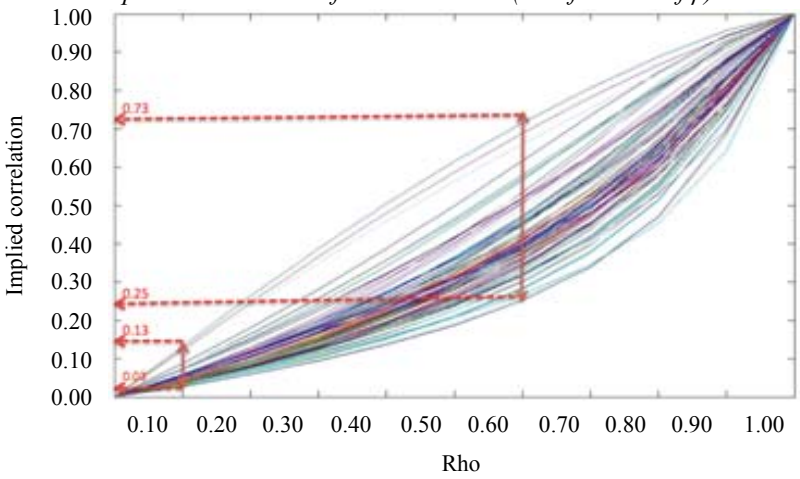


The paper now focuses on how implied correlations behave as functions of true parameter ρ (i.e., irrespective of data). From Figures 3a and 3b, it can observed that for any given ρ, there is a high variability in correlations among all the pairs of observations. For example, when ρ=0.1, the implied correlations vary from 0.03 to 0.13; while for ρ=0.6, they vary from 0.25 to 0.73. From Figure 3a, this can be observed from the red arrows as marked. As ρ increases, the implied correlations of all locations increases monotonically, which matches the behaviour of autoregressive models in a time series, i.e., correlation increases with the autoregressive parameter. However, as observed in Figure 3b, the most unintuitive behaviour is that as ρ changes, there are many lines that *cross each other*, implying the inconsistency of ranking of relative implied correlations. For example, when ρ=0.2 the correlation (Missouri, Arkansas) =0.17 and correlation (Tennessee, Arkansas) =0.24. However, when ρ=0.7, then correlation (Missouri, Arkansas) =0.33 and correlation (Tennessee, Arkansas) = 0.26. Wall (2004) reported the same phenomenon. Therefore, the implied correlations of SAR model with first-order neighbour *W* matrix do exhibit some unintuitive and impractical behaviour. The marked area in the Figure 3b shows the crossing of implied correlation lines.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

Figure 3a

Implied correlations of the SAR model (as a function of ρ)

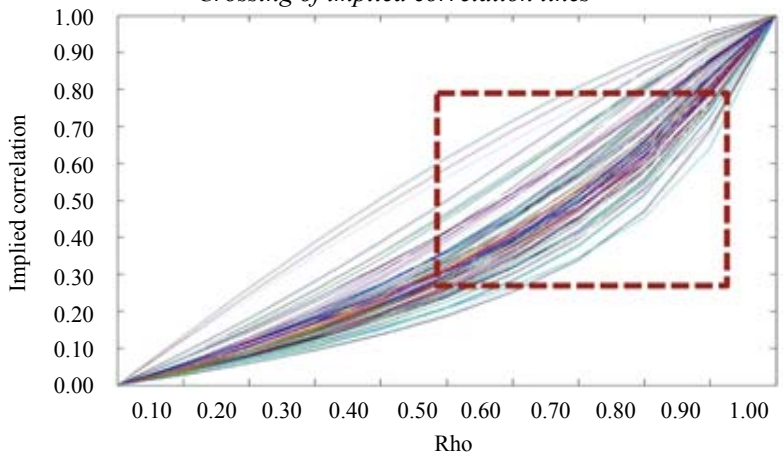


Numerical optimization

It is a general understanding that the weight matrix captures the "spatial dependence" of the observations as Ord (1975) stated that the (i, j)th element of W "represents the degree of possible interaction of location j on location i". However, each element of $(I-\rho W)^{-1}(I-\rho W)^{t-1}$ provides the correlation structure of y. As evident from Wall (2004) and from Fig. 3a, if one expresses the spatial dependence in terms of neighbourhood matrix W, then the covariance from $(I-\rho W)^{-1}(I-\rho W)^{t-1}$ does not have a strong connection to the spatial correlation.

Figure 3b

Crossing of implied correlation lines



The choice of spatial weights is a central component of spatial models as it imposes a priori structure on spatial dependence. Although the existing literature contains an implicit acknowledgement of the issues of choosing an appropriate weight matrix, most empirical

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

studies treat *W* to be a known, fixed and arbitrary spatial weight matrix (Giacomini and Granger 2004). It is proposed to construct the weight matrix using *past* time-series data to remove the odd features of implied correlations discussed previously.

Suppose the dependent variable y_i is observed over n locations, where i=1…n for t=1,…,T in the *past* T periods. Given y_i for T periods, we estimate the variance-covariance matrix$V(y) = \Sigma$, whose (i,j)th element is given by $\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j)$, where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{y}_j = \frac{1}{T}\sum_{t=1}^{T} y_{jt}$. Our objective is to investigate the implied correlation structure of a SAR model at the current time, therefore, construction of the weight matrix based on *past* T periods helps us to avoid the endogeneity issue.

We solve the following system for *W*

$$\sigma^2(I - W)^{-1}(I - W)^{-1'} = \Sigma,$$

We can take $\sigma^2 = 1$, which will have no consequence for our solution to W. Also, since W is row standardised, the solution will be invariant to ρ. Therefore, without loss of generality we solve

$$(I - W)^{-1}(I - W)^{-1'} = \Sigma,$$

i.e.,

$$(WW') - (W + W') = \Sigma^{-1} - I. \qquad (4)$$

We need to find *W* that solves the equation (4) subject to

i) $w_{ii} = 0$

ii) $w_{ij} \geq 0$

iii)$\sum_j w_{ij} = 1$

ii) and iii) imply the range of w_{ij}, i.e., $0 \leq w_{ij} \leq 1$. Alternatively, our objective is to find a solution to a constrained system of nonlinear equations:

$$F(w) = I + (w * w') - (w + w') - \Sigma^{-1} = 0, \quad \mathrm{w} \in \mathrm{W}, \qquad (5)$$

where $W \subseteq R^{m+}$ is a nonempty, closed and convex set and $F: \mathcal{O} \longrightarrow R^m$ is a given mapping defined on an open neighbourhood $\mathcal{O}$ of the set *W*. Here $m = n^2$, where *n* is the number of locations. W^* is the set of solutions to (5). To solve (5) we consider the related optimization problem:

min $f(w)$ subject to the constraints as above, where $f(w) \coloneqq ||F(w)||^2$, and ||. || is the Euclidean norm.

The Levenberg (1944) and Marqaurdt (1963) algorithm (LM) that interpolates between the Gauss-Newton algorithm and method of gradient descent is used, as in many cases, the LM algorithm is more robust than Gauss-Newton as it finds a solution even if it starts very far off from the optimal values. It is an iterative procedure where in each step *w* is replaced by *w+d*. To determine the increment vector *d*, the function *F(w+d)* are approximated by their linearisation using Taylor Theorem i.e., $F(w + d) \approx F(w) + J * d$, where

$J = \partial F(w)/\partial w$ is the gradient of *F* with respect to *w*. At its minimum, the gradient of *f* with respect to *d* will be zero. The above 1st order approximation gives

$f(w + d) \coloneqq ||F(w + d)||^2 \approx ||F(w) + J * d||^2$.

Taking derivative with respect to *d* and setting the result equal to zero gives

$(J^T J)d = -J^T F(w)$, where *J* is the Jacobian term. This gives us a set of linear equations that can be solved for the increment vector *d*. The Levenberg-Marquardt contribution is to replace this equation by a '*damped version'*,

$$\left(J^T J + \mu * diag\left(J^T J\right)\right) d = -J^T F(w).$$

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

The main difference between Gauss-Newton and the LM algorithm is in terms of normal equations. In the LM algorithm, the normal equations are modified in such a way that the increment vector *d* is always rotated towards the direction of steepest descent.

In a more formal way, LM type method for this system of equations generates a sequence $\{w^k\}$ by setting $w^{k+1} = (w^k + d^k)$, where d^k is the solution to the linearised sub-problem:

$$min\ \theta^k(d) = ||F(w^k) + J_k d||^2 + \mu_k||d||^2, s.t \qquad w^k + d \in W \qquad (6$$

Here, J_k is an approximation of Jacobian of $F'(w^k)$ and μ_k is the positive parameter. Note that $\theta^k(.)$ is a strictly convex quadratic function; hence the solution d^k of (6) always exists uniquely. Since our constraints are of box constraints type, any iterate w^k can be projected easily into the feasible region *W*. The feasible region of *W* is such that any $w \in W$ has the structure defined by the above constraints. Therefore, we set $w^{k+1} = P_W(w^k + d_u^k),\ k = 0,1, \ldots.$, where P_W is the projection matrix and d_u^k is the unique solution to the *unconstrained* sub-problem:

$$min\ \theta^k(d_u),\ d_u \in R^m.$$

We call this the projected LM method since the unconstrained step gets projected onto the feasible region *W*. The projected version of LM algorithm needs significantly less time per iteration since the strict convexity of the function $\theta^k(.)$ ensures that a global minimum is reached at d_u^k if and only if $\nabla\ \theta^k\left({d_u}^k\right) = 0$, i.e., if and only if ${d_u}^k$ is the unique solution of the system of linear equations [For detailed discussion on the Levenberg-Marquardt Method, see Numerical Optimization by Nocedal and Wright (2006)]:

$$\left({J_k}^T J_k + \mu_k diag\left({J_k}^T J_k\right)\right) d_u = -{J_k}^T F(w^k). \qquad (7)$$

The step-by-step algorithm is as follows:

S1) Choose $w^0 \in W, \mu > 0,\ v > 1, \gamma > 0$ and set $k = 0$, tolerance = $1e - 10$.

S2) If $F(w^k) <$ tolerance, then Stop, otherwise go to S3.

S3) Compute $J_k = F'(w^k)$.

S4) $Set\ \mu_k = \mu/v^k$ and compute ${d_u}^k$ as a solution to (7).

S5) If $||F(P_W(w^k + d_u^k)|| \leq \gamma||F(w^k)||$, then set $w^{k+1} = P_W(w^k + d_u^k)$, update k to k+1 and go to S2; Otherwise go to S6.

S6) Set $\mu_k = \mu * v^k$ and compute ${d_u}^k$ as a solution to (7).

S7) If $||F(P_W(w^k + d_u^k)|| \leq \gamma||F(w^k)||$, then set $w^{k+1} = P_W(w^k + d_u^k)$, update k to k+1 and go to S2.

Note, if any k^{th} iteration comes to S6, then for k+1th iteration onwards, it will flow as S2 → S3 → S6 → S7. This is due to the choice of dampening factor as suggested by Marquardt (1963). If there is no reduction in residual by setting $\mu_k = \mu/v^k$, then the dampening factor is increased by successive multiplication by v until a better point is found with the new dampening factor $\mu_k = \mu * v^k$ for some k. However, if the use of $\mu_k = \mu/v^k$ results in the reduction of residuals, then this is taken as a new value of μ and the process continues. In other words, as μ_k gets small, the algorithm approaches the Gauss-Newton algorithm, if μ_k becomes large with successive iterations, it approaches the steepest gradient algorithm. The Levenberg-Marquardt technique is a "blending" between these two extremes, and uses a steepest descent type method until our objective function approaches a minimum, and then gradually switches to the quadratic rule. It tries to guess the proximity to a minimum by how the error is changing. The intuition is simple; i.e., if error is increasing, then our quadratic approximation is not working well and we are likely not near

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

a minimum, so we should increase μ_k in order to blend more towards simple gradient descent. Conversely, if error is decreasing, our approximation is working well, and we expect that we are getting closer to a minimum so μ_k is decreased to bank more on the Hessian. The algorithm we used is very similar to the projected LM algorithm of Kanzow-Yamashita-Fukushima (2002). As long as *F* is affine and twice continuously differentiable, any accumulation point of the sequence $\{w^k\}$ generated by our algorithm is a stationary point of (6).

Application of the proposed solutions

The SAR model for the year 1992 is estimated using the proposed weight matrix in Section 5.1. In order to avoid endogeneity problem, the *W* matrix using the data on *C* (Cigarette sales) from 46 states for the period 1963-1991. The model is the same as in Equation (3).

Table 3 provides the estimates of the SAR model using the standard *W* matrix, and the numerically solved *W* using the Levenberg-Marquardt algorithm.

Table 3

Estimation Results of Model (3) (Standard errors are in parentheses)

Parameters	SAR(*W*=Constructed using the LM algorithm)	SAR(*W*=row-standardised first-order contiguity)
β_1	–1.10 (0.29)	–1.15 (0.29)
β_2	0.18 (0.29)	0.27 (0.30)
β_3	0.55 (0.17)	0.74 (0.15)
P	0.45 (0.16)	0.28 (0.14)
σ^2	0.03	0.04
Log Likelihood	26.37	25.78
R^2	0.27	0.18

It is clear that the estimated SAR model using the proposed *W* matrix is as good as (if not better than) the standard *W* in terms of log likelihood values.

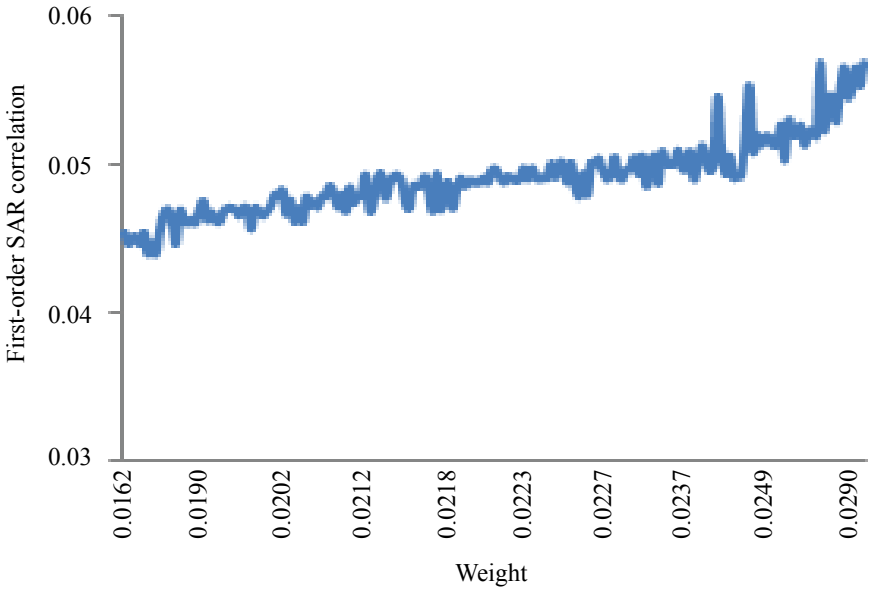
Figure 4 plots the first-order implied correlation as a function of weights from the estimated *W*. Out of 46*46=2116 pairs of locations, the 188 first-order neighbour correlations are plotted. First, arrange the weights of 188 pairs of first-order neighbours are arranged in ascending order, and then the implied correlations are sorted out in ascending order as well. From Figure 4, note that the implied correlations have a very slow increasing trend with weights. There is also little variation in contrast to Figure 2 (where number of neighbour increases means weight decreases), which displayed much higher variation.

Implied correlations of the SAR model for different values of ρ are plotted in Figure 5. In contrast to Figure 3, now for each value of ρ, the band of variation of implied correlations is very narrow in Figure 5. For example, when ρ=0.1, the implied correlations vary only between 0.004 and 0.006; while for ρ=0.6, they vary from 0.09 to 0.11. Also now there is no crossing, and thus the inconsistency of the ranking of implied correlations seen in Figure 3, is absent in Figure 5.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

Figure 4

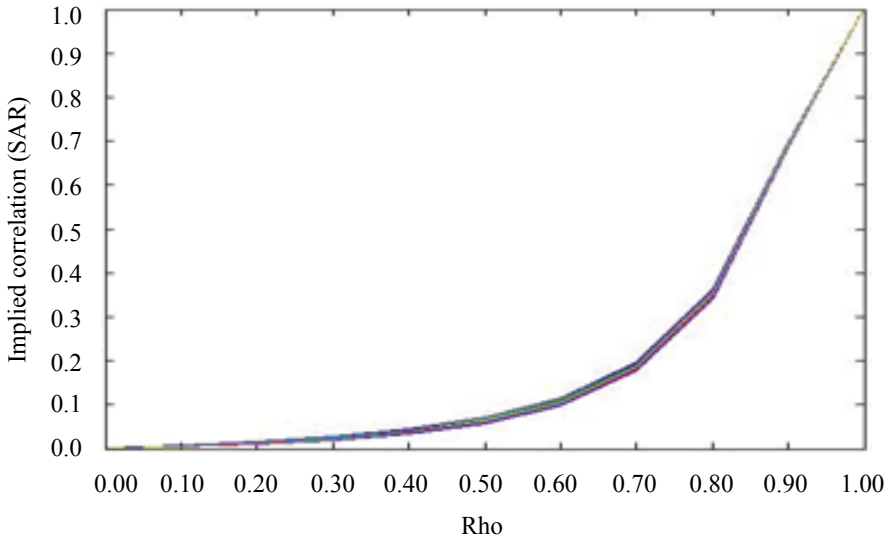
First-order correlations of SAR model (W=Constructed)



How implied correlations behave as function of ρ is examined in Figure 5.

Figure 5

Implied correlations of SAR model (as a function of ρ)

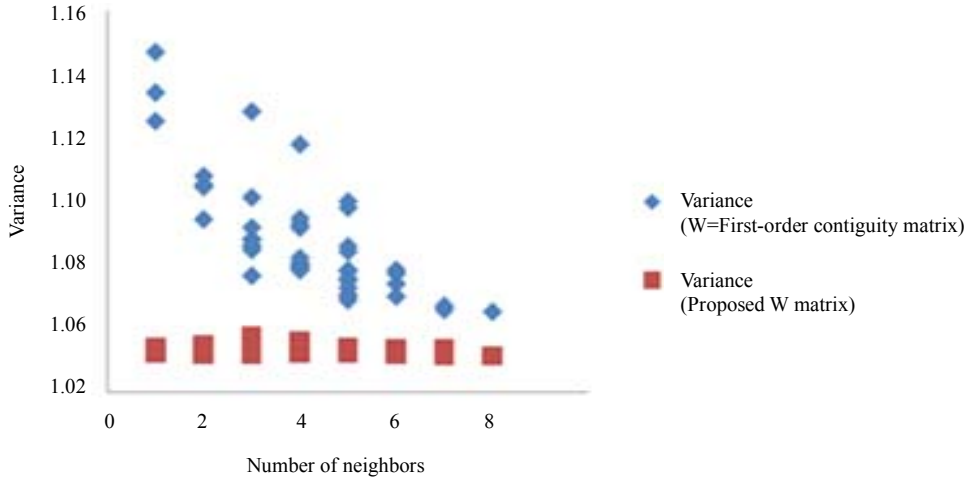


REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

To address the implied heterogeneity of the SAR model, in Figure 6, the 46 diagonal elements of Σ as a function of the number of first-order neighbours are plotted. Using the first-order contiguity matrix leads to substantial variation of implied variances of y_i (that decreases with the number of neighbours). In contrast, the proposed *W* matrix leads to much less variation.

Figure 6

Variance comparison of 46 US States



Conclusion

The paper first demonstrates the unintuitive and impractical nature of the implied correlations implied by the estimated SAR models with row-standardised neighbourhood matrix. It then proposes a simple methodology for estimation of spatial weight matrix that yields very intuitive results in terms of implied correlations. Finally, the proposed methodology is illustrated using real data on cigarette sales. Although the methodology is applied only for the SAR model, it can be easily extended to the CAR model. For CAR, Var(y) $=\sigma^2(I-\rho W)^{-1}$, which is a variation of Equation (2). Therefore, the procedure can be applied to yield a *W* that is consistent with the underlying dependence structure. Another interesting extension that the authors would like to address in future is the influence of *W* on the impact (direct and indirect) factor, which can be obtained from Equation (1) as $\frac{\partial y_i}{\partial x_{jk}} = g_k(i,j)$, where $g_k = (I-\rho W)^{-1}\beta_k, i,j = 1,2,\dots,n$ and $k = 1,2,\dots,K$. This has a simpler structure than the implied covariance matrix in Equation (2), and thus, as the referee conjectured, the two *W* matrices (binary contiguity and the "estimated" one) may give similar results.

REFERENCES

Besag, J. E. (1974): Spatial interaction and the statistical analysis of lattice data *Journal of the Royal Statistical Society* 36 (2): 192–236.

Besag, J. E. – Kooperberg, C. (1995): On conditional and intrinsic autoregressions *Biometrika* 82 (4):733–746.

REGIONAL STATISTICS, 2014, VOL 4, No 1: 3–15; DOI: 10.15196/RS04101

Baltagi, B. H. – Griffin, J. M. – Xiong, W. (2000): To Pool or Not to Pool: Homogeneous Versus Heterogeneous Estimators Applied to Cigarette Demand *The Review of Economics and Statistics* 82 (1): 117–126.

Baltagi, B. H. – Levin, D. (1986): Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered *The Review of Economics and Statistics* 68 (1): 148–155.

Cliff, A. – Ord, J. (1980): *Spatial Processes: Models and Applications* Pion Limited, NewYork.

Elhorst, J. (2005): Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels *Geographical Analysis* 37 (1): 85–106.

Giacomini, R. – Granger, C.W.J. (2005): Aggregation of space-time processes *Journal of Econometrics* 118 (1-2): 7–26.

Haining, R. (1990): *Spatial Data Analysis in the Social and Environmental Sciences* Cambridge University Press, Cambridge.

Kanzow, C. – Yamashita, N. – Fukushima, M. (2002): Levenberg-Marquardt methods for constrained nonlinear equations *Journal of Computational and Applied Mathematics* 172 (2): 375–397.

Levenberg, K. (1944): A Method for the Solution of Certain Non-Linear Problems in Least Squares, *Quarterly of Applied Mathematics* 2 (2): 164–168.

Marquandt, D. (1963): An Algorithm for least-squares estimation of nonlinear parameters *SIAM Journal on Applied Mathematics* 11 (2): 431–441.

Martellosio, F. (2009): Correlation properties of spatial autoregressions *Personal RePEc Archive, Paper* No. 17254.

Nocedal, J. – Wright, S. J. (2006): *Numerical Optimization* 2nd Edition Springer, New York.

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

Wall, M. M. (2004): A close look at the spatial structure implied by the CAR and SAR models *Journal of Statistical Planning and Inference* 121: 311–324.

Whittle, P. (1954): On stationary processes in the plane *Biometrika* 41 (3-4): 434–449.