Nonlinear Models with Spatial Data

William Greene
Stern School of Business, New York University

Washington D.C.

July 12, 2013

Y=1[New Plant Located in County]

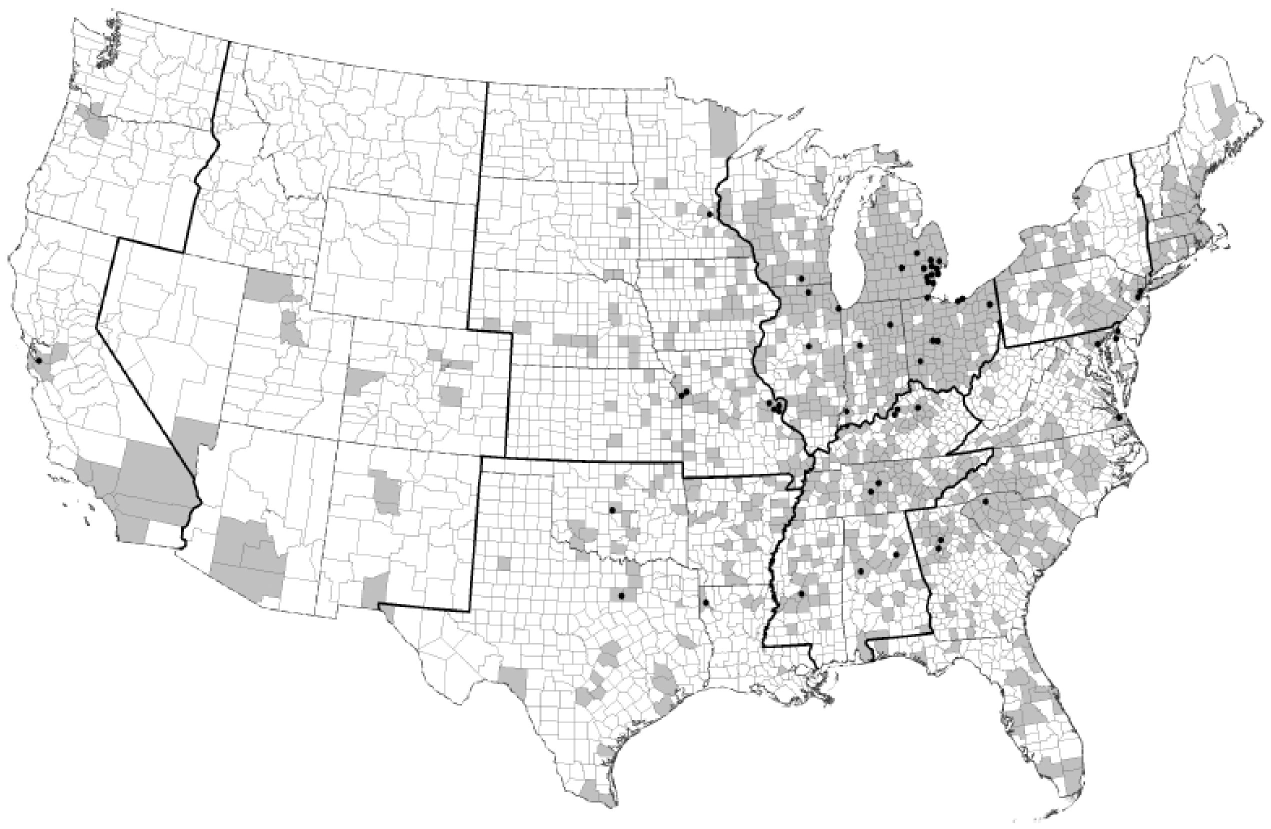


Figure 1. Counties with existing plants.

Klier and McMillen: Clustering of Auto Supplier Plants in the United States. JBES, 2008

Outcome Models for Spatial Data

❶ Spatial Regression Models

❷ Estimation and Analysis

❸ Nonlinear Models and Spatial Regression

❹ Nonlinear Models: Specification, Estimation

- Discrete Choice: Binary, Ordered, Multinomial, Counts
- Sample Selection
- Stochastic Frontier

Spatial Autocorrelation

$(\mathbf{x} - \mu\mathbf{i}) = \lambda\mathbf{W}(\mathbf{x} - \mu\mathbf{i}) + \boldsymbol{\varepsilon}$, N observations on a spatially arranged variable

$\mathbf{W}$ = contiguity matrix; $\mathbf{W}_{ii} = 0$

λ = spatial autocorrelation parameter, $-1 < \lambda < 1$.

$E[\boldsymbol{\varepsilon}]=\mathbf{0}$, $\quad Var[\boldsymbol{\varepsilon}]=\sigma_\varepsilon^2\mathbf{I}$

Spatial "moving average" form

$(\mathbf{x} - \mu\mathbf{i}) = [\mathbf{I} - \lambda\mathbf{W}]^{-1}\boldsymbol{\varepsilon}$

$E[\mathbf{x}]=\mu\mathbf{i}$, $Var[\mathbf{x}]=\sigma_\varepsilon^2[(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}$

Spatial Autocorrelation in Regression

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})\boldsymbol{\varepsilon}. \quad w_{ii} = 0.$$

$$E[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \quad Var[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \sigma_\varepsilon^2\mathbf{I}$$

$$E[\mathbf{y} \mid \mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$$

$$Var[\mathbf{y} \mid \mathbf{X}] = \sigma_\varepsilon^2(\mathbf{I} - \lambda\mathbf{W})(\mathbf{I} - \lambda\mathbf{W})'$$

A Generalized Regression Model

$$\hat{\boldsymbol{\beta}} = \left\{\mathbf{X}'\left((\mathbf{I} - \lambda\mathbf{W})(\mathbf{I} - \lambda\mathbf{W})'\right)^{-1}\mathbf{X}\right\}^{-1}\mathbf{X}'\left((\mathbf{I} - \lambda\mathbf{W})(\mathbf{I} - \lambda\mathbf{W})'\right)^{-1}\mathbf{y}$$

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left((\mathbf{I} - \lambda\mathbf{W})(\mathbf{I} - \lambda\mathbf{W})'\right)^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

$\hat{\lambda}$ = The subject of much research

Bell and Bockstael (2000) Spatial Autocorrelation in Real Estate Sales

$$
\begin{aligned}
\ln \textit{Price} = \alpha &+ \beta_1 \ln \textit{Assessed value (LIV)} \\
&+ \beta_2 \ln \textit{Lot size (LLT)} \\
&+ \beta_3 \ln \textit{Distance in km to Washington, DC (LDC)} \\
&+ \beta_4 \ln \textit{Distance in km to Baltimore (LBA)} \\
&+ \beta_5 \textit{\% land surrounding parcel in publicly owned space (POPN)} \\
&+ \beta_6 \textit{\% land surrounding parcel in natural privately owned space (PNAT)} \\
&+ \beta_7 \textit{\% land surrounding parcel in intensively developed use (PDEV)} \\
&+ \beta_8 \textit{\% land surrounding parcel in low density residential use (PLOW)} \\
&+ \beta_9 \textit{Public sewer service (1 if existing or planned, 0 if not) (PSEW)} \\
&+ \varepsilon.
\end{aligned}
$$

(Land surrounding the parcel is all parcels in the GIS data whose centroids are within 500 meters of the transacted parcel.) For the full model, the specification is

$$
\mathbf{y} = \mathbf{X}\beta + \varepsilon,
$$

$$
\varepsilon = \lambda \mathbf{W}\varepsilon + \mathbf{v}.
$$

Agreed Upon Objective: Practical Obstacles

- Problem: Maximize logL involving sparse ($\mathbf{I}$-$\lambda\mathbf{W}$)
- Inaccuracies in determinant and inverse
- Kelejian and Prucha (1999) moment based estimator of λ
- Followed by FGLS

Spatial Autoregression in a Linear Model

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

$$E[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \mathbf{0}, \quad \text{Var}[\boldsymbol{\varepsilon} \mid \mathbf{X}] = \sigma_\varepsilon^2 \mathbf{I}$$

$$\mathbf{y} = [\mathbf{I} - \lambda \mathbf{W}]^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= [\mathbf{I} - \lambda \mathbf{W}]^{-1}\mathbf{X}\boldsymbol{\beta} + [\mathbf{I} - \lambda \mathbf{W}]^{-1}\boldsymbol{\varepsilon}$$

$$E[\mathbf{y} \mid \mathbf{X}] = [\mathbf{I} - \lambda \mathbf{W}]^{-1}\mathbf{X}\boldsymbol{\beta}$$

$$\text{Var}[\mathbf{y} \mid \mathbf{X}] = \sigma_\varepsilon^2 [(\mathbf{I} - \lambda \mathbf{W})'(\mathbf{I} - \lambda \mathbf{W})]^{-1}$$

Estimators: Various forms of generalized least squares.

Maximum likelihood | $\boldsymbol{\varepsilon} \sim \text{Normal}[\mathbf{0}, \Sigma]$

Complications of the Generalized Regression Model

❶ Potentially very large N – GIS data on agriculture plots

❷ Estimation of λ. There is no natural residual based estimator

❸ Complicated covariance structures – no simple transformations

Panel Data Application

E.g., N countries, T periods

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + \varepsilon_{it}$$

$\boldsymbol{\varepsilon}_t = \lambda \mathbf{W}\boldsymbol{\varepsilon}_t + \mathbf{v}_t =$ N observations at time t.

Similar assumptions

Candidate for SUR or Spatial Autocorrelation model.

Spatial Autocorrelation in a Panel

Spatial Lags in Health Expenditures

Moscone, Knapp, and Tosetti (2007) investigated the determinants of mental health expenditure over six years in 148 British local authorities using two forms of the spatial correlation model to incorporate possible interaction among authorities as well as unobserved spatial heterogeneity. The models estimated, in addition to pooled regression and a random effects model, were as follows. The first is a model with **spatial lags:**

$$\mathbf{y}_t = \gamma_t \mathbf{i} + \rho \mathbf{W} \mathbf{y}_t + \mathbf{X}_t \beta + \mathbf{u} + \varepsilon_t,$$

where $\mathbf{u}$ is a 148×1 vector of random effects and $\mathbf{i}$ is a 148×1 column of ones. For each local authority,

$$y_{it} = \gamma_t + \rho(\mathbf{w}_i' \mathbf{y}_t) + \mathbf{x}_{it}' \beta + u_i + \varepsilon_{it},$$

where $\mathbf{w}_i'$ is the ith row of the contiguity matrix, $\mathbf{W}$. Contiguities were defined in $\mathbf{W}$ as one if the locality shared a border or vertex and zero otherwise. (The authors also experimented with other contiguity matrices based on "sociodemographic" differences.) The second model estimated is of **spatial error correlation**

$$\mathbf{y}_t = \gamma_t \mathbf{i} + \mathbf{X}_t \beta + \mathbf{u} + \varepsilon_t,$$

$$\varepsilon_t = \lambda \mathbf{W} \varepsilon_t + \mathbf{v}_t.$$

Analytical Environment

- ❶ Generalized linear regression
- ❷ Complicated disturbance covariance matrix
- ❸ Estimation platform: Generalized least squares or maximum likelihood (normality)
- ❹ Central problem, estimation of λ

Foundations and Trends® in Econometrics > Vol 4 > Issue 1–2

Estimation of Spatial Panels

Lee, Lung-fei [1] Yu, Jihai [2]

Table of contents

Outcomes in Nonlinear Settings

❶ Land use intensity in Austin, Texas – Discrete Ordered
Intensity = 1,2,3,4

❷ Land Usage Types in France, 1,2,3 – Discrete Unordered

❸ Oak Tree Regeneration in Pennsylvania – Count
Number = 0,1,2,… (Many zeros)

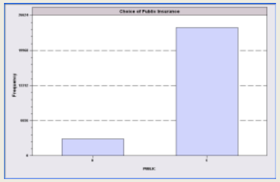
❹ Teenagers in the Bay area:
physically active = 1 or physically inactive = 0 – Binary

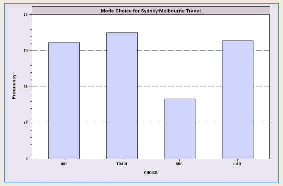
❺ Pedestrian Injury Counts in Manhattan – Count

❻ Efficiency of Farms in West-Central Brazil – Nonlinear
Model (Stochastic frontier)

❼ Catch by Alaska trawlers in a nonrandom sample

Nonlinear Outcomes



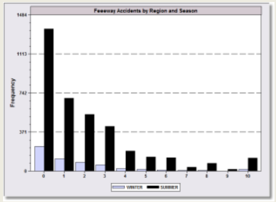


❶ Discrete revelation of choice indicates latent underlying preferences

① Binary choice between two alternatives

② Unordered choice among multiple choices ④

③ Ordered choice revealing underlying strength of preferences



❷ Counts of events

❸ Stochastic frontier and efficiency

❹ Nonrandom sample selection

Modeling Discrete Outcomes

❶ "Dependent Variable" typically labels an outcome

- No quantitative meaning
- Conditional relationship to covariates

❷ No "regression" relationship in most cases.

- Models are often not conditional means.
- The "model" is usually a probability

❸ Nonlinear models – usually not estimated by any type of linear least squares

Nonlinear Spatial Modeling

❶ Discrete outcome y_{it} = 0, 1, …, J for some finite or infinite (count case) J.

- i = 1,…,n
- t = 1,…,T

❷ Covariates $\mathbf{x}_{it}$

❸ Conditional Probability (y_{it} = j)
= a function of $\mathbf{x}_{it}$.

Regional Science and Urban Economics 40 (2010) 292–298

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

ELSEVIER

Modeling spatial discrete choice

Oleg A. Smirnov

Department of Economics, University of Toledo, 2801 W Bancroft St., Toledo, OH 43606-3390, USA

Issues in Spatial Discrete Choice

- **A series of Issues**
  - ❶ Spatial dependence between alternatives: Nested logit
  - ❷ Spatial dependence in the LPM: Solves some practical problems. A bad model
  - ❸ Spatial probit and logit: Probit is generally more amenable to modeling
  - ❹ Statistical mechanics: Social interactions – not practical
  - ❺ Autologistic model: Spatial dependency between outcomes or utillities. See below
  - ❻ Variants of autologistic: The model based on observed outcomes is incoherent ("selfcontradictory")
  - ❼ Endogenous spatial weights
  - ❽ Spatial heterogeneity: Fixed and random effects. Not practical.
- **The model discussed below**

Two Platforms

❶ Random Utility for Preference Models
Outcome reveals underlying utility

- Binary: $u^* = \theta'\mathbf{x}$ $y = 1$ if $u^* > 0$
- Ordered: $u^* = \theta'\mathbf{x}$ $y = j$ if $\mu_{j-1} < u^* < \mu_j$
- Unordered: $u^*(j) = \theta'\mathbf{x}_j$, $y = j$ if $u^*(j) > u^*(k)$

❷ Nonlinear Regression for Count Models
Outcome is governed by a nonlinear regression

- $E[y|\mathbf{x}] = g(\theta,\mathbf{x})$

Maximum Likelihood Estimation
Cross Section Case: Binary Outcome

- Random Utility: $y^* = \theta'\mathbf{x} + \varepsilon$
- Observed Outcome: $y = 1$ if $y^* > 0$,
  0 if $y^* \leq 0$.
- Probabilities: $P(y=1|x) = \text{Prob}(y^* > 0|\mathbf{x})$
  $= \text{Prob}(\varepsilon > -\theta'\mathbf{x})$
  $P(y=0|x) = \text{Prob}(y^* \leq 0|\mathbf{x})$
  $= \text{Prob}(\varepsilon \leq -\theta'\mathbf{x})$
- Likelihood for the sample = joint probability
  $= \prod_{i=1}^{n} \text{Prob}(y=y_i|\mathbf{x}_i)$
- Log Likelihood $= \sum_{i=1}^{n} \log\text{Prob}(y=y_i|\mathbf{x}_i)$

Cross Section Case: n observations

$$\text{Prob}\begin{pmatrix} y_1 = j \mid \mathbf{x}_1 \\ y_2 = j \mid \mathbf{x}_2 \\ \ldots \\ y_n = j \mid \mathbf{x}_n \end{pmatrix} = \text{Prob}\begin{pmatrix} \varepsilon_1 \;\le\; \text{or} > \theta'\mathbf{x}_1 \\ \varepsilon_2 \;\le\; \text{or} > \theta'\mathbf{x}_2 \\ \ldots \\ \varepsilon_n \;\le\; \text{or} > \theta'\mathbf{x}_n \end{pmatrix} = \begin{pmatrix} \text{Prob}(\varepsilon_1 \;\le\; \text{or} > \theta'\mathbf{x}_1) \\ \text{Prob}(\varepsilon_2 \;\le\; \text{or} > \theta'\mathbf{x}_2) \\ \ldots \\ \text{Prob}(\varepsilon_n \;\le\; \text{or} > \theta'\mathbf{x}_n) \end{pmatrix}$$

We operate on the marginal probabilities of n observations

$$\text{LogL}(\theta|\mathbf{X},\mathbf{y}) = \sum_{i=1}^{n} \log F\left[(2y_i - 1)\theta'\mathbf{x}_i\right]$$

- Probit $\quad F(t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)dt = \int_{-\infty}^{t} \phi(t)dt$
- Logit $\quad F(t) = \Lambda(t) = \frac{\exp(t)}{1+\exp(t)}$

How to Induce Correlation

❶ Joint distribution of multiple observations

❷ Correlation of unobserved heterogeneity

❸ Correlation of latent utility

Bivariate Counts

❶ Intervening variable approach

$Y_1 = X_1 + Z$, $Y_2 = Y_2 + Z$; All 3 Poisson distributed

Only allows positive correlation.

Limited to two outcomes

❷ Bivariate conditional means

$\lambda_1 = \exp(\mathbf{x}'\beta_1 + \varepsilon_1)$, $\lambda_2 = \exp(\mathbf{x}'\beta_2 + \varepsilon_2)$, $Cor(\varepsilon_1,\varepsilon_2)=\rho$

$|Cor(y1,y2)| << |\rho|$ (Due to residual variation)

❸ Copula functions – Useful for bivariate. Less so if > 2.

Spatially Correlated Observations Correlation Based on Unobservables

$$
\begin{aligned}
y_1 &= \theta'\mathbf{x}_1 + u_1 \\
y_2 &= \theta'\mathbf{x}_2 + u_2 \\
&\ldots \\
y_n &= \theta'\mathbf{x}_n + u_n
\end{aligned}
\qquad
\begin{pmatrix} u_1 \\ u_2 \\ \ldots \\ u_n \end{pmatrix} = (\mathbf{I} - \rho\mathbf{W}) \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix} \sim f\left[\begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, (\mathbf{I} - \rho\mathbf{W})(\mathbf{I} - \rho\mathbf{W})' \right]
$$

$\mathbf{W}$ = the usual spatial weight matrix.

In the cross section case, $\mathbf{W} = \mathbf{0}$. Now, it is a full matrix.

The joint probably is a single n fold integral.

Spatially Correlated Observations
Correlated Utilities

$$\begin{pmatrix} y_1^* \\ y_2^* \\ \ldots \\ y_n^* \end{pmatrix} = \rho \mathbf{W} \begin{pmatrix} y_1^* \\ y_2^* \\ \ldots \\ y_n^* \end{pmatrix} + \begin{pmatrix} \theta'\mathbf{x}_1 + \varepsilon_1 \\ \theta'\mathbf{x}_2 + \varepsilon_2 \\ \ldots \\ \theta'\mathbf{x}_n + \varepsilon_n \end{pmatrix} = \left(\mathbf{I} - \rho \mathbf{W}\right)^{-1} \begin{pmatrix} \theta'\mathbf{x}_1 + \varepsilon_1 \\ \theta'\mathbf{x}_2 + \varepsilon_2 \\ \ldots \\ \theta'\mathbf{x}_n + \varepsilon_n \end{pmatrix}$$

$\mathbf{W}$ = the usual spatial weight matrix.

In the cross section case, $\mathbf{W} = \mathbf{0}$. Now, it is a full matrix. The joint probably is a single n fold integral.

Log Likelihood

❶ In the unrestricted spatial case, the log likelihood is one term,

❷ LogL = log Prob($y_1|\mathbf{x}_1$, $y_2|\mathbf{x}_2$, … ,$y_n|\mathbf{x}_n$)

❸ In the discrete choice case, the probability will be an n fold integral, usually for a normal distribution.

A Theoretical Behavioral Conflict

$$\begin{pmatrix} y_1^* \\ y_2^* \\ \ldots \\ y_n^* \end{pmatrix} = \rho\mathbf{W}\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} + \begin{pmatrix} \theta'\mathbf{x}_1 + \varepsilon_1 \\ \theta'\mathbf{x}_2 + \varepsilon_2 \\ \ldots \\ \theta'\mathbf{x}_n + \varepsilon_n \end{pmatrix}$$

$$y_i = 1[y_i^* > 0]$$

$$y_1 = 1[\rho(w_{12}y_2 + w_{13}y_3 + \ldots) + \theta'\mathbf{x}_1 + \varepsilon_1 > 0]$$

$$y_2 = 1[\rho(w_{21}y_1 + w_{23}y_3 + \ldots) + \theta'\mathbf{x}_2 + \varepsilon_2 > 0] \text{ etc.}$$

The model based on observables is more reasonable.

There is no reduced form unless is **W** lower triangular.

This model is not identified. (It is "incoherent.")

This model extends readily to the model with a spatially lagged dependent variable. To do so, we must reinterpret (1) as the underlying latent variable explaining the *propensity* to have $d = 1$. As the propensity to have $d = 1$ increases for nearby observations, the propensity increases for observation i also. This assumption is different from a model in which the discrete variable d depends directly on neighboring values of d, that is, where $d = \rho Wd + X\beta + \varepsilon$. It is also different from a model in which the value of the underlying variable depends on neighboring values of d, so that $y = \rho Wd + X\beta + \varepsilon$. These models are not algebraically consistent.

See Maddala (1983)

From Klier and McMillen (2012)

The assumption that the latent variable depends on spatially lagged values of the latent variable may be disputable in some settings. In our example, we are assuming that the propensity to locate a new supplier plant in a county depends on the propensity to locate plants in nearby counties, and it does *not* depend simply on whether new plants have located nearby. The assumption is reasonable in this context because of the forward-looking nature of plant location decisions. Having other plants

LogL for an Unrestricted BC Model

$$\text{LogL}(\theta|\mathbf{X},\mathbf{y})=$$

$$\log\int_{-\infty}^{\theta'\mathbf{x}_n}\ldots\int_{-\infty}^{\theta'\mathbf{x}_1}\phi_n\left[\begin{pmatrix} q_1\varepsilon_1 \\ q_2\varepsilon_2 \\ \ldots \\ q_n\varepsilon_n \end{pmatrix}\middle|\begin{pmatrix} 1 & q_1q_2w_{12} & \ldots & q_1q_nw_{1n} \\ q_1q_2w_{21} & 1 & \ldots & q_2q_nw_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ q_nq_1w_{n1} & q_nq_2w_{n2} & \ldots & 1 \end{pmatrix}\right]d\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}$$

$q_i = -1$ if $y_i = 0$ and $+1$ if $y_i = 1$ $=$ $2y_i - 1$

- One huge observation - n dimensional normal integral.
- Not feasible for any reasonable sample size.
- Even if computable, provides no device for estimating sampling standard errors.

Solution Approaches for Binary Choice

- ❶ Distinguish between private and social shocks and use pseudo-ML
- ❷ Approximate the joint density and use GMM with the EM algorithm
- ❸ Parameterize the spatial correlation and use copula methods
- ❹ Define neighborhoods – make **W** a sparse matrix and use pseudo-ML
- ❺ Others …

JOURNAL OF REGIONAL SCIENCE, VOL. 32, NO. 3, 1992, pp. 335–348

PROBIT WITH SPATIAL AUTOCORRELATION

Daniel P. McMillen*

Department of Economics, University of Oregon, Eugene, OR 97403

ABSTRACT. Commonly-employed spatial autocorrelation models imply heteroskedastic errors, but heteroskedasticity causes probit to be inconsistent. This paper proposes and illustrates the use of two categories of estimators for probit models with spatial autocorrelation. One category is based on the EM algorithm, and requires repeated application of a maximum-likelihood estimator. The other category, which can be applied to models derived using the spatial expansion method, only requires weighted least squares.

Spatial autocorrelation in the heterogeneity

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i + \rho\sum_{j\neq i} w_{ij}\varepsilon_j$$

$$y_i = 1[y_i^* > 0],\ \text{Prob}\left[y_i = 1\right] = \Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sqrt{\text{Var}\left(\varepsilon_i + \rho\sum_{j\neq i} w_{ij}\varepsilon_j\right)}}\right]$$

or

$$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + u_i$$

$$y_i = 1[y_i^* > 0]\text{Prob}\left[y_i = 1\right] = \Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sqrt{\text{Var}(u_i)}}\right] = \Phi\left[\frac{\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right]$$

$$\sigma_i^2 = 1 + \rho^2\sum_{j\neq i} w_{ij}^2$$

Heteroscedastic Probit

Estimation and Inference

MLE: $\log L = \sum_{i=1}^{n} \log\Phi\left[\dfrac{(2y_i\text{-}1)\boldsymbol{\beta}'\mathbf{x}_i}{\sigma_i}\right]$

$$\sqrt{n}\left(\hat{\gamma}_{MLE} - \gamma\right) \rightarrow \mathbf{H}^{-1}(\boldsymbol{\gamma})\mathbf{S}(\boldsymbol{\gamma}) \quad \mathbf{S}=\text{Score vector}$$

implies the algorithm, Newton's Method.

EM algorithm essentially replaces **H** with **X′X** during iterations.

(Slightly more involved for the heteroscedasticity. LHS variable in the EM iterations is the score vector.)

To compute the asymptotic covariance, we need Var[$\mathbf{S}(\gamma)$]

Observations are (spatially) correlated! How to compute it?

GMM

Pinske, J. and Slade, M., (1998) "Contracting in Space: An Application of Spatial Statistics to Discrete Choice Models," Journal of Econometrics, 85, 1, 125-154.
Pinkse, J. , Slade, M. and Shen, L (2006) "Dynamic Spatial Discrete Choice Using One Step GMM: An Application to Mine Operating Decisions", Spatial Economic Analysis, 1: 1, 53 — 99.

$$\mathbf{y}^*=\mathbf{X}\beta+\varepsilon,\quad \varepsilon = \rho\mathbf{W}\boldsymbol{\varepsilon}+\mathbf{u}$$
$$= [\mathbf{I}-\rho\mathbf{W}]^{-1}\mathbf{u}$$
$$= \mathbf{A}u$$

Cross section case: $\rho=0$

Probit Model: FOC for estimation of θ is based on the generalized residuals $\hat{u}_i = y_i - E[\varepsilon_i \mid y_i]$

$$\sum_{i=1}^{n} \mathbf{x}_i \left(\frac{(y_i - \Phi(\beta'\mathbf{x}_i))\phi(\beta'\mathbf{x}_i)}{\Phi(\beta'\mathbf{x}_i)[1-\Phi(\beta'\mathbf{x}_i)]} \right) = \mathbf{0}$$

Spatially autocorrelated case: Moment equations are still valid. Complication is computing the variance of the moment equations, which requires some approximations.

GMM Approach

❶ Spatial autocorrelation induces heteroscedasticity that is a function of ρ

❷ Moment equations include the heteroscedasticity and an additional instrumental variable for identifying ρ.

❸ LM test of   = 0 is carried out under the null hypothesis that = 0.

❹ Application: Contract type in pricing for 118 Vancouver service stations.

GMM

$$\mathbf{y}^* = \mathbf{X}\beta + \varepsilon,\quad \varepsilon = \rho\mathbf{W}\varepsilon + \mathbf{u}$$
$$= [\mathbf{I} - \rho\mathbf{W}]^{-1}\mathbf{u}$$
$$= \mathbf{A}u$$

Autocorrelated Case: $\rho \neq 0$

Probit Model: FOC for estimation of θ is based on the generalized residuals $\hat{u}_i = y_i - E[\varepsilon \mid y_i]$

$$\sum_{i=1}^{n} \mathbf{z}_i \left(\frac{\left(y_i - \Phi\left[\frac{\beta'\mathbf{x}_i}{a_{ii}(\rho)}\right]\right)\phi\left[\frac{\beta'\mathbf{x}_i}{a_{ii}(\rho)}\right]}{\Phi\left[\frac{\beta'\mathbf{x}_i}{a_{ii}(\rho)}\right]\left(1 - \Phi\left[\frac{\beta'\mathbf{x}_i}{a_{ii}(\rho)}\right]\right)} \right) = \mathbf{0}$$

Requires at least K + 1 instrumental variables.

Extension to Dynamic Choice Model

Spatial Economic Analysis

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t724921264

Dynamic Spatial Discrete Choice Using One-step GMM: An Application to Mine Operating Decisions

Joris Pinkse; Margaret Slade; Lihong Shen

Pinske, Slade, Shen (2006)

Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Clustering of Auto Supplier Plants in the United States

Thomas Klier[a] & Daniel P McMillen[b]
[a] Federal Reserve Bank of Chicago, Research Department, Chicago, IL 60604
[b] Department of Economics (MC 144), University of Illinois at Chicago, Chicago, IL 60607
Published online: 01 Jan 2012.

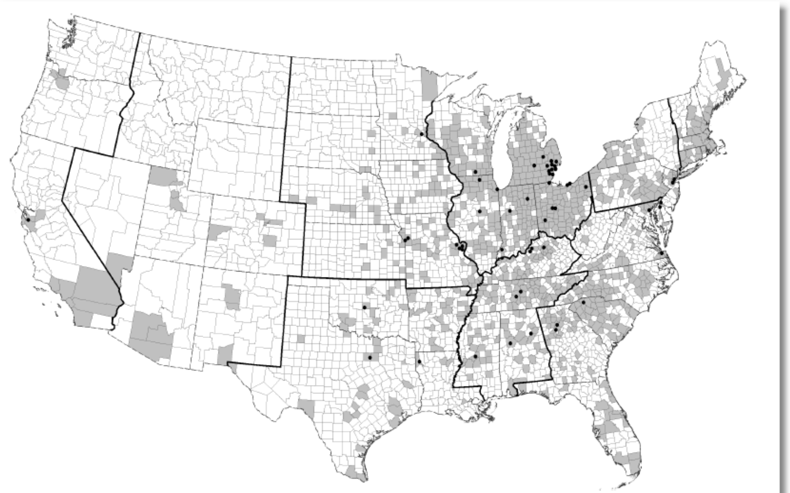
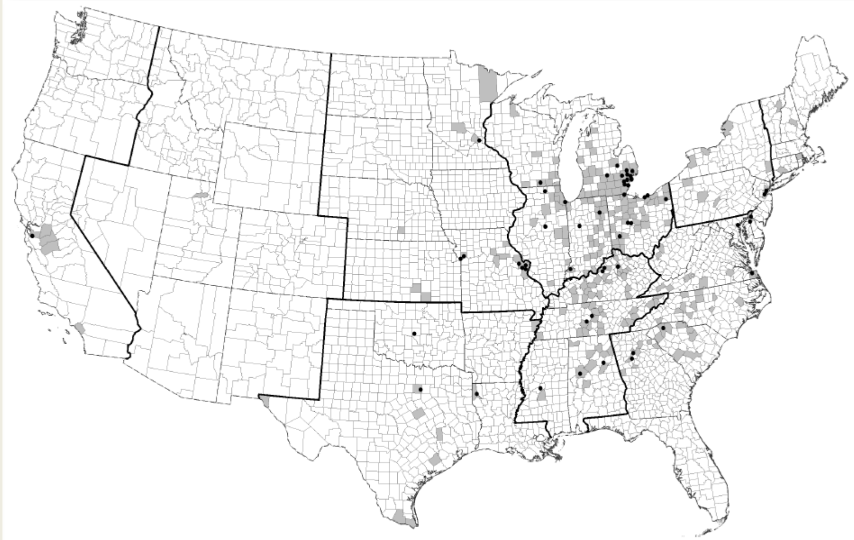


Figure 1. Counties with existing plants.



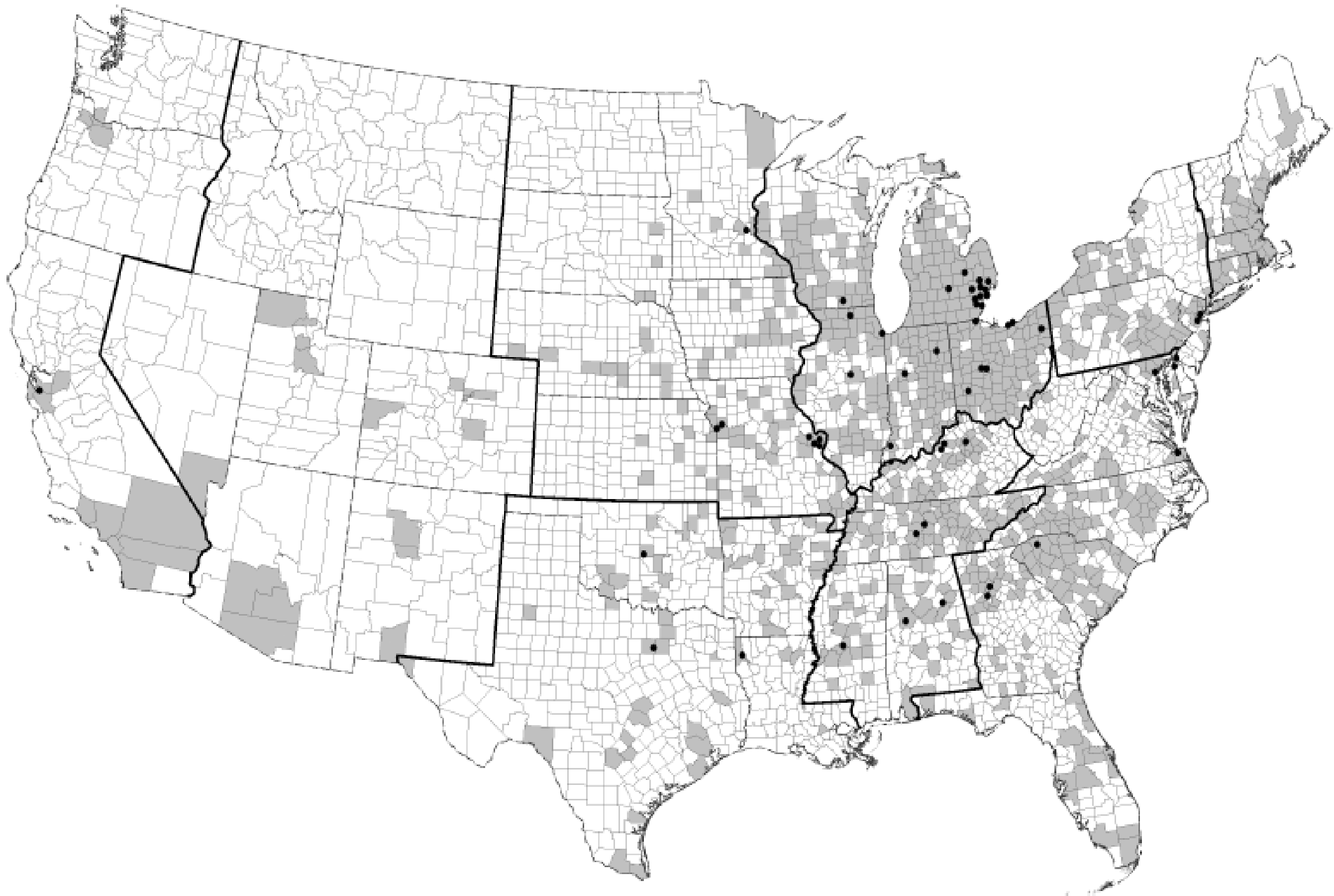


Figure 1. Counties with existing plants.

Spatial Logit Model

$$\mathbf{y}^* = \mathbf{X}\beta + \mathbf{e},\ \ \mathbf{e} = \rho\mathbf{W}\mathbf{e} + \varepsilon = (\mathbf{I}\text{-}\rho\mathbf{W})^{-1}\varepsilon$$

$$\mathbf{d} = \mathbf{1}[\mathbf{y}^* > \mathbf{0}],\ \text{Var}[\mathbf{e}] = \left[(\mathbf{I}\text{-}\rho\mathbf{W})'(\mathbf{I}\text{-}\rho\mathbf{W})\right]^{-1} = \Sigma,\ \Sigma_{ii} = \sigma_i^2$$

$$\text{Prob}(y_i = 1) = \Lambda\left(\frac{\beta'\mathbf{x}_i}{\sigma_i}\right) = \Lambda\left(\beta'\mathbf{x}_i^*\right) = \Lambda_i$$

Iterated 2SLS (GMM)

Generalized residual $u_i = d_i - \Lambda_i$

Instruments $\mathbf{Z}$

Criterion: $q = \mathbf{u}(\beta,\rho)'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{u}(\beta,\rho)$

Algorithm

$$u_i = d_i - \Lambda_i$$

$$\mathbf{g}_i = \begin{pmatrix} \partial u_i / \partial \beta \\ \partial u_i / \partial \rho \end{pmatrix} = \begin{pmatrix} -\Lambda_i(1-\Lambda_i)\mathbf{x}_i^* \\ \Lambda_i(1-\Lambda_i)\dfrac{\boldsymbol{\beta}'\mathbf{x}_i^*}{\sigma_i^2} A_{ii} \end{pmatrix}, \quad \mathbf{A}=(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\rho\mathbf{W})^{-1}$$

$$\mathbf{G} = [\mathbf{g}_1', \mathbf{g}_2', \ldots \mathbf{g}_n']'$$

Iterated 2SLS (GMM)

1. Logit estimation of $\boldsymbol{\beta} \,|\, \rho=0$, $\mathbf{G}_0$
2. $\mathbf{u}_k = (\mathbf{d} - \hat{\Lambda}_k)$, $\hat{\mathbf{G}}_k = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{G}_k$
3. $\Delta_k = \left(\hat{\mathbf{G}}_k'\hat{\mathbf{G}}_k\right)^{-1}\hat{\mathbf{G}}_k'\mathbf{u}_k$
4. $\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\rho} \end{pmatrix}_{k+1} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\rho} \end{pmatrix}_k + \boldsymbol{\Delta}_k$ until $\boldsymbol{\Delta}_k$ is sufficiently small.

LM Test?

- If $\rho = 0$, $g_\rho = 0$ because $\mathbf{A}_{ii} = 0$
- At the initial logit values, $\mathbf{g}_\beta = \mathbf{0}$
- Thus, if $\rho = 0$, $\mathbf{g} = 0$
- How to test $\rho = 0$ using an LM style test.
- Same problem shows up in RE models
- But, here, ρ is in the interior of the parameter space!

Pseudo Maximum Likelihood

❶ Maximize a likelihood function that approximates the true one

❷ Produces consistent estimators of parameters

❸ How to obtain standard errors?

❹ Asymptotic normality? Conditions for CLT are more difficult to establish.

Pseudo MLE

$$\mathbf{y}^* = \mathbf{X}\theta + \varepsilon, \quad \varepsilon = \rho\mathbf{W}\varepsilon + \mathbf{u}$$

$$= [\mathbf{I} - \rho\mathbf{W}]^{-1}\mathbf{u}$$

$$= \mathbf{A}u$$

Autocorrelated Case: $\rho \neq 0$

$$y_i^* = \mathbf{x}_i'\theta + \varepsilon_i + \sum_{j \neq i} \rho W_{ij}$$

$$y_i = 1[y_i^* > 0]. \quad \text{Var}[y_i^*] = 1 + \rho^2 \sum_{j \neq i} W_{ij}^2 = a_{ii}(\rho)$$

Implies a heteroscedastic probit.

Pseudo MLE is based on the marginal densities.

How to obtain the asymptotic covariance matrix?

[See Wang, Iglesias, Wooldridge (2013)]

Heteroscedastic Probit Approach

Estimation and Inference

MLE: $\log L = \sum_{i=1}^{n} \log\Phi\left[\frac{(2y_i\text{-}1)\beta'\mathbf{x}_i}{\sigma_i}\right]$

$$\sqrt{n}\left(\hat{\gamma}_{MLE} - \gamma\right) \rightarrow \mathbf{H}^{-1}(\gamma)\mathbf{S}(\gamma) \quad \mathbf{S}\text{=Score vector}$$

implies the algorithm, Newton's Method.

EM algorithm essentially replaces **H** with **X′X** during iterations. (Slightly more involved for the heteroscedasticity. LHS variable in the EM iterations is the score vector.)

To compute the asymptotic covariance, we need Var[$\mathbf{S}(\gamma)$]

Observations are (spatially) correlated! How to compute it?

Covariance Matrix for Pseudo MLE

$$\mathbf{V} = \mathbf{A}(\text{data},\hat{\theta})\ \mathbf{B}(\text{data},\hat{\theta})\ \mathbf{A}(\text{data},\hat{\theta})$$

$\mathbf{A}(\text{data},\hat{\theta}) =$ Negative inverse of Hessian

$\mathbf{B}(\text{data},\hat{\theta}) =$ Covariance matrix of scores.

How to compute $\mathbf{B}(\text{data},\hat{\theta})$

Terms are not independent in a spatial setting.

'Pseudo' Maximum Likelihood

Smirnov, A., "Modeling Spatial Discrete Choice," Regional Science and Urban Economics, 40, 2010.

Spatial Autoregression in Utilities

$$\mathbf{y}^* = \rho \mathbf{W}\mathbf{y}^* + \mathbf{X}\theta + \varepsilon, \quad \mathbf{y} = 1(\mathbf{y}^* > \mathbf{0}) \text{ for all n individuals}$$

$$\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\theta + (\mathbf{I} - \rho\mathbf{W})^{-1}\varepsilon$$

$$(\mathbf{I} - \rho\mathbf{W})^{-1} = \sum_{t=0}^{\infty} (\rho\mathbf{W})^t \text{ assumed convergent}$$

$$= \mathbf{A}$$

$$= \mathbf{D} + \mathbf{A} - \mathbf{D} \quad \text{where } \mathbf{D} = \text{diagonal elements}$$

$$\mathbf{y}^* = \mathbf{AX}\theta + \underbrace{\mathbf{D}\varepsilon}_{\text{Private}} + \underbrace{(\mathbf{A} - \mathbf{D})\varepsilon}_{\text{Social}}$$

Suppose individuals ignore the social "shocks." Then

$$\text{Prob}[y_i = 1 \text{ or } 0 \mid \mathbf{X}] = F\left[(2y_i - 1)\frac{\Sigma_{j=1}^{n} a_{ij}(\rho)\theta'\mathbf{x}_j}{d_i}\right], \text{ probit or logit.}$$

Pseudo Maximum Likelihood

❶ Bases correlation in underlying utilities

❷ Assumes away the correlation in the reduced form

❸ Makes a behavioral assumption

❹ Requires inversion of ($\mathbf{I}$-$\rho\mathbf{W}$)

❺ Computation of ($\mathbf{I}$-$\rho\mathbf{W}$) is part of the optimization process - ρ is estimated with θ .

❻ Does not require multidimensional integration (for a logit model, requires no integration)

Copula Method and Parameterization

Bhat, C. and Sener, I., (2009) "A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units," Journal of Geographical Systems, 11, 243–272

Basic Logit Model

$y_i^* = \beta'\mathbf{x}_i + \varepsilon_i, \quad y_i = 1[y_i^* > 0]$ (as usual)

Rather than specify a spatial weight matrix, we assume $[\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]$ have an n-variate distribution.

Sklar's Theorem represents the joint distribution in terms of the continuous marginal distributions, $\Lambda(\varepsilon_i)$ and a copula function $C[u_1 = \Lambda(\varepsilon_1), u_2 = \Lambda(\varepsilon_2), ..., u_n = \Lambda(\varepsilon_n) \mid \theta]$

Copula Representation

A particularly appealing approach to constructing a multivariate logistic distribution for spatial correlation analysis is to allow pairwise correlation across observational units (see Karunaratne and Elston, 1998 for such a pairwise correlation structure):

$$\Lambda(V_1 < v_1, V_2 < v_2, \ldots, V_q < v_q, \ldots, V_Q < v_Q) = \left[\prod_{q=1}^{Q} \Lambda_q(v_q)\right] \times \left[1 + \sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} \theta_{qk} \cdot \big(1 - \Lambda_q(v_q)\big)\big(1 - \Lambda_k(v_k)\big)\right], \qquad (7)$$

where θ_{qk} is the dependence parameter between V_q and V_k $(-1 \leq \theta_{qk} \leq 1)$, $\theta_{qk} = \theta_{kq}$ for all q and k, and $\Lambda_q(v_q) = \frac{1}{1+e^{-v_q}}$.

Model

3 The binary choice model with spatial correlation

Consider that the data (z_q, x_q) for $q = 1, 2, \ldots, Q$ are generated by the following latent variable framework:

$$z_q^* = \beta' x_q + \varepsilon_q$$

$$z_q = \begin{cases} 0 & \text{if } z_q^* < 0 \\ 1 & \text{if } z_q^* \geq 0 \end{cases} \tag{10}$$

where z_q^* is an unobserved propensity variable, β is a vector of coefficients to be estimated, and ε_q is a logistically distributed idiosyncratic error term with a scale parameter of σ_q (this allows spatial heteroscedasticity).[5] Define $V_q = \varepsilon_q / \sigma_q$, where V_q is standard logistic distributed. Let the V_q terms ($q = 1, 2, \ldots, Q$) follow the standard multivariate logistic distribution in Eq. 7. Also, let d_q be the actual observed value of z_q in the sample. Then, the probability of the observed vector of choices $(d_1, d_2, d_3, \ldots, d_Q)$ can be written, after some algebraic manipulations, as:

Likelihood

$$P(z_1 = d_1, z_2 = d_2, \ldots, z_Q = d_Q) = \left[\prod_{q=1}^{Q} \frac{e^{\left(\frac{\beta' x_q}{\sigma_q}\right)\cdot d_q}}{1 + e^{\left(\frac{\beta' x_q}{\sigma_q}\right)}}\right]$$

$$\times \left[1 + \sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} (-1)^{d_q + d_k} \cdot \theta_{qk} \left\{1 - \frac{e^{\left(\frac{\beta' x_q}{\sigma_q}\right)\cdot d_q}}{1 + e^{\left(\frac{\beta' x_q}{\sigma_q}\right)}}\right\} \left\{1 - \frac{e^{\left(\frac{\beta' x_k}{\sigma_k}\right)\cdot d_k}}{1 + e^{\left(\frac{\beta' x_k}{\sigma_k}\right)}}\right\}\right] \qquad (11)$$

Parameterization

$$\theta_{qk} = \pm\left[\frac{(e^{\delta})' s_{qk}}{1 + (e^{\delta})' s_{qk}}\right]$$

The parameter σ_q in Eq. 11 is next parameterized as:

$$\sigma_q = g(\lambda' \varpi_q) = \exp(\lambda' \varpi_q),$$

where ϖ_q includes variables specific to pre-defined “neighborhoods” (or other groupings) of observational units and individual-related factors

Table 1 Estimation results for teenagers' weekday physical activity participation choice

Variables	Binary (aspatial) logit model		Copula-based spatially correlated and heteroscedastic model	
	Parameter	*t* statistic	Parameter	*t* statistic
Constant	–5.534	–7.45	–3.211	–3.56
Individual demographics				
Male	0.238	1.18	0.259	2.22
Caucasian	0.722	2.42	0.320	1.82
Hispanic	0.457	0.95	0.336	1.72
Driver's license	0.661	3.02	0.309	2.23
Household demographics				
Household size	0.562	5.40	0.275	2.73
Single parent family	1.264	2.95	1.070	3.35
Presence of bicycle	–0.266	–0.93	0.168	1.35
Household location and season variables				
San Francisco County	1.309	1.84	0.341	1.36
Summer	0.816	3.94	0.450	3.28
Fall	4.265	8.47	2.459	3.37

Table 1 Estimation results for teenagers' weekday physical activity participation choice

Variables	Binary (aspatial) logit model		Copula-based spatially correlated and heteroscedastic model	
	Parameter	t statistic	Parameter	t statistic
(Spatial) heteroscedasticity variables				
Single parent family	–	–	–2.177	–3.95
Presence of bicycle	–	–	–0.305	–1.23
Fraction of multi-family dwelling units	–	–	–0.982	–2.02
Spatial correlation variables (δ) in the θ parameter				
Inverse of distance between zonal centroids	–	–	3.862	1.81
Number of observations	722		722	
Log-likelihood at convergence	–318.323		–308.273	

Other Approaches

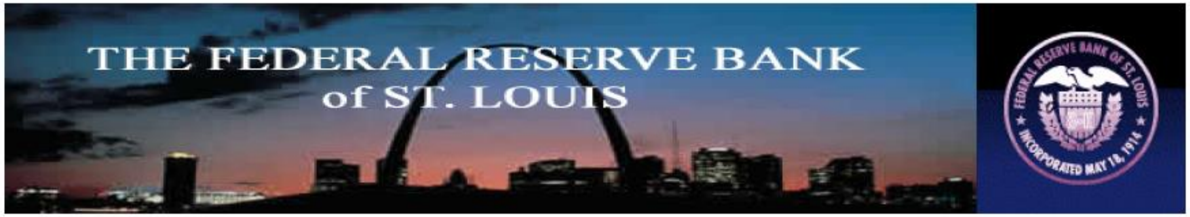
Case A (1992) Neighborhood influence and technological change. Economics 22:491–508
Beron KJ, Vijverberg WPM (2004) Probit in a spatial context: a monte carlo analysis. In: Anselin L, Florax RJGM, Rey SJ (eds) Advances in spatial econometrics: methodology, tools and applications. Springer, Berlin

❶ Case (1992): Define “regions” or neighborhoods. No correlation across regions. Produces essentially a panel data probit model. (Wang et al. (2013))

❷ Beron and Vijverberg (2003): Brute force integration using GHK simulator in a probit model.

❸ Lesage: Bayesian - MCMC

❹ Others. See Bhat and Sener (2009).



WORKING PAPER SERIES

A Spatial Analysis of State Banking Regulation

Thomas A. Garrett
Gary A. Wagner
and
David C. Wheelock

Working Paper 2003-044C
http://research.stlouisfed.org/wp/2003/2003-044.pdf

December 2003
Revised February 2005

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

ELSEVIER

Journal of Econometrics

Volume 172, Issue 1, January 2013, Pages 77–89

Partial maximum likelihood estimation of spatial probit models ☆

Honglin Wang[a], Emma M. Iglesias[b], Jeffrey M. Wooldridge[c]

[a] Hong Kong Institute for Monetary Research, 55/F, Two International Finance Centre, 8 Finance Street, Central, Hong Kong

[b] Department of Applied Economics II. Facultad de Economía y Empresa. University of A Coruña, Campus de Elviña, 15071. A Coruña, Spain

[c] Department of Economics, Michigan State University, 101 Marshall-Adams Hall, East Lansing, MI 48824-1038, USA

See also Arbia, G., "Pairwise Likelihood Inference for Spatial Regressions Estimated on Very Large Data Sets" Manuscript, Catholic University del Sacro Cuore, Rome, 2012.

Partial MLE

Observation 1

$$\begin{bmatrix} y_1^* = \mathbf{x}_1'\theta + \varepsilon_1 + \sum_{j\neq 1}\rho W_{1j}\varepsilon_j \\ y_1 = 1[y_1^* > 0] \quad \mathrm{Var}[y_1^*] = 1 + \rho^2 \sum_{j\neq 1} W_{1j}^2 = a_{11}(\rho) \end{bmatrix}$$

Observation 2

$$\begin{bmatrix} y_2^* = \mathbf{x}_2'\theta + \varepsilon_2 + \sum_{j\neq 2}\rho W_{2j}\varepsilon_j \\ y_2 = 1[y_2^* > 0] \quad \mathrm{Var}[y_2^*] = 1 + \rho^2 \sum_{j\neq 2} W_{2j}^2 = a_{22}(\rho) \end{bmatrix}$$

Covariance of y_1^* and y_2^* $=$ $a_{12}(\rho)$

Bivariate Probit

- ❶ Pseudo MLE
- ❷ Consistent
- ❸ Asymptotically normal?
  - Resembles time series case
  - Correlation need not fade with ‘distance’
- ❹ Better than Pinske/Slade Univariate Probit?
- ❺ How to choose the pairings?

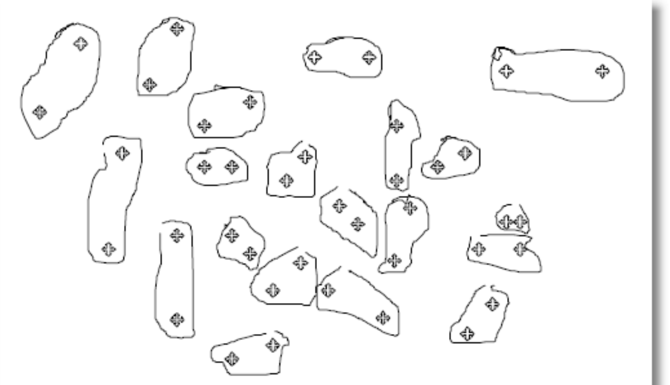


Fig. 1. $2n$ observations $\Longrightarrow$ n groups.

Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models (pages 19–35)

James P. LeSage

Article first published online: 3 SEP 2010 | DOI: 10.1111/j.1538-4632.2000.tb00413.x

geographical analysis

Geographical Analysis

January 2000

Volume 32, Issue 1
Pages 1–93

Previous Issue | Next Issue

Lesage methods - MCMC

- SEM Model…
- Bayesian MCMC
- Data augmentation for unobserved y
- Quirks about sampler for rho.

Ordered Probability Model

$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$, we assume **x** contains a constant term

$y = 0$ if $y^* \leq 0$

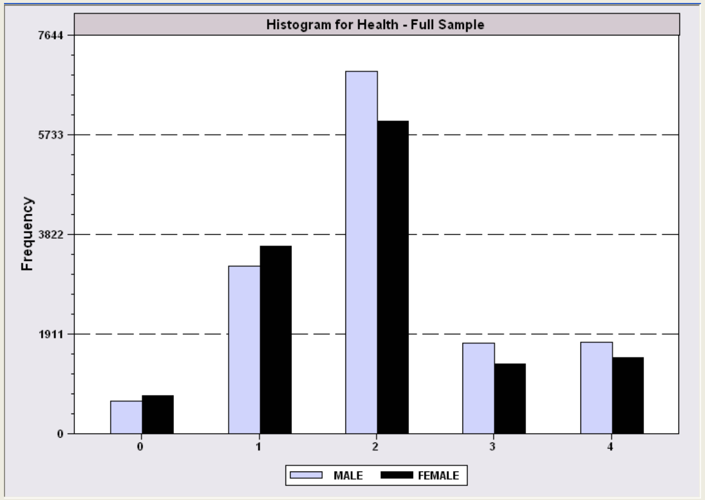
$y = 1$ if $0 < y^* \leq \mu_1$

$y = 2$ if $\mu_1 < y^* \leq \mu_2$

$y = 3$ if $\mu_2 < y^* \leq \mu_3$

...

$y = J$ if $\mu_{J-1} < y^* \leq \mu_J$



In general: $y = j$ if $\mu_{j-1} < y^* \leq \mu_j$, $j = 0,1,...,J$

$\mu_{-1} = -\infty,\ \mu_o = 0,\ \mu_J = +\infty,\ \mu_{j-1} < \mu_{j,}\ j = 1,...,J$

A Spatial Ordered Choice Model

Wang, C. and Kockelman, K., (2009) Bayesian Inference for Ordered Response Data with a Dynamic Spatial Ordered Probit Model, Working Paper, Department of Civil and Environmental Engineering, Bucknell University.

Core Model: Cross Section

$y_i^* = \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i$, y_i = j if $\mu_{j-1} < y_i^* \leq \mu_j$, $\text{Var}[\varepsilon_i] = 1$

Spatial Formulation: There are R regions. Within a region

$y_{ir}^* = \boldsymbol{\beta}'\mathbf{x}_{ir} + u_i + \varepsilon_{ir}$, y_{ir} = j if $\mu_{j-1} < y_{ir}^* \leq \mu_j$

Spatial heteroscedasticity: $\text{Var}[\varepsilon_{ir}] = \sigma_r^2$

Spatial Autocorrelation Across Regions

$\mathbf{u} = \rho\mathbf{W}\mathbf{u} + \mathbf{v}$, $\mathbf{v} \sim N[\mathbf{0}, \sigma_v^2\mathbf{I}]$

$\mathbf{u} = (\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{v} \sim N[\mathbf{0}, \sigma_v^2\{(\mathbf{I}-\rho\mathbf{W})'(\mathbf{I}-\rho\mathbf{W})\}^{-1}]$

The error distribution depends on 2 parameters, σ_v^2 and ρ

Estimation Approach: Gibbs Sampling; Markov Chain Monte Carlo

Dynamics in latent utilities added as a final step: y*(t)=f[y*(t-1)].

An Ordered Probability Model

$y^* = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$, we assume **x** contains a constant term

$y = 0$ if $y^* \leq 0$

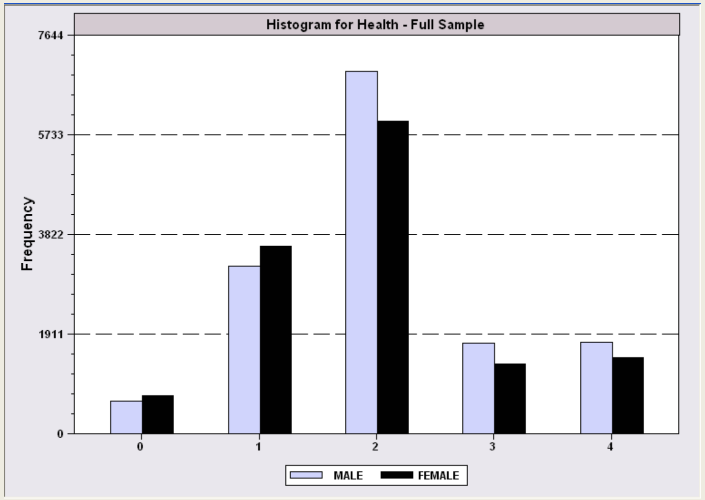
$y = 1$ if $0 < y^* \leq \mu_1$

$y = 2$ if $\mu_1 < y^* \leq \mu_2$

$y = 3$ if $\mu_2 < y^* \leq \mu_3$

...

$y = J$ if $\mu_{J-1} < y^* \leq \mu_J$



In general: $y = j$ if $\mu_{j-1} < y^* \leq \mu_j$, $j = 0,1,...,J$

$\mu_{-1} = -\infty,\ \mu_o = 0,\ \mu_J = +\infty,\ \mu_{j-1} < \mu_{j},\ j = 1,...,J$

OCM for Land Use Intensity

Table 1 Data Description for Land Development Intensity Level Analysis

Variable	Description
INTLV	Development intensity level
ELEVTN	Average elevation of the 300m grid cell (km)
SLOPE	Average slope of the 300m grid cell (%)
NSCHOOL	Number of K-12 schools in the neighborhood
POP	Population (thousand) in the neighborhood
WORKER	Number of workers (thousand) living in the neighborhood
INC	Average household income (thousand dollars) in the neighborhood
EMPTT	Travel time to nearest major (top 15) employer (hours)
CBDTT	Travel time to CBD (hours)
AIRTT	Travel time to nearest airfield (hours)
RDTT	Travel time to nearest highway (hours)

OCM for Land Use Intensity

Table 2 Summary Statistics for Land Development Intensity Analysis

	Variable	Minimum	Maximum	Mean	Std. Deviation
Constant through Years	ELEVTN	0.136	0.390	0.251	0.061
	SLOPE	0.034	17.328	2.699	2.196
	NSCHOOL	0.000	7.000	1.208	1.377
1983	INTLV	0.000	3.000	0.826	0.774
	POP	0.225	37.531	4.632	7.298
	WORKER	0.121	19.997	2.408	3.918
	INC	17.330	88.941	45.368	15.109
	EMPTT	0.004	1.115	0.453	0.223
	CBDTT	0.000	0.358	0.154	0.070
	AIRTT	0.005	0.784	0.345	0.157
	RDTT	0.002	0.498	0.111	0.093

Estimated Dynamic OCM

Table 3 Estimation Results for Model of Land Development Intensity Levels

Variable	Mean	Std. Dev.	t-stat.
POP	-0.024	0.036	-0.668
WORKER	0.089	0.067	1.327
INC	0.019	0.002	9.143
EMPTT	-0.232	0.130	-1.778
CBDTT	-4.365	0.851	-5.126
AIRTT	-2.867	0.248	-11.550
RDTT	2.309	0.385	6.001
NSCHOOL	0.039	0.017	2.305
ELEV	-0.239	0.696	-0.343
SLOPE	-0.034	0.010	-3.394
λ	0.561	0.019	30.005
ρ	0.857	0.074	11.612
σ^2	0.871	0.222	3.931
γ_1	-0.834	0.011	-77.231
γ_2	2.235	0.031	71.393
γ_3	4.361	0.034	130.167

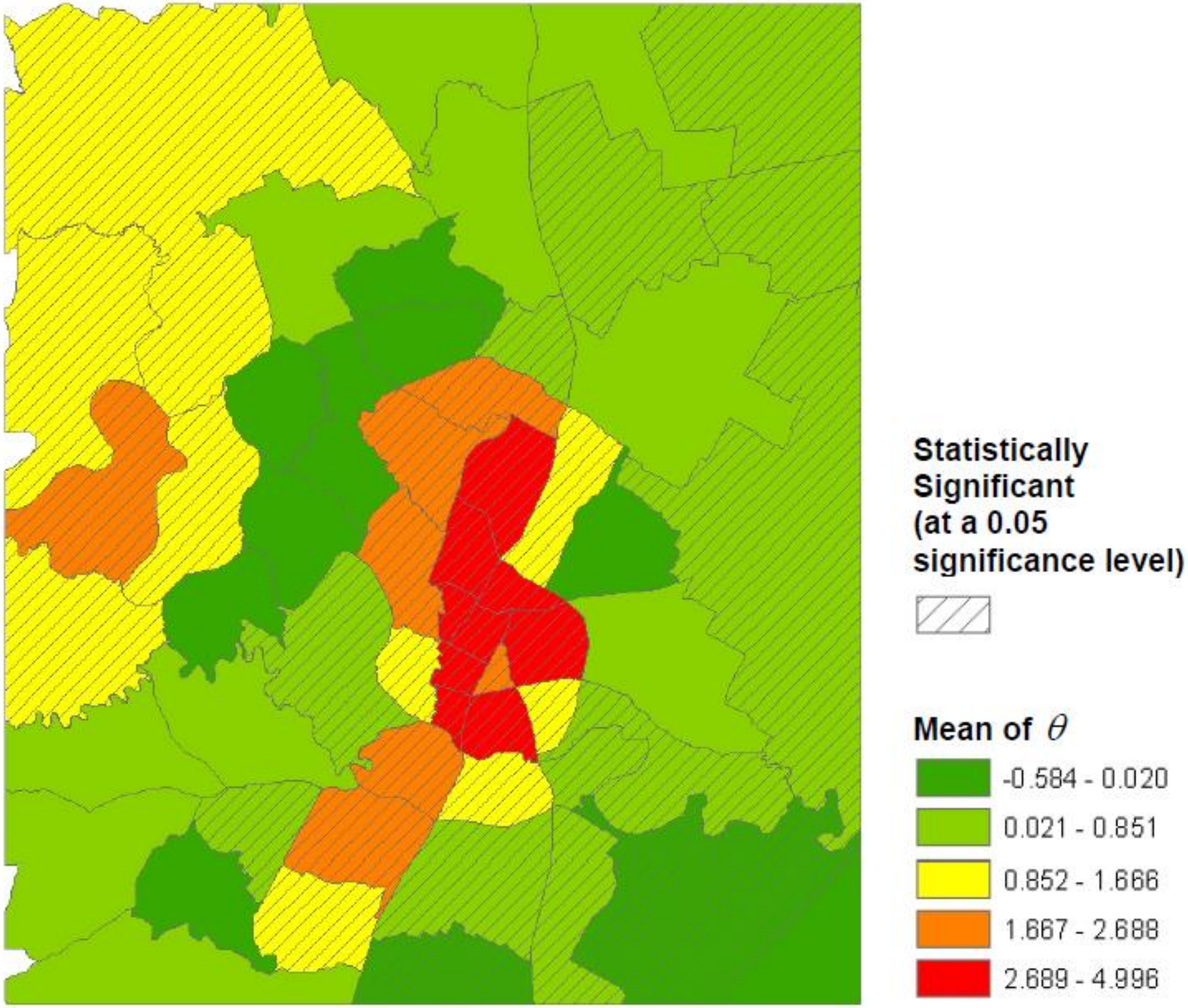


Figure 2 Distribution of Region-Specific Error Term Estimates (θ) for Land Development Intensity Levels

doi:10.1111/j.1435-5957.2009.00249.x

Application of the dynamic spatial ordered probit model: Patterns of land development change in Austin, Texas

Xiaokun Wang[1], **Kara M. Kockelman**[2]

[1] Bucknell University, Department of Civil Engineering, 701 Moore Avenue, Lewisburg, PA 17837, USA (e-mail: cara.wang@bucknell.edu)
[2] University of Texas, 1 University Station, ECJ Suite 6.9, C1761, Austin, TX 78712, USA (e-mail: kkockelm@mail.utexas.edu)

Received: 12 February 2008 / Accepted: 13 December 2008

APPLICATION OF THE DYNAMIC SPATIAL ORDERED PROBIT MODEL: PATTERNS OF OZONE CONCENTRATION IN AUSTIN, TEXAS

Xiaokun (Cara) Wang
Assistant Professor
Department of Civil and Environmental Engineering
Bucknell University
Lewisburg, PA 17837, USA

Kara M. Kockelman
(corresponding author)
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

This latent utility difference is influenced by many factors

$$y^*_{ikt} = X'_{ikt}\beta + \varepsilon_{ikt}$$

$$\varepsilon_{ikt} = \lambda_k \sum_{\substack{j=1 \\ j\neq i}}^{N} w_{ij}\varepsilon_{jkt} + \theta_{ikt}$$

$$\theta_{ikt} = \rho\theta_{ikt-1} + \eta_{ikt}$$

Bayesian Estimation

The joint posterior distribution for all parameters can be written as follows:

$$p(y^*, \beta, \lambda_k, \rho, \mathrm{B}|Y, X) \propto p(Y|y^*)\pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\beta)\pi(\lambda_k)\pi(\rho)\pi(\mathrm{B}) \quad (16)$$

As is standard in Bayesian estimation, the conditional posterior distributions of all parameters can be derived by extracting only items that contain them, as follows:

$$p(\beta| \ldots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\, \pi(\beta) \quad (17)$$

$$p(y^*| \ldots) \propto p(Y|y^*)\pi(y^*|\beta, \lambda_k, \rho, \mathrm{B}) \quad (18)$$

Data Augmentation

$$p(\rho| \ldots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\rho) \quad (19)$$

$$p(\mathrm{B}| \ldots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\mathrm{B}) \quad (20)$$

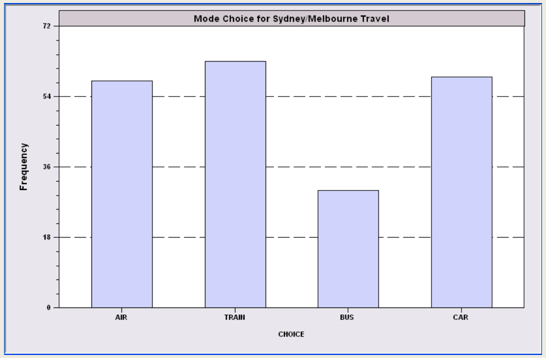
$$p(\lambda_k| \ldots) \propto \pi(y^*|\beta, \lambda_k, \rho, \mathrm{B})\pi(\lambda_k) \quad (21)$$

Unordered Multinomial Choice

Core Random Utility Model

- Underlying Random Utility for Each Alternative

  U(i,j) = $\beta_j' \boldsymbol{x}_{ij} + \varepsilon_{ij}$, i = individual, j = alternative



- Preference Revelation

  Y(i) = j if and only if U(i,j) > U(i,k) for all k $\neq$ j

- Model Frameworks

  Multinomial Probit: $[\varepsilon_1,\ldots,\varepsilon_J] \sim N[0,\Sigma]$

  Multinomial Logit: $[\varepsilon_1,\ldots,\varepsilon_J] \sim$ iid type 1 extreme value

Spatial Multinomial Probit

Chakir, R. and Parent, O. (2009) "Determinants of land use changes: A spatial multinomial probit approach, Papers in Regional Science, 88, 2, 328-346.

Utility Functions, land parcel i, usage type j, date t

$$U(i,j,t)=\boldsymbol{\beta}'_{jt}\mathbf{x}_{ijt} + \theta_{ik} + \varepsilon_{ijt}$$

Spatial Correlation at Time t

$$\theta_{ij} = \rho \sum_{l=1}^{n} w_{il}\theta_{lk}$$

Modeling Framework: Normal / Multinomial Probit

Estimation: MCMC - Gibbs Sampling

Table 1. Variables description

Variable	Description	Scale	Source of the data
land use	land use (= 1 if agriculture, 2 if urban, 3 if forest and 0 if no-use)	Parcel	TERUTI survey
NRSEC90	number of second homes	County	INSEE population census
aver	average rain	County	The Climate Database of Europe at the resolution of 50 km
avesl	average slope	County	The Digital Elevation Model of Europe at the resolution of 1 km
REV	average household income	County	Income tax survey *Impôt sur le revenu des communes*
whyd	wheat yield	Region	AGRESTE
grpop	population growth between 1990 and 1999	County	INSEE population census
network	travel time to the nearest highway	County	Microsoft Autoroute 2007
TEXT1	Soil quality 0, if coarse texture (clay <18% and sand > 65 %); 1, otherwise	County	The French Soil map at the scale of 1/1,000,000

Table 2. Descriptive statistics of explanatory variables

Variable	mean	std	min	max
grpop	12.83	17.23	–11.40	149.00
NRSEC90	211.84	1,103.54	6.00	8,949.00
aver	2.38	0.53	1.01	3.63
avesl	2.49	1.31	0.24	7.09
REV	14,320.08	4,513.06	5,102.07	48,469.43
whyd	106.17	14.74	80.00	130.00
network	21.88	15.55	1	61
text1	0.48	0.50	0	1.00

Table 6. Estimation results for spatial multinomial probit model

Variable	Mean	s.d.	2.5%	97.5%
ρ	0.632†	0.004	0.627	0.639
σ_ϕ	1.327†	0.031	1.278	1.388
σ_{11}	1.000	−1.000	1.000	1.000
σ_{12}	0.975†	0.007	0.961	0.991
σ_{13}	1.008†	0.006	0.996	1.018
σ_{21}	0.975†	0.007	0.961	0.991
σ_{12}	0.966†	0.014	0.938	0.997
σ_{23}	0.984†	0.012	0.966	1.004
σ_{31}	1.008†	0.006	0.996	1.018
σ_{32}	0.984†	0.012	0.966	1.004
σ_{33}	1.025†	0.013	1.003	1.045

Note: † Numerical Standard Errors (NSE) less than 1%.

The population growth has a significant and negative effect on urban land use suggesting that counties with a higher population growth rate tend to be in suburban areas. This result confirms the findings of Carrion-Flores and Irwin (2004) that suggest that new urban development is less likely to be located in densely developed areas. This is what they call a ‘congestion effect’: higher population density decreases the attractiveness of areas that are already substantially developed.

Random Parameters Models

Incorporating Spatial Dependencies in Random Parameter Discrete Choice Models

Abolfazl (Kouros) Mohammadian
Department of Civil and Materials Engineering
University of Illinois at Chicago
842 W. Taylor St.
Chicago, Illinois 60607-7023
Tel: (312) 996-9840
Fax: (312) 996-2426
Email: kouros@uic.edu

Murtaza Haider
School of Urban Planning and Dept. of Civil Engineering
McGill University
815 Sherbrooke Street West, Suite 400
Montreal, Quebec, Canada
Tel: (514) 398-4079
Fax: (514) 398-8376
Email: murtaza.haider@mcgill.ca

Pavlos S. Kanaroglou
School of Geography and Geology
McMaster University
Hamilton, Ontario, Canada
Tel: (905) 525-9140 Ext. 23525
Fax: (905) 546-0463
Email: pavlos@mcmaster.ca

Paper submitted for presentation at the
84th Annual Transportation Research Board Meeting
January 2005
Washington D.C.

Tracking Land Cover Change in a Mixed Logit Model: Recognizing Temporal and Spatial Effects

Xiaokun Wang
Graduate Student Researcher
The University of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
wangxk@mail.utexas.edu

Kara M. Kockelman
Clare Boothe Luce Associate Professor of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
kkockelm@mail.utexas.edu
Phone: 512-471-0210
FAX: 512-475-8744
(Corresponding Author)

To be presented at the 85th Annual Meeting of the Transportation Research Board and under consideration for publication by *Transportation Research Record*

Each decision–maker in this study is assumed to hold a parcel of land and is planning to start a housing project. Developers are faced with the decision of what type of residential units to build (i.e., detached, semi-detached, condo, or townhouse). It can be postulated that this decision is influenced, to some extend at least, by nearby housing development projects. In other words, the existing housing stock, as well as the location factors will affect the future housing developments in the same neighborhood. This implies that the unobserved attributes of the neighborhood tend to be correlated.

$$U_{in} = V_{in} + \varepsilon_{in} = (\sum \beta_i X_{in} + \sum_{s=1}^{S} \rho_{nsi} y_{si}) + \varepsilon_{in}$$

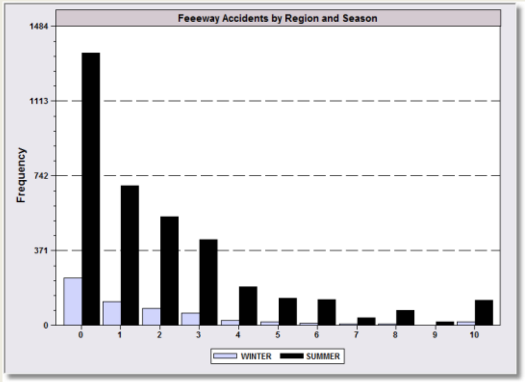
zero otherwise. ρ can be modeled similar to an impedance function. In spatial statistics, it usually takes the form of a negative exponential function of the distance separating the two decision-makers (D_{ns}).

$$\rho_{nsi} = \lambda \exp(-\frac{D_{ns}}{\gamma}) \qquad [3]$$

$$P_{nt}(i \mid \beta_n) = \frac{\exp(\alpha_{in} + \gamma_i W_n + \beta_{in} X_{\text{int}} + \sum_{s=1}^{S} \rho_{\sin} y_{si} + \varepsilon_{\text{int}})}{\sum_{j \in C_{nt}} \exp(\alpha_{jn} + \gamma_j W_n + \beta_{jn} X_{jnt} + \sum_{s=1}^{S} \rho_{\sin} y_{si} + \varepsilon_{jnt})} \qquad [8]$$

Canonical Model

Rathbun, S and Fei, L (2006) “A Spatial Zero-Inflated Poisson Regression Model for Oak Regeneration,” Environmental Ecology Statistics, 13, 2006, 409-426



Poisson Regression

$y = 0,1,\ldots$

$$\text{Prob}[y = j|\mathbf{x}] = \frac{\exp(-\lambda)\lambda^j}{j!}$$

Conditional Mean $\lambda = \exp(\boldsymbol{\beta}'\mathbf{x})$

Signature Feature: Equidispersion

Usual Alternative: Various forms of Negative Binomial

Spatial Effect: Filtered through the mean

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \theta_i)$$

$$\theta_i = \rho\sum_{m=1}^{n} w_{im}\theta_m + \varepsilon_i$$

Canonical Model for Counts

Rathbun, S and Fei, L (2006) "A Spatial Zero-Inflated Poisson Regression Model for Oak Regeneration," Environmental Ecology Statistics, 13, 2006, 409-426

Poisson Regression

$y = 0,1,...$

$$\text{Prob}[y = j|\mathbf{x}] = \frac{\exp(-\lambda)\lambda^{j}}{j!}$$

Conditional Mean $\lambda = \exp(\boldsymbol{\beta}'\mathbf{x})$

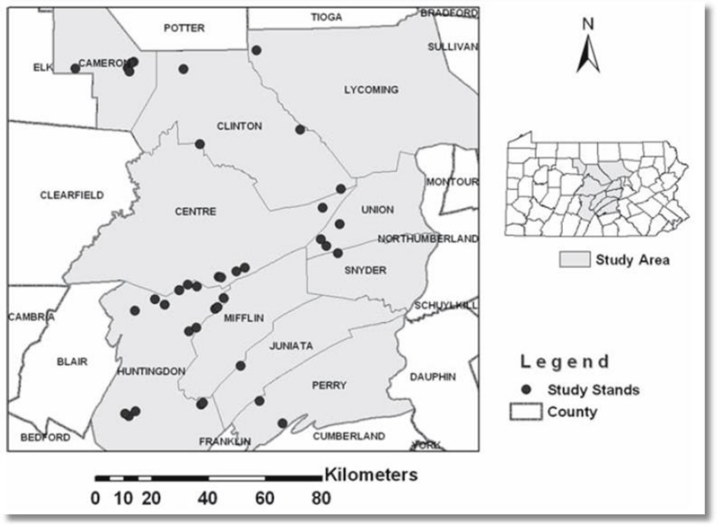
Signature Feature: Equidispersion

Usual Alternative: Negative Binomial

Spatial Effect: Filtered through the mean

$$\lambda_i = \exp(\boldsymbol{\beta}'\mathbf{x}_i + \theta_i)$$

$$\theta_i = \rho\sum_{m=1}^{n} w_{im}\theta_m + \varepsilon_i$$



Zero Inflation

❶ There are two states
- Always zero
- Zero is one possible value, or 1,2,...

❷ Prob(0) = Prob(state 1) +
Prob(state 2) P(0|state 2)

A Spatial Multivariate Count Model for Firm Location Decisions

Chandra R. Bhat
Dept of Civil, Architectural and Environmental Engineering
The University of Texas at Austin

Rajesh Paleti
Parsons Brinckerhoff

Palvinder Singh
Parsons Brinckerhoff

A Blend of Ordered Choice and Count Data Models

Numbers of firms locating in Texas counties: Count data (Poisson)

On Accommodating Spatial Dependence in Bicycle and Pedestrian Injury Counts by Severity Level

Chandra R. Bhat
Dept of Civil, Architectural and Environmental Engineering
The University of Texas at Austin

Rajesh Paleti
Parsons Brinckerhoff

Sriram Narayanamoorthy
Dept. of Civil, Architectural and Environmental Engineering
The University of Texas at Austin

Bicycle and pedestrian injuries in census tracts in Manhattan. (Count data and ordered outcomes)

Kriging

FORECASTING NETWORK DATA:
SPATIAL INTERPOLATION OF TRAFFIC COUNTS USING TEXAS DATA

Xiaokun Wang
Assistant Professor
Department of Civil and Environmental Engineering
Bucknell University

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

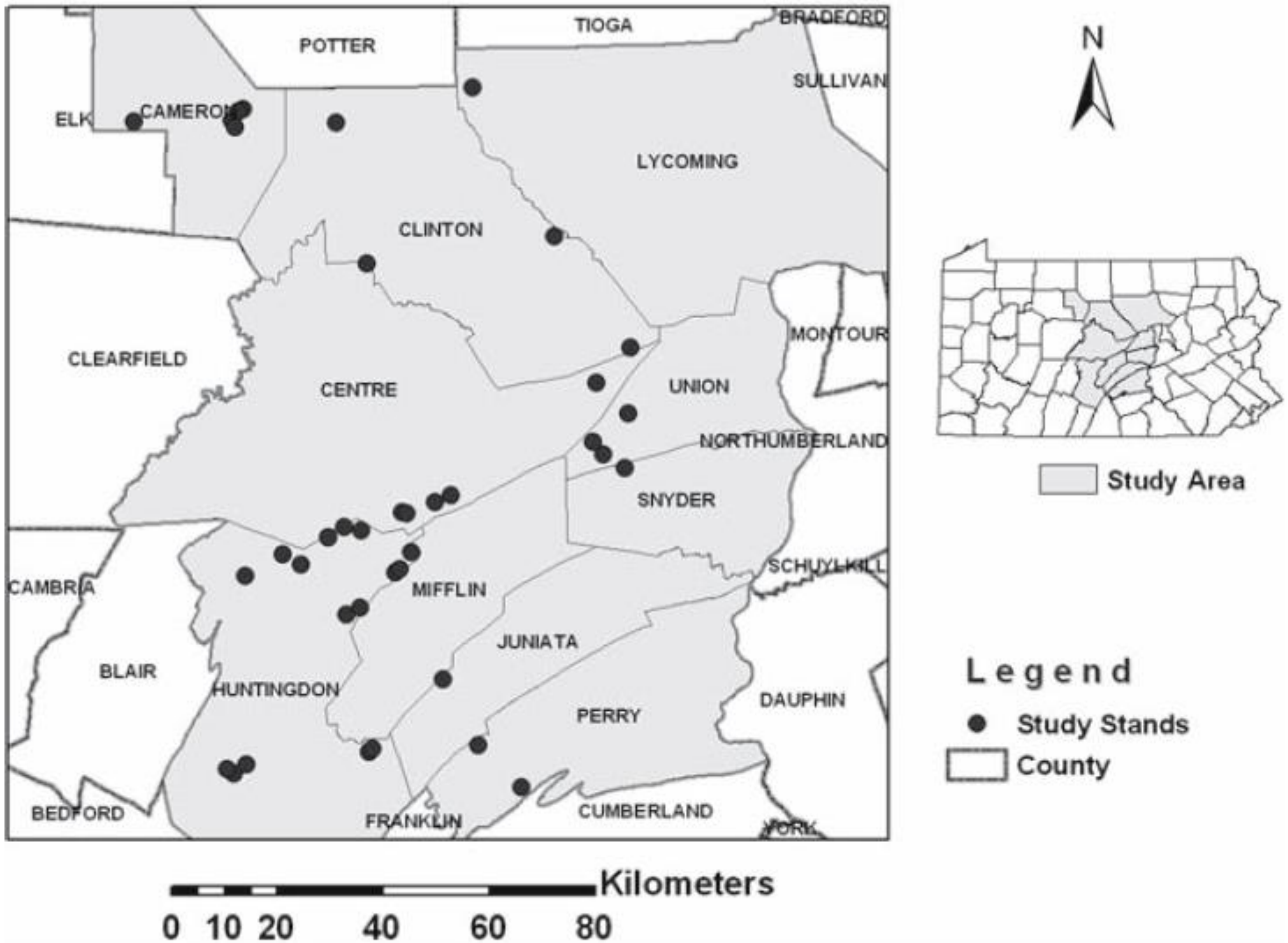


Fig. 1 Locations of the 38 mixed-oak stands in central Pennsylvania

The method of Kriging was first developed by Georges Matheron (1963), based on the Master's thesis of Krige (1951), a South African mining engineer who used a prototype of this technique to predict ore reserves. After several decades' development, Kriging has become a core geostatistics tool and is now used in many topic areas. For example, Bayraktar and Turalioglu (2005) used Kriging for air quality analysis, Emerson (2005) applied Kriging to natural resource analysis, and Zimmerman et al. (1998) used Kriging for water studies. Such methods can be used to predict count values at unmeasured locations while assessing the errors of these predictions. They rely on the notion that unobserved factors are autocorrelated over space, and the levels of autocorrelation decline with distance. Meanwhile, the values to be predicted may depend on several observable causal factors (e.g., number of lanes, posted speed limit, and facility type). These create a "trend" estimate, $\mu(s)$; so, in general, spatial variables can be defined as follows:

$$Z_i(s) = \mu_i(s) + \varepsilon_i(s) \qquad (1)$$

where $Z_i(s)$ is the variable of interest (actual traffic count here) and s gives location (x,y coordinates) of site i. $Z_i(s)$ is composed of a deterministic trend $\mu_i(s)$ and a random error component $\varepsilon_i(s)$. The various $\varepsilon(s)$ values are correlated over space. Features of "trend" (often called "drift" in other studies), or the expected value of $Z(s)$, result in three types of Kriging: If

Spatial Autocorrelation in a Sample Selection Model

Flores-Lagunes, A. and Schnier, K., “Sample selection and Spatial Dependence,” Journal of Applied Econometrics, 27, 2, 2012, pp. 173-204.

- Alaska Department of Fish and Game.
- Pacific cod fishing eastern Bering Sea – grid of locations
- Observation = ‘catch per unit effort’ in grid square
- Data reported only if 4+ similar vessels fish in the region
- 1997 sample = 320 observations with 207 reported full data

Spatial Autocorrelation in a Sample Selection Model

Flores-Lagunes, A. and Schnier, K., "Sample selection and Spatial Dependence," Journal of Applied Econometrics, 27, 2, 2012, pp. 173-204.

- LHS is catch per unit effort = CPUE
- Site characteristics: MaxDepth, MinDepth, Biomass
- Fleet characteristics:
  - Catcher vessel (CV = 0/1)
  - Hook and line (HAL = 0/1)
  - Nonpelagic trawl gear (NPT = 0/1)
  - Large (at least 125 feet) (Large = 0/1)

Spatial Autocorrelation in a Sample Selection Model

$$y_{i1}^* = \alpha_0 + \mathbf{x}_{i1}'\boldsymbol{\alpha} + u_{i1} \qquad u_{i1} = \delta \sum_{j \neq i} c_{ij} u_{j1} + \varepsilon_{i1}$$

$$y_{i2}^* = \beta_0 + \mathbf{x}_{i2}'\boldsymbol{\beta} + u_{i2} \qquad u_{i2} = \gamma \sum_{j \neq i} c_{ij} u_{j2} + \varepsilon_{i2}$$

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right], \quad (??\ \sigma_1 = 1??)$$

Observation Mechanism

$y_{i1} = 1\left[y_{i1}^* > 0\right]$ Probit Model

$y_{i2} = y_{i2}^*$ if $y_{i1} = 1$, unobserved otherwise.

Spatial Autocorrelation in a Sample Selection Model

$$\mathbf{u}_1 = \delta \mathbf{C} \mathbf{u}_1 + \boldsymbol{\varepsilon}_1$$

$\mathbf{C}$ = Spatial weight matrix, $\mathbf{C}_{ii} = 0.$

$\mathbf{u}_1 = [\mathbf{I} - \delta \mathbf{C}]^{-1} \boldsymbol{\varepsilon}_1 = \boldsymbol{\Omega}^{(1)} \boldsymbol{\varepsilon}_1$, likewise for $\mathbf{u}_2$

$$y_{i1}^* = \alpha_0 + \mathbf{x}_{i1}' \boldsymbol{\alpha} + \sum_{j=1}^{N} \omega(\delta)_{ij}^{(1)} \varepsilon_{i1}, \ \text{Var}[u_{i1}] = \sigma_1^2 \sum_{j=1}^{N} \left(\omega(\delta)_{ij}^{(1)} \right)^2$$

$$y_{i2}^* = \beta_0 + \mathbf{x}_{i2}' \boldsymbol{\beta} + \sum_{j=1}^{N} \omega(\gamma)_{ij}^{(2)} \varepsilon_{i2}, \ \text{Var}[u_{i2}] = \sigma_1^2 \sum_{j=1}^{N} \left(\omega(\gamma)_{ij}^{(2)} \right)^2$$

$$\text{Cov}[u_{i1}, u_{i2}] = \sigma_{12} \sum_{j=1}^{N} \omega(\delta)_{ij}^{(1)} \omega(\gamma)_{ij}^{(2)}$$

Spatial Weights

$$c_{ij} = \frac{1}{d_{ij}^2},$$

$d_{ij} =$ Euclidean distance

Band of 7 neighbors is used

Row standardized.

Two Step Estimation

❶ Probit estimated by Pinske/Slade GMM

$$\lambda_i = \frac{\sum_{j=1}^{N} \omega(\delta)_{ij}^{(1)} \omega(\gamma)_{ij}^{(2)}}{\sqrt{\sum_{j=1}^{N}\left(\omega(\delta)_{ij}^{(1)}\right)^2}} \frac{\phi\left[\dfrac{\alpha_0 + \mathbf{x}'_{i1}\boldsymbol{\alpha}}{\sqrt{\sigma_1^2 \sum_{j=1}^{N}\left(\omega(\delta)_{ij}^{(1)}\right)^2}}\right]}{\Phi\left[\dfrac{\alpha_0 + \mathbf{x}'_{i1}\boldsymbol{\alpha}}{\sqrt{\sigma_1^2 \sum_{j=1}^{N}\left(\omega(\delta)_{ij}^{(1)}\right)^2}}\right]}$$

❷ Spatial regression with included IMR in second step

(*) GMM procedure combines the two steps in one large estimation.

	Selection equation[c]			Employing probit[d]	
	Heckit	Spheck-E	Spheck-O	Spheck-E	Spheck-O
Constant	−0.104	−0.135	−0.098	4.971	5.141***
	(0.648)	(19.720)	(0.756)	(19.221)	(1.048)
Max. depth	0.179*	0.197	0.232**	0.321	0.285***
	(0.092)	(1.201)	(0.116)	(1.894)	(0.108)
Min. depth	−0.093	−0.088	−0.133*	−0.073	−0.051
	(0.068)	(0.572)	(0.073)	(0.888)	(0.061)
Biomass	0.005	0.005	0.007	0.200***	0.162***
	(0.078)	(0.842)	(0.072)	(0.038)	(0.037)
Dum CV	−0.739***	−0.696	−0.736***	0.020	0.024
	(0.183)	(1.883)	(0.257)	(9.657)	(0.537)
Dum HAL	0.650***	0.475	0.581**	0.931	1.072***
	(0.202)	(3.281)	(0.269)	(6.543)	(0.411)
Dum NPT	0.073	0.078	0.080	−0.474	−0.381**
	(0.261)	(4.654)	(0.313)	(0.834)	(0.178)
Dum Large	−0.078	−0.098	−0.088	0.494	0.399***
	(0.176)	(2.063)	(0.177)	(0.770)	(0.092)
IMR				2.205	2.543**
				(16.327)	(1.083)
Lag biomass	−0.043	−0.043	−0.069		
	(0.080)	(1.064)	(0.066)		
SAE parameter (γ)				0.947***	0.872***
				(0.141)	(0.070)
SAE parameter (δ)		0.392	0.203		
		(1.020)	(0.133)		

Spatial Stochastic Frontier

Production function model

$y = \boldsymbol{\beta}'\mathbf{x} + \varepsilon$

$y = \boldsymbol{\beta}'\mathbf{x} + v - u$

v = unexplained noise = N[0,1]

u = inefficiency > 0; efficiency = exp(-u)

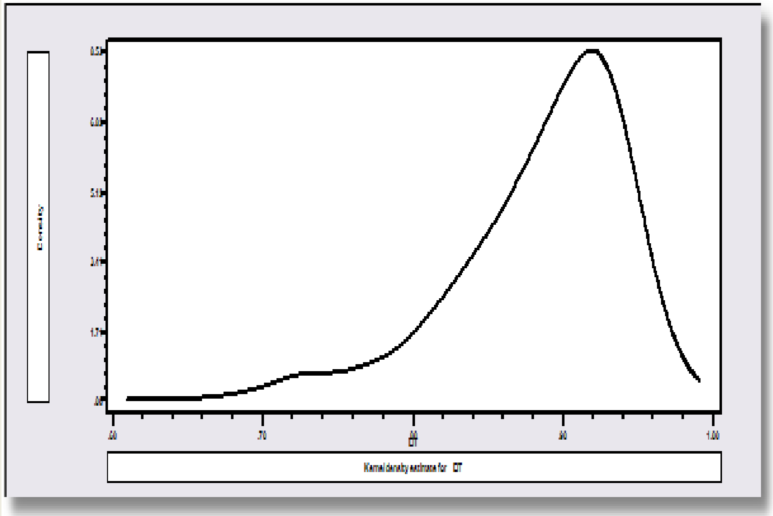
Object of estimation is u, not $\boldsymbol{\beta}$

Not a linear regression. Fit by MLE or MCMC.

247 Spanish Dairy Farms, 6 Years

	FARM	YEAR	COWS	LAND	MILK	LABOR	FEED
1 »	1	93	15.3	8	73647	2	33435.7
2 »	1	94	18.1	8	91260	2	36869
3 »	1	97	17.1	7	110419	2	51013.6
4 »	1	96	17.3	8	111454	2	50711.6
5 »	1	95	17.8	8	118498	2	54153.6
6 »	1	98	19.5	7.2	131197	2	59038.7
7 »	2	93	20.3	9	118149	2	53875.9
8 »	2	94	20.3	10.4	127742	2	51991

```
-----[ Tests vs. No Inefficiency ]-----
LR test for inefficiency vs. OLS v only
Deg. freedom for sigma-squared(u):    1
Deg. freedom for heteroscedasticity:  0
Deg. freedom for truncation mean:     0
Deg. freedom for inefficiency model:  1
LogL when sigma(u)=0          809.67610
Chi-sq=2*[LogL(SF)-LogL(LS)] =   26.024
Kodde-Palm C*: 95%: 2.706,  99%:  5.412
LM test for sigma(u) = 0 based on ols e
Chi-sq[1]=(N/6)*[m3/s^3]^2       21.665
Wald tests based on MLEs shown in table
--------+--------------------------------------------------------------
        |                  Standard            Prob.       95% Confidence
     YIT|  Coefficient       Error       z    |z|>Z*          Interval
--------+--------------------------------------------------------------
        |Deterministic Component of Stochastic Frontier Model
Constant|    11.7014***     .00447  2614.87  .0000    11.6926    11.7101
      X1|     .58369***     .01887    30.93  .0000     .54670     .62068
      X2|     .03555***     .01113     3.20  .0014     .01375     .05736
      X3|     .02256*       .01281     1.76  .0783    -.00256     .04768
      X4|     .44948***     .01035    43.42  .0000     .42919     .46977
        |Variance parameters for compound error
  Lambda|    1.50164***     .08748    17.17  .0000    1.33019    1.67310
   Sigma|     .18710***     .00011  1698.90  .0000     .18688     .18732
--------+--------------------------------------------------------------
```



A True Random Effects Model

Spatial Stochastic Frontier Models; Accounting for Unobserved Local Determinants of Inefficiency
Schmidt, Moriera, Helfand, Fonseca; Journal of Productivity Analysis, 2009.

$$y_{ij} = \text{Output of farm j in municipality i in Center-West Brazil}$$

$$y_{ij} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{ij} + v_{ij} - u_{ij}$$

$$(\alpha_1,...,\alpha_n) = \text{conditionally autoregressive based on neighbors}$$

$$\alpha_i - \alpha_k \text{ is smaller when municipalities i and k are closer together}$$

A Spatial Autoregressive Production Frontier Model for Panel Data: With an Application to European Countries

Anthony J. Glass[*], Karligash Kenjegalieva[†] and Robin Sickles[‡§]

Estimating Efficiency Spillovers with State Level Evidence for Manufacturing in the U.S.

Anthony Glass[a,*], Karligash Kenjegalieva[a], Robin C. Sickles[b]

[a] *School of Business and Economics, Loughborough University, Leics, LE11 3TU, UK*
[b] *Department of Economics, Rice University, Houston, U.S., and School of Business and Economics, Loughborough University, Leics, LE11 3TU, UK*

$$C_{it} = \kappa + \alpha_i + \tau_t + TL\,(h, q, t)_{it} + \lambda \sum_{j=1}^{N} w_{ij}\,C_{jt} + z_{it}\phi + \varepsilon_{it},$$
$$i = 1, \ldots, N;\; t = 1, \ldots, T.$$

Consider the following Cliff-Ord type production function for panel data:

$$\begin{aligned} y_{it} &= X_{it}\beta + \lambda \sum_{j=1}^{N} w_{ij} y_{jt} + \varepsilon_{it} \\ i &= 1, ..., N;\; t = 1, ..., T, \end{aligned}$$

Estimation by Maximum Likelihood

LeSage (2000) on Timing

"The Bayesian probit and tobit spatial autoregressive models described here have been applied to samples of 506 and 3,107 observations. The time required to produce estimates was around 350 seconds for the 506 observations sample and 900 seconds for the case infolving 3,107 observations. … (inexpensive Apple G3 computer running at 266 Mhz.)"

Time and Space (In Your Computer)

```
-----------------------------------------------------------------------------
Binomial Probit Model
Dependent variable                    IP
Log likelihood function      -4134.21620
Restricted log likelihood    -4283.16647
Chi squared [   6](P= .000)    297.90053
Significance level                .00000
McFadden Pseudo R-squared       .0347757
Estimation based on N =   6350, K =    7
Inf.Cr.AIC  =   8282.4 AIC/N =     1.304
--------+--------------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
      IP|  Coefficient     Error       z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
        |Index function for probability
Constant|    -1.76287***     .56235    -3.13  .0017    -2.86506    -.66067
      IM|      .44303        .39455     1.12  .2615     -.33028    1.21635
    IMUM|      .67753***     .25128     2.70  .0070      .18502    1.17003
   FDIUM|     3.19779***     .43002     7.44  .0000     2.35496    4.04061
      SP|     1.15566***     .14063     8.22  .0000      .88002    1.43129
    PROD|    -4.70073***     .55281    -8.50  .0000    -5.78422   -3.61725
LOGSALES|      .18656***     .05378     3.47  .0005      .08114     .29197
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 14, 2013 at 02:47:14 PM
-----------------------------------------------------------------------------
Elapsed time:     0 hours,  0 minutes,  .148 seconds.
```

Efficient Spatial Econometric Model Estimation with Very Large Datasets Using the Maximum Likelihood Coding Technique

Giuseppe Arbia, University "G. d'Annunzio" of Pescara and "Catholic University of the Sacred Hearth", Rome

and

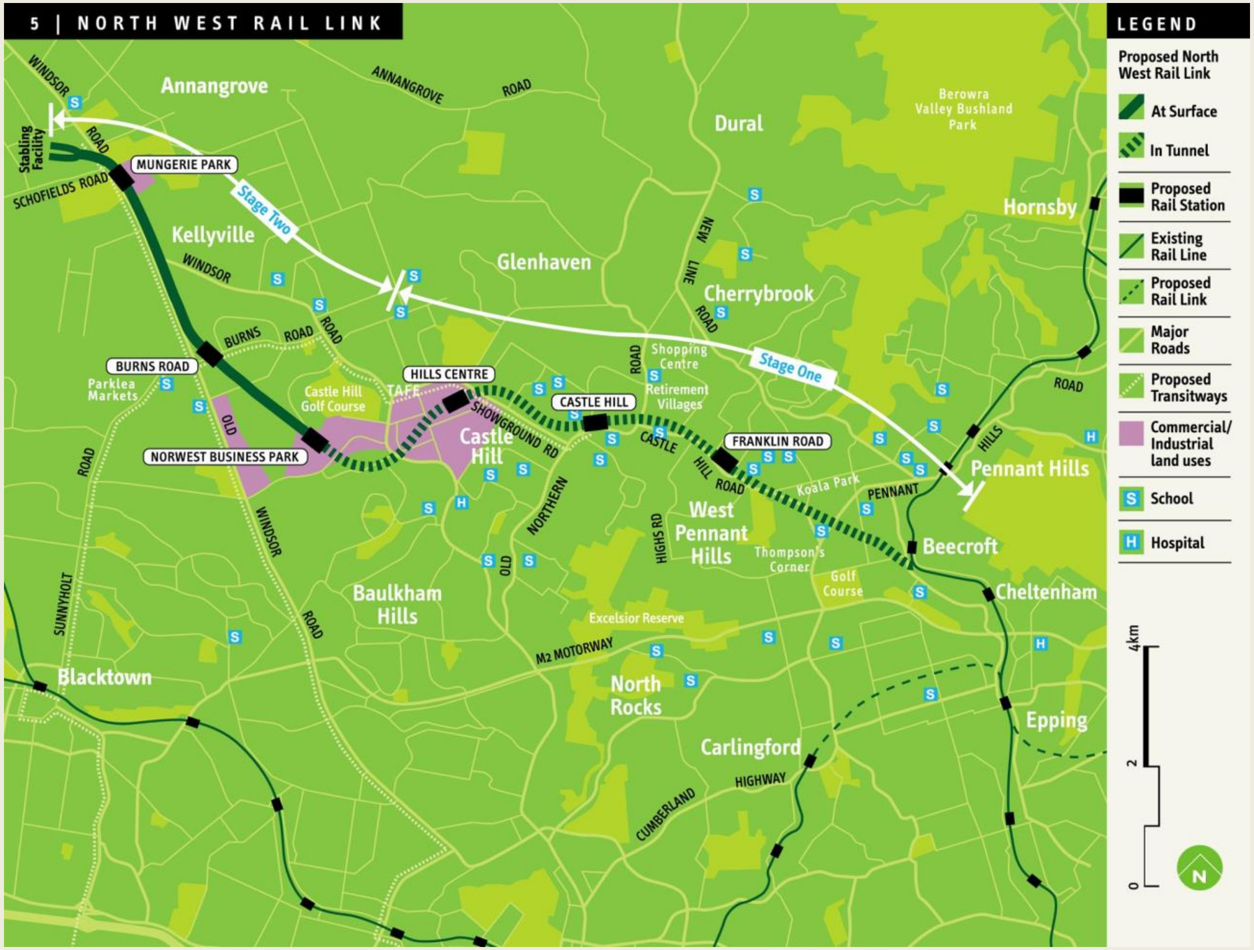
Pedro Amaral, University of Cambridge (UK)

Computation time is still an issue in econometrics !!

In the spatial econometric modeling when we follow a maximum likelihood approach, the problem originates mainly from the inversion of the matrix

$$(I - \rho W)$$


The VII World Conference of the Spatial Econometrics Association


5 | NORTH WEST RAIL LINK
LEGEND
Proposed North West Rail Link
At Surface
In Tunnel
Proposed Rail Station
Existing Rail Line
Proposed Rail Link
Major Roads
Proposed Transitways
Commercial/ Industrial land uses
School
Hospital
4km
2
0
N
Annangrove
Dural
Berowra Valley Bushland Park
Hornsby
Kellyville
Glenhaven
Cherrybrook
MUNGERIE PARK
Stage Two
Stage One
Stabling Facility
SCHOFIELDS ROAD
BURNS ROAD
Parklea Markets
Castle Hill Golf Course
TAFE
HILLS CENTRE
CASTLE HILL
Shopping Centre
Retirement Villages
FRANKLIN ROAD
NORWEST BUSINESS PARK
Castle Hill
SHOWGROUND RD
West Pennant Hills
Koala Park
Pennant Hills
Beecroft
Cheltenham
Thompson's Corner
Golf Course
Baulkham Hills
Excelsior Reserve
M2 MOTORWAY
North Rocks
Carlingford
Epping
Blacktown
HIGHWAY
CUMBERLAND

Thank you

- ① ② ③ ④ ⑤
- ❶ ❷ ❸ ❹ ❺ ❻