GENERALIZED MAXIMUM ENTROPY ESTIMATION

OF A FIRST ORDER SPATIAL AUTOREGRESSIVE MODEL

by

Thomas L. Marsh

and

Ron C. Mittelhammer

Draft
June 2004

Associate Professor, School of Economic Sciences, Washington State University, Pullman, Washington 99164, tl_marsh@wsu.edu. Regents Professor, School of Economic Sciences, and Department of Statistics, Washington State University, Pullman, Washington 99164. We are indebted to Kelly Pace, James LeSage, and anonymous reviewers for insightful comments and recommended corrections. Any errors are solely the responsibility of the authors.

Abstract: We formulate generalized maximum entropy estimators for the general linear model and the censored regression model when there is first order spatial autoregression in the dependent variable. Monte Carlo experiments are provided to compare the performance of spatial entropy estimators relative to classical estimators. Finally, the estimators are applied to an illustrative model allocating agricultural disaster payments.

1.0 Introduction

In this paper we examine the use of generalized maximum entropy estimators for linear and censored regression models when the data generating process is afflicted by first order spatial autoregression in the dependent variable. Generalized maximum entropy (GME) estimators of regression models in the presence of spatial autocorrelation are of interest because they 1) offer a systematic way of incorporating prior information on parameters of the model,[1] 2) are straightforwardly applicable to non-normal error distributions,[2] and 3) are robust for ill-posed and ill-conditioned problems (Golan, Judge, and Miller 1996).[3] Prior information in the form of parameter restrictions arise naturally in the context of spatial models because spatial autoregressive coefficients are themselves inherently bounded. The development of estimators with finite sample justification across a wide range of sampling distributions and an investigation of their performance relative to established asymptotically justified estimators provides important insight and guidance to applied economists regarding model and estimator choice.[4]

Various econometric approaches have been proposed for accommodating spatial autoregression in linear regression models and in limited dependent variable models. In the case of the linear regression model, Cliff and Ord (1981) provide a useful introduction to spatial statistics. Anselin (1988) provides foundations for spatial effects in econometrics, discussing least squares, maximum likelihood, instrumental variable, and method of moment estimators to account for spatial correlation issues in the linear regression model. More recently, generalized two stage least squares and generalized moments estimators have been examined by Kelejian and Prucha (1998, 1999). Meanwhile, Lee (2002, 2003) examined asymptotic properties of least squares estimation for mixed regressive, spatial autoregressive models and two-stage least squares estimators for a spatial model with autoregressive disturbances.

In the case of the limited dependent variable model, most research has focused on the binary regression model and to a lesser extent the censored regression model. Besag (1972) introduced the auto-logistic model and motivated its use on plant diseases. The auto-logistic model incorporated spatial correlation into the logistic model by conditioning the probability of occurrence of disease on its presence in neighboring quadrants (see also Cressie 1991). Poirier and Ruud (1988) investigated a probit model with dependent observations and proved consistency and asymptotic normality of maximum likelihood estimates. McMillen (1992) illustrated the use of a spatial autoregressive probit model on urban crime data with an Expectation-Maximization (EM) algorithm. At the same time, Case (1992) examined regional influence on the adoption of agricultural technology by applying a variance normalizing transformation in maximum likelihood estimator to correct for spatial autocorrelation in a probit model. Marsh, Mittelhammer, and Huffaker (2000) also applied this approach to correct for spatial autocorrelation in a probit model by geographic region while examining an extensive data set pertaining to disease management in agriculture.

Bayesian estimation has also played an important role in spatial econometrics. LeSage (1997) proposed a Bayesian approach using Gibbs sampling to accommodate outliers and nonconstant variance within linear models. LeSage (2000) extended this to limited dependent variable models with spatial dependencies, while Smith and LeSage (2002) applied a Bayesian probit model with spatial dependencies to the 1996 presidential election results. Although Bayesian estimation is well-suited for representing uncertainty with respect to model parameters, it can also require extensive Monte Carlo sampling when numerical estimation techniques are required, as is often the case in non-normal, non-conjugate prior model contexts. In comparison, GME estimation also enforces restrictions on parameter values, is arguably no more difficult to specify, and does not require the use of Monte Carlo sampling in the estimation phase of the analysis.[5]

The principle of maximum entropy has been applied in a variety of modeling contexts, including applications to limited dependent variable models. However, to date, GME or other information theoretic estimators have not been applied to spatial regression models.[6] Golan, Judge, and Miller (1996, 1997) proposed estimation of both the general linear model and the censored regression model based on the principle of generalized maximum entropy in order to deal with small samples or ill-posed problems. Adkins (1997) investigated properties of a GME estimator of the binary choice model using Monte Carlo analysis. Golan, Judge, and Perloff (1996) applied maximum entropy to recover information from multinomial response data, while Golan, Judge, and Perloff (1997) recovered information with censored and ordered multinomial response data using generalized maximum entropy. Golan, Judge, and Zen (2001) proposed entropy estimators for a censored demand system with nonnegativity constraints and provided asymptotic results. These studies provide the basic foundation from which we define spatial entropy estimators for the general linear model and the censored regression model when there is first order spatial autoregression in the dependent variable.

The current paper proceeds as follows. First, we motivate GME estimation of the general linear model (GLM) and then investigate generalizations to spatial GME-GLM estimators. Second, Monte Carlo experiments are provided to benchmark the mean squared error loss of the spatial GME-GLM estimators relative to ordinary least squares (OLS) and maximum likelihood (ML) estimators. We also examine the sensitivity of the spatial GME estimators to user-supplied supports and their performance across a range of spatial autoregressive coefficients. Third, Golan, Judge, and Perloff's (1997) GME estimator of the censored regression model (i.e., Tobit) is extended to a spatial GME-Tobit estimator, and additional Monte Carlo experiments are presented to investigate the sampling properties of the method. Finally, the spatial entropy GLM and Tobit approaches are applied empirically to the estimation of a simultaneous Tobit model of agricultural disaster payment allocations across political regions.

2.0 Spatial GME-GLM Estimator

2.1 Data Constrained GME-GLM

Following the maximum entropy principle, the entropy of a distribution of probabilities $\mathbf{p} = (p_1, ..., p_M)'$, $\sum_{m=1}^{M} p_m = 1$, is defined by $H(\mathbf{p}) = -\mathbf{p}' \ln \mathbf{p} = -\sum_{m=1}^{M} p_m \ln p_m$ (Shannon, 1948). The value of $H(\mathbf{p})$, which is a measure of the uncertainty in the distribution of probabilities, reaches a maximum when $p_{\mathrm{m}} = M^{-1}$ for m=1,...,M characterizing the uniform distribution. Generalizations of the entropy function that have been examined elsewhere in the econometrics and statistics literature include the Cressie-Read power divergence statistic (Imbens et al., 1998), Kullback-Leibler Information Criterion (Kullback, 1959), and the α-entropy measure (Pompe, 1994). For example, the well known Kullback-Leibler cross-entropy extension, $-\sum_{m=1}^{M} p_m \ln(p_m/q_m)$, is a discrepancy measure between distributions **p** and **q** where **q** is a reference distribution. In the event that the reference distribution is uniform, then the maximum entropy and cross-entropy functions coincide. We restrict our analysis to the maximum entropy objective function due to its efficiency and robustness properties (Imbens et al., 1998), and its current universal use within the context of GME estimation applications (Golan, Judge, and Miller 1996).

To motivate the maximum entropy estimator, it is informative to revisit the least squares estimator.[7] Consider the general linear model

(1) $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with $\mathbf{Y}$ a $N \times 1$ dependent variable vector, $\mathbf{X}$ a $N \times K$ matrix of explanatory variables, $\boldsymbol{\beta}$ a $K \times 1$ vector of parameters, and $\boldsymbol{\varepsilon}$ a $N \times 1$ vector of disturbance terms. The standard least squares optimization problem is to $\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{N} \varepsilon_i^2 \text{ subject to } Y_i - X_{i.}\boldsymbol{\beta} = \varepsilon_i, \forall i \right\}$. The objective is to minimize the quadratic sum of squares function for $\boldsymbol{\beta} \in R^K$ subject to the data constraint in (1).

There are fundamental differences between the least squares and maximum entropy approaches. First, the maximum entropy approach is based on the entropy objective function $H(\mathbf{p})$ instead of the quadratic sum of squares objective function. Instead of minimizing the sum of squared errors, the entropy approach selects the $\mathbf{p}$ closest to the uniform distribution given the data constraint in (1).[8] Second, the maximum entropy approach provides a means of formalizing more adhoc methods researchers have commonly employed to impose a priori restrictions on regression parameters and ranges of disturbance outcomes. To do so the unknown parameter vector $\boldsymbol{\beta}$ is reparameterized as $\beta_k = \sum_{j=1}^{J} s_{kj}^{\boldsymbol{\beta}} p_{kj}^{\boldsymbol{\beta}} \in \left[s_{k1}^{\boldsymbol{\beta}}, s_{kJ}^{\boldsymbol{\beta}} \right]$ onto a user defined discrete support space $s_{k1}^{\boldsymbol{\beta}} \le s_{k2}^{\boldsymbol{\beta}} \le \cdots \le s_{kJ}^{\boldsymbol{\beta}}$ for $J \ge 2$ with a $(J \times 1)$ vector of unknown weights $\mathbf{p}_k^{\beta} = \left(p_{k1}^{\beta}, \ldots p_{kJ}^{\beta}\right)'$ $\forall k = 1,\ldots,K$. The discrete support space includes the lower truncation point $s_{k1}^{\boldsymbol{\beta}}$, the upper truncation point $s_{kJ}^{\boldsymbol{\beta}}$, and *J*-2 remaining intermediate support points. For instance, consider a discrete support space with *J*=2 support points $\left\{s_{k1}^{\boldsymbol{\beta}}, s_{kJ}^{\boldsymbol{\beta}}\right\} = \{-1,1\}$ for β_k that has only lower and upper truncation points and allows no intermediate support points. The reparameterized expression yields $\beta_k = (1 - p_{k2}^{\beta})(-1) + p_{k2}^{\beta}(1)$ with a single unknown p_{k2}^{β}.[9] Likewise the unknown error vector $\boldsymbol{\varepsilon}$ is reparameterized as $\varepsilon_i = \sum_{m=1}^{M} s_{im}^{\boldsymbol{\varepsilon}} p_{im}^{\boldsymbol{\varepsilon}} \in \left[s_{i1}^{\boldsymbol{\varepsilon}}, s_{iM}^{\boldsymbol{\varepsilon}}\right]$ such that $s_{i1}^{\boldsymbol{\varepsilon}} \le s_{i2}^{\boldsymbol{\varepsilon}} \le \cdots \le s_{iM}^{\boldsymbol{\varepsilon}}$ with a $(M \times 1)$ vector of unknown weights $\mathbf{p}_i^{\varepsilon} = \left(p_{i1}^{\varepsilon}, \ldots p_{iM}^{\varepsilon}\right)'$ $\forall i = 1,\ldots,N$.[10] In practice, discrete support spaces for both the parameters and errors are supplied by the user based on economic or econometric theory or other prior information. Third, unlike least squares, there is a bias-efficiency tradeoff that arises in GME when parameter support spaces are specified in terms of bounded intervals. A disadvantage of bounded intervals is that they will generally introduce bias into the GME estimator for finite samples unless the intervals happen to be centered on the true values of the parameters. An advantage of restricting parameters to finite intervals is that they can lead to increases in efficiency by lowering parameter estimation variability. The underlying idea is that the bias introduced by bounded parameter intervals in the GME estimator can be more-than compensated for by substantial decreases in variability, leading to notable increases in overall estimation efficiency.

The data constrained GME estimator of the general linear model (hereafter GME-

D) is defined by the following constrained maximum entropy problem (Golan, Judge, and Miller 1996):[11]

(2a)
$$\max_{\mathbf{p}} \left\{ -(\mathbf{p})' \ln(\mathbf{p}) \right\}$$

subject to

(2b)
$$\mathbf{Y} = \mathbf{X}\left(\mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}\right) + \left(\mathbf{S}^{\boldsymbol{\varepsilon}}\mathbf{p}^{\boldsymbol{\varepsilon}}\right)$$

(2c)
$$\mathbf{1}'\mathbf{p}_{\mathbf{k}}^{\boldsymbol{\beta}} = 1 \ \forall k, \mathbf{1}'\mathbf{p}_{\mathbf{i}}^{\boldsymbol{\varepsilon}} = 1 \ \forall i$$

(2d)
$$\mathbf{p} = vec\left(\mathbf{p}^{\boldsymbol{\beta}}, \mathbf{p}^{\boldsymbol{\varepsilon}}\right) > [\mathbf{0}]$$

In matrix notation, the unknown parameter vector $\boldsymbol{\beta}$ and error vector $\boldsymbol{\varepsilon}$ are reparameterized as $\boldsymbol{\beta} = \mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}$ and $\boldsymbol{\varepsilon} = \mathbf{S}^{\boldsymbol{\varepsilon}}\mathbf{p}^{\boldsymbol{\varepsilon}}$ from known matrices of user supplied discrete support points $\mathbf{S}^{\boldsymbol{\beta}}$ and $\mathbf{S}^{\boldsymbol{\varepsilon}}$ and an unknown $(KJ + NM) \times 1$ vector of weights $\mathbf{p} = vec\left(\mathbf{p}^{\boldsymbol{\beta}}, \mathbf{p}^{\boldsymbol{\varepsilon}}\right)$. The $KJ \times 1$ vector $\mathbf{p}^{\boldsymbol{\beta}} = vec\left(\mathbf{p}_{\mathbf{1}}^{\boldsymbol{\beta}}, ..., \mathbf{p}_{\mathbf{K}}^{\boldsymbol{\beta}}\right)$ and the $NM \times 1$ vector $\mathbf{p}^{\boldsymbol{\varepsilon}} = vec\left(\mathbf{p}_1^{\boldsymbol{\varepsilon}}, ..., \mathbf{p}_N^{\boldsymbol{\varepsilon}}\right)$ consist of $J \times 1$ vectors $\mathbf{p}_{\mathbf{k}}^{\boldsymbol{\beta}}$ and $M \times 1$ vectors $\mathbf{p}_{\mathbf{i}}^{\boldsymbol{\varepsilon}}$, each having nonnegative elements summing to unity. The matrices $\mathbf{S}^{\boldsymbol{\beta}}$ and $\mathbf{S}^{\boldsymbol{\varepsilon}}$ are $K \times KJ$ and $N \times NM$ block-diagonal matrices of support points for the unknown $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ vectors. For example, consider the support matrix for the β vector,

(3)
$$\mathbf{S}^{\boldsymbol{\beta}} = \begin{pmatrix} \left(\mathbf{s}_1^{\boldsymbol{\beta}}\right)' & 0 & \cdots & 0 \\ 0 & \left(\mathbf{s}_2^{\boldsymbol{\beta}}\right)' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\mathbf{s}_K^{\boldsymbol{\beta}}\right)' \end{pmatrix}$$

Here, $\mathbf{s}_k^{\boldsymbol{\beta}} = \left(s_{k1}^{\boldsymbol{\beta}}, ..., s_{kJ}^{\boldsymbol{\beta}}\right)'$ is a $J \times 1$ vector such that $s_{k1}^{\boldsymbol{\beta}} \le s_{k2}^{\boldsymbol{\beta}} \le \cdots \le s_{kJ}^{\boldsymbol{\beta}}$ where

$\beta_k = \sum_{j=1}^{J} s_{kj}^{\boldsymbol{\beta}} p_{kj}^{\boldsymbol{\beta}} \in \left[s_{k1}^{\boldsymbol{\beta}}, s_{kJ}^{\boldsymbol{\beta}} \right] \; \forall k = 1,...,K$. Given this reparameterization, the empirical distribution of estimated weights $\hat{\mathbf{p}}$ (and subsequently $\hat{\boldsymbol{\beta}}$) are determined by the entropy objective function subject to constraints of the model and user supplied supports.

The choice of support points $\mathbf{S}^{\varepsilon}$ depends inherently on the properties of the underlying error distribution. In most but not all circumstances, error supports have been specified to be symmetric and centered about the origin. Excessively wide truncation points reflect uncertainty about information in the data constraint and correspond to solutions $\hat{\mathbf{p}}$ that are more uniform, implying the $\hat{\beta}_k$ approach the average of the support points. Given ignorance regarding the error distribution, Golan, Judge, and Miller (1996) suggest calculating a sample scale parameter and using the three-sigma rule to determine error bounds. The three-sigma rule for random variables states that the probability for a unimodal random variable falling away from its mean by more than three standard deviations is at most 5% (Vysochanskii and Petunin 1980; Pukelsheim 1994). The three-sigma rule is a special case of Vysochanskii and Petunin's bound for unimodal distributions. Letting Y be a real random variable with mean μ and variance σ^2, the bound is given by

$$\Pr\left(\left|Y-\mu\right| \geq r\right) \leq \frac{4}{9}\frac{\sigma^2}{r^2}$$

where $r>0$ is the half length of an interval centered at μ. For $r = 3\sigma$ it yields the three-sigma rule and more than halves the Chebyshev bound. In more general terms the above bound can yield a j-sigma rule with $r = j\sigma$ for $j=\{1,2,3, \ldots\}$.

The generalized maximum entropy formulation in (2) incorporates inherently a dual entropy loss function that balances estimation precision in coefficient estimates and predictive accuracy subject to data, adding up, and nonnegativity constraints.[12] The specification of (2) leads to first order optimization conditions that are different from the standard least squares estimator with the notable difference that the first order conditions for GME-D do not require orthogonality between right hand side variables and error terms.[13] Mittelhammer and Cardell (1998) have provided regularity conditions, derived the first order conditions, asymptotic properties, and asymptotic test statistics for the data-constrained GME-D estimator. They also identified a more computationally efficient approach with which to solve entropy based problems that does not expand with sample size and provides the basis for the optimization algorithms used in the current study.

2.2 Spatial GME Estimators

The first order spatial autoregressive model can be expressed as (see Cliff and Ord 1981; Anselin 1988)

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{4}$$

where **W** is a $N \times N$ spatial proximity matrix structuring the lagged dependent variable. For instance, the elements of the proximity matrix $\mathbf{W}=\{w_{ij}^*\}$ may be defined as a standardized joins matrix where $w_{ij}^* = w_{ij} \big/ \sum_j w_{ij}$ with w_{ij}=1 if observations i and j are from an adjoining spatial region (for $i \neq j$) and w_{ij}=0 otherwise. In (4), **u** is a ($N\times 1$) vector of iid error terms, while ρ is an unknown scalar spatial autoregressive parameter to be estimated.

In general, the ordinary least squares (OLS) estimator applied to (4) will be inconsistent. The maximum likelihood estimator of (4), for the case of normally distributed errors, is discussed in Anselin (1988). Consistent generalized two stage least squares and generalized method of moments estimators of (4) are discussed in Kelejian and Prucha (1998, 1999). Lee (2002, 2003) examined consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models and investigated best two-stage least squares estimators for a spatial model with autoregressive disturbances.

In the subsections below, we introduce both moment and data constrained GME estimators of the spatial regression model. Doing so enables us to investigate the finite sample performance of the two estimators relative to one another and relative to other competing estimators.

2.2.1 Normalized Moment Constrained Estimator

Consider the normalized (by sample size) moment constraint relating to the GLM given by

(5)
$$\mathbf{X}'\left[\left(\mathbf{I}-\rho\mathbf{W}\right)\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right]\big/N=\mathbf{X}'\mathbf{u}/N$$

The spatial GME method for defining the estimator of the unknown parameters $\boldsymbol{\beta}$ and ρ in the spatial autoregressive model using the normalized moment information in (5) (hereafter GME-N) is represented by the following constrained maximum entropy problem:

(6a)
$$\max_{\mathbf{p}}\left\{-\left(\mathbf{p}\right)'\ln\left(\mathbf{p}\right)\right\}$$

subject to

(6b)
$$\mathbf{X}'\left[\left(\mathbf{I}-\left(\mathbf{S}^{\rho}\mathbf{p}^{\rho}\right)\mathbf{W}\right)\mathbf{Y}-\mathbf{X}\left(\mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}\right)\right]\big/N=\left(\mathbf{S}^{\mathbf{u}}\mathbf{p}^{\mathbf{u}}\right)$$

(6c)
$$\mathbf{1}'\mathbf{p}_{\mathbf{k}}^{\boldsymbol{\beta}}=1\ \forall k,\mathbf{1}'\mathbf{p}^{\rho}=1,\mathbf{1}'\mathbf{p}_{\mathbf{i}}^{\mathbf{u}}=1\ \forall i$$

(6d)
$$\mathbf{p}=vec\left(\mathbf{p}^{\boldsymbol{\beta}},\mathbf{p}^{\rho},\mathbf{p}^{\mathbf{u}}\right)>[\mathbf{0}]$$

where $\mathbf{p} = vec\left(\mathbf{p}^{\boldsymbol{\beta}}, \mathbf{p}^{\rho}, \mathbf{p}^{\mathbf{u}}\right)$ is a $(KJ + J + NM) \times 1$ vector of unknown parameters. In (6), we define $\boldsymbol{\beta} = \mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}$ as before and $\mathbf{X}'\mathbf{u}/N = \mathbf{S}^{\mathbf{u}}\mathbf{p}^{\mathbf{u}}$. The additional term $\mathbf{S}^{\rho}$ is a $1 \times J$ vector of support points for the spatial autoregressive coefficient ρ. The $J \times 1$ vector $\mathbf{p}^{\rho}$ is an unknown weight vector having nonnegative elements that sum to unity and are used to reparameterize the spatial autoregressive coefficient as $\rho = \mathbf{S}^{\rho}\mathbf{p}^{\rho} = \sum_{j=1}^{J} s_j^{\rho} p_j^{\rho}$. The lower and upper truncation points can be selected to bound ρ and additional support points can be specified to recover higher moment information about it. The GME-N estimator when $\rho = 0$ is defined in Golan, Judge, and Miller (1996, 1997) along with its asymptotic properties.[14]

2.2.2 Data Constrained Estimator

As an alternative way of accounting for simultaneity in (4), we follow Theil (1971) and Zellner (1998) by specifying a data constraint for use in the GME method as

(7) $$\mathbf{Y} = (\rho\mathbf{W})(\mathbf{Z}\boldsymbol{\pi}) + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}^*$$

where $\mathbf{u}^*$ is a $N \times 1$ vector of appropriately transformed residuals. The above expression is derived using the spatial model in (4) and from substitution of an unrestricted reduced form equation based on

$$\mathbf{Y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u}$$

which can be expressed as

$$\mathbf{Y} = \lim_{t \to \infty} \sum_{j=0}^{t} (\rho\mathbf{W})^j \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u}, \quad -1<\rho<1$$

for eigenvalues of **W** less than or equal to one in absolute value (Kelejian and Prucha 1998). The reduced form model can be approximated by

(8) $$\mathbf{Y} \approx \sum_{j=0}^{t} (\rho\mathbf{W})^j \mathbf{X}\boldsymbol{\beta} + \mathbf{v} = \mathbf{Z}\boldsymbol{\pi} + \mathbf{v}$$

where $\mathbf{Z}\boldsymbol{\pi} = \sum_{j=0}^{t} (\mathbf{W}^j\mathbf{X})(\rho^j\boldsymbol{\beta})$ is a partial sum of the infinite series for some upper index value *t*. Here, **Z** can be interpreted as a $N \times L$ matrix of instruments consisting of $\{\mathbf{X}, \mathbf{WX}, ..., \mathbf{W}^t\mathbf{X}\}$, $\boldsymbol{\pi}$ is a $L \times 1$ vector of unknown and unrestricted parameters, and **v** is a $N \times 1$ vector of reduced form residuals. Zellner (1998) refers to equation (7) as a nonlinear in parameters (NLP) form of the simultaneous equations model.

The spatial GME method for defining the estimator of the unknown parameters $\boldsymbol{\beta}, \boldsymbol{\pi},$ and ρ in the combined models (7) and (8) (hereafter GME-NLP) is represented by the following constrained maximum entropy problem:

(9a) $$\max_{\mathbf{p}} \left\{ -(\mathbf{p})' \ln(\mathbf{p}) \right\}$$

subject to

(9b) $$\mathbf{Y} = \left[(\mathbf{S}^{\rho}\mathbf{p}^{\rho}) \mathbf{W} \right] \left[\mathbf{Z}(\mathbf{S}^{\pi}\mathbf{p}^{\pi}) \right] + \mathbf{X}(\mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}) + (\mathbf{S}^{\mathbf{u}^*}\mathbf{p}^{\mathbf{u}^*})$$

(9c) $$\mathbf{Y} = \mathbf{Z}(\mathbf{S}^{\pi}\mathbf{p}^{\pi}) + (\mathbf{S}^{v}\mathbf{p}^{v})$$

(9d) $$\mathbf{1}'\mathbf{p}^{\boldsymbol{\beta}}_{\mathbf{k}} = 1 \ \forall k, \mathbf{1}'\mathbf{p}^{\pi}_{\ell} = 1 \ \forall \ell, \mathbf{1}'\mathbf{p}^{\rho} = 1, \mathbf{1}'\mathbf{p}^{\mathbf{u}^*}_{\mathbf{i}} = 1 \ \forall i, \mathbf{1}'\mathbf{p}^{\mathbf{v}}_{\mathbf{i}} = 1 \ \forall i,$$

(9e) $$\mathbf{p} = vec\left(\mathbf{p}^{\boldsymbol{\beta}}, \mathbf{p}^{\pi}, \mathbf{p}^{\rho}, \mathbf{p}^{\mathbf{u}^*}, \mathbf{p}^{\mathbf{v}}\right) > [\mathbf{0}]$$

where $\mathbf{p} = vec\left(\mathbf{p}^{\boldsymbol{\beta}}, \mathbf{p}^{\pi}, \mathbf{p}^{\rho},\ \mathbf{p}^{\mathbf{u}^*}, \mathbf{p}^{\mathbf{v}}\right)$ is a $(KJ + LJ + J + 2NM) \times 1$ vector of unknown parameters. Including the reduced form model in (9c) is necessary to identify the reduced form parameters. In effect this is a one-step estimator in which the reduced and structural form parameters, as well as the spatial correlation coefficients, are estimated concurrently.[15]

2.3 *Monte Carlo Experiments - Spatial GLM*

The linear component $\mathbf{X\beta}$ is specified as

$$\mathbf{X\beta} = [1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4]\begin{bmatrix} 2 \\ 1 \\ -1 \\ 3 \end{bmatrix} \tag{10}$$

where the values of x_{i2}, $i = 1,...,N$ are iid outcomes from Bernoulli(.5), and the values of the pair of explanatory variables x_{i3} and x_{i4} are generated as iid outcomes from

$$N\left(\begin{pmatrix} \mu_3 \\ \mu_4 \end{pmatrix}, \begin{pmatrix} \sigma^2_{x_3} & \sigma_{x_3,x_4} \\ \sigma_{x_3,x_4} & \sigma^2_{x_4} \end{pmatrix}\right) = N\left(\begin{pmatrix} 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right) \tag{11}$$

which are then truncated at ±3 standard deviations. The disturbance terms u_i^* are drawn iid from a $N(0,\sigma^2)$ distribution that is truncated at ±3 standard deviations and $\sigma^2 = 1$. Thus, the true support of the disturbance distribution in this Monte Carlo experiment is truncated normal, with lower and upper truncation points located at -3 and +3, respectively.

Additional information in the form of user supplied supports for the structural and reduced form parameters and the error terms must be provided for the GME estimators. Supports for the structural parameters are specified as $\mathbf{s}_k^{\boldsymbol{\beta}} = (-20,0,20)'$, $\forall k$. Supports of the reduced form model are $\mathbf{s}_1^{\pi} = (-100,0,100)'$ for the intercept and $\mathbf{s}_l^{\pi} = (-20,0,20)'$, $\forall l > 1$. Error supports for the GME estimator are defined using the *j*-sigma rule.

Specifically, in the experiments for the general linear model the supports are

$\mathbf{s}_i^{\mathbf{u}^*} = \mathbf{s}_i^{\mathbf{v}} = \left(-j\hat{\sigma}_y, 0, j\hat{\sigma}_y\right)' \; \forall i$ where $\hat{\sigma}_y$ is the sample standard deviation of the dependent variable with either j=3 defining a 3-sigma rule or j=5 defining a 5-sigma rule.

Next we specify the model assumptions for the spatial components (i.e., proximity matrices and spatial autoregressive coefficients with supports) of the regression model. Seven true spatial autoregressive coefficients $\mathbf{\Xi} = \{-0.75, -0.50, -0.25, 0, 0.25, 0.50, 0.75\}$ were selected bounded between -1 and 1 for the experiments. Supports for the spatial autoregressive coefficients are specified by using narrower eigenvalue bounds as $\mathbf{s}^{\rho} = \left(1/\xi_{\min}, 0, 1/\xi_{\max}\right)'$ and wider bounds $\mathbf{s}^{\rho} = \left(-20, 0, 20\right)'$ coinciding with supports on the structural parameters.[16] The spatial weight matrix **W** is constructed using the Harrison and Rubinfeld (1978) example discussed in Gilley and Pace (1996). This example is based on the Boston SMSA with 506 observations (one observation per census tract). The elements of the weight matrix are defined as $w_{ij} = \max\left\{1 - \frac{d_{ij}}{d_{\max}}, 0\right\}$ where d_{ij} is the distance in miles (converted from the latitude and longitude for each observation). A nearest neighbor weight matrix was constructed with elements given the value 1 for observations within $d_{\max}$ and 0 otherwise. The weight matrix was then row normalized bounding its eigenvalues ξ between -1 and 1. The reduced form model in (8) is approximated at t=1 with a matrix of instrumental variables **Z** defined by $\{\mathbf{X}, \mathbf{WX}\}$.

Table 1 presents the assumptions underlying twenty-one experiments that compare the performance of GME to the maximum likelihood (ML) estimator under normality of the GLM with a first order spatial autoregressive dependent variable.[17]

Experiments 1-7 maintain support $\mathbf{s}^{\rho} = \left(1/\xi_{min}, 0, 1/\xi_{max}\right)'$ for the autoregressive parameters in $\boldsymbol{\Xi}$ with the 3-sigma rule bounding the error terms. Experiments 8-14 deviate from the first seven by widening supports on the spatial autoregressive parameter to $\mathbf{s}^{\rho} = \left(-20, 0, 20\right)'$, while retaining the 3-sigma rule. Experiments 15-21 maintain support $\mathbf{s}^{\rho} = \left(1/\xi_{min}, 0, 1/\xi_{max}\right)'$ on the spatial autoregressive parameter, but expand the error supports to the 5-sigma rule. Experiments 8-14 and experiments 15-21 test the sensitivity of the GME estimators to the support spaces on the spatial autoregressive parameter and the error terms, respectively.

Other pertinent details are that all of the Monte Carlo results are based on 500 replications with *N*=506 observations consistent with the dimension of the weighting matrix **W**. It is noteworthy to point out that convergence of the GME spatial estimator [using the general objective function optimizer OPTMUM in GAUSS (Aptech Systems Inc.)] occurred within a matter of seconds for each replication across all the Monte Carlo experiments. Appendix A discusses computational issues and provides derivations of the gradient and Hessian for the GME-NLP estimator in (9). For completeness, we also include performance measures of OLS in the Monte Carlo analysis.

2.3.1 *Experiment Results*

Table 2 contains the mean squared error loss (MSEL) from experiments 1-21 for: (a) the regression coefficients $\boldsymbol{\beta}$ from the spatial ML and GME estimators and the OLS estimator and (b) the spatial autoregressive coefficient ρ estimated by the spatial ML and

GME estimators. Tables 3-7 report the estimated average bias (e.g., $\text{bias}\left(\hat{\boldsymbol{\beta}}\right) = \text{E}[\hat{\boldsymbol{\beta}}]\text{-}\boldsymbol{\beta}$) and variance (e.g., $\text{var}\left(\hat{\beta}_k\right)$) of $\beta_1, \beta_2, \beta_3, \beta_4$, and ρ, respectively.

Consider the MSEL for regression coefficients $\boldsymbol{\beta}$. In cases with nonzero spatial autoregressive parameters, both GME-NLP and ML tend to exhibit smaller MSEL than either OLS or GME-N (with OLS having the largest MSEL values). Except for GME-NLP, the MSEL values are larger for positive than negative spatial autoregressive parameters. Meanwhile, ML had the smallest MSEL values for negative spatial autoregressive parameters equal to -0.75 and -0.50. For GME-NLP, widening the autoregressive support space in experiments 8-14 decreased (increased) the MSEL relative to experiments 1-7 for negative (positive) spatial values in $\boldsymbol{\Xi}$. Experiments 15-21, with the expanded error supports for GME-NLP, exhibited increased MSEL relative to experiments 1-7 across nonzero spatial autoregressive values in $\boldsymbol{\Xi}$. At $\rho = 0$, the MSEL is lowest for OLS followed by GME-NLP, ML and GME-N, respectively.

Turning to the MSEL for the spatial autoregressive parameter ρ, both GME-NLP and ML exhibited smaller MSEL than GME-N. ML had the smallest MSEL values for spatial autoregressive parameters -0.75 and -0.50. In experiments 1-7 GME-NLP had smaller MSEL than ML except for -0.75 and -0.50. Widening the error support space in experiments 15-21 tended to increase the MSEL of the spatial autoregressive parameter for GME-NLP.

Next, we compare the estimated average bias and variance for ML and GME-NLP across $\beta_1, \beta_2, \beta_3, \beta_4$, and ρ in tables 3-7, respectively. The obvious differences between ML and GME-NLP arise for the intercept (Table 3). In most cases, ML exhibited smaller absolute bias and larger variance on the intercept coefficient relative to GME-NLP. In tables 4-6 there are no obvious systematic differences in the average bias and variance for the remaining parameters. Turning to the spatial autoregressive parameter (Table 7), ML tended to have smaller absolute bias and larger variance relative to GME-NLP for nonzero values of ρ. Narrower support spaces for the spatial autoregressive parameter in experiments 1-7 and 15-21 yielded positive (negative) bias for negative (positive) spatial values in $\boldsymbol{\Xi}$. Comparing across the experiments, performance of the GME-NLP estimator exhibited sensitivity to the different parameter and error support spaces.

Overall, the Monte Carlo results for the spatial regression model indicate that the data constrained GME-NLP estimator dominated the normalized moment constrained GME-N estimator in MSEL. As a result, we focus on the data constrained GME-NLP and not the moment constrained GME-N estimator when investigating the censored Tobit model.

3.0 Spatial GME Tobit Estimator

3.1 GME-Tobit Model

Consider a Tobit model with censoring of the dependent variable at zero

$$(12) \qquad \begin{cases} Y_i = Y_i^* & \text{if } \mathrm{Y}_i^* = \mathbf{X}_{\mathrm{i.}}\boldsymbol{\beta} + \varepsilon_i > 0 \\ Y_i = 0 & \text{if } \mathrm{Y}_i^* = \mathbf{X}_{\mathrm{i.}}\boldsymbol{\beta} + \varepsilon_i \le 0 \end{cases}$$

where Y_i^* is an unobserved latent variable. Traditional estimation procedures are discussed in Judge et al (1988). Golan, Judge, and Perloff (1997) reorder the observations and rewrite (12) in matrix form as

$$(13) \qquad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\varepsilon}_1 > \mathbf{0} \\ \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}_2 \le \mathbf{0} \end{bmatrix}$$

where the subscript 1 indexes the observations associated with N_1 positive elements, the subscript 2 indexes the observations associated with N_2 zero elements of **Y**, and $N_1 + N_2 = N$. In the GME method, the estimator of the unknown $\boldsymbol{\beta}$ in the Tobit model formulation is given by $\boldsymbol{\beta} = \mathbf{S}^{\beta}\mathbf{p}^{\beta}$, where $\mathbf{p}^{\beta} = vec\left(\mathbf{p}_1^{\beta},...,\mathbf{p}_K^{\beta}\right)$ is a $KJ \times 1$ vector, and $\boldsymbol{\varepsilon} = \mathbf{S}^{\varepsilon}\mathbf{p}^{\varepsilon}$, where $\mathbf{p}^{\varepsilon} = vec\left(\mathbf{p}^{\varepsilon_1}, \mathbf{p}^{\varepsilon_2}\right)$ is a $NM \times 1$ vector, with both $\mathbf{p}^{\beta}$ and $\mathbf{p}^{\varepsilon}$ being derived from the following constrained maximum entropy problem:

(14a) $$\max_{\mathbf{p}} \left\{ -\left(\mathbf{p}\right)' \ln\left(\mathbf{p}\right) \right\}$$

subject to

(14b) $$\begin{cases} \mathbf{Y}_1 = \mathbf{X}_1\left(\mathbf{S}^{\beta}\mathbf{p}^{\beta}\right) + \mathbf{S}^{\varepsilon_1}\mathbf{p}^{\varepsilon_1} \\ \mathbf{0} \geq \mathbf{X}_2\left(\mathbf{S}^{\beta}\mathbf{p}^{\beta}\right) + \mathbf{S}^{\varepsilon_2}\mathbf{p}^{\varepsilon_2} \end{cases}$$

(14c) $$\mathbf{1}'\mathbf{p}_{\mathbf{k}}^{\beta} = 1,\ \mathbf{1}'\mathbf{p}_{\mathbf{i}}^{\varepsilon_1} = 1,\ \mathbf{1}'\mathbf{p}_{\mathbf{i}}^{\varepsilon_2} = 1$$

(14d) $$\mathbf{p} = vec\left(\mathbf{p}^{\beta}, \mathbf{p}^{\varepsilon_1}, \mathbf{p}^{\varepsilon_2}\right) > [\mathbf{0}]$$

where $\mathbf{p} = vec\left(\mathbf{p}^{\beta}, \mathbf{p}^{\varepsilon_1}, \mathbf{p}^{\varepsilon_2}\right)$ is a $(KJ + NM) \times 1$ vector of unknown support weights. Under general regularity conditions, Golan, Judge, and Perloff (Proposition 4.1, 1997) demonstrate that the GME-Tobit model is a consistent estimator.

3.2 Spatial GME-Tobit Model

Assume a Tobit model with censoring of the dependent variable at zero and first order spatial autoregressive process in the dependent variable as

(15) $$\begin{cases} Y_i = Y_i^* & \text{if } \mathrm{Y}_i^* = \left[\left(\rho\mathbf{W}\right)\mathbf{Y}^*\right]_i + \mathbf{X}_{i.}\boldsymbol{\beta} + u_i > 0 \\ Y_i = 0 & \text{if } \mathrm{Y}_i^* = \left[\left(\rho\mathbf{W}\right)\mathbf{Y}^*\right]_i + \mathbf{X}_{i.}\boldsymbol{\beta} + u_i \leq 0 \end{cases}$$

where $\mathbf{X}_i$ denotes the ith row of the matrix $\mathbf{X}$ and $\left[(\rho\mathbf{W})\mathbf{Y}^*\right]_i$ denotes the ith row of $(\rho\mathbf{W})\mathbf{Y}^*$. If $\rho \neq 0$ then this is a Tobit model with a first order spatial autoregressive process in the dependent variable. Equation (15) reduces to the standard Tobit formulation in (14) if $\rho = 0$.

A spatial GME approach for defining the spatial Tobit estimator of the unknown parameters $\boldsymbol{\beta}, \pi,$ and ρ of the spatial autoregressive model in (15) is represented by:

(16a)
$$\max_{\mathbf{p}} \left\{-(\mathbf{p})' \ln(\mathbf{p})\right\}$$

subject to

(16b)
$$\begin{cases} \mathbf{Y}_1 = \left[\left(\left(\mathbf{S}^{\rho}\mathbf{p}^{\rho}\right)\mathbf{W}\right)\mathbf{Z}\right]_1 \left(\mathbf{S}^{\pi}\mathbf{p}^{\pi}\right) + \mathbf{X}_1\left(\mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}\right) + \left(\mathbf{S}^{\mathbf{u}_1^*}\mathbf{p}^{\mathbf{u}_1^*}\right) \\ \mathbf{0} \geq \left[\left(\left(\mathbf{S}^{\rho}\mathbf{p}^{\rho}\right)\mathbf{W}\right)\mathbf{Z}\right]_2 \left(\mathbf{S}^{\pi}\mathbf{p}^{\pi}\right) + \mathbf{X}_2\left(\mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}\right) + \left(\mathbf{S}^{\mathbf{u}_2^*}\mathbf{p}^{\mathbf{u}_2^*}\right) \end{cases}$$

(16c)
$$\mathbf{Y} = \mathbf{Z}\left(\mathbf{S}^{\pi}\mathbf{p}^{\pi}\right) + \left(\mathbf{S}^{v}\mathbf{p}^{v}\right)$$

(16d)
$$\mathbf{1}'\mathbf{p}_k^{\beta} = 1\ \forall k, \mathbf{1}'\mathbf{p}_k^{\pi} = 1\ \forall k, \mathbf{1}'\mathbf{p}^{\rho} = 1, \mathbf{1}'\mathbf{p}_i^{u^*} = 1\ \forall i, \mathbf{1}'\mathbf{p}_i^{v} = 1\ \forall i$$

(16e)
$$\mathbf{p} = vec\left(\mathbf{p}^{\boldsymbol{\beta}}, \mathbf{p}^{\pi}, \mathbf{p}^{\rho}, \mathbf{p}^{\mathbf{u}^*}, \mathbf{p}^{\mathbf{v}}\right) > [\mathbf{0}]$$

where subscripts 1 and 2 represent partitioned submatrices or vectors corresponding to the ordering of $\mathbf{Y}$ discussed in (13). Apart from the censoring and reordering of the data, the spatial Tobit estimator in (16) has the same structural components as the estimator in (9).

3.3 Iterative Spatial GME-Tobit Model

Breiman et. al. (1993) discussed iterative least squares estimation of the censored regression model that coincides with ML and the EM algorithm under normality, but the method does not necessarily coincide with ML or EM with nonnormal errors. Each iteration involves two steps: an expectation step and a re-estimation step.[18] Following this approach, Golan, Judge, and Perloff (1997) suggested using the initial estimates from optimizing (14) defined by $\hat{\boldsymbol{\beta}}^{(0)} = \mathbf{S}^{\boldsymbol{\beta}}\hat{\mathbf{p}}^{\boldsymbol{\beta}(0)}$ to predict $\hat{\mathbf{Y}}_2^{(1)}$ and then redefine $\mathbf{Y}^{(1)} = vec\left(\mathbf{Y}_1, \hat{\mathbf{Y}}_2^{(1)}\right)$ in re-estimating and updating $\hat{\boldsymbol{\beta}}^{(1)} = \mathbf{S}^{\boldsymbol{\beta}}\hat{\mathbf{p}}^{\boldsymbol{\beta}(1)}$.[19] In the empirical exercises below, we follow this approach to obtain the *i*th iterated estimate $\hat{\boldsymbol{\beta}}^{(i)} = \mathbf{S}^{\boldsymbol{\beta}}\hat{\mathbf{p}}^{\boldsymbol{\beta}(i)}$ of the spatial GME-Tobit model in equation (16).

3.4 *Monte Carlo Experiments - Spatial Tobit Model*

The sampling experiments for the Tobit model follow closely those in Golan, Judge, and Perloff (1997) and Paarsch (1984). The explanatory variables and coefficients of the model are defined as

(17)
$$\mathbf{X\beta} = \left[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\right]\begin{bmatrix} 2 \\ 1 \\ -3 \\ 2 \end{bmatrix}$$

where the x_{il}, $i = 1,\ldots,N$, $l = 1,\ldots,4$ are generated iid from $N(\bar{x},1)$ for $\bar{x} = 0$ and 2.

Increasing the mean reduced the percent of censored observations from approximately 50% across the sampling experiments for $\bar{x} = 0$ to less than 20% for $\bar{x} = 2$.[20] The disturbance terms u_i are drawn iid from a $N(0,1)$ distribution.

The structural and reduced form error supports for the GME estimator are defined using a variation of Pukelsheim's 3-sigma rule. Here, the supports are defined as $\left(-3\hat{\sigma}_y, 0, 3\hat{\sigma}_y\right)'$ where $\hat{\sigma}_y = \sqrt{(y_{\max} - \hat{y}_{\min})/12}$ (Golan, Judge, and Perloff 1997). Supports for the spatial autoregressive coefficient are specified as

$\mathbf{s}^{\rho} = \left(1/\xi_{\min}, 0, 1/\xi_{\max}\right)'$. Supports for the structural parameters are specified as

$\mathbf{s}_k^{\boldsymbol{\beta}} = \left(-20, 0, 20\right)'$, $\forall k$. Supports of the reduced form model are $\mathbf{s}_1^{\pi} = \left(-100, 0, 100\right)'$ for

the intercept and $\mathbf{s}_l^{\pi} = \left(-20, 0, 20\right)'$, $\forall l > 1$.

The Monte Carlo results are based on 500 replications and the true spatial autoregressive parameters are specified as $\rho = -0.5$ and 0.5. As before, the spatial weight matrix **W** is constructed using the Harrison and Rubinfeld (1978) example with *N*=506 observations and the reduced form model in (8) is approximated at *t*=1.

3.4.1 ***Experiment Results***

Table 8 reports the MSEL for the regression coefficients $\boldsymbol{\beta}$ and the spatial autoregressive coefficient ρ for the standard ML estimator of the Tobit model (ML-Tobit), as well as the non-iterative (GME-NLP) and iterative (IGME-NLP) spatial entropy estimators.[21] Tables 9-12 provide estimated average bias and variance of the regression coefficients $\boldsymbol{\beta}$ for each estimator. Table 13 reports estimated average bias and variance for the spatial autoregressive coefficient ρ from the non-iterative and iterative spatial GME estimators.

Evidence from the Monte Carlo simulations indicates that ML-Tobit and IGME-NLP estimator outperformed the non-iterative GME-NLP estimator in MSEL for the regression coefficients $\boldsymbol{\beta}$. When the percent of censored observations increased from less than 20% (experiments 2 and 4) to approximately 50% (experiments 1 and 3), ML-Tobit tends to outperform IGME-NLP in MSEL. Alternatively, when the censoring is less than 20% (experiments 2 and 4), then IGME-NLP outperforms ML-Tobit. Regarding the spatial autoregressive parameter, the IGME-NLP estimator outperformed the non-iterative GME-NLP estimator in MSEL.

Finally, in terms of average bias and variance estimates in tables 9-12, in most cases the absolute bias for regression coefficients $\boldsymbol{\beta}$ is less for IGME-NLP than ML-Tobit with censoring less than 20% (experiments 2 and 4). Both GME estimators exhibited positive bias in the estimated spatial coefficients across the four experiments (Table 13). Interestingly, the IGME-NLP estimator reduced the absolute bias and variance relative to the non-iterative GME-NLP estimator.

4.0 Illustrative Application: Allocating Agricultural Disaster Payments

Agricultural disaster relief in the U.S. has commonly taken one of three forms - emergency loans, crop insurance, and direct disaster payments (U.S. GAO). Of these, direct disaster payments are considered the least efficient form of disaster relief (Goodwin and Smith, 1995).

Direct disaster payments from the government provide cash payments to producers who suffer catastrophic losses, and are managed through the USDA's Farm Service Agency (FSA). The bulk of direct disaster funding is used to reimburse producers for crop and feed losses rather than livestock losses. Direct disaster payments approached $30 billion during the 1990s FSA, by far the largest of the three disaster relief programs. Unlike the crop insurance program which farmers use to manage their risk, it is usually legislators who decide whether or not a direct payment should be made to individual farmers after a disaster occurs.

The amount of disaster relief available through emergency loans and crop insurance is determined by contract, whereas direct disaster relief is determined solely by legislators, and only after a disaster occurs. Politics thus plays a much larger role in determining the amount of direct disaster relief than it does with emergency loans and crop insurance. Legislators from a specific state find it politically harmful not to subsidize farmers who experienced a disaster, given the presence of organized agriculture interest groups within that state (Becker, 1983; Gardner 1987).

4.1 Modeling Disaster Relief

Several important econometric issues arise in allocating agricultural disaster payments. First, there is potential simultaneity between disaster relief and crop insurance payments. Second, and more importantly for current purposes, regional influences of natural disasters and subsequent political allocations may have persistent spatial correlation effects across states. Ignoring either econometric issue can lead to biased and inconsistent estimates and faulty policy recommendations.

Consider the following simultaneous equations model with spatial components:

(18) $$\mathbf{Y}_d = (\rho\mathbf{W})\mathbf{Y}_d + \delta\mathbf{Y}_c + \mathbf{X}_d\boldsymbol{\beta} + \mathbf{u}$$

(19) $$\mathbf{Y}_c = \mathbf{Z}\pi_c + \mathbf{v}_c$$

(20) $$\mathbf{Y}_d = \mathbf{Z}\pi_d + \mathbf{v}_d$$

where the dependent variable $\mathbf{Y}_d$ denotes disaster payments, and is censored because some states do not receive any direct agriculture disaster relief in certain years (in the context of the Tobit model, $\mathbf{Y}_d = 0$ if $\mathbf{Y}_d^* \leq 0$ and $\mathbf{Y}_d = \mathbf{Y}_d^*$ if $\mathbf{Y}_d^* > 0$). In (18), $\mathbf{Y}_c$ denotes crop insurance payments (non-censored) and $\mathbf{X}_d$ are exogenous variables including measures of precipitation, political interest group variables, year specific dummy variables, and number of farms. In the reduced form models (19) and (20), the exogenous variables include per capita personal income, farm income, the number of farm acres, total crop values, geographical census region, income measures, year specific dummy variables, and number of farms and political factors.[22] The parameters to be estimated are δ and $\boldsymbol{\beta}$ structural coefficients, as well the reduced form coefficients π_c and π_d.

The data were collected from the FSA and other sources. A complete list and description of all direct disaster relief programs are available through the FSA. The FSA data set maintains individual farmer transactions of all agricultural disaster payments in the U.S. For the purposes of the current study, FSA aggregated the transactions across programs and individuals to obtain annual levels of disaster payments for each of the 48 contiguous states from 1992 to 1999. A list of selected variables and definitions are provided in Table 14. For this application the elements of the proximity matrices for each time period $\mathbf{W} = \left\{ w_{ij}^{*} \right\}$ are defined as a standardized joins matrix where $w_{ij}^{*} = w_{ij} \Big/ \sum_{j} w_{ij}$ with w_{ij}=1 if observations i and j (for $i \neq j$) are from adjoining states (e.g., Kansas and Nebraska) and w_{ij}=0 otherwise (e.g., Kansas and Washington). To account for the time series cross-sectional nature of the data, the full proximity matrix used in modeling all of the observed data was defined as a block diagonal matrix such that $\mathbf{W}_1 = \left(\mathbf{I}_T \otimes \mathbf{W} \right)$ where $\mathbf{W}$ is the joins matrix defined above and $\mathbf{I}_T$ is an $T \times T$ identity matrix with T=8 representing the 8 years of available data.

The analysis proceeded in several steps. First, to simplify the estimation process and focus on the spatial components of the model, predicted crop insurance values were obtained from the reduced form model. Then predicted values crop insurance values were used in the disaster relief model.[23] Second, equations (18) and (20) were jointly estimated with the iterative GME-NLP estimator. For further comparison, the model was estimated with ML-Tobit and the spatial ML estimator under normality.

Supports were specified as $\mathbf{s}_i^{\beta} = \mathbf{s}_j^{\pi} = (-1000, 0, 1000)'$ for the structural and reduced form parameters and $\mathbf{s}^{\rho} = \left(1/\hat{\xi}_{\min}, 0, 1/\hat{\xi}_{\max}\right)'$ for the spatial parameter with estimated eigenvalues $\hat{\xi}_{\min} = -0.7$ and $\hat{\xi}_{\max} = 1$. Effectively, structural coefficient supports were selected to allow political coefficients to range between -$1 billion to $1 billion. Because of the inherent unpredictability and political nature of disaster allocations, circumstances arose in the data that yielded relatively large residuals for selected states in specific years. Rather than removing outlying observations, we chose to expand the error supports to a 5-sigma rule.

Table 15 presents structural and spatial coefficient estimates and asymptotic t-values for the estimates of the disaster relief model for ML-Tobit, ML estimator under normality, and GME-NLP. Results from the latter two estimators demonstrate that the spatial autoregressive coefficient is significant and positive. The structural parameters of the disaster relief model are consistent with the findings of Garrett, Marsh, and Marshall (2003), who applied a maximum likelihood estimator of the simultaneous Tobit model and discuss implications of results.[24]

5.0 Conclusions

In this paper we extended generalized maximum entropy (GME) estimators for the general linear model (GLM) and the censored Tobit model to the case where there is first order spatial autoregression present in the dependent variable. Monte Carlo experiments were conducted to analyze finite sample size behavior and estimators were compared using mean squared error loss (MSEL) of the regression coefficients for the spatial general linear model and censored regression model. For spatial autoregression in the dependent variable of the standard regression model, the data constrained GME estimator outperformed a specific normalized moment constrained GME estimator in MSEL. In most cases, maximum likelihood (ML) exhibited smaller absolute bias and larger variance relative to GME. However, across the experiments, the GME estimators exhibited sensitivity to the different parameter and error support spaces. For the censored regression model, when there is first order spatial autoregression in the dependent variable, the iterative data constrained GME estimator exhibited reduced absolute bias and variance relative to the non-iterative GME estimator. Interestingly, the relative performance of GME was sensitive to the degree of censoring for the given experiments. When censoring was less than 20%, the iterative GME estimator outperformed ML-Tobit in MSEL. Alternatively, when the percent of censored observations increased to approximately 50%, ML-Tobit tended to outperform iterative GME in MSEL. Finally, we provided an illustrative application of the spatial GME estimators in the context of a model allocating agricultural disaster payments using a simultaneous Tobit framework.

The GME estimators defined in this paper provides a conceptually new approach to estimating spatial regression models with parameter restrictions imposed. The method is computationally efficient and robust for given support points. Future research could include examining spatial autoregressive errors, other spatial GME estimators (e.g., alternative moment constraints), further analysis of the impacts of user supplied support points, or further assessments of the relative benefits of iterative versus non-iterative GME estimators. Overall, the results suggest that continued investigation of GME estimators for spatial autoregressive models could yield additional findings and insight useful to applied economists.

References

Adkins, L. 1997. "A Monte Carlo Study of a Generalized Maximum Entropy Estimator of the Binary Choice Model." In *Advances in Econometrics: Applying Maximum Entropy to Econometric Problems*, Volume 12, edited by T. B. Fomby and R. C. Hill. Greenwich, Connecticut: JAI Press Inc., 183-200.

Aptech Systems, Inc. 1996. GAUSS: Optimization Application Module. Maple Valley, Washington.

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

Becker, Gary S. 1983. "A Theory of Competition Among Pressure Groups for Political Influence." *Quarterly Journal of Economics* 98: 371-400.

Besag, J. E. 1972. "Nearest-Neighbour Systems and the Auto-logistic Model for Binary Data." *Journal of the Royal Statistical Society*, Ser. B, 34, 75-83.

Breiman, L., Y. Tsur, and A. Zemel. 1993. "On a Simple Estimation Procedure for Censored Regression Moodels with Known Error Distributions," *Annals of Statistics*, 21:1711-1720.

Case, A. C. 1992. "Neighborhood Influence and Technological Change." *Regional Science and Urban Economics*, 22: 491-508.

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

Cressie, N. A. C. 1991. *Statistics for Spatial Data*. New York: John Wiley & Sons.

Fraser, I. 2000. "An Application of Maximum Entropy Estimation: The Demand For Meat in the United Kingdom." *Applied Economics* 32: 45-59.

Gardner, Bruce L. 1987. "Causes of Farm Commodity Programs." *Journal of Political Economy* 95: 290-310.

Garrett, T. A., T. L. Marsh, and M.I. Marshall. 2003. "Political Allocation of Agriculture Disaster Payments in the 1990s." Forthcoming *International Review of Law and Economics*.

Garrett, T. A. and R. S. Sobel. 2003. "The Political Economy of FEMA Disaster Payments," *Economic Inquiry*.

Gilley, O.W., and R. Kelley Pace. 1996. "On the Harrison and Rubinfeld Data," *Journal of Environmental Economics and Management*, 31: 403-405.

Golan, A., Judge, G. G. and Miller, D., 1996. *Maximum Entropy Econometrics: Robust Information with Limited Data*. New York: John Wiley and Sons.

Golan, A., Judge, G. G. and Perloff, J. 1996. “A Maximum Entropy Approach to Recovering Information from Multinomial Response Data.” *Journal of the American Statistical Association*, 91:841-853.

Golan, A., Judge, G. G. and Miller, D. 1997. “The Maximum Entropy Approach to Estimation and Inference: An Overview.” In *Advances in Econometrics: Applying Maximum Entropy to Econometric Problems*, Volume 12, edited by T. B. Fomby and R. C. Hill. Greenwich, Connecticut: JAI Press Inc., 3-24.

Golan, A., Judge, G. G. and Perloff, J. 1997. “Estimation and inference with censored and ordered multinomial data.” *Journal of Econometrics*, 79: 23-51.

Golan, A., Judge, G. G. and Zen, E.Z. 2001. “Estimating a Demand System with Nonnegativity Contraints: Mexican Meat Demand.” *The Review of Economics and Statistics*, 83: 541-550.

Goodwin, Barry K. and Vincent H. Smith. 1995. *The Economics of Crop Insurance and Disaster Aid.* The AEI Press, Washington, D.C.

Harrison, David, and Daniel L. Rubinfeld. 1978. "Hedonic Housing Prices and the Demand for Clean Air," *Journal of Environmental Economics and Management*, 5:81-102.

Imbens G, Spady R, Johnson P. 1998. Information Theoretic Approaches to Inference in Moment Condition Models. *Econometrica* 66: 333-57.

Judge, G. G., Hill, R. C., William, E. G., Lutkepohl, H., and Lee, T. 1988. *Introduction to the Theory and Practice of Econometrics*. John Wiley and Sons, New York.

Kelejian, H. H. and I. R. Prucha. 1998. “A Generalized Spatial Two-Stage Least Sqaures Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Distrubances,” *Journal of Real Estate Finance and Economics*, 17:99-121.

Kelejian, H. H. and I. R. Prucha. 1999. “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, 40: 509-533.

Kullback S. 1959. *Information Theory and Statistics*. John Wiley & Sons: New York.

Lee, L. 2002. “Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models,” *Econometric Theory* 18: 252-277.

Lee, L. 2003. "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances," *Econometric Reviews* 22: 307-335.

LeSage, J. P. 1997. "Bayesian estimation of Spatial Autoregressive Models," *International Regional Science Review* 20:113-129.

LeSage, J. P. 2000. "Bayesian estimation of Limited Dependent Spatial Autoregressive Models," *Geographical Analysis* 32:19-35.

Malinvaud E. 1980. *Statistical Methods of Econometrics*. Third Ed, North-Holland: Amsterdam.

Marsh, T. L., R. C. Mittelhammer, and N. S. Cardell. 2003. "A Generalized Maximum Entropy Estimator of the Simultaneous Linear Statistical Model." Working paper, Washington State University.

Marsh, T. L., R. C. Mittelhammer, and R. G. Huffaker. 2000. "Probit with Spatial Correlation by Plot: PLRV Net Necrosis." *Journal of Agricultural, Biological, and Environmental Statistics*, 5: 22-36.

McMillen, D. P. 1992. "Probit with Spatial Correlation." *Journal of Regional Science*, 32: 335-348.

McLachlan, G. J. and T. Krishnan. 1997. *The EM Algorithm and Extensions*. New York: John Wiley & Sons.

Mittelhammer, R. C. and N. S. Cardell. 1998. "The Data-Constrained GME Estimator of the GLM: Asymptotic Theory and Inference." Mimeo, Washington State University.

Mittelhammer, R., Judge, G. and Miller, D., 2000. *Econometric Foundations*. New York: Cambridge University Press.

Paarsch, H. J. 1984. "A Monte Carlo Comparison of Estimators for Censored Regression Models." *Journal of Econometrics* 24:197-213.

Preckel, P. V. 2001. "Least Squares and Entropy: A Penalty Perspective," *American Journal of Agricultural Economics*, 83: 366-367.

Poirier, D. J. and P. A. Ruud. 1988. "Probit with Dependent Observations." *The Review of Economic Studies*, 55:593-614.

Pompe B. 1994. On Some Entropy Measures in Data Analysis. *Chaos, Solitons, and Fractals* 4: 83-96.

Pukelsheim, F. 1994. "The Three Sigma Rule." *The American Statistician* 48: 88-91.

Shannon CE. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27.

Smith, T. E. and LeSage, J. P. 2002. "A Bayesian Probit Model with Spatial Dependencies," Working Paper.

Smith, Richard J. and Richard W. Blundell. 1986. "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply." *Econometrica* 54: 679-685.

Theil H. 1971. *Principles of Econometrics*. John Wiley & Sons: New York.

Vysocahnskii, D. F. and Petunin, Y. I. 1980. "Justification of the 3σ Rule for Unimodal Distributions," *Theory of Probability and Mathematical Statistics*, 21: 25-36.

Zellner A. 1994. Bayesian and non-bayesian estimation using balanced loss functions. In *Statistical Decision Theory and Related Topics*, Gupta S, Berger J (eds). Springer Verlag: New York.

Zellner A. 1998. The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators Bayesian and Non-Bayesian Approaches. *Journal of Econometrics* 83: 185-212.

Table 1. Monte Carlo Experiments for Spatial Regression Model.

Experiment	$j\hat{\sigma}_y$	True ρ	$\mathbf{s}^{\rho\prime}$
1	3	-0.75	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
2	3	-0.5	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
3	3	-0.25	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
4	3	0.00	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
5	3	0.25	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
6	3	0.50	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
7	3	0.75	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
8	3	-0.75	(-20, 0, 20)
9	3	-0.5	(-20, 0, 20)
10	3	-0.25	(-20, 0, 20)
11	3	0.00	(-20, 0, 20)
12	3	0.25	(-20, 0, 20)
13	3	0.50	(-20, 0, 20)
14	3	0.75	(-20, 0, 20)
15	5	-0.75	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
16	5	-0.5	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
17	5	-0.25	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
18	5	0.00	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
19	5	0.25	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
20	5	0.50	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$
21	5	0.75	$\left(\xi_{\min}^{-1},0,\xi_{\max}^{-1}\right)$

Table 2. MSEL Results for OLS and Spatial ML and GME Estimators.

					MSEL($\hat{\boldsymbol{\beta}}$)			
True ρ	OLS	ML	GME-NLP	GME-N	GME-NLP	GME-N	GME-NLP	GME-N
-0.75	46.85445	0.73342	3.89496	7.86198	1.27085	12.41126	10.81546	8.15715
-0.50	27.66546	0.94232	2.62515	7.46430	1.12270	12.62844	6.32048	7.72633
-0.25	9.94484	1.35587	1.20896	7.13155	1.20733	12.89162	1.94408	7.34293
0.00	0.06413	1.87715	0.67812	7.12143	1.44093	13.20860	0.12045	7.21240
0.25	26.64269	2.99540	1.80136	8.20153	2.24008	13.58534	1.88375	7.84772
0.50	236.98749	4.06597	3.13564	12.89220	4.74866	14.01641	5.25398	10.37058
0.75	2127.57748	8.97558	3.09602	24.41890	18.19564	14.49347	4.39171	12.68538
Exp.			1-7	1-7	8-14	8-14	15-21	15-21

				MSEL($\hat{\rho}$)			
True ρ	ML	GME-NLP	GME-N	GME-NLP	GME-N	GME-NLP	GME-N
-0.75	0.00775	0.04582	0.39984	0.01486	2.23989	0.13166	0.39945
-0.50	0.00816	0.02348	0.14877	0.00989	1.69301	0.05685	0.14845
-0.25	0.00838	0.00730	0.01995	0.00761	1.21968	0.01162	0.01977
0.00	0.00744	0.00266	0.00977	0.00579	0.81743	0.00048	0.01003
0.25	0.00673	0.00416	0.10799	0.00532	0.48493	0.00477	0.11016
0.50	0.00417	0.00339	0.27891	0.00509	0.22847	0.00608	0.29199
0.75	0.00236	0.00085	0.37947	0.00488	0.06030	0.00129	0.46872
Exp.		1-7	1-7	8-14	8-14	15-21	15-21

Table 3. Bias and Variance Estimates for Spatial ML and GME Estimators of β_1.

True ρ	Bias				Variance			
	ML	GME-NLP	GME-NLP	GME-NLP	ML	GME-NLP	GME-NLP	GME-NLP
-0.75	0.08108	-1.78936	0.11754	-1.82650	0.71281	0.67862	1.24219	0.79184
-0.50	0.00365	-1.45051	-0.29581	-1.36924	0.92817	0.50703	1.02128	0.52741
-0.25	0.07222	-0.83051	-0.41072	-0.76081	1.33798	0.50662	1.02452	0.51563
0.00	0.22323	0.05655	-0.59280	0.00105	1.81573	0.66343	1.07579	0.64257
0.25	0.42844	0.89485	-0.90295	1.18259	2.79831	0.98712	1.41168	0.47163
0.50	0.41828	1.27590	-1.69444	2.19136	3.87680	1.49351	1.86405	0.43703
0.75	1.00396	0.30056	-3.94049	1.95238	7.95493	2.99275	2.65450	0.56532
Exp.		1-7	8-14	15-21		1-7	8-14	15-21

Table 4. Bias and Variance Estimates for Spatial ML and GME Estimators of β_2.

True ρ	Bias				Variance			
	ML	GME-NLP	GME-NLP	GME-NLP	ML	GME-NLP	GME-NLP	GME-NLP
-0.75	-0.00153	0.00195	-0.00256	-0.00505	0.00853	0.00874	0.00884	0.00790
-0.50	0.00137	0.00268	-0.00129	-0.00968	0.00858	0.00849	0.00792	0.00800
-0.25	0.00436	0.00448	0.00148	-0.00028	0.00731	0.00725	0.00895	0.00694
0.00	-0.00502	-0.00500	-0.00074	0.00118	0.00664	0.00659	0.00849	0.00729
0.25	-0.00255	-0.00137	-0.00528	-0.00311	0.00819	0.00820	0.00781	0.00782
0.50	-0.00295	-0.00091	-0.00168	0.00945	0.00818	0.00816	0.00815	0.00861
0.75	0.00258	0.00376	-0.00328	0.00387	0.00714	0.00720	0.00801	0.00829
Exp.		1-7	8-14	15-21		1-7	8-14	15-21

Table 5. Bias and Variance Estimates for Spatial ML and GME Estimators of β_3.

True ρ	Bias				Variance			
	ML	GME-NLP	GME-NLP	GME-NLP	ML	GME-NLP	GME-NLP	GME-NLP
-0.75	0.00067	-0.00293	-0.00252	-0.00486	0.00284	0.00291	0.00290	0.00255
-0.50	-0.00401	-0.00510	0.00181	0.00226	0.00249	0.00249	0.00315	0.00263
-0.25	0.00358	0.00397	0.00049	0.00385	0.00281	0.00280	0.00260	0.00262
0.00	0.00037	0.00100	-0.00120	-0.00227	0.00240	0.00239	0.00272	0.00270
0.25	0.00205	0.00232	0.00007	-0.00034	0.00263	0.00261	0.00269	0.00276
0.50	0.00144	0.00140	0.00242	-0.00059	0.00316	0.00315	0.00276	0.00292
0.75	-0.00057	0.00025	0.00249	0.00056	0.00263	0.00268	0.00266	0.00281
Exp.		1-7	8-14	15-21		1-7	8-14	15-21

Table 6. Bias and Variance Estimates for Spatial ML and GME Estimators of β_4.

True ρ	Bias				Variance			
	ML	GME-NLP	GME-NLP	GME-NLP	ML	GME-NLP	GME-NLP	GME-NLP
-0.75	-0.00006	0.01181	0.00815	0.01267	0.00267	0.00271	0.00302	0.00268
-0.50	0.00032	0.00633	0.00189	0.00306	0.00304	0.00310	0.00285	0.00285
-0.25	0.00085	0.00341	-0.00063	-0.00166	0.00252	0.00249	0.00257	0.00252
0.00	-0.00028	0.00229	0.00300	0.00152	0.00252	0.00248	0.00253	0.00272
0.25	-0.00277	0.00191	-0.00113	0.01507	0.00269	0.00266	0.00255	0.00279
0.50	0.00069	0.00738	-0.00179	0.02398	0.00285	0.00284	0.00255	0.00270
0.75	0.00057	0.00133	-0.01178	0.02221	0.00294	0.00303	0.00286	0.00300
Exp.		1-7	8-14	15-21		1-7	8-14	15-21

Table 7. Bias and Variance Estimates for Spatial ML and GME Estimators of ρ.

True ρ	Bias				Variance			
	ML	GME-NLP	GME-NLP	GME-NLP	ML	GME-NLP	GME-NLP	GME-NLP
-0.75	-0.00927	0.19577	-0.01730	0.20061	0.00767	0.00750	0.01456	0.00879
-0.50	0.00010	0.13804	0.02749	0.13101	0.00816	0.00442	0.00914	0.00456
-0.25	-0.00692	0.06472	0.03321	0.06142	0.00833	0.00312	0.00651	0.00300
0.00	-0.01404	-0.00426	0.03750	-0.00011	0.00724	0.00264	0.00438	0.00253
0.25	-0.02020	-0.04403	0.04419	-0.06095	0.00632	0.00222	0.00337	0.00105
0.50	-0.01375	-0.04254	0.05490	-0.07476	0.00399	0.00158	0.00207	0.00049
0.75	-0.01627	-0.00504	0.06443	-0.03342	0.00210	0.00082	0.00073	0.00017
Exp.		1-7	8-14	15-21		1-7	8-14	15-21

Table 8. Monte Carlo Results for GME-NLP Estimators and the ML-Tobit Estimator for the Censored Regression Model.

Exp.	True ρ	X_{ij}	MSEL($\hat{\boldsymbol{\beta}}$)			MSEL($\hat{\rho}$)	
			ML-Tobit	GME-NLP	IGME-NLP	GME-NLP	IGME-NLP
1	-0.5	N(0,1)	0.01956	4.08584	0.02150	3.01889	0.19875
2	-0.5	N(2,1)	0.10631	1.27839	0.04942	0.23721	0.01880
3	0.5	N(0,1)	0.04673	4.37441	0.06812	0.83522	0.11481
4	0.5	N(2,1)	0.90989	0.03844	0.01253	0.00129	0.00076

Table 9. Bias and Variance Estimates for Spatial ML-Tobit and GME Estimators of β_1.

Exp.	True ρ	Bias			Variance		
		ML-Tobit	GME-NLP	IGME-NLP	ML-Tobit	GME-NLP	IGME-NLP
1	-0.5	0.01733	-0.95542	0.01128	0.00463	0.00613	0.00508
2	-0.5	-0.16694	-0.47565	-0.05550	0.00267	0.00440	0.00308
3	0.5	0.03134	-0.98231	-0.08957	0.01015	0.00901	0.00712
4	0.5	0.49925	-0.08200	-0.03823	0.00458	0.00278	0.00218

Table 10. Bias and Variance Estimates for Spatial ML-Tobit and GME Estimators of β_2.

Exp.	True ρ	Bias			Variance		
		ML-Tobit	GME-NLP	IGME-NLP	ML-Tobit	GME-NLP	IGME-NLP
1	-0.5	0.00317	-0.48003	-0.00070	0.00400	0.00560	0.00483
2	-0.5	-0.16371	-0.21208	-0.01208	0.00227	0.00397	0.00285
3	0.5	0.01145	-0.49760	-0.04791	0.00600	0.00649	0.00527
4	0.5	0.48135	-0.04128	-0.02194	0.00405	0.00243	0.00206

Table 11. Bias and Variance Estimates for Spatial ML-Tobit and GME Estimators of β_3.

Exp.	True ρ	Bias			Variance		
		ML-Tobit	GME-NLP	IGME-NLP	ML-Tobit	GME-NLP	IGME-NLP
1	-0.5	-0.02003	1.42290	-0.01343	0.00534	0.00635	0.00580
2	-0.5	-0.11426	0.87512	0.17769	0.00395	0.00595	0.00346
3	0.5	-0.04041	1.46910	0.13497	0.01657	0.01200	0.01029
4	0.5	0.39492	0.11093	0.02858	0.00398	0.00356	0.00223

Table 12. Bias and Variance Estimates for Spatial ML-Tobit and GME Estimators of β_4.

Exp.	True ρ	Bias			Variance		
		ML-Tobit	GME-NLP	IGME-NLP	ML-Tobit	GME-NLP	IGME-NLP
1	-0.5	0.01180	-0.94536	0.00896	0.00475	0.00615	0.00539
2	-0.5	-0.16558	-0.47223	-0.04889	0.00228	0.00400	0.00283
3	0.5	0.02742	-0.98393	-0.09566	0.01051	0.00798	0.00776
4	0.5	0.50640	-0.07952	-0.03470	0.00393	0.00260	0.00210

Table 13. Bias and Variance Estimates for Spatial ML-Tobit and GME Estimators of ρ.

Exp.	True ρ	Bias		Variance	
		GME-NLP	IGME-NLP	GME-NLP	IGME-NLP
1	-0.5	1.68403	0.31373	0.18292	0.10032
2	-0.5	0.48177	0.12239	0.00511	0.00382
3	0.5	0.86468	0.25800	0.08755	0.04825
4	0.5	0.02749	0.01678	0.00053	0.00048

Table 14. Definitions of Selected Variables for Disaster Relief Model (*N*=384).

Variables	Definition
(+) Percent change in precipitation[a]	To capture periods of increased wetness, one variable contains positive percent changes in precipitation; 0 otherwise.
(-) Percent change in precipitation[a]	Periods of relatively dryer weather are reflected in another variable containing negative percent changes in precipitation; 0 otherwise.
Percent change in low temperature[a]	For extreme or severe freezes, the annual percent change in low temperature.
Crop Insurance	These payments include both government and private insurance payments from the Crop Insurance program, and are computed from subtracting total farmer payments (which equals total insurance premiums plus a federal subsidy) from total indemnity payments.
Secretary of Agriculture [b]	1 if secretary of agriculture from a specific state; 0 otherwise
House Agriculture Subcommittee [b]	1 if state represented on House Agriculture Committee, subcommittee on General Farm Commodities, Resource Conservation, and Credit; 0 otherwise
Senate Agriculture Subcommittee [b]	1 if state represented on Senate Agriculture Committee, subcommittee on Research, Nutrition, and General Legislation; 0 otherwise
House Appropriations Subcommittee [b]	1 if state represented on House Appropriations Committee, subcommittee on Agriculture, Rural Development, Food and Drug Administration, and Related Agencies; 0 otherwise
Senate Appropriations Subcommittee [b]	1 if state represented on Senate Appropriations Committee, subcommittee on Agriculture, Rural Development, and Related Agencies; 0 otherwise
Income Measures, Farm Acres, Number of Farms	U.S Bureau of the Census' Bureau of Economic Analysis
Crop Values	USDA's National Agricultural Statistics Service
Electoral[c]	Represents a measure of electoral importance.
Census Regions[d]	1 if state in a specific Census Region; 0 otherwise
Year Dummies	1 if a specific year from 1992 to 1999; 0 otherwise

[a] For each state, average annual precipitation data were gathered over the period 1991 to 1999 from the National Oceanic Atmospheric Administration's (NOAA) National Climatic Data Center.
[b] From the *Almanac of American Politics.*
[c] Garrett and Sobel (2003).
[d] New England: Connecticut, Vermont, Massachusetts, Maine, Rhode Island, New Hampshire; Mid Atlantic: New Jersey, New York, Pennsylvania; East North Central: Michigan, Indiana, Illinois, Wisconsin, Ohio; West North Central: North Dakota, Minnesota, Nebraska, South Dakota, Iowa, Missouri, Kansas; South Atlantic: West Virginia, Delaware, South Carolina, North Carolina, Maryland, Florida, Virginia, Georgia; East South Central: Kentucky, Mississippi, Alabama, Tennessee; West South Central: Arkansas, Oklahoma, Texas, Louisiana; Mountain: Montana, Colorado, New Mexico, Arizona, Wyoming, Nevada, Idaho, Utah; Pacific: Oregon, Washington, California.

Table 15. Results for Disaster Payment Model.

Variable	ML-Tobit Coefficients	ML-Tobit T-values	ML-Spatial Coefficients	ML-Spatial T-values	IGME-NLP Coefficients	IGME-NLP T-values
Constant	-115.382	-6.297	-72.316	-5.629	-90.619	-4.467
(+) Percent change in precipitation	0.702	1.681	0.646	1.942	0.494	1.098
(-) Percent change in precipitation	-1.154	-1.871	-0.965	-1.977	-0.803	-1.205
Percent change in low temperature	2.012	0.846	1.517	0.785	2.405	0.920
Crop Insurance	0.525	3.979	0.483	4.515	-0.061	-0.610
Number of Farms	0.001	4.943	0.001	4.218	0.001	4.837
Secretary of Agriculture	118.065	3.087	99.854	3.245	94.813	2.422
House Agriculture Subcommittee	58.767	4.868	34.661	3.487	43.208	3.165
Senate Agriculture Subcommittee	43.360	2.677	28.651	2.181	49.152	2.882
House Appropriations Subcommittee	37.800	2.836	34.670	3.216	43.568	3.118
Senate Appropriations Subcommittee	34.707	2.566	16.080	1.454	16.020	1.099
1992	130.860	6.023	55.577	3.409	71.967	2.794
1993	119.890	5.544	43.327	2.630	68.487	2.601
1994	106.055	4.830	43.453	2.716	47.366	1.882
1995	57.935	2.701	18.163	1.162	32.043	1.361
1997	4.283	0.192	-0.751	-0.100	-1.082	-0.045
1998	71.128	3.315	26.025	1.662	36.086	1.507
1999	135.895	6.036	50.456	2.979	72.546	2.634
ρ	---	---	0.547	13.065	0.488	5.093

Appendix A.

Conditional Maximum Value Function

Define the conditional entropy function by conditioning on the $(L+K+1)\times 1$ vector $\theta = \boldsymbol{\tau} = vec\left(\boldsymbol{\tau}^{\pi}, \boldsymbol{\tau}^{\boldsymbol{\beta}}, \boldsymbol{\tau}^{\rho}\right)$ yielding

(A.1)
$$F(\tau) = \max_{\mathbf{p}:\theta=\boldsymbol{\tau}} \{-\sum_{k,j} p_{kj}^{\boldsymbol{\beta}} \ln\left(p_{kj}^{\boldsymbol{\beta}}\right) - \sum_{k,j} p_{kj}^{\pi} \ln\left(p_{kj}^{\pi}\right) - \sum_{j} p_{j}^{\rho} \ln\left(p_{j}^{\rho}\right) - \sum_{i,m} p_{im}^{\mathbf{u}^*} \ln\left(p_{im}^{\mathbf{u}^*}\right) - \sum_{i,m} p_{im}^{\mathbf{v}} \ln\left(p_{im}^{\mathbf{v}}\right)\}$$

The optimal value of $\mathbf{p}_i^{\mathbf{u}^*} = \left(p_{i1}^{\mathbf{u}^*}, \ldots, p_{iM}^{\mathbf{u}^*}\right)'$ in the conditionally-maximized entropy function is given by

(A.2)
$$\mathbf{p}_i^{u^*}(\tau) = \underset{p_{i\ell}^{\mathbf{u}^*}:\sum_{\ell=1}^{M} p_{i\ell}^{\mathbf{u}^*}=1, \sum_{\ell=1}^{M} s_{\ell}^{u^*} p_{i\ell}^{\mathbf{u}^*}=u_i^*(\boldsymbol{\tau})}{\arg\max} \left(-\sum_{\ell=1}^{M} p_{i\ell}^{\mathbf{u}^*} \ln(p_{i\ell}^{\mathbf{u}^*})\right),$$

which is the maximizing solution to the Lagrangian

(A.3)
$$L_{p_i^{u^*}} = -\sum_{\ell=1}^{M} p_{i\ell}^{u^*} \ln(p_{i\ell}^{u^*}) + \phi_i^{u^*}\left(\sum_{\ell=1}^{M} p_{i\ell}^{u^*} - 1\right) + \gamma_i^{u^*}\left(\sum_{\ell=1}^{M} s_{i\ell}^{u^*} p_{i\ell}^{u^*} - u_i^*(\tau)\right).$$

The optimal value of $p_{i\ell}^{u^*}$ is

(A.4)
$$p_{i\ell}^{u^*}\left(\gamma_i^{u^*}(u_i^*(\tau))\right) = \frac{e^{\gamma_i^{u}(u_i^*(\tau)) s_{i\ell}^{u^*}}}{\sum_{m=1}^{M} e^{\gamma_i^{u^*}(u_i^*(\tau)) s_{im}^{u^*}}}, \ \ell = 1, \ldots, M,$$

where $\gamma_i^u(u_i^*(\boldsymbol{\tau}))$ is the optimal value of the Lagrangian multiplier γ_i^u. The optimal value of $\mathbf{p}_k^{\boldsymbol{\beta}} = (p_{k1}^{\boldsymbol{\beta}}, \ldots, p_{kJ}^{\boldsymbol{\beta}})'$ in the conditionally-maximized entropy function is given by

(A.5)
$$\mathbf{p}_k^{\boldsymbol{\beta}}(\tau_k^{\beta}) = \underset{p_k^{\boldsymbol{\beta}}:\sum_{\ell=1}^{J} p_{k\ell}^{\boldsymbol{\beta}}=1, \sum_{\ell=1}^{J} s_{k\ell}^{\beta} p_{k\ell}^{\boldsymbol{\beta}}=\tau_k^{\beta}}{\arg\max} \left(-\sum_{\ell=1}^{J} p_{k\ell}^{\boldsymbol{\beta}} \ln(p_{k\ell}^{\boldsymbol{\beta}})\right),$$

which is the maximizing solution to the Lagrangian

(A.6) $$L_{p_k} = -\sum_{\ell=1}^{J} p_{k\ell}^{\beta} \ln(p_{k\ell}^{\beta}) + \phi_k^{\beta}\left(\sum_{\ell=1}^{J} p_{k\ell}^{\beta} - 1\right) + \eta_k^{\beta}\left(\sum_{\ell=1}^{J} s_{k\ell}^{\beta} p_{k\ell}^{\beta} - \tau_k^{\beta}\right).$$

The optimal value of $p_{k\ell}^{\beta}$ is then

(A.7) $$p_{k\ell}^{\beta}(\tau_k^{\beta}) = \frac{e^{\eta_k^{\beta}(\tau_k^{\beta}) s_{k\ell}^{\beta}}}{\sum_{m=1}^{J} e^{\eta_k^{\beta}(\tau_k^{\beta}) s_{km}^{\beta}}}, \ell = 1, \ldots, J,$$

where $\eta_k^{\beta}(\tau_k^{\beta})$ is the optimal value of the Lagrangian multiplier η_k^{β}. Likewise the optimal values of $\mathbf{p}^{\pi}, \mathbf{p}^{v}, \mathbf{p}^{\rho}$ can be derived. Substituting optimal values for $\mathbf{p} = vec\left(\mathbf{p}^{\beta}, \mathbf{p}^{\pi}, \mathbf{p}^{\rho}, \mathbf{p}^{u}, \mathbf{p}^{v}\right)$ into the conditional entropy function (A.1) yields

(A.8)

$$F(\tau) = -\sum_k \left[\eta_k^{\pi}\left(\tau_k^{\pi}\right)\tau_k^{\pi} - \ln\left(\sum_j \exp\left(\eta_k^{\pi}\left(\tau_k^{\pi}\right) p_{kj}^{\pi}\right)\right)\right] - \sum_k \left[\eta_k^{\beta}\left(\tau_k^{\beta}\right)\tau_k^{\beta} - \ln\left(\sum_j \exp\left(\eta_k^{\beta}\left(\tau_k^{\beta}\right) p_{kj}^{\beta}\right)\right)\right]$$
$$- \left[\eta^{\rho}\left(\tau^{\rho}\right)\tau^{\rho} - \ln\left(\sum_j \exp\left(\eta^{\rho}\left(\tau^{\rho}\right) p_j^{\rho}\right)\right)\right] - \sum_i \left[\gamma_i^u(u_i^*(\boldsymbol{\tau}))u_i^* - \ln\left(\sum_m \exp\left(\gamma_i^u(u_i^*(\boldsymbol{\tau})) p_{im}^{u^*}\right)\right)\right]$$
$$- \sum_i \left[\gamma_i^v(v_i(\boldsymbol{\tau}))v_i - \ln\left(\sum_m \exp\left(\gamma_i^v(v_i(\boldsymbol{\tau})) p_{im}^{v}\right)\right)\right]$$

Computational Issues

Following the computationally efficient approach of Mittelhammer and Cardell (1998), the conditional entropy function [equation (A.8)] was maximized. Note that the constrained maximization problem in (9) requires estimation of $(KJ + LJ + J + 2NM) \times 1$ unknown parameters. Solving (9) for $(KJ + LJ + J + 2NM)$ unknowns is not computationally practical as the sample size, *N*, grows larger. In contrast, maximizing

(A.8) requires estimation of only $(L+K+1)$ unknown coefficients for any positive value of *N*.

The GME-NLP estimator uses the reduced and structural form models as data constraints with a dual objective function as part of its information set. To completely specify the GME-NLP model, support points (upper and lower truncation and intermediate) for the individual parameters, support points for each error term, and $(L+K+1)$ starting values for the parameter coefficients are supplied by the user. In the Monte Carlo analysis and empirical application, the model was estimated using the unconstrained optimizer OPTIMUM in the econometric software GAUSS. We used 3 support points for each parameter and error term. To increase the efficiency of the estimation process the analytical gradient and Hessian were coded in GAUSS and called in the optimization routine. This also offered an opportunity to empirically validate the derivation of the gradient and Hessian (provided below). Given suitable starting values the optimization routine generally converged within seconds for the empirical examples discussed above. Moreover, solutions were robust to alternative starting values.

Gradient

The gradient vector $\nabla = vec\left(\nabla_{\pi}, \nabla_{\boldsymbol{\beta}}, \nabla_{\rho}\right)$ of $F(\tau)$ is

$$\text{(A.9)} \quad \nabla = -\begin{pmatrix} \eta^{\pi}\left(\boldsymbol{\tau}^{\pi}\right) \\ \eta^{\beta}\left(\boldsymbol{\tau}^{\beta}\right) \\ \eta^{\rho}\left(\tau^{\rho}\right) \end{pmatrix} + \begin{pmatrix} \mathbf{Z}' & \left[(\rho\mathbf{W})\mathbf{Z}\right]' \\ [\mathbf{0}] & \left[\mathbf{X}\right]' \\ 0 & \left[(\mathbf{W})\mathbf{Z}\pi\right]' \end{pmatrix} \begin{pmatrix} \gamma^{v} \\ \gamma^{u} \end{pmatrix}$$

Hessian

The Hessian matrix $\mathbf{H}(\boldsymbol{\tau}) = \dfrac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau} \partial \boldsymbol{\tau}'}$ is composed of submatrices

(A.10.1)

$$\frac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^{\pi} \partial \boldsymbol{\tau}^{\pi\prime}} = -\left[(\rho \mathbf{W})\mathbf{Z}\right]' \left[\frac{\partial \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))}{\partial \mathbf{u}(\boldsymbol{\tau})} \odot \left[(\rho \mathbf{W})\mathbf{Z}\right]\right] \text{-} \mathbf{Z}' \left[\frac{\partial \gamma^{\mathbf{v}}(\mathbf{v}(\boldsymbol{\tau}))}{\partial \mathbf{v}(\boldsymbol{\tau})} \odot \mathbf{Z}\right] - \frac{\partial \eta^{\pi}(\boldsymbol{\tau}^{\pi})}{\partial \boldsymbol{\tau}^{\pi\prime}}$$

(A.10.2)

$$\frac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^{\pi} \partial \boldsymbol{\tau}^{\beta\prime}} = -\left[(\rho \mathbf{W})\mathbf{Z}\right]' \left[\frac{\partial \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))}{\partial \mathbf{u}(\boldsymbol{\tau})} \odot [\mathbf{X}]\right]$$

(A10.3)

$$\frac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^{\pi} \partial \boldsymbol{\tau}^{\rho\prime}} = -\left[(\rho \mathbf{W})\mathbf{Z}\right]' \left[\frac{\partial \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))}{\partial \mathbf{u}(\boldsymbol{\tau})} \odot \left[(\mathbf{W})\mathbf{Z}\pi\right]\right] + \left[(\mathbf{W})\mathbf{Z}\right]' \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))$$

(A.10.4)

$$\frac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^{\boldsymbol{\beta}} \partial \boldsymbol{\tau}^{\boldsymbol{\beta}\prime}} = -\frac{\partial \eta^{\boldsymbol{\beta}}(\boldsymbol{\tau}^{\boldsymbol{\beta}})}{\partial \boldsymbol{\tau}^{\boldsymbol{\beta}\prime}} - [\mathbf{X}]' \left[\frac{\partial \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))}{\partial \mathbf{u}(\boldsymbol{\tau})} \odot [\mathbf{X}]\right]$$

(A.10.5)

$$\frac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^{\boldsymbol{\beta}} \partial \boldsymbol{\tau}^{\rho\prime}} = -[\mathbf{X}]' \left[\frac{\partial \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))}{\partial \mathbf{u}(\boldsymbol{\tau})} \odot \left[(\mathbf{W})\mathbf{Z}\pi\right]\right]$$

(A.10.6)

$$\frac{\partial^2 F(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}^{\rho} \partial \boldsymbol{\tau}^{\rho}} = -\frac{\partial \eta^{\rho}(\tau^{\rho})}{\partial \tau^{\rho}} - \left[(\mathbf{W})\mathbf{Z}\pi\right]' \left[\frac{\partial \gamma^{\mathbf{u}}(\mathbf{u}(\boldsymbol{\tau}))}{\partial \mathbf{u}(\boldsymbol{\tau})} \odot \left[(\mathbf{W})\mathbf{Z}\pi\right]\right]$$

In the above equations, the notation $\odot$ implies a Hadamard product (element by element multiplication) and the derivatives of the Lagrangian multipliers are defined as

(A.11.1) $$\frac{\partial \eta_k^{\ell}(\tau_k^{\ell})}{\partial \tau_k^{\ell}} = \left(\sum_{j=1}^{J} \left(s_{kj}^{\ell}\right)^2 p_{kj}^{\ell} - \left(\tau_k^{\ell}\right)^2 \right)^{-1} \quad \text{for } \ell \in \{\pi, \boldsymbol{\beta}, \rho\}$$

(A.11.2) $$\frac{\partial \gamma_i^{u}\left(u(\tau)\right)}{\partial u_i(\tau)} = \left(\sum_{\ell=1}^{J} \left(s_{i\ell}^{u}\right)^2 p_{i\ell}^{u}(\gamma_i^{u}(u_i(\tau))) - u_i^2(\tau) \right)$$

(A.11.3) $$\frac{\partial \gamma_i^{v}\left(v(\tau)\right)}{\partial v_i(\tau)} = \left(\sum_{\ell=1}^{J} \left(s_{i\ell}^{v}\right)^2 p_{i\ell}^{v}(\gamma_i^{v}(v_i(\tau))) - v_i^2(\tau) \right)$$

In the equations (A.9)-(A.11) the superscript * was dropped in the notation of the structural equations residuals for simple convenience. Given the above derivations (A.1)-(A.11), the asymptotic properties of the GME-NLP estimator follow from Theorems 1 and 2 in Marsh, Mittelhammer, and Cardell (2003).

Endnotes

[1] Standard econometric methods of imposing parameter restrictions on coefficients are constrained maximum likelihood or Bayesian regression. For additional discussion of these methods see Mittelhammer, Judge, and Miller (2000) or Judge et al (1988).
[2] Indeed, other spatial estimation procedures do not require normality for large sample properties. Examples include two stage least squares and generalized moments estimators in Kelejian and Prucha (1998, 1999).
[3] A problem may be described as ill-posed because of non-stationarity or because the number of unknown parameters to be estimated exceeds the number of data points. A problem may be described as ill-conditioned if the parameter estimates are highly unstable. An example of an ill-conditioned problem in empirical application is collinear data (Fraser 2000).
[4] Zellner (1998), and others, have discussed limitations of asymptotically justified estimators in finite sample situations and the lack of research on estimators that have small sample justification. See Anselin (1988) for further motivation and discussion regarding finite sample justified estimators.
[5] Moreover, and in contrast to constrained maximum likelihood or the typical parametric Bayesian analysis, GME does not require explicit specification of the distributions of the disturbance terms or of the parameter values. However, both the coefficient and the disturbance support spaces are compact in the GME estimation method, which may not apply in some idealized empirical modeling contexts.
[6] Mittelhammer, Judge, and Miller (2000) provide an introduction to information theoretic estimators for different econometric models and their connection to maximum entropy estimators.
[7] Preckel (2001) interpreted entropy as a penalty function over deviations. In the absence of strong prior knowledge, for the general linear model with symmetric penalty, Preckel argues there is little or no gain from using GME over generalized ordinary least squares.
[8] In this manner, one chooses the **p** that could have been generated in the greatest number of ways consistent with the given information (Golan, Judge, and Miller 1996).
[9] Specifying a support with J=2 is an effective means to impose bounds on a coefficient. The number of support points can be increased to reflect or recover higher moment information of the coefficient. Uncertainty about support points can be incorporated using a cross-entropy extension to GME.
[10] For empirical purposes, disturbance support spaces can always be chosen in a manner that provides a reasonable approximation to the true disturbance distribution because upper and lower truncation points can be selected sufficiently wide to contain the true disturbances of regression models (Malinvaud, 1980). Additional discussion about specifying supports is provided ahead and is available in Golan, Judge, and Miller (1996). For notational convenience it is assumed that each coefficient has J support points and each error has M support points.
[11] In contrast to the pure data constraint in (2b), the GME estimator could have been specified with the moment constraint $\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{X}\boldsymbol{\beta} + \mathbf{X}'\boldsymbol{\varepsilon}$. Data and moment constrained GME estimators are further discussed below and Monte Carlo results are provided for both.
[12] A dual loss function combined with the flexibility of parameter restrictions and user-supplied supports can provide an estimator that is robust in small samples or in ill-posed problems. For further discussion of dual loss functions see Zellner (1994).
[13] In other words, GME-D relaxes the orthogonality condition between **X** and the error term. See appendix materials for further insight into the gradient derivation.
[14] An alterative specification of the moment constraint would be to replace equation (6b) with
$\mathbf{X}'\left[\left(\mathbf{I}-\left(\mathbf{S}^{\rho}\mathbf{p}^{\rho}\right)\mathbf{W}_1\right)\mathbf{Y}-\mathbf{X}\left(\mathbf{S}^{\boldsymbol{\beta}}\mathbf{p}^{\boldsymbol{\beta}}\right)\right]\Big/N = \mathbf{X}'\left(\mathbf{S}^{\mathbf{u}}\mathbf{p}^{\mathbf{u}}\right)/N$.
[15] This specification is based on a GME estimator of the simultaneous equations model introduced by Marsh, Mittelhammer, and Cardell (2003), who demonstrated properties of consistency and asymptotic normality of the estimator. Historically, Theil (1971) used equation (7) to motivate two stage least squares and three stage least squares estimators. The first stage is to approximate E[**Y**] by applying OLS to the unrestricted reduced form model and thereby obtaining predicted values of **Y**. Then, using the predicted values to replace E[**Y**], the second stage is to estimate the structural model with OLS.

[16] Anselin (1988) suggests that the spatial coefficient is bounded by $1/\xi_{min} < \rho < 1/\xi_{max}$, where ξ_{min} and ξ_{max} are the minimum and maximum eigenvalues of the standardized spatial weight matrix, respectively.

[17] The log-likelihood function of the ML estimator for the spatial model with $A = I - \rho W$ is $\ln L = -(N/2)\ln(\pi) - (1/2)\ln|\Omega| + \ln|A| - (1/2)v'v$ (Anselin 1988).

[18] For further information on EM approaches also see McLachlan and Krishnan (1997). Both McMillen (1992) and LeSage (2000) follow similar approaches where the censored estimation problem is reduced to a noncensored estimation problem using ML and Bayesian estimation, respectively.

[19] The process of updating can continue iteratively to construct $\hat{Y}_2^{(i)}$ and then $\hat{\beta}^{(i)}$. The notation (*i*) in the superscript of the variables indicates the i[th] iteration with *i*=0 representing initial values.

[20] The different percent of censoring was incorporated into the Monte Carlo analysis to better sort out the performance capabilities of GME in presence of both the censoring and spatial processes.

[21] For the iterative GME-NLP estimator, convergence was assumed when the absolute value of the sum of the difference between the estimated parameter vectors from the i[th] iteration and i-1[st] iteration was less than 0.0005.

[22] The reduced form models are specified as in (8) and approximated at *t*=1.

[23] Garrett, Marsh, and Marshall (2003) explicitly investigated the simultaneous effects between disaster relief and crop payments. Alternatively, we choose to proxy crop insurance payments with predicted values and then focus on spatial effects of the disaster relief model.

[24] See Smith and Blundell (1986) regarding inference for maximum likelihood estimators of the simultaneous Tobit model.