

Specification and Estimation of a Spatially and Temporally Autocorrelated Seemingly Unrelated Regression Model: Application to Crash Rates in China

Xiaokun Wang

Graduate Student Researcher
The University of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
wangxk@mail.utexas.edu Phone:
512-471-8270
FAX: 512-475-8744

Kara M. Kockelman

Associate Professor & William J. Murray Jr. Fellow Department of
Civil, Architectural and Environmental Engineering The University
of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
kkockelm@mail.utexas.edu

The following paper is a pre-print and the final publication can be found in
Transportation 34 (3):281-300, 2007.
Presented at the 86th Annual Meeting of the Transportation Research Board,
January 2007

ABSTRACT

In transportation studies, variables of interest are often influenced by similar factors and have correlated latent terms (errors). In such cases, a seemingly unrelated regression (SUR) model is normally used. However, most studies ignore the potential temporal and spatial autocorrelations across observations, which may lead to inaccurate conclusions. In contrast, the SUR model proposed in this study also considers the spatial and temporal correlations across observations, making the model more behaviorally convincing and applicable to circumstances where a three-dimensional correlation exists, across time, space and equations. An example of crash rates in Chinese cities is used. The results show that incorporation of spatial and temporal effects significantly improves the model. Moreover, investment in transportation infrastructure is estimated to have statistically significant effects on reducing severe crash rates, but with an elasticity of only -0.078. It is also observed that though vehicle ownership is associated with higher crash per capita rates, elasticities for severe and non-severe crashes are just 0.13 and 0.18 respectively; much lower than one. The techniques illustrated in this study should contribute to future studies requiring multiple equations in the presence of temporal and spatial effects.

KEY WORDS: Spatial econometrics, seemingly unrelated regression, spatial and temporal autocorrelation, random effects, crash rates

INTRODUCTION

In transportation studies and other regional analyses, many variables relate. If dependent variables are co-dependent, a simultaneous equation model (SEM) is appropriate. In other cases, these variables are simply correlated in their regression error terms, and a seemingly unrelated regression (SUR) approach becomes more reasonable. Some transportation examples include: (1) trade flows by different industries (Egger and Pfaffermayr 2004), (2) sales impacts of highway bypasses on different industry sectors across cities and over time (Srinivasan and Kockelman 2002), and (3) travel demand induced by road capacity increases (Noland, 2001). In addition, observations are often panel data with spatial interaction: The same units are observed for multiple periods, and nearby units tend to have stronger correlations. Previous models are incapable of including all these correlated effects. Thus, the primary motivation for – and the most important contribution of – this study are an econometric model and estimation techniques that recognize all these effects.

In this study, correlations across equations are specified in a general way: temporal correlation across observations is assumed to be a random-effect, and spatial effects are incorporated via a spatial autoregressive (SAR) component in the error term. The estimation techniques are a mixture of generalized least squares (GLS) and maximum likelihood estimation (MLE) and can handle these complicated correlation patterns. The overall methodology is an important extension to existing studies and may serve as a useful tool for future work.

The model is applied for analysis of city-level severe and non-severe crash rates (per capita) across China. While numerous studies on vehicle crash rates and counts using disaggregate (segment-based) data can efficiently disclose the effects of geometric design, such models can miss the effects of broader, non-design policies and trends. In contrast, aggregated city-level data can better illuminate the effects of regional policies. Such continuous response variables also lend themselves to more convenient linear regression methods (rather than integer crash counts, for example). Of course, the cost of this convenience is that the aggregation process obscures information on driver, road and vehicle characteristics. Such covariates are absorbed into the models' unobserved terms. For the same city, these unobserved, error terms for crash rates of different severities are interrelated. Meanwhile, neighboring cities tend to have similar topography, educational attributes, weather patterns, industry and reporting rates; they also may have similar traffic volumes on national highways. All these factors are influential and tend to be spatially correlated; yet, as uncontrolled variables, all are absorbed into the error term. In addition, for panel data, it is natural to consider the temporal correlations in these unobserved effects (for the same city, over time). Hence, the methodology used here is very well suited to such situations.

Another reason for analyzing Chinese crash rates is the rapid, recent motorization taking place there. From 1994 to 2004, the total number of motor vehicles in China (including motorcycles) has increased from 19.5 million to 107.8 million (Su, 2005), with an average increase over 16% per year. As a point of comparison, in 2004 there were 242 million motor vehicles in the U.S. (also including motorcycles). In 1994 there were 201.8 million such vehicles in the U.S., suggesting an average growth rate of just 1.8 % per year (FHWA, 1994-2004). Meanwhile, due to the unbalanced regional economic development, some areas in China have been experiencing much more rapid motorization than others. For example, of the 169 cities in the sample data, the highest motorization rate (from 2001 to 2002) is 60% while the lowest one is -3%. The standard deviation of this speed is as high as 12%. Such variability provides an analysis of transportation issues related to motorization in developing countries.

The following sections discuss previous studies on related topics and offer a detailed explanation of the model specification and estimation techniques. The data set's statistics and potential data quality problems are explored. The estimation results are examined and compared to other, more restrictive models, highlighting the superiority of the model proposed here.

LITERATURE REVIEW

Zellner (1962) first proposed seemingly unrelated regression (SUR) in order to analyze multiple equations with correlated error terms. Each equation may have a different set of explanatory variables; however, in referring to responses of the same set of observational units, the errors of these equations are likely to correlate. Originally applied in micro-economic contexts, SUR is now used broadly in many research areas. Anselin (1988) first extended an SUR model to a spatial environment. By incorporating spatial autoregression into the error term, the model exhibits spatial autocorrelations across observations. Anselin's focus was a Lagrange multiplier test for spatial SUR, and he did not explore estimation techniques. Nevertheless, the likelihood function and some information matrix elements presented in his work provide important inspiration for this study. Interestingly, Anselin's work originally was designed for panel data analysis, where each equation stands for a period and the correlation across periods is in a general form with the assumption that the data is a short panel (i.e., T is small compared to N). For most regional and transportation studies, this assumption of short panel holds true. However, computational burdens can increase dramatically with the number of periods. Therefore, in most present studies, the temporal dependence is often structured by decomposing the error term into an individual time-invariant effect plus a time variant term.

Elhorst (2003) first provided a comprehensive illustration of how to combine panel data with spatial analysis. In his study, "spatial dependence" can be either spatial error autocorrelation or a spatially lagged dependent variable. The panel nature of the data can be recognized using fixed or random effects, and fixed or random coefficients. As Elhorst concludes, fixed-effects models are suitable for short panels and coefficient estimates can be inconsistent for larger time periods (T). In comparison, the random-effects model is restricted by an assumption of zero correlation between the individual effects and explanatory variables. However, when the cross sectional dimension (N) is large (while T is fixed), the random-effects model may be preferable. The fixed-coefficients model is quite similar to Anselin's model (1988), except that Elhorst treated every spatial unit as a separate equation. In this way the spatial correlation structure is not restricted; however, as Elhorst notes, due to the large number of parameters, such a model is useful only when the cross-sectional dimension is small. As for a random-coefficients model, in addition to computational issues, it may become "asymptotically suspect" for large panels (when N is large relative to T).

Another important contribution to spatial panel data analysis is Kapoor et al.'s work (2004), where a mixture of generalized moments estimators and feasible generalized least square (FGLS) is used. The most attractive feature of their method is that it is computationally feasible for large panels. However, in many circumstances (e.g., for evaluating estimators' asymptotic properties or for various hypothesis tests), it may be necessary to calculate the full information matrix. For this reason, Elhorst's MLE method remains preferred.

In terms of crash rate analyses, many studies exist. For example, Aarts and Shagen (2006) analyzed the effect of driving speed on crash rates, Kweon and Kockelman (2005) studied the impact of roadway design and speed limits on crash rates, and Ivan (2004) explored the relation between traffic volumes and crash rates. Most studies define crash rate as the crash count divided

by VMT; here, however, as in Noland and Karlaftis (2005), crash rate is defined as the ratio of crash count to population.

Sophisticated crash rate analysis using a Chinese database is quite limited, due primarily to a lack of data. As Yi and Ran (2003) pointed out, Chinese traffic accident data acquisition, communications, and analysis systems need improvement. Among the few such studies, Qin et al. (2004) investigated 1085 crashes occurring in 1998 and 1999 on a single 198-mile highway section and concluded that crash rates (though not necessarily crash severities) increase with traffic volume, and safety policies can have a significant effect on reducing such rates. A city-based aggregate study of Chinese crash rates remains absent, and is addressed by this work.

Furthermore, no previous study of crash rates has simultaneously explored the temporal, spatial and cross-equation correlations which exist. The methodology proposed in the following section fills this void, offering new opportunities for spatial analysts. By recognizing error terms that correlate over three dimensions, this model can be viewed as a panel-based extension of Anselin's (1988) spatial SUR model, as well as an SUR extension to Elhorst's (2003) spatial-random effects model. Thus, the work provides a new model for use with panel data and a system of seemingly-unrelated equations. Of course, the methods are applicable for any type of data, not simply various crash rates.

METHODOLOGY

In this study, spatial effects are incorporated via autocorrelation in spatial error terms. A spatial lag model (with a spatially lagged dependent variable) can be specified and estimated in a similar way but is not discussed here, due to space limitations. Interested readers may refer to Elhorst's (2003) study for a comparison of these two model types. Derivation of a panel spatial lag SUR model via modifications similar to those discussed here should be relatively straightforward.

Here, panel effects are incorporated via random effects, and the reasons for and limitations of a random effect assumption are discussed below. In "model estimation", a three-step method involving a (FGLS) regression and MLE is introduced. A comparison with previous studies demonstrates that these estimation techniques can be viewed as an extension of and a counterpart to Kapoor et al.'s (2004) work (which is based on a generalized method of moments).

Model Specification

In general, a spatial SUR model for panel data can be described as follows: If the problem under study is composed of G equations (each potentially having a different set of explanatory variables) for N individuals and the study relies on balanced panel data (such that each individual is represented T times, for each of the G equations), the system can be specified as follows:

$$y_{git} = X_{git}\beta_g + \xi_{git}, \quad g = 1, 2, \dots, G, i = 1, 2, \dots, N, t = 1, 2, \dots, T \quad (1)$$

where y_{git} denotes the dependent variable value of the i^{th} individual in period t in equation g , X_{git} is a $1 \times K_g$ vector of explanatory variables, β_g is a $K_g \times 1$ vector of parameters, and ξ_{git} is a scalar error term.

Stacking the observations (first by equation, then by time, and finally by individual), the system can be denoted as $Y = X\beta + \xi$, where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_G \end{bmatrix}, \text{ where } Y_g = \begin{bmatrix} Y_{g1} \\ Y_{g2} \\ \vdots \\ Y_{gT} \end{bmatrix} \text{ and each } Y_{gt} = \begin{bmatrix} y_{g1t} \\ y_{g2t} \\ \vdots \\ y_{gNt} \end{bmatrix} \quad (2)$$

Here Y is of dimension $GTN \times 1$; X is a diagonal block matrix of dimension $GTN \times \sum_g K_g$, with each sub-matrix X_g (located along the main diagonal) representing an $TN \times K_g$ matrix; β is $\sum_g K_g \times 1$; and ξ is $GTN \times 1$.

For standard, a-spatial panel data, the error term can be decomposed into two parts: an individual-specific (or time-constant) error term and a time-variant error term:

$$\xi_{git} = a_{gi} + \varepsilon_{git} \quad (3)$$

In this study, a random effects model is assumed to be appropriate, implying that individual effects are independent of explanatory variables. Thus,

$$E(\varepsilon | a, X) = 0 \quad (4)$$

$$E(a | X) = 0 \quad (5)$$

The individual-specific error a and the idiosyncratic error ε also are assumed to be homoscedastic (within each equation), so that:

$$E(a_{gi} \cdot a_{gi}) = \sigma_{ag}^2 \text{ (or } \sigma_{agg} \text{, in a more generalized form), for all } g, i \quad (6)$$

$$E(\varepsilon_{git} \cdot \varepsilon_{git}) = \sigma_{eg}^2 \text{ (or } \sigma_{egg} \text{), for all } g, i, t \quad (7)$$

It is also assumed that, by incorporating the individual-specific error, the idiosyncratic errors will be serially uncorrelated:

$$E(\varepsilon_{git} \cdot \varepsilon_{gis}) = 0, \text{ for all } g, i \text{ and } t \neq s \quad (8)$$

For an SUR model, the correlations between equations also need to be considered:

$$E(a_{gi} \cdot a_{hi}) = \sigma_{agh}, \text{ for all } i \text{ and } g \neq h \quad (9)$$

$$E(\varepsilon_{git} \cdot \varepsilon_{hit}) = \sigma_{egh}, \text{ for all } i, t \text{ and } g \neq h \quad (10)$$

It should be noted that, as compared to a fixed-effects model, the assumption of zero correlation between individual effects and explanatory variables is restrictive and can be unrealistic at times. However, as Anselin (1999, pp 14) notes, “since the estimation of the spatial process models requires asymptotics in the cross-sectional domain ($N \rightarrow \infty$)”, fixed effects will lead to inconsistent estimation due to the incidental parameter problem¹ – and this is incompatible with spatial processes. Moreover, a standard fixed-effects approach cannot give coefficient estimates for time-constant variables, whereas random-effects models can. In light of all these facts, a random-effect models is specified here.

In a *spatial* error model, the error terms are spatially autocorrelated:

$$\xi_{g \bullet t} = \lambda_g W_g \xi_{g \bullet t} + a_{g \bullet} + \varepsilon_{g \bullet t} \text{ or } \xi_{g \bullet t} = (I_N - \lambda_g W_g)^{-1} (a_{g \bullet} + \varepsilon_{g \bullet t}), \text{ for all } g, t \quad (11)$$

¹ An incidental parameters problem means that, for a fixed-effects model, when T is small, estimators of the constant terms do not converge, leading to inconsistent estimators of all coefficients. [See Lancaster (2000).]

where W_g is the weight matrix for equation g and λ_g is the corresponding autocorrelation coefficient. W_g is an $N \times N$ exogenous matrix that reflects the spatial dependence pattern across observations, with zero-valued diagonal elements. Normally, W_g is row standardized to sum to 1 and λ_g is restricted so that $-1 < \lambda_g < 1$. In this way, the stationary autocorrelation is assured, without loss of (much) generality. Thus, in the spatial SUR model of panel data, the error term can be expressed as follows:

$$\xi = H^{-1} [a + \varepsilon] \quad (12)$$

$$\text{where } H = \begin{bmatrix} I_T \otimes H_1 & 0 & \cdots & 0 \\ 0 & I_T \otimes H_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I_T \otimes H_G \end{bmatrix} \quad \text{with each } H_g = I_N - \lambda_g W_g \quad (13)$$

Therefore, as Anselin (1988) shows, the inverse of variance-covariance matrix is $\Omega^{-1} = H'(\Sigma^{-1} \otimes I_N)H$, where $\Sigma \otimes I_N$ is the variance-covariance of the composite error term $(a + \varepsilon)$, with

$$\Sigma = A \otimes l_T l_T' + B \otimes I_T \quad (14)$$

$$\text{and } A = \begin{bmatrix} \sigma_{a1}^2 & \sigma_{a12} & \cdots & \sigma_{a1G} \\ \sigma_{a21} & \sigma_{a2}^2 & \cdots & \sigma_{a2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{aG1} & \sigma_{aG2} & \cdots & \sigma_{aG}^2 \end{bmatrix}, \quad B = \begin{bmatrix} \sigma_{e1}^2 & \sigma_{e12} & \cdots & \sigma_{e1G} \\ \sigma_{e21} & \sigma_{e2}^2 & \cdots & \sigma_{e2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{eG1} & \sigma_{eG2} & \cdots & \sigma_{eG}^2 \end{bmatrix} \quad (15)$$

and where l_T is a $T \times 1$ vector of 1's.

If the error terms are assumed to be normally distributed, following Anselin's model specification, the log-likelihood function (without the constant term) is

$$L = -\frac{1}{2} \ln |\Omega| - \frac{1}{2} (Y - X\beta)' \Omega^{-1} (Y - X\beta) \quad (16)$$

and some basic mathematical manipulations result in the following:

$$L = -\frac{N}{2} \ln |\Sigma| + T \sum_g \ln |H_g| - \frac{1}{2} (Y - X\beta)' H' (\Sigma^{-1} \otimes I_n) H (Y - X\beta) \quad (17)$$

Model Estimation

The parameters are intertwined in the above log-likelihood function, so ordinary regression methods are not feasible. The model can be estimated using a three-step method: first, β can be estimated using a generalized least squares model (GLS), conditional on A, B and λ . Then A and B can be estimated conditional on β and λ . These first two steps are iterated until the optimal A, B and β are found (conditional on λ). The third step is to substitute the values of estimated A, B and β values and to maximize the concentrated log-likelihood function over λ . The estimated λ then re-enters the estimation of A, B and β . This procedure is iterated until convergence. In short, first $L(\beta|\lambda, A, B)$ and $L(A, B|\lambda, \beta)$ are iteratively maximized over A, B

and β , in order to maximize $L(\beta, A, B, |\lambda)$. Then, $L(\lambda | \beta, A, B)$ and $L(\beta, A, B, |\lambda)$ are iteratively maximized to find the complete MLE parameter.

It should be noted that the log-likelihood function specified in Equations (11) through (17) can be further simplified by following the second part of Magnus' (1982) lemma : Let

$M_1 = \frac{1}{T} l_T l_T'$, (which has rank 1) and $M_2 = I_T - \frac{1}{T} l_T l_T'$ (which has rank $T - 1$); then, Σ can be expressed as

$$\Sigma = A \otimes l_T l_T' + B \otimes I_T = (B + TA) \otimes M_1 + B \otimes M_2 \quad (18)$$

According to Magnus (1982),

$$|\Sigma| = |B + TA| |B|^{T-1} \quad \text{and} \quad (19)$$

$$\Sigma^{-1} = (B + TA)^{-1} \otimes M_1 + B^{-1} \otimes M_2 \quad (20)$$

Thus, the log-likelihood function (Equation 17) can be expressed as

$$\begin{aligned} L = & -\frac{N}{2} \ln |B + TA| - \frac{N(T-1)}{2} \ln |B| + T \sum_g \ln |H_g| \\ & - \frac{1}{2} (HY - HX \beta)' \left((B + TA)^{-1} \otimes M_1 \otimes I_N \right) (HY - HX \beta) \\ & - \frac{1}{2} (HY - HX \beta)' \left(B^{-1} \otimes M_2 \otimes I_N \right) (HY - HX \beta) \end{aligned} \quad (21)$$

Using the above decomposition of the determinant and inverse of Σ , a data transformation trick (as inspired by Elhorst et al.'s (2003) work) can be applied as follows:

Step 1. Estimate β conditional on A, B and λ (Maximizing $L(\beta | \lambda, A, B)$)

Since $M_1 \otimes I_N$ denotes an average of the $(HY - HX \beta)$ values over time for each equation, and $M_2 \otimes I_n$ denotes each observation's deviation from this average (over time), if one lets

$P'P = (B + TA)^{-1}$ and $Q'Q = B^{-1}$, one can transform the data by making

$$Y^* = (P \otimes I_{NT}) \overline{HY} + (Q \otimes I_{NT}) HY - (Q \otimes I_{NT}) \overline{HY} = (Q \otimes I_{NT}) HY + ((P - Q) \otimes I_{NT}) \overline{HY} \quad (22)$$

$$X^* = (P \otimes I_{NT}) \overline{HX} + (Q \otimes I_{NT}) HY - (Q \otimes I_{NT}) \overline{HX} = (Q \otimes I_{NT}) HX + ((P - Q) \otimes I_{NT}) \overline{HX} \quad (23)$$

(where bars indicate averages over time). In this way, the regression resembles a standard linear regression, with transformed data:

$$\hat{\beta} = \left(X^{*'} X^* \right)^{-1} \left(X^{*'} Y^* \right) \quad (24)$$

Step 2. Estimate A and B conditional on β and λ (Maximizing $L(A, B | \lambda, \beta)$)

To simplify the following expressions, we denote $\hat{e} = H(Y - X \hat{\beta})$, which can be interpreted as the transformed residuals, or more strictly, the spatial-autocorrelated transformed residuals. Then the last part in Equation (21) (conditional on both β and λ) is simply

$$\begin{aligned} & -\frac{1}{2} \hat{e}' \left(B^{-1} \otimes M_2 \otimes I_N \right) \hat{e} . \text{ This term is actually a scalar that equals its trace, so:} \\ & \hat{e}' \left(B^{-1} \otimes M_2 \otimes I_N \right) \hat{e} = \text{tr} \left(\hat{e}' \left(B^{-1} \otimes M_2 \otimes I_N \right) \hat{e} \right) \end{aligned} \quad (25)$$

Based on properties of direct product, the above expression can be further manipulated to

$$tr\left(\tilde{e}'(I_G \otimes M_2 \otimes I_N)' \cdot (B^{-1} \otimes I_{NT}) \cdot (I_G \otimes M_2 \otimes I_N) \tilde{e}\right) \quad (26)$$

This $(I_G \otimes M_2 \otimes I_N) \tilde{e}$ can be denoted as $\tilde{\tilde{e}}$. As previously discussed, $M_2 \otimes I_N$ denotes an observation's deviation from its average value over time for each equation. Thus, this $\tilde{\tilde{e}}$ is simply the transformed residual's individual deviation from its time mean. Thus, Equation (25) can be further simplified as

$$\tilde{e}'(B^{-1} \otimes M_2 \otimes I_n) \tilde{e} = tr\left(\tilde{\tilde{e}}' \cdot (B^{-1} \otimes I_{NT}) \cdot \tilde{\tilde{e}}\right) = tr\left((B^{-1} \otimes I_{NT}) \cdot \tilde{\tilde{e}} \tilde{\tilde{e}}'\right) \quad (27)$$

Using $\tilde{\Pi}$ (of dimension $GNT \times GNT$) to denote the matrix $\tilde{\tilde{e}} \tilde{\tilde{e}}'$, or the variance-covariance matrix of the demeaned transformed error terms, Equation (25) can be further simplified as

$$\tilde{e}'(B^{-1} \otimes M_2 \otimes I_n) \tilde{e} = tr\left((B^{-1} \otimes I_{NT}) \cdot \tilde{\Pi}\right) = tr\left(B^{-1} \cdot \tilde{\Theta}\right) \quad (28)$$

where $\tilde{\Theta}$ is a $G \times G$ matrix. Each element of $\tilde{\Theta}$ is the trace of a $NT \times NT$ sub-block matrix in $\tilde{\Pi}$'s corresponding position:

$$\tilde{\Theta}_{gh} = tr \left[\begin{array}{cccc} \tilde{\Pi}_{(g-1)NT+1, (h-1)NT+1} & \tilde{\Pi}_{(g-1)NT+1, (h-1)NT+2} & \cdots & \tilde{\Pi}_{(g-1)NT+1, hNT} \\ \tilde{\Pi}_{(g-1)NT+2, (h-1)NT+1} & \tilde{\Pi}_{(g-1)NT+2, (h-1)NT+2} & \cdots & \tilde{\Pi}_{(g-1)NT+2, (h-1)NT+2} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\Pi}_{gNT, (h-1)NT+1} & \tilde{\Pi}_{gNT, (h-1)NT+2} & \cdots & \tilde{\Pi}_{gNT, hNT} \end{array} \right], \text{ for all } g, h \quad (29)$$

Similarly, $\tilde{e}'((B+TA)^{-1} \otimes M_1 \otimes I_n) \tilde{e}$ can be simplified as $tr\left((B+TA)^{-1} \cdot \bar{\Theta}\right)$

where $\bar{\Theta}$ also is a $G \times G$ matrix with each element being the trace of the corresponding sub-block matrix of $\bar{\Pi}$, which comes from the transformed residuals' individual mean (over time). Thus Equation (21) can be finally expressed as

$$L = -\frac{N}{2} \ln|B+TA| - \frac{N(T-1)}{2} \ln|B| + T \sum_g \ln|H_g| - \frac{1}{2} tr\left((B+TA)^{-1} \cdot \bar{\Theta}\right) - \frac{1}{2} tr\left(B^{-1} \cdot \tilde{\Theta}\right) \quad (30)$$

which gives one immediate optimal solutions for A and B:

$$\frac{\partial L}{\partial A} = -\frac{NT}{2} (B+TA)^{-1} + \frac{T}{2} (B+TA)^{-1} \bar{\Theta} (B+TA)^{-1} \quad (31)$$

$$\frac{\partial L}{\partial B} = -\frac{N}{2} (B+TA)^{-1} - \frac{N(T-1)}{2} B^{-1} + \frac{1}{2} (B+TA)^{-1} \bar{\Theta} (B+TA)^{-1} + \frac{1}{2} B^{-1} \tilde{\Theta} B^{-1} \quad (32)$$

resulting in the optimal set:

$$B = \frac{1}{N(T-1)} \tilde{\Theta} \quad (33)$$

$$A = \frac{1}{NT} \bar{\Theta} - \frac{1}{NT(T-1)} \tilde{\Theta} \quad (34)$$

By iterating steps 1 and 2, the optimal values for A, B and β can be obtained conditional on H.

Step 3. Estimate λ conditional on A, B and β (Maximizing $L(\lambda|\beta, A, B)$)

The optimized A, B, β from the first two steps then are substituted into the log-likelihood function, and the only parameters left to be obtained are $\lambda_g, g = 1, 2, \dots, G$. The optimal λ_g cannot be derived analytically and can be found only by using a nonlinear optimization tool (such as Matlab and GAUSS). One can iteratively maximize equation (21) via $L(\lambda|\beta, A, B)$ and $L(\beta, A, B|\lambda)$ until convergence, in order to obtain the maximum *unconditional* likelihood.

Covariance of Estimates

In order to obtain (asymptotic) estimates of the variance-covariance matrix for all parameters of interest, one can construct an information matrix. This can be used to conduct various hypothesis tests on parameters in addition to estimate uncertainty in estimations. As Anselin (1988) notes, the information matrix for the maximum likelihood estimators can be expressed as

$$\Psi_{ij} = \frac{1}{2} tr \left[\left(\frac{\partial \Omega^{-1}}{\partial \theta_i} \right) \Omega \left(\frac{\partial \Omega^{-1}}{\partial \theta_j} \right) \right] \quad (35)$$

where θ_i stands for the i^{th} parameter in the estimation. In this model, the information matrix is block diagonal between the elements of β and all other parameters. The part for β is the usual $X'\Omega^{-1}X$. The elements of the information matrix for λ , A and B (which are composed of σ_g) need to be derived with some mathematical manipulations. Due to the space limitations, only the final results are shown here:

$$\Psi(\lambda_g, \lambda_g) = T tr(D_g)^2 + tr \left[E^{gg} (B+TA)^{-1} E^{gg} (B+TA) + (T-1) E^{gg} B^{-1} E^{gg} B \right] \cdot tr(D_g D_g) \quad (36)$$

$$\Psi(\lambda_g, \lambda_h) = tr \left[E^{gh} (B+TA)^{-1} E^{gh} (B+TA) + (T-1) E^{gh} B^{-1} E^{gh} B \right] \cdot tr(D_g D_h), \text{ for all } g \neq h \quad (37)$$

$$\Psi(\lambda_g, \sigma_{amn}) = T \cdot tr \left[E^{gg} (B+TA)^{-1} E^{nm} \right] \cdot tr(D_g) \quad (38)$$

$$\Psi(\lambda_g, \sigma_{emm}) = tr \left[E^{gg} \left((B+TA)^{-1} + (T-1) \cdot B^{-1} \right) E^{mm} \right] \cdot tr(D_g) \quad (39)$$

$$\Psi(\sigma_{agh}, \sigma_{amn}) = \frac{NT^2}{2} \cdot tr \left[(B+TA)^{-1} E^{gh} (B+TA)^{-1} E^{nm} \right] \quad (40)$$

$$\Psi(\sigma_{agh}, \sigma_{emm}) = \frac{NT}{2} \cdot tr \left[(B+TA)^{-1} E^{gh} (B+TA)^{-1} E^{mm} \right] \quad (41)$$

$$\Psi(\sigma_{emm}, \sigma_{enn}) = \frac{N}{2} \cdot tr \left[(B+TA)^{-1} E^{mm} (B+TA)^{-1} E^{mm} + (T-1) B^{-1} E^{mm} B^{-1} E^{mm} \right] \quad (42)$$

where $D_g = W_g H_g^{-1}$ and E^{gh} is a $G \times G$ matrix with its (g, h) and (h, g) elements equal to one and zeros elsewhere, and g, h, m and n index equations 1 through G .

DATA DESCRIPTION

The data used for this study come from a transportation survey conducted in China, over a four-year period (1999 - 2002) for 491 cities. However, only 169 cities have valid data for all four years. Therefore, this study uses only these 169 cities as its sample data. Among these 169 observations, there are 9 provincial capitals, 62 big cities, 75 medium cities and 23 small cities. It should be noted that here the “big” or “small” are not indicators of city sizes cities, but simply

translations of their administrative status.² These administrative levels exhibit a strong correlation with each city's degree of urbanization, total GDP and industrial composition. In other words, small cities may have a population and area larger than those of provincial capitals; however, they normally contain more rural and suburban lands. Controlling for these administrative levels may compensate for the lack of several socio-economic factors. Other factors controlled in the model include average income, population and roadway densities, yearly investment in transportation per capita, average vehicle ownership and vehicle type proportions (such as car versus light duty truck fractions).

Summary statistics for all four years' data are shown in Table 1. As can be seen, the data exhibit great variability. In addition, Figure 1 shows the locations of the cities, together with their total populations.

To ensure that a random-effects spatial SUR model is suitable and necessary for this sample data, a preliminary data check was carried out: For each year and each severity level, an OLS model was estimated and the regression errors analyzed. The error correlations existing between equations and across space are shown in Figure 2. The values for Moran's I (Cliff and Ord, 1972) are calculated based on the weight matrix described in the next section. These statistics suggest that there are potential correlations across equations and that the error terms of non-severe crash rates spatially cluster. Moreover, Figure 2 shows strong temporal correlations of the regression errors. All these features indicate that a spatial SUR model for these panel data is quite meaningful.

It should be mentioned that crash data everywhere suffer from underreporting issues, particularly for non-severe crashes (see, e.g., Blincoe et al. 2002). This may be especially severe in China, and rates of reporting may differ across police departments and thus cities. According to China's national standards (MPS, 1991), severe crashes are those in which there is at least one fatality, or three incapacitating injuries, or property damage of at least (approx.) \$4,000 U.S. dollars. However, the number of severe crashes and the number of fatalities in the data set are almost equal, though crashes with two or more deaths are far less likely than those with one death (so one would expect far fewer deaths than severe crashes). Apparently, the "severe" crashes in the dataset are very severe. Vehicle insurance was not mandatory during the survey period (though it is now); so reporting may have been less common for that reason. Unfortunately, there is no reliable data source for estimating a precise underreporting rate in this dataset. (Blincoe et al. (2002) suggest that underreporting rates range from 48% for property-damage-only crashes to 0% for fatal crashes in the U.S.)

In addition to underreporting issues, variable definitions can differ from those in use elsewhere. For example, vans are classified as mini-buses or buses in China, rather than as light-duty vehicles, or LDTs (as in the United States). In this dataset, LDTs are almost exclusively pickup trucks. The definition of motorcycles includes mopeds, and in some sampled cities, almost all "motorcycles" are actually mopeds. Additionally, the definition of "investment on transportation infrastructure construction" is not standardized across cases. Some cities include investment on traffic signals while others do not. This also obscures the definition of "investment on traffic management" because any non-construction investment in transportation is defined as an investment in transportation management. The standard definition of "arterial roads" is almost the same as the American definition, but a preliminary data check shows that most sampled cities

² A more formal translation is as follows: 1) Provincial capitals are capitals of autonomous regions and municipalities directly under the Central Government; 2) Big cities are those specifically designated in the State Plan, 3) Medium cities: are important at the regional level, and 4) Small cities are important at the county-level.

also count expressways and highways as arterial roads. Finally, average vehicle ownership is not collected from household census data, but from registered vehicle lists, which include all commercial and public transit vehicles. To the extent that some vehicles are not registered, average vehicle ownership values may be biased low.

RESULTS

In spatial econometrics, weight matrix elements W_g generally are monotonically decreasing functions of distance. To obtain the weight matrix that best fits the model, the first step of model estimation sought a proper function. Twelve functions, $d_{ij}^{-J/4}$ ($J=1, 2, \dots, 12$), were tested, where d_{ij} is the Euclidean distance. The resulting weight matrices were standardized (so that each row's values sum to one). Figure 3 shows log-likelihood values (i.e., Equation [16] plus constant terms) of the random-effects spatial SUR model resulting from these different weight-matrix functions. The function $d_{ij}^{-3/4}$ offers the maximum log-likelihood value and thus was selected.

Then, four models, ranging from a simple, but restrictive case to the most general case, were estimated and compared. The first model is an OLS model, implying that no correlations exist temporally or spatially, or across equations. The second model only considers temporal correlations by using a standard random-effects model (Greene, 2002). The third recognizes both temporal and cross-equation correlations but ignores spatial autocorrelation. The fourth allows both temporal and spatial correlations, as well as cross-equation correlation (across the two severity rates).

All four models were estimated using code programmed in Matlab. Their results are shown in Table 2. In this example, recognizing temporal correlations (model 2, for panel effects) most improves the model likelihood. This is followed by model 3's permission of correlations across equations. The recognition of spatial autocorrelation (model 4) also brings statistically significant improvement: the associated likelihood ratio (LR) test statistic is $-2(-277.461+268.171) = 18.58$, which exceeds the 95% critical value (of 5.99).

The absolute value of the log-likelihood ratio index (LRI) is not always meaningful for continuous models because there is a chance that the log-likelihood value is positive. However, because it provides a relation to a constants-only model, it is used here to compare the four models' performance. (Alternatively, readers can derive AIC based on the reported log-likelihood values.) Based on these LRI values, an OLS model is no better than a constants-only model, and the specification improves as more correlation patterns are allowed. Model 4 offers the greatest improvement from a constants-only model (and the best goodness of fit among these four models, as expected). In terms of parameter estimates, the OLS results generally fall near the other models' estimates. However, some OLS parameter estimates do differ quite a lot from those in the fourth, preferred model's results. The OLS-estimated coefficient for average vehicle ownership, for example, is more than double the final model's value. For several statistically insignificant variables, such as income for non-severe crash rates, the OLS model's estimate carries a sign opposed to that in the other three models. While not dramatic, the differences between models 2 through 4 are noticeable, especially the changes in t-statistics. Depending on the significance level needed for model selection, incorporation of spatial autocorrelation may lead to different conclusions regarding the effects of certain variables.

In considering the preferred model (model 4), the severe crash rate error term does not offer evidence of spatial correlation at the 90% level. However, non-severe crash rates exhibit statistically and practically positive correlations over space, with a coefficient (lambda) of 0.588.

This confirms the preliminary data check undertaken using Moran's I statistics, as shown in Figure 2.

It also can be observed that most explanatory variables are not statistically significant in the final model's estimates. This may result from the aggregate data, which obscures site-specific information, yielding greater uncertainty. Thus, more of a dependent variable's variation is explained by errors, or latent variables. Nevertheless, several factors do appear to have significant effects on crash rates:

Both crash rates are estimated to fall with population density, perhaps as travel distances fall (due to congestion and shorter trip lengths). However, this variable's effects are not estimated to be practically significant: *ceteris paribus*, every 14,000 more persons (approximately one standard deviation) per square kilometer (a 96% increase in average current densities) is associated with severe and non-severe crash rate reductions of just 0.07 and 0.24 (21% and 14% of current values), respectively

Investment in transportation infrastructure construction has a statistically significant effect on severe crash rates, with a negative coefficient of $-6.3E-4$. When evaluated at their means, the elasticity of investment in transportation infrastructure is only -0.078 . This means that every \$10 more in annual spending per capita (about 25% more of the current investment value and accounting for 1.6% of the sample's average per-capital income) is expected to result in 0.0063 fewer severe crashes each year, for every 1,000 persons, even though more infrastructure may mean greater car ownership and VMT. This reduction is about 1.9% of the current (severe) crash rates. Its effect on non-severe crash rates is estimated to be positive, but not statistically significant. Therefore, if the average cost of a severe crash were to be valued at \$1.6 million or more, increasing China's transportation investment may be expected to have net-positive safety effects. However, at present, the monetary value of life in China is felt to be much lower than \$1.6 million (Jin, 1999). Nevertheless, many feel that there exist other, economic benefits of transportation investment (e.g., Kim et al., 2004), which could tip the balance in favor of greater investment.

A city's reported percentage of arterial roads has a strong effect on severe crash rates. If a city's fraction of arterial roadways were to increase by 10%, average severe crash rates also are expected to rise by about 10%: from the current 0.336 to 0.367 (per 1,000 persons per year). This suggests that, in order to improve safety in this sample of Chinese cities, it may be best to construct more local streets and collectors, rather than higher-speed arterial roads. (Of course, the effect of this same variable may be quite different in studies of U.S. or other cities.)

Average vehicle ownership also is predicted to have significant positive effects on crash rates of both severity levels, as one would expect (since crash rates are per capita, rather than per VMT). Everything else constant, if average vehicle ownership were to increase by just over 500 percent, from the current 0.17 to 0.8 (the U.S. average), severe crash rates are predicted to increase by 0.154, or 50%. And non-severe crash rates are expected to increase by 1.145 (per 1,000 persons), or almost 70% of their current average³. These results suggest that motorization has a strong, but far from one-to-one impact on the traffic safety situation and residents' exposure to crash risk. Since China is now experiencing a rapid motorization process, through vehicle purchase and ownership, traffic safety agencies would do well to anticipate the best ways of avoiding serious death tolls.

³ Of course, a 500% increase in vehicle ownership is well outside the sample data range of values, so the model may not be appropriate for such extrapolation. This case is simply offered as an example.

CONCLUSIONS

This study specifies a spatial SUR model for analyzing panel data by incorporating random effects and spatial patterns in model error terms. A three-step estimation method, a mixture of FGLS and MLE methods, was used to estimate crash rates across Chinese cities. This model performs significantly better than models that do not recognize correlations across observations and across equations.

The study also reveals the effects of several influential factors. For example, it is estimated that population density decreases both types of crash rates and that investment decreases severe crash rates, but neither effects is practically significant. A higher fraction of arterial roads is associated with higher severe crash rates. Average vehicle ownership has a positive impact on crash rates (per capita), as one would expect, suggesting that the motorization of China may significantly increase traffic losses per capita in this rapidly developing country. Valuable model extensions are likely to include a non-linear specification, incorporation of spatial lag terms, and an ability to handle unbalanced panel data.

The methodology contributes to multivariate modeling when temporal and spatial effects exist. It is highly applicable to transportation issues in general, particularly those involving regional land use, travel, and demographics. Moreover, the empirical results for crash rates serve as a valuable reference for researchers interested in transportation issues in countries experiencing a rapid motorization process.

ACKNOWLEDGEMENTS

The Benjamin H. Stevens Graduate Fellowship in Regional Science provided financial support for this study. The authors are grateful for the Institute of Transportation Engineering at Tsinghua University's provision of data sets. Dr. Huapu Lu was generous in sharing data and resolving data issues. The authors also thank Annette Perrone for her administrative and editing assistance.

REFERENCES

- Aarts, L., and I. van Schagen. 2006. Driving speed and the risk of road crashes: A review. *Accident Analysis and Prevention* 38 (2):215-24.
- Anselin, L. 1988. A test for spatial auto-correlation in seemingly unrelated regressions. *Economics Letters* 28 (4):335-41.
- Anselin, L. 1999. *Spatial Econometrics*. [Accessed 2006 May, 10; http://www.csiss.org/learning_resources/content/papers/baltchap.pdf].
- Blincoe, L., A. Seay, E. Zaloshnja, T Miller., E. Romano, S. Luchter, and R. Sicer. 2002. *The Economic Impact of Motor Vehicle Crashes, 2000*. Washington, D.C: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Cliff, A. and J.K.Ord. 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4:267-84.
- Egger, P., and M. Pfaffermayr. 2004. Distance, trade and FDI: A Hausman-Taylor SUR approach. *Journal of Applied Econometrics* 19 (2):227-46.
- Elhorst, J. P. 2003. Specification and estimation of spatial panel data models. *International Regional Science Review* 26 (3):244-68.
- ESRI. 2006. *ArcGIS Desktop Help 9.1*. [Accessed 2006 May, 10; <http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?TopicName=welcome>]

- Federal Highway Administration. 1994-2004. Highway Statistics Publications, [Accessed 2006 May, 10; <http://www.fhwa.dot.gov/policy/ohpi/hss/hsspubs.htm>].
- Greene, W. H. 2002. *Econometric Analysis*. 5th ed. Upper Saddle River, New Jersey: Prentice Hall.
- Ivan, J. N. 2004. New approach for including traffic volumes in crash rate analysis and forecasting. In *Statistical Methods and Safety Data Analysis and Evaluation*.
- Jin, L. 1999. Value of life: No need to avoid the assessment. *Chongqing Environment Science* 4: 47-52.
- Kapoor, M., H. H. Kelejian, and I. R. Prucha. 2004. Panel data models with spatially correlated error components. [Accessed 2006 May, 10; http://www.isb.edu/faculty/Working_Papers_pdfs/Panel_Data_Models.pdf].
- Kelejian, H. H., and I. R. Prucha. 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* 118 (1-2):27-50.
- Kim, E., J.D.G. Hewings, and C. Hong. 2004. An Application of an Integrated Transport Network- Multiregional CGE Model: a Framework for the Economic Analysis of Highway Projects. *Economic Systems Research* 16(3): 235-258.
- Kweon, Y. J., and K. M. Kockelman. 2005. Safety effects of speed limit changes: Use of panel models, including speed, use, and design variables. *Transportation Research Record* No. 1908: 148-158.
- Su, J. 2005. Status analysis and countermeasures of road traffic safety in China. Presented at International Road Safety Seminar, Beijing, China. [Accessed June, 15 2006; <http://www.piarc.org/exec/link/library/download.htm?site=en&objectId=1247>].
- Magnus, J. R. 1982. Multivariate error-components analysis of linear and non-linear regression-models by maximum-likelihood. *Journal of Econometrics* 19 (2-3):239-285.
- Ministry of Public Security of the People's Republic of China. (MPS).1991. Notice on revised classification of traffic accidents. [Accessed 2006 May, 20; <http://www.szlaw.org/2004/6-14/16749.htm>].
- Noland, R. B. 2001. Relationships between highway capacity and induced vehicle travel. *Transportation Research* 35A (1): 47-72
- Noland, R. B., and M. G. Karlaftis. 2005. Sensitivity of crash models to alternative specifications. *Transportation Research* 41E (5):439-58.
- Qin, L., C. Shao, and Ying Wang. 2004. Analysis of freeway traffic accidents in China. *Proceedings of the Conference on Traffic and Transportation Studies, ICTTS* 4: 165-172.
- Srinivasan, S., and K. Kockelman. 2002. The Impacts of Bypasses on Small- and Medium-Sized Communities: An Econometric Analysis. *Journal of Transportation and Statistics* 5 (2):57-69.
- Yi, P., and B. Ran. 2003. Streamlining Chinese highway accident data acquisition, communications, and analysis. *Transportation Research Record* No. 1846:31-8.
- Zellner, A. 1962. An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *Journal of the American Statistical Association* 57:348-68.

LIST OF TABLES

TABLE 1 Summary Statistics

TABLE 2 Estimation Results for Different Models

LIST OF FIGURES

FIGURE 1 Locations and populations of sampled cities.

FIGURE 2 Regression errors with ordinary least squares (OLS) models.

FIGURE 3 Log-likelihood values with different weight matrix functions.

Table 1. Summary Statistics

Variable	Description	Min	Max	Mean	SD
SEV	Severe crash rate (per thousand persons per year)	0	6.444	0.3357	0.4961
NSEV	Non-severe crash rate (per thousand persons per year)	0	9.504	1.729	1.783
CAP	Indicator for provincial capital	0	1	0.0533	0.2247
BIG	Indicator for big city	0	1	0.3669	0.4823
MED	Indicator for medium city	0	1	0.4438	0.4972
SMALL	Indicator for small city (used as base condition for city type)	0	1	0.1361	0.3431
POPDENS	Population density (in thousands of persons per square kilometer)	2.75E-01	123.0	14.74	14.01
ROADDENS	Roadway density = total centerline length of roads/ developed area (km/km ²)	0.0967	16.77	4.956	3.083
ART_FRXN	Arterial roads fraction versus total road length	0.0464	0.858	0.3944	0.1796
CAR	Fraction of registered vehicles that are cars	4.69E-04	0.557	0.1370	0.1122
BUS	Fraction of registered vehicles that are buses	2.91E-04	0.249	0.0254	0.0273
LDT	Fraction of registered vehicles that are LDTs	0	0.348	0.0672	0.0589
HDT	Fraction of registered vehicles that are HDTs	0	0.380	0.0674	0.0532
MOTOR	Fraction of registered vehicles that are motorcycles (including mopeds)	1.12E-01	0.976	0.6278	0.1893
OTHER	Fraction of other types of registered motor vehicles (most likely to be farm vehicles, and used held as base vehicle type)	0	0.550	0.0751	0.0861
VEHOWN	Average vehicle ownership (veh/capita)	6.45E-03	0.949	0.1729	0.1513
INCOME	Annual personal income, in thousands of dollars (\$1000)	0.0975	8.197	0.7176	0.5804
INFINV	Yearly investment in transportation infrastructure construction (dollars per capita)	0	539.7	41.50	54.53
MGMINV	Yearly investment in transportation management (dollars per capita)	0	40.34	1.631	3.073

Table 2. Estimation Results for Different Models

	MODEL1 OLS				MODEL2 Panel-NonSUR-Nonspatial				MODEL3 Panel-SUR-Nonspatial				MODEL4 Panel-SUR-Spatial			
	Severe		Non-severe		Severe		Non-severe		Severe		Non-severe		Severe		Non-severe	
	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.	Coef.	t-stat.
CONS	2.30E-02	0.034	0.346	0.345	0.101	0.410	0.619	0.625	0.126	0.515	0.751	0.759	0.125	0.503	0.813	0.797
CAP	7.74E-03	0.016	0.651	1.311	4.69E-03	0.024	0.707	1.063	6.80E-03	0.034	0.707	1.059	-9.08E-03	-0.046	0.627	0.948
BIG	6.99E-02	0.245	0.218	0.729	6.80E-02	0.574	0.207	0.521	7.12E-02	0.602	0.210	0.526	6.43E-02	0.546	0.183	0.462
MED	0.105	0.384	0.355	1.277	0.105	0.916	0.370	0.977	0.106	0.928	0.372	0.977	9.84E-02	0.866	0.341	0.901
POPDENS	-4.50E-03	-0.645	-9.47E-03	-1.305	-5.00E-03	-1.712	-1.66E-02	-1.703	-4.89E-03	-1.682	-1.70E-02	-1.744	-4.98E-03	-1.720	-1.67E-02	-1.723
ROADDENS	9.99E-03	0.296	2.49E-02	0.717	1.13E-02	0.798	5.41E-02	1.150	1.07E-02	0.762	5.49E-02	1.163	9.59E-03	0.685	4.79E-02	1.024
ART_FRXN	0.328	0.609	0.212	0.388	0.329	1.455	0.284	0.379	0.326	1.447	0.281	0.374	0.306	1.363	0.202	0.270
CAR	0.159	0.188	0.445	0.338	0.130	0.452	0.728	0.579	0.096	0.336	0.646	0.514	0.117	0.406	0.560	0.445
BUS	0.303	0.186	-1.853	-0.558	0.374	0.704	1.381	0.555	0.279	0.528	1.331	0.536	0.306	0.576	0.920	0.372
LDT	6.58E-02	0.060	-1.259	-0.689	0.107	0.292	-0.669	-0.409	6.89E-02	0.189	-0.671	-0.412	6.15E-02	0.169	-1.007	-0.625
HDT	0.349	0.299	2.824	1.330	2.17E-02	0.057	-0.563	-0.319	1.79E-02	0.047	-0.972	-0.552	5.77E-02	0.151	-0.428	-0.247
MOTOR	7.25E-02	0.123	0.157	0.158	2.25E-02	0.115	0.615	0.692	-1.18E-02	-0.061	0.506	0.571	-4.20E-03	-0.022	0.393	0.449
VEHOWN	0.284	0.790	4.135	6.423	0.153	1.299	1.662	3.052	0.190	1.612	1.552	2.859	0.245	1.972	1.817	3.026
INCOME	-4.59E-03	-0.067	0.102	0.701	-1.90E-03	-0.087	-6.06E-02	-0.579	2.07E-05	0.001	-5.48E-02	-0.524	1.37E-03	0.063	-5.50E-02	-0.529
INFINV	-5.63E-04	-0.855	7.84E-04	0.504	-6.67E-04	-3.205	-1.82E-04	-0.179	-6.78E-04	-3.264	-3.31E-04	-0.326	-6.31E-04	-3.017	7.60E-05	0.075
MGMINV	1.44E-03	0.158	2.98E-03	0.118	1.97E-03	0.691	1.27E-02	0.897	1.79E-03	0.627	1.31E-02	0.924	2.35E-03	0.822	1.74E-02	1.246
Log-likelihood	-1007.821				-363.420				-277.461				-268.171			
LRI	0.084				0.670				0.748				0.756			
Lamda	0	---	0	---	0	---	0	---	0	---	0	---	0.239	1.173	0.588	4.697
Var (a)	0		0		0.216		2.269		0.214		2.288		0.212		2.249	
Cov(a)	0				0				0.318				0.304			
Var(e)	0.230		2.611		0.022		0.563		0.022		0.562		0.022		0.537	
Cov(e)	0				0				0.053				0.052			

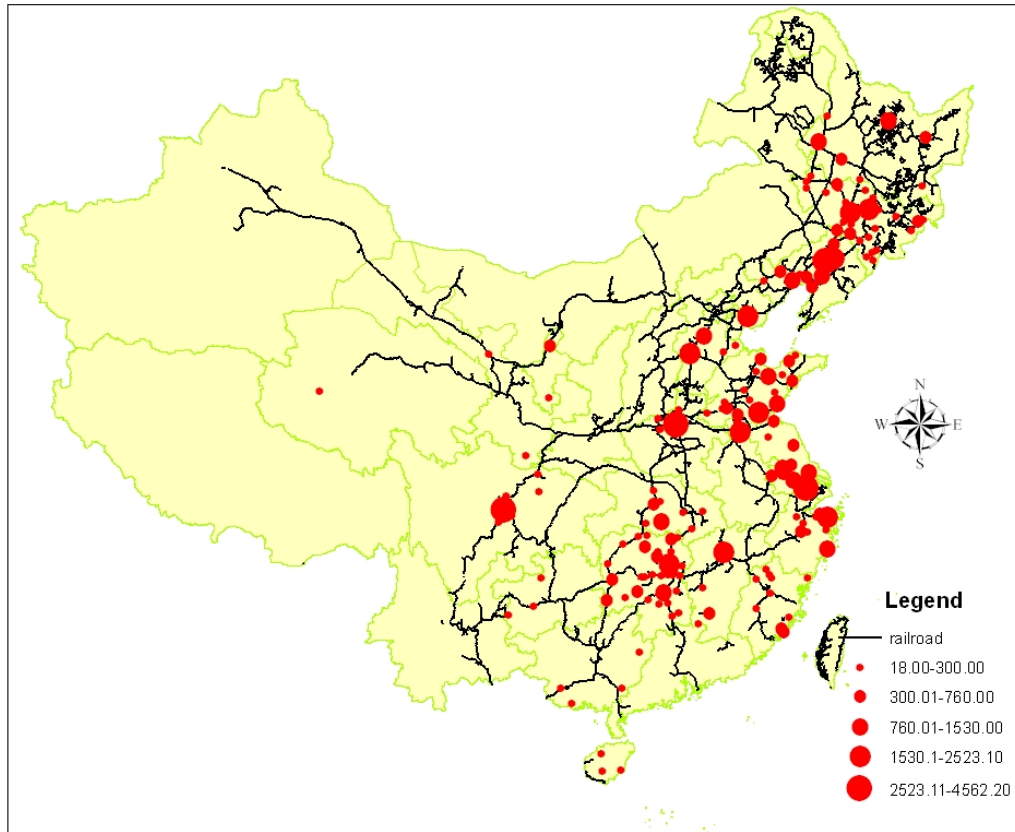


Figure 1. Locations and populations of sampled cities.

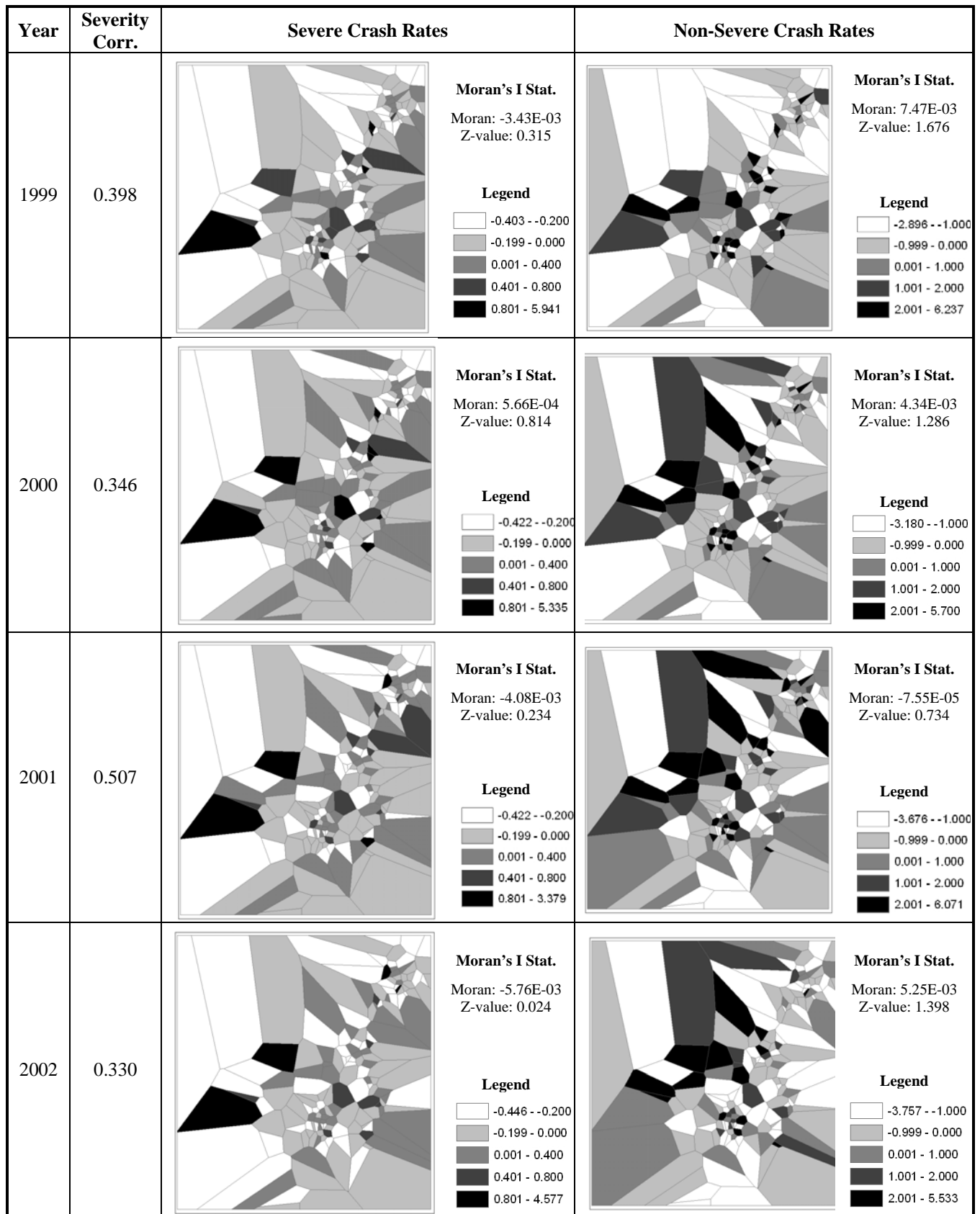


Figure 2. Regression errors with ordinary least squares (OLS) models.

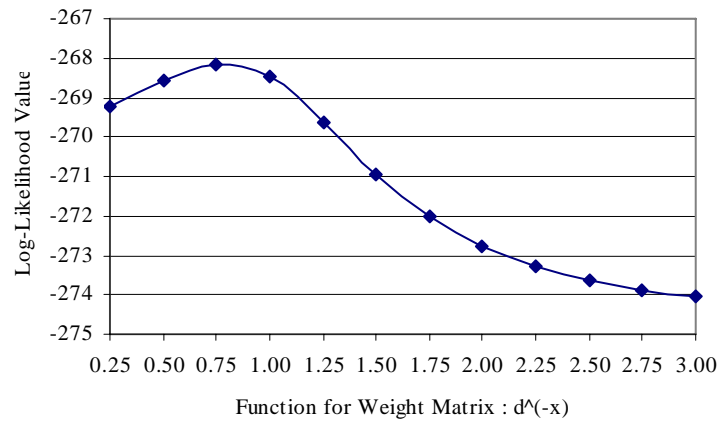


Figure 3. Log-likelihood values with different weight matrix functions.