THE IMPROBABLE NATURE OF THE IMPLIED CORRELATION MATRIX FROM SPATIAL REGRESSION MODELS

Monalisa Sen and Anil K. Bera

REAL 11-T-03   April, 2011

The Improbable Nature of the Implied Correlation Matrix from Spatial Regression Models[1]

Monalisa Sen and Anil K. Bera

Department of Economics

University of Illinois at Urbana Champaign

ABSTRACT: Spatial lag dependence in a regression model is similar to the inclusion of a serially autoregressive term for the dependent variable in 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 auto correlation 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. We illustrate this using cigarette sales data (1963-92) of 46 US states. We observe that two "close" neighbors can have very low implied correlations compared to distant neighbors 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 behavior of implied correlation gets corrected to a large extent. Keeping this in mind, we explore 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. The results using our suggested procedure are very encouraging.

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

1. INTRODUCTION

The key idea of modeling of spatial data is that a set of locations can characterize the dependence among the observations. One of the many general ways to do this is to define a neighborhood structure based on the shape of lattice. Among others, are measuring 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 into account are simultaneously 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 regular lattice these models resemble the well understood stationary time series model defined on the integers. On irregular lattice, however, which is most common in economic applications, the effect that the exogenously defined arbitrary neighborhood structure and spatial correlation parameter have on implied covariance structure is not well understood. Wall (2004) was probably the first to do 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 inquires, see for instance Martellosio (2009).

In this paper we highlight the problem of implied structure of the SAR model in case of irregular lattice and suggest a possible solution. Although our proposal is for the SAR model, it can be easily extended to the CAR model. Section 2 provides a summary of the existing literature. In Section 3, we set up the notations and discuss the implied correlation problem arising from the SAR model. Section 4 presents empirical example using cigarette sales data on 46 US states, and highlights the unintuitive and impractical behavior of the implied correlation structure when the usual neighborhood matrix is used. Our findings match with the results of Wall (2004). Section 5 first gives the basic idea behind our W matrix construction and then we estimate W using Levenberg-Marquardt non-linear optimization procedure. In Section 6, we demonstrate how our 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.

2. SUMMARY OF RELEVANT PREVIOUS WORK

Although the implied correlation structures of the spatial models have such 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 parameter ρ and the implied pairwise correlations among the scores of various states when W was based on first-order neighbors. 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 neighbors. Martellosio (2009) shed some further light on how correlation structure of the SAR model depends on W and ρ. He showed that implied correlation between two spatial units depends on particular type of walks connecting the units. When $|\rho|$ is small, the correlation is largely 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 correlation between pair of locations as ρ changes and also how the sign of correlation depends on the length of the shortest walk (in graph theoretic sense) from one location to another.

3. THE SAR MODEL AND THE IMPLIED CORRELATION PROBLEM

Let $\{y(A_i): A_i \in A_1 \dots A_n\}$ be a Gaussian random process where $A_1 \dots A_n$ are n different locations. The value of the variable y in location A_i depends on the values in its neighboring

locations A_j. One way to model this dependence is by the simultaneous autoregressive (SAR) model:

$$y = \rho W y + X\beta + \varepsilon , \tag{1}$$

where y is a n×1 vector observation on the dependent variable, ρ is the spatial autoregressive parameter, $W \equiv \quad w_{ij} \quad$ is n×n spatial weight matrix representing degree of potential interactions between neighboring 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 standardized weight matrix W. One common way to do this is to define $W = (w_{ij})$ is

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

The other ways to define the neighborhood structure W is to express weights as functions of the distance between two points or as functions of length of borders. For ease of interpretation, the weight matrix is often standardized such that the elements of each row sum to one. This ensures that all the weights are between 0 and 1, and facilitates the interpretation of operations with the weight matrix as an averaging of neighborhood 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 ρ must lie in the interval $1/\omega_{min}$ to $1/\omega_{max}$, where ω_{min} and ω_{max} are, respectively the smallest and largest eigen values of W [Cliff and Ord (1980)]. For a row standardized matrix, the largest eigen value is +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

$$Var(y) = \sigma^2 (I - \rho W)^{-1} (I - \rho W)^{-1\prime} . \tag{2}$$

Using (2), we can calculate the pair-wise correlations $corr\ y_i, y_j\ = \rho_{ij}$, $i,\ j=1,2,\dots\ ,\ n,\ i{\neq}j$. However, given ρ and W, these implied ρ_{ij} values can be very hard to interpret in a practical way.

To illustrate the implied correlation problem, we first provide some analytical results under two extreme cases of weight matrix.

Case 1: All units are neighbors of each other: Cases which may be consistent with this are the ones in which all cross sectional units interact in a confined space. Such a matrix was considered by Case (1992) in a panel data study of the adoption of new technologies by farmers, and by Lee (1999) in a study of the properties of least squares estimators in linear spatial models, and also by Kelijian and Prucha (2002) to evaluate the properties of 2SLS and OLS estimators of SAR models. Here

$$w_{ij} = \begin{cases} \dfrac{1}{n-1} & for\ i \neq j \\ 0 & for\ i = j. \end{cases}$$

Therefore, the weight matrix can be expressed as: $W\ =\ \frac{1}{n-1}\ [J-I]$, n>1, where J is the n x n matrix of ones.

It can be verified that $(I - \rho W)^{-1} = \delta_1 J + \delta_2 I$, where $\delta_1 = \frac{\rho}{n-1+\rho\ (1-\rho)}$ and $\delta_2 = \frac{n-1}{n-1+\rho}$. Assuming $\sigma^2 = 1$ we obtain

$$Var(Y\ )=(I\text{-}\rho W)^{-1}(I\text{-}\rho W)^{-1}= (n\delta_1^2 + 2\delta_1\delta_2)J + \delta_2{}^2 I. \qquad (3)$$

Given this variance-covariance structure, it can be seen that correlation matrix goes to I matrix as $n \rightarrow \infty$. Thus, when each unit is neighbor of each other, in the limit the correlation matrix does not depend on ρ!

Case II: Here each unit has only two neighbors. For instance, when n=4 we have

$$W = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$, which is a tridiagonal Toeplitz form. The implied correlation

matrix (using (2)) is also tridiagonal Toeplitz, and is given by

$$\text{Corr}(y) = \begin{pmatrix} 1 & \frac{\beta}{\alpha} & \frac{\gamma}{\alpha} & \frac{\beta}{\alpha} \\ \frac{\beta}{\alpha} & 1 & \frac{\beta}{\alpha} & \frac{\gamma}{\alpha} \\ \frac{\gamma}{\alpha} & \frac{\beta}{\alpha} & 1 & \frac{\beta}{\alpha} \\ \frac{\beta}{\alpha} & \frac{\gamma}{\alpha} & \frac{\beta}{\alpha} & 1 \end{pmatrix}$$,

where $\alpha = \frac{(2a^2-1)^2+2a^2+4a^4}{(4a^2-1)^2}, \beta = \frac{2(2a^2-1)+4a^4}{a(4a^2-1)^2}, \gamma = \frac{-4(2a^2-1)+2a^4}{a^2(4a^2-1)^2}$ and

$a = -0.5\rho$.

Each element of the inverse of such tridiagonal matrix is non zero (El-Shehaway,El-Shreff,Al-Henaway (2008)). Here units 1 and 3 are not connected ($w_{13} = w_{31} = 0$) directly, but we have a non zero implied correlation. In spatial context it implies that even if two units are "not" neighbors of each other, they can have very high non- zero implied spatial correlations. This can be interpreted as the spill-over effects from neighbors.

These examples are somewhat artificial. Therefore in the next section, using the widely applied cigarette sales data on 46 States, we demonstrate that a single value of ρ can imply a large band of values of ρ_{ij} with the same w_{ij} values. Our findings confirm the results of Wall (2004).

4. AN EMPIRICAL EXAMPLE

In order to analyze the spatial interaction and implied correlation structure of a SAR model we consider the 1963-1992 cigarette sales data on 46 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. The underlying model is:

$$\log C = \alpha + \rho W \log C + \beta_1 \log P + \beta_2 \log Y + \beta_3 \log Pn + \epsilon, \tag{4}$$

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 neighboring state. This last variable is a proxy for the 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 use the conventional form of row-standardized first-order neighborhood weight matrix, and in Table 1, present the estimation results based on 1992 cross-section data for the 46 states.

Table 1: Estimation Results of Model (3) (Standard errors are in parentheses)

Parameters	OLS	SAR(W=row standardized 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(0.15)
ρ	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 behavior of the implied correlation structure from the estimated SAR model, in Figure 1, we display the histogram of all the implied *first-order* neighbor correlations and notice 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, though she used *different* data (1999 US statewide average SAT verbal scores) and model. The common feature between Wall's and our situations is the W matrix, more specifically, Maine has only one neighbor, i.e., New Hampshire, and Tennessee and Missouri have 7 and 8 neighbors, respectively. Also the qualitative nature of the histograms of Wall (in her Figure 3 with ρ =0.60)

and ours are very similar. Therefore, we can say that implied correlation is simply a function of the first-order neighbors each state has.

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

Figure 1: HISTOGRAM OF IMPLIED CORRELATIONS
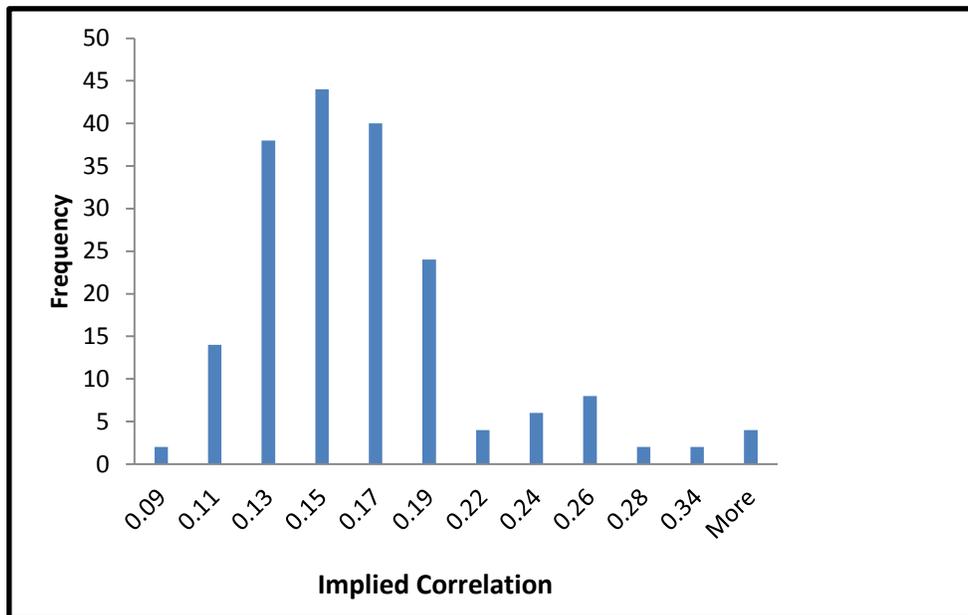
Our relative ranking of the states using implied spatial correlation almost coincides with that of Wall (2004). These two datasets have no connection economically; and ranking of implied spatial correlation is determined by the priory fixed weight matrix.

Table 2: IMPLIED CORRELATION BETWEEN FIRST-ORDER NEIGHBORS OF MISSOURI AND TENNESSEE

| Missouri		Tennessee	
1st order neighbors	Implied correlation	1st order neighbors	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 neighbors is not that simple. If number of neighbors is less, then implied correlation is strong. There is a band in which the implied correlations vary for a given number of neighbors, and we observe less heterogeneity for extreme number of neighbors.

Figure 2: IMPLIED CORRELATIONS OF SAR MODEL
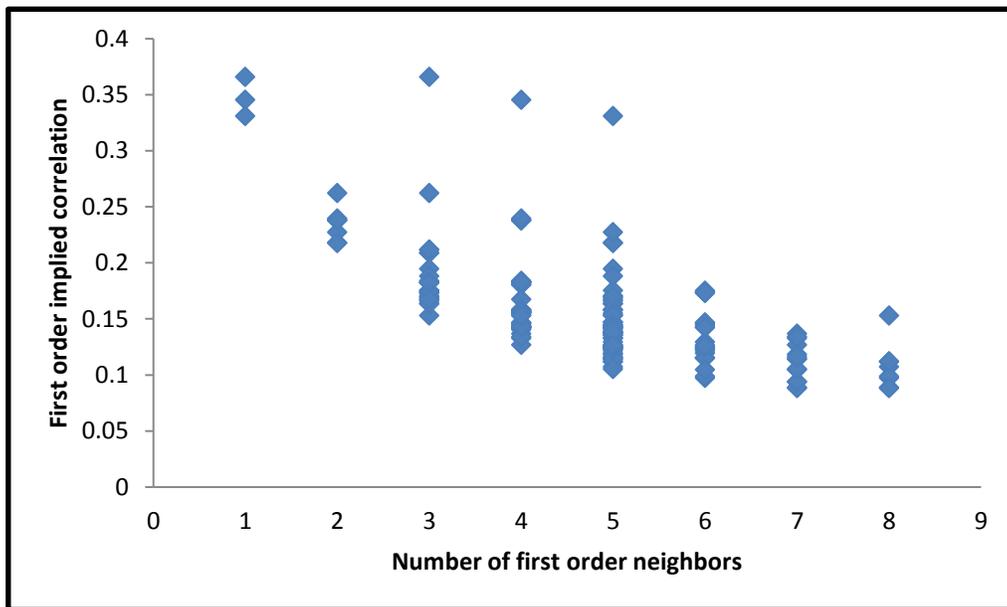
Now we focus on how implied correlations behave as functions of true parameter ρ (i.e., irrespective of data). From Figure 3, we observe that for any given ρ, there is a high variability in correlations. For example, when ρ=0.1, the implied correlations vary from 0.03 to 0.13; while for ρ=0.6, they vary from 0.25 to 0.73. As ρ increases the implied correlations of all locations increases monotonically, which matches the behavior of autoregressive models in time series, i.e., correlation increases with the autoregressive parameter. However, the most unintuitive behavior is that as ρ changes, there are many lines that *cross each other*, implying the inconsistency of ranking of 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, 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 neighbor *W* matrix do exhibit some unintuitive and impractical behavior.

Figure 3: IMPLIED CORRELATIONS OF SAR MODEL (as a function of ρ)
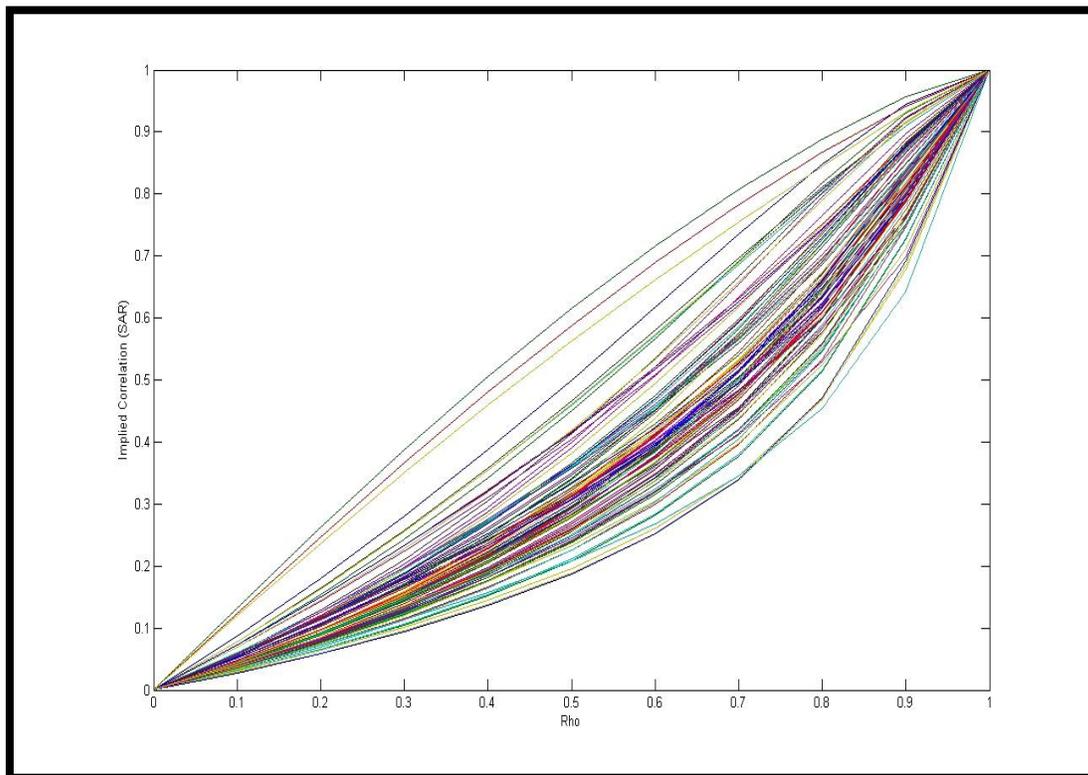
5. NUMERICAL OPTIMIZATION

It is a general understanding that the weight matrix captures the "spatial-links" 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 $(I-\rho W)^{-1}(I-\rho W)^{-1'}$ provides the correlation structure of y. As evident from Wall (2004) and from our illustration above, if one expresses the spatial dependence in terms of neighborhood matrix W, then the covariance from $(I-\rho W)^{-1}(I-\rho W)^{-1'}$ does not have a logical connection to the spatial correlation.

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 studies treat W known, fixed and arbitrary spatial weight matrix (Giacomini and Granger 2004). We propose to construct the weight matrix using *past* time series data to remove the odd features of implied correlations discussed above.

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

We solve the following system for W

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

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

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

i.e.,

$$WW' - W + W' = \Sigma^{-1} - I. \qquad (5)$$

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

i) $w_{ii} = 0$

ii) $w_{ij} \geq 0$

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

Alternatively, our objective is to find a solution to the constrained system of nonlinear equations:

$$F(w) = I + (w * w') - w + w' - \Sigma^{-1} = 0, \quad w \in W, \qquad (6)$$

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 neighborhood $\mathcal{O}$ of the set W. Here $m = n^2$, where n is the number of locations. We denote by W^* the set of solutions to (5). To solve (6) we minimize:

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

We employ Levenberg-Marqaurdt (LM) algorithm that interpolates between Gauss-Newton algorithm and method of gradient descent. In many cases, 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 d, the function $F(w+d)$ are approximated by their linearization 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

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

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 which can be solved for the increment vector d. Levenberg-Marquardt contribution is to replace this equation by a '*damped version*',

$$J^T J + \mu * diag \, J^T J \quad d = -J^T F \, w \; .$$

The main difference between Gauss-Newton and LM algorithm is in terms of normal equations. In 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 subproblem:

$$min \; \theta^k \; d \;\; = \;\; F \; w^k \; + J_k d \mid^2 + \mu_k \; d \mid^2, s.t \qquad w^k + d \in W \; . \tag{7}$$

Here, J_k is an approximation of Jacobian of $F' \; w^k$ and μ_k is the positive parameter. Note that θ^k is a strictly convex quadratic function, hence the solution d^k of (6) always exists uniquely. Since our constraints is 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* subproblem:

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

We call this 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 θ^k ensures that d_u^k is a global minimum of this function if and only if $\nabla \; \theta^k \; d_u{}^k \; = 0$, i.e., if and only if $d_u{}^k$ is the unique solution of the system of linear equations [for detailed discussion on Levenberg- Marquardt Method, see Nocedal and Wright (2006)]:

$$J_k{}^T J_k \; + \mu_k diag \, J_k{}^T J_k \quad d_u = -J_k{}^T F \; w^k \tag{8}$$

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 (8).

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 (8).

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 $S2 \rightarrow S3 \rightarrow S6 \rightarrow S7$. 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 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 technique invented by Levenberg-Marquardt involves "blending" between these two extremes. It 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 how close we are to a minimum by how our 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 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 (7).

6. APPLICATION OF THE PROPOSED SOLUTIONS

We estimate the SAR model (4) for the year 1992 using our proposed weight matrix. In order to avoid endogeneity problem, we construct our W matrix using the data on C (Cigarette sales) from 46 states for the period 1963-1991.

Table 3 provides the estimates of the SAR model using the standard W matrix and our numerically solved W using Levenberg-Marquardt algorithm. It is clear that the estimated SAR model using our proposed W matrix is equally good compared to that with the standard W in terms of log-likelihood value.

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

	SAR(W=Constructed using the LM algorithm)	SAR(W=row standardized 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)
ρ	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

In Figure 4 we plot the first-order implied correlation as a function of weights from our estimated W. Out of 46 x 46=2116 pairs of locations, we only plots the 188 first-order neighbor correlations. We first arrange the weights of 188 pairs of first-order neighbors in ascending order, and then the implied correlations are sorted out in ascending order as well. From the Figure 4 we note that the implied correlations have very slow increasing trend with weights. Also there is little variation. This is in contrast to Figure 2 (where number of neighbor increases means weight decreases) which displayed much higher variation.

Next we focus on how implied correlations behave as function of ρ.

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 $\rho=0.1$, the implied correlations vary only in between 0.004 and 0.006; while for $\rho=0.6$ 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.

Figure 4: IMPLIED CORRELATIONS OF SAR MODEL (W=Constructed)
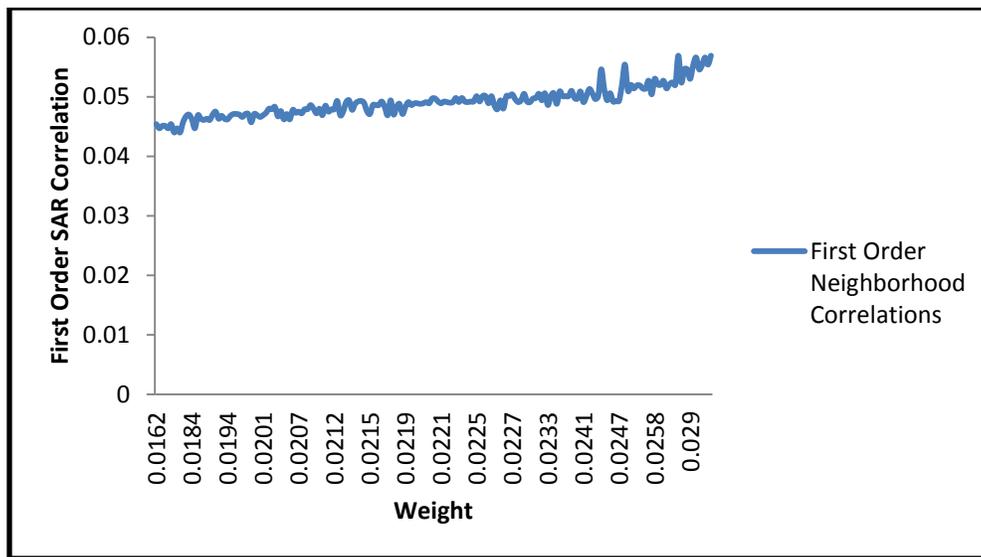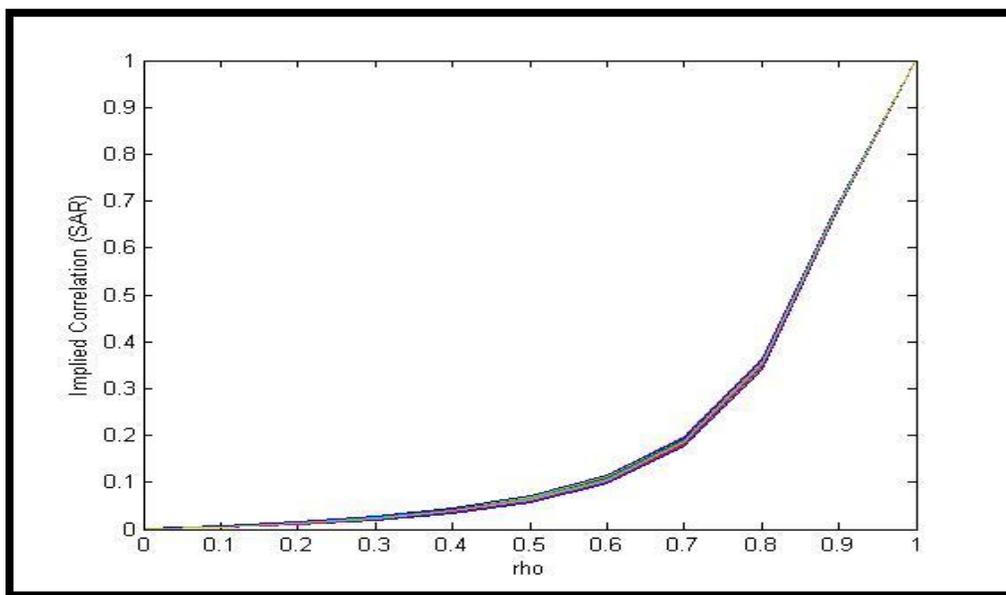

Figure 5: IMPLIED CORRELATIONS OF SAR MODEL (as a function of ρ)

Finally, to address the implied heterogeneity of SAR model, in Figure 6, we plot the 46 diagonal elements of Σ as a function of the number of first-order neighbors. Using the first-order contiguity matrix leads to substantial variation of implied variances of y_i (which decreases with the number of neighbors). In contrast, our proposed W matrix hardly produces any implied heterogeneity.

Figure 6: VARIANCE COMPARISON OF 46 US STATES
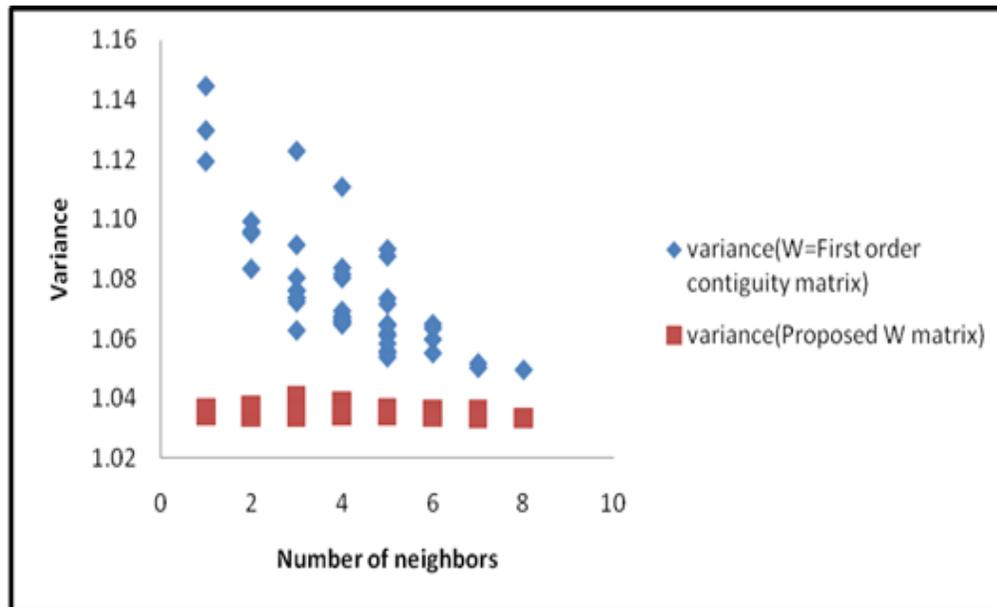
6. CONCLUSION

We first demonstrate that the unintuitive and impractical nature of the implied correlations implied by the estimated SAR models with row standardized neighborhood matrix. We propose a simple methodology for estimation of spatial weight matrix. Our procedure yields very intuitive results in terms of implied correlations and variances. Our proposed methodology is illustrated using the cigarette sales data. Although we apply our proposed method only to the SAR model, it can be easily extended to the CAR model. For CAR, $\text{Var}(y) = \sigma^2(I-\rho W)^{-1}$, which is a variation of (5), and therefore, one can apply the LM procedure to construct a more meaningful weight matrix.

REFERENCES

ANSELIN, L., BERA, A., (1998), "Spatial dependence in linear regression models with an introduction to spatial econometrics", *Handbook of Applied Economic Statistics,* ed. by Aman Ullah and David E.A. Giles, Marcel Dekker, 237-289.

BESAG, J. E., (1974), "Spatial interaction and the statistical analysis of lattice data", *Journal of the Royal Statistical Society*, 36, 192-236.

BESAG, J. E., KOOPERBERG, C., (1995), "On conditional and intrinsic autoregressions", *Biometrika*, 82,733-746.

BALTAGI, B. H., (2001), *Econometric Analysis of Panel Data*, 2nd ed. Chichester, UK Wiley.

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, 117–126.

BALTAGI, B. H., and 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,* 48, 148–155.

CASE, A., (1991), "Spatial patterns in household demands", *Econometrica*, 59, 953-966.

CRESSIE, N., (1993), *Statistics for Spatial Data*, Wiley, New York.

CLIFF, A., ORD, J., (1980), *Spatial Processes: Models and Applications*, Pion Limited.

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

EL-SHEHAWEY, M.A., EL-SHREEF, GH.A., AL-HENAWY, (2008), **"**Analytical inversion of general periodic tridiagonal matrices", *Journal of Mathematical Analysis and Applications Analysis*, 345, 123-134.

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, 375-397.

KELIJIAN, H, H., PRUCHA, I., (2002), "2SLS and OLS in a spatial autoregressive model with equal spatial weights", *Regional Science and Urban Economics*, 32, 691-707.

LEE, L.F., (1999), "Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models", Department of Economics, HKUST, Hong Kong.

LEVENBERG,K., (1944), "A method for the solution of certain non-linear problems in least squares", *The Quarterly of Applied Mathematics*, 2,164–168.

MARQUANDT,D., (1963), "An algorithm for least-squares estimation of nonlinear parameters", *SIAM Journal on Applied Mathematics*, 11, 431–441.

MARTELLOSIO, F., (2009), ''Some correlation properties of spatial autoregressions", Munich Personal RePEc Archive, Paper No. 17254.

NOCEDAL, J; WRIGHT, S,J., (2006), *Numerical Optimization ,* 2nd Edition. Springer.

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, 434-449.