# Conditional Autoregressions with Doubly Stochastic Weight Matrices

R. Kelley Pace[1]
LREC Endowed Chair of Real Estate
Department of Finance
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803-6308
OFF: (225)-578-6256, FAX: (225)-578-6095
kelley@pace.am, www.spatial-statistics.com

James P. LeSage
University of Toledo
Department of Economics
Toledo, OH 43606
jlesage@spatial-econometrics.com

February 13, 2003

[1]We would like to gratefully acknowledge the research support received from the National Science Foundation (BCS-0136193 and BCS-0136229).

**Abstract**

A conditional spatial autoregression (CAR) specifies dependence via a weight matrix. Employing a doubly stochastic weight matrix allows users to interpret the CAR prediction rule as a semiparametric prediction rule and as BLUP with smoothing in addition to other benefits. We examine standard and doubly stochastic weight matrices in the context of an illustrative data set to demonstrate feasibility and statistical performance advantages.

KEYWORDS: spatial autoregression, spatial statistics, maximum likelihood, sparse matrices, doubly stochastic, spatial data mining.

# I. Introduction

In terms of spatial autoregressions, many statisticians prefer conditional spatial autoregressions (CAR) over simultaneous spatial autoregressions (SAR). Unlike SAR, CAR can achieve minimum mean squared prediction error and maximum entropy in some circumstances (Cressie (1993, p. 408-410)). Despite these benefits of CAR, SAR continues to be more widely used in several areas, most notably in economics and quantitative geography (Anselin (1988, p. 33)). Part of the popularity of SAR may lie in its intuitive interpretation as a semiparametric estimator.[1]

We show that specifying a conditional autoregression with a doubly stochastic weight matrix (CARDS) also leads to an attractive semiparametric interpretation. At the same time, CARDS remains the best linear unbiased predictor (BLUP) with smoothing. In addition, the use of doubly stochastic weight matrices reveals some interesting connections and simplifications. For example, maximum likelihood implies the existence of an intuitively appealing set of instrumental variables and CARDS residuals sum to zero, like OLS.

We an empirical illustrations with over 50,000 observations to show the feasibility of CARDS computations and its ability to outperform the standard CAR (CARS) specification in an applied example.

# II. CAR and Doubly Stochastic Weight Matrices

To explore the connection between CAR and doubly stochastic matrices, section A examines standard and doubly stochastic scalings of a basic spatial weight matrix, Section B discusses the CAR estimator and the interpretative benefits of using doubly stochastic weight matrices, while Section C applies both the standard and doubly stochastic CAR estimators to data on cash gifts.

[1] This statement pertains to SAR when used with "row-standardized" or row-stochastic weight matrices. This is one of the most common forms in practice.

## A. Spatial Weight Matrices

Weight matrices provide the means of specifying spatial dependence in spatial lattice models. A weight matrix $C$ is a non-negative, $n$ by $n$ matrix, with zeros on the diagonal and $C_{ij} > 0$, if observation $j$ influences observation $i$. In addition, CAR requires a symmetric weight matrix. Various weight matrices that meet these requirements can be constructed using nearest neighbors, Delaunay triangles, or decreasing functions of distance.

Given a vector $u$, $z = Cu$ gives for each element $z_i$ a combination of the elements of $u$ with the weights supplied by the $i$th row of $C$ (*i.e.*, $z_i = C_{i1}u_1 + C_{i2}u_2 + \ldots C_{in}u_n$). The restriction that a series of weights sum to one is a frequent theme in statistics. Imposing such a restriction means that each row (and each column due to symmetry) of the weight matrix must sum to 1. This property, along with non-negativity, leads to a doubly stochastic weight matrix, $C_D$.

In terms of computation, imposition of the doubly stochastic restriction proceeds iteratively. Beginning with a symmetric candidate matrix $A^{(t)}$, one can measure the row (or column sums) of $A^{(t)}$ in a vector $r^{(t)}$ and create $R_{ii}^{(t)} = r_i^{(t)}$ for $i = 1 \ldots n$. Next compute $A^{(t+1)} = R^{(t)^{-\frac{1}{2}}} A^{(t)} R^{(t)^{-\frac{1}{2}}}$ and iterate until $R \simeq I$. This last instance of $A$ can serve as $C_D$, the doubly stochastic weight matrix. Sparse matrices require $O(n)$ operations, but dense matrices require $O(n^2)$ operations to perform the doubly stochastic reweighting for each iteration. If the matrix is overly sparse, such as a symmetric first-order neighbor matrix, the algorithm can fail to converge. See Bapat and Raghavan (1997, p. 261-263) for more on doubly stochastic scaling algorithms, some of which may improve speed at the cost of implementation complexity relative to the simple approach described above.

Note, the first iteration $A^{(t+1)}$ in the sequence of scaled matrices converging to the doubly stochastic matrix has the same eigenvalues as the row-stochastic version of this matrix (Ord (1975)). Let $C_S$ represent $A^{(t+1)}$. This spatial weight matrix has been used with CAR (*e.g.*, Pace and Barry (1997)) and so we label $C_S$ as the standard weight matrix. Both doubly stochastic and standard weight matrices have minimum eigenvalues greater than or equal to $-1$ and maximum eigenvalues of 1.

A number of approaches lead to spatial weight matrices. In the empirical examples we rely on symmetric nearest neighbor weight matrices (based on a Euclidean metric). The use of $m$ nearest neighbors, where $N^{(1)}, N^{(2)}, \ldots, N^{(m)}$ represent a sequence of $m$ individual nearest neighbor weight matrices, provides a flexible way of constructing a spatial weight matrix. The individual neighbor matrices $N^{(r)}$ contain a single 1 in each row, with all other entries equal to zero. The first individual neighbor matrix contains the very nearest neighbor to each observation, the second individual neighbor matrix contains the second nearest neighbor to each observation, and so forth. We first symmetricize these individual neighbor matrices by defining $N_s^{(r)} = N^{(r)} + N^{(r)\prime}$ for $r = 1 \ldots m$. These are then combined using geometrically declining weights, defined by setting $A_N = \rho^1 N_s^{(1)} + \rho^2 N_s^{(2)} + \ldots + \rho^m N_s^{(m)}$, where the parameter $\rho \in (0, 1]$ imposes a geometric decline in influence over neighbors. As previously defined, the first round of the scaling algorithm produces the matrix $C_S$, and iterating the algorithm until convergence results in the doubly stochastic matrix $C_D$.

## B. The CAR Model

The conditional autoregressive model is a particular specification of the linear multivariate normal model $y = X\beta + \varepsilon$ where $y$ represents the $n$ by 1 vector of observations on the dependent variable, $X$ represents an $n$ by $k$ matrix of observations on the independent variables, $\beta$ represents a $k$ by 1 vector of parameters, and the $n$ by 1 vector of errors $\varepsilon$ follows a multivariate normal density (Cressie (1993, p. 407-408)). Specifically, $\varepsilon$ is distributed as $N(0, \Omega)$, where the $n$ by $n$ variance-covariance matrix $\Omega = \sigma^2 (I - \phi C)^{-1}$ and both $\sigma^2$, $\phi$ represent scalar parameters. One can write the profile log-likelihood as,

$$L(\phi) = \kappa + \left(\frac{1}{2}\right) \ln |I - \phi C| - \left(\frac{n}{2}\right) \ln \left( \left(y - X\widetilde{\beta}\right)' (I - \phi C) \left(y - X\widetilde{\beta}\right) \right)$$

where $\widetilde{\beta} = (X' (I - \phi C) X)^{-1} X' (I - \phi C) y$ and $\kappa$ represents a constant (Pace and Barry (1997)).

The best linear unbiased prediction (BLUP) rule for the linear multivariate normal model when applied to the sample observations is $\widetilde{y} = X\widetilde{\beta} + \Psi\Omega^{-1}(y - X\widetilde{\beta})$ where $\sigma^2\Psi$ specifies the covariance among observations. If the observations have no individual error or measurement error, $\Psi = \Omega$ and the BLUP rule yields the exact interpolator $\widetilde{y} = y$. However, in many cases the observations exhibit individual or idiosyncratic errors. Following the development of Christensen (1991, p. 273-276), one can decompose the overall variance-covariance into a component reflecting the pure spatial association among observations and another component reflecting idiosyncratic or measurement error, resulting in $\sigma^2\Omega = \sigma^2\Psi + \sigma^2 I$. If the observations have idiosyncratic or measurement errors $\sigma^2 I$, the purely spatial component equals $\sigma^2\Psi = \sigma^2\Omega - \sigma^2 I$. Substituting this specification of $\sigma^2\Psi$ into the BLUP rule yields $\widetilde{y} = X\widetilde{\beta} + \widetilde{\phi}C(y - X\widetilde{\beta})$, resulting in the usual CAR prediction rule (Cressie (1993, p. 408)).

In addition, the CAR prediction rule has exactly the same form as the semiparametric model (or partially linear model) prediction rule $\widehat{y} = X\widehat{\beta} + \widehat{\phi}S(y - X\widehat{\beta})$ where $\widehat{y}$ are the semiparametric predictions, $\widehat{\beta}$ are the semiparametric estimates, $\widehat{\phi}$ is a shrinkage parameter, and $S$ is an $n$ by $n$ smoother matrix. For CARDS the smoother matrix equals the doubly stochastic weight matrix ($S = C_D$). For CARDS the smoother matrix $S$ is symmetric, shrinking, and constant preserving (Buja *et al.* (1984, p. 465)).[2] As Hastie and Tibshirani (1990, p. 46) state, "First note that any reasonably smoother is constant preserving, that is $S_\lambda \mathbf{1} = \mathbf{1}$,

[2] Alternatively, let $S(\phi) = \phi S$. The smoother matrix $S(\phi)$ is symmetric, shrinking, and constant shrinking since $S(\phi) = \phi[\mathbf{1}]$.

where $\mathbf{1}$ is an $n$-vector of ones." Projection matrices, one of the main types of smoother matrices, share these same properties. In contrast, $C_S$, while symmetric and shrinking, is not a constant preserving smoother matrix.

The semiparametric prediction rule allows users to interpret the predictions in terms of the conditional mean of the linear part $(X\widehat{\beta})$ and a nonparametric smooth of the residual part $(S(y - X\widehat{\beta}))$. As many users prefer to interpret statistical procedures in terms of the mean part of the model as opposed to the covariance part of the model, the semiparametric approach has appeal.

Simultaneous autoregressions (SAR) provide an alternate to CAR (Ord (1975)). Let $D$ denote the SAR spatial weight matrix (possibly asymmetric). The SAR estimator has the form $\tilde{\beta}_{SAR} = (X'\Omega_{SAR}X)^{-1} X'\Omega_{SAR}\, y$ where $\Omega_{SAR} = \sigma^2 (I - \alpha D)' (I - \alpha D)$ and $\alpha$ represents a scalar parameter. The SAR prediction rule is $\tilde{y}_{SAR} = X\tilde{\beta}_{SAR} + \tilde{\alpha} D \left(y - X\tilde{\beta}_{SAR}\right)$. In its usual form with a row-stochastic weight matrix $D$, the SAR prediction rule also has the same form as a semiparametric prediction rule with a constant preserving smoother matrix.[3] This easily interpreted prediction rule may explain why the SAR specification with a row-stochastic weight matrix has been used more often than the standard CAR specification in the geography and economics literature.

While SAR with row-stochastic weight matrices does match the semiparametric prediction rule, it is not the BLUP rule. Only CAR with a doubly stochastic weight matrix (CARDS) can have the interpretive benefits of both the semiparametric prediction rule and the optimality of the BLUP rule.

Instrumental variable regression methods play an important role in econometrics (Bowden and Turkington (1990)). These techniques provide a means for simultaneous equations estimation, errors-in-variables techniques, and other substantive problems. The use of a doubly stochastic weight matrix provides for a highly intuitive set of instrumental variables. Let $Z = (I - \phi C_D)\, X$ represent a set of instru-

[3] Compare equation 6.3.9 and 6.3.12 in Cressie (1993, p. 406-407).

ments. The instruments, $Z$, are simply the independent variables less a shrinkage parameter times the corresponding spatial averages of the independent variables and it seems reasonable that partially differencing independent variables with their spatial averages can reduce or eliminate the correlation between the transformed variables and the residuals. The resulting IV estimates $\tilde{\beta}_{IV} = (Z'X)^{-1} Z'y$ equal the CAR estimates (Ripley (1981, p. 90), Bowden and Turkington (1990, p. 12)). In fact, $Z'\tilde{e} = 0$, where $\tilde{e} = y - X\tilde{\beta}_{IV}$ for CAR and this condition is necessary to establish consistency in instrumental variable regressions. In contrast, let $e_{CAR} = (I - \phi C)\,\tilde{e}$ and, as Ripley (1981, p. 91) showed, $X'\tilde{e}_{CAR} = 0$. Both $Z'\tilde{e} = 0$ and $X'\tilde{e}_{CAR} = 0$ are the same mathematically, but the former shows orthogonality between the instruments and the ordinary residuals while the latter shows orthogonality between the regressors and the CAR residuals. These various relations demonstrates some of the potential interpretive benefits from using doubly stochastic weight matrices.

The use of doubly stochastic matrices leads to an additional interpretive benefit. Suppose the design matrix contains a vector of ones in the first column or $X_1 = \mathbf{1}$. Since $Z_1 = (I - \phi C)\,\mathbf{1} = (1 - \phi)\mathbf{1}$ and since $Z_1'\tilde{e} = 0$, the residuals sum to 0 just as with OLS. Of course, in the case where the model only includes the intercept term, the CAR intercept estimate equals $(1 - \phi)\beta_{OLS}$ and thus shows a simple trade-off between modeling spatial and non-spatial estimation.

## C. Cash Contributions

While CAR using a doubly stochastic weight matrix (CARDS) has interpretative benefits over the traditional CAR specifications, can CARDS outperform standard CAR (CARS) on actual data? As an empirical illustration of using CARS versus CARDS we examine cash contribution expenditure data by census tract for 1999. Political parties and eleemosynary organizations have a great interest in maximizing cash contributions. Accurate predictions of such giving behavior as a function of observable characteristics and space could assist in devising marketing plans and other optimizing behavior.

To examine this issue, we use 25 variables from the 1990 census on race, population, house prices, and income as explanatory variables along with 1999 total consumer expenditures, and 51 state dichotomous variables to predict 1999 cash contributions. We included the state dichotomous variable to allow for variation in giving incentives across states arising from tax code differences. This results in 54,584 complete data observations that we use to fit ln(cash contributions) to the explanatory variables via CAR.[4] We maximized the profile log-likelihood using the three non-linear parameters $\phi$, $m$, and $\rho$ over a grid of values $\phi = 0.000, 0.001, \ldots, 0.999$, $m = 20, 25, 30$, $\rho = 0.8, 0.9, 1.0$. Table 1 shows the resulting profile log-likelihoods for both CARS and CARDS. The CARS specification results in a log-likelihood of $-56,975.8$ while the CARDS specification results in a log-likelihood of $-56,958.0$, a difference of 17.8 in favor of CARDS over CARS. For both CARS and CARDS the estimated number of neighbors was 30 and the optimal rate of geometric decay was 0.9. Both CARS and CARDS yield large spatial autoregressive parameters of 0.960 and 0.965. Thus, space plays an important role in predicting the amount of gifts.

# III. Conclusion

The use of conditional autoregressions with doubly stochastic weight matrices (CARDS) provides a number of benefits. Specifically, CARDS predictions possess the optimality of BLUP with smoothing along with a natural semiparametric interpretation. In addition, CARDS estimates have both maximum likelihood and instrumental variable interpretations. Finally, the constant preserving nature of

[4] We use the techniques in Pace and Barry (1997) to quickly compute the necessary log-determinants of the 54,584 by 54,584 over a grid of values for $\phi$. These commands, along with those to estimate CAR, are in the Spatial Statistics toolbox 1.1 at www.spatial-statistics.com.

doubly stochastic matrices allow the CARDS residuals to sum to 0 as with OLS. Empirically, CARDS demonstrated its potentially to produce higher likelihoods in a large empirical examples. In conclusion, the use of CAR with doubly stochastic weight matrices (CARDS) can improve both statistical interpretation as well as empirical performance.

## References


Anselin, Luc (1988), *Spatial Econometrics: Methods and Models*, Dorddrecht: Kluwer Academic Publishers.

Bapat, R.B., and T.E.S. Raghavan (1997), *Nonnegative Matrices and Applications*, Cambridge: Cambridge University Press.

Bowden, Roger and Darrell Turkington (1990), *Instrumental Variables*, Cambridge: Cambridge University Press.

Buja, A., T. Hastie, and R. Tibshirani (1989), "Linear Smoothers and Additive Models (with discussion)," *Annals of Statistics*, Volume 17, p. 453-555.

Christensen, Ronald (1991), *Linear Models for Multivariate, Time Series, and Spatial Data*, New York: Springer-Verlag, 1991.

Cressie, Noel (1993), *Statistics for Spatial Data*, Revised Edition, New York: John Wiley & Sons.

Hastie, Trevor, and Robert Tibshirani (1990), *Generalized Additive Models,* London: Chapman and Hall.

Ord, J.K. (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, Volume 70, p. 120-126.

Pace, R. Kelley, and Ronald Barry (1997), "Fast CARs," *Journal of Statistical Computation and Simulation*, Volume 59, p. 123-147.

Ripley, Brian (1981), *Spatial Statistics*, New York: Wiley.

**Table 1 – Optimal Log-likelihoods for Contributions Across US Census Tracts Using both Standard (CARS) and Doubly Stochastic Weight Matrices (CARDS)**

| $m$ | $\rho$ | $\alpha_s$ | $\alpha_d$ | CARS Loglik | CARDS Loglik |
|---|---|---|---|---|---|
| 20 | 1.0 | 0.965 | 0.969 | $-57,167.9696$ | $-57,141.5586$ |
| 20 | 0.9 | 0.943 | 0.948 | $-57,061.4971$ | $-57,044.7399$ |
| 20 | 0.8 | 0.894 | 0.901 | $-57,301.7640$ | $-57,280.0609$ |
| 25 | 1.0 | 0.976 | 0.980 | $-57,182.2572$ | $-57,156.2483$ |
| 25 | 0.9 | 0.954 | 0.959 | $-57,009.5600$ | $-56,992.1429$ |
| 25 | 0.8 | 0.899 | 0.905 | $-57,282.3381$ | $-57,260.1137$ |
| 30 | 1.0 | 0.984 | 0.987 | $-57,200.3278$ | $-57,176.2576$ |
| 30 | 0.9 | 0.960 | 0.965 | $-56,975.8089$ | $-56,957.9824$ |
| 30 | 0.8 | 0.900 | 0.907 | $-57,275.2575$ | $-57,252.7954$ |