Spatial Discrete Choice Models

Professor William Greene
Stern School of Business, New York University

SPATIAL ECONOMETRICS ADVANCED INSTITUTE
University of Rome
May 23, 2011
Spatial Correlation

Alternative Hypotheses of SA

Positive Spatial Autocorrelation

- like values tend to \textit{cluster} in space
- neighbors are similar
- compatible with diffusion
Spatiotially Autocorrelated Data

Per Capita Income in Monroe County, New York, USA
The Hypothesis of Spatial Autocorrelation

Random or Clustered?
Spatial Discrete Choice Modeling: Agenda

1. Linear Models with Spatial Correlation
2. Discrete Choice Models
3. Spatial Correlation in Nonlinear Models
   • Basics of Discrete Choice Models
   • Maximum Likelihood Estimation
4. Spatial Correlation in Discrete Choice
   • Binary Choice
   • Ordered Choice
   • Unordered Multinomial Choice
   • Models for Counts
Linear Spatial Autocorrelation

\[(\mathbf{x} - \mu_i) = \lambda \mathbf{W}(\mathbf{x} - \mu_i) + \mathbf{\varepsilon},\]  
\[N \text{ observations on a spatially arranged variable}\]

\[\mathbf{W} = \text{'contiguity matrix';} \quad \mathbf{W}_{ii} = 0\]

\[\mathbf{W} \text{ must be specified in advance. It is not estimated.}\]

\[\lambda = \text{ spatial autocorrelation parameter, } -1 < \lambda < 1.\]

\[E[\mathbf{\varepsilon}] = \mathbf{0}, \quad \text{Var}[\mathbf{\varepsilon}] = \sigma^2_\varepsilon \mathbf{I}\]

\[(\mathbf{x} - \mu_i) = [\mathbf{I} - \lambda \mathbf{W}]^{-1} \mathbf{\varepsilon} = \text{Spatial "moving average" form}\]

\[E[\mathbf{x}] = \mu_i, \quad \text{Var}[\mathbf{x}] = \sigma^2_\varepsilon [(\mathbf{I} - \lambda \mathbf{W})'(\mathbf{I} - \lambda \mathbf{W})]^{-1}\]
Testing for Spatial Autocorrelation

Moran’s $I$ Spatial Autocorrelation Statistic

- cross-product statistic

$$I = \left(\frac{N}{S_0}\right) \sum_i \sum_j w_{ij} \cdot z_i \cdot z_j / \sum_i z_i^2$$

with $z_i = x_i - \mu$ and $S_0 = \sum_i \sum_j w_{ij}$
**Spatial Autocorrelation in Real Estate Sales**

Bell and Bockstael analyzed the problem of modeling spatial autocorrelation in large samples. This is likely to become an increasingly common problem with GIS (geographic information system) data sets. The central problem is maximization of a likelihood function that involves a sparse matrix, \((I - \lambda W)\). Direct approaches to the problem can encounter severe inaccuracies in evaluation of the inverse and determinant. Kelejian and Prucha (1999) have developed a moment-based estimator for \(\lambda\) that helps to alleviate the problem. Once the estimate of \(\lambda\) is in hand, estimation of the spatial autocorrelation model is done by FGLS. The authors applied the method to analysis of a cross section of 1,000 residential sales in Anne Arundel County, Maryland, from 1993 to 1996. The parcels sold all involved houses built within one year prior to the sale. GIS software was used to measure attributes of interest.

The model is

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

(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

\[
y = X\beta + \varepsilon, \\
\varepsilon = \lambda W\varepsilon + v.
\]
Spatial Autocorrelation

\[ y = X\beta + \lambda W\varepsilon. \]

\[ \mathbb{E}[\varepsilon | X] = 0, \quad \text{Var}[\varepsilon | X] = \sigma^2 \mathbf{I} \]

\[ \mathbb{E}[y | X] = X\beta \]

\[ \text{Var}[y | X] = \lambda^2 \sigma^2 \mathbf{W}\mathbf{W}' \]

A Generalized Regression Model
Spatial Autoregression in a Linear Model

\[
y = \lambda Wy + X\beta + \epsilon.
\]

\[
E[\epsilon | X] = 0, \quad \text{Var}[\epsilon | X] = \sigma^2 \mathbf{I}
\]

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

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

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

\[
\text{Var}[y | X] = \sigma^2 \varepsilon [(\mathbf{I} - \lambda \mathbf{W})'(\mathbf{I} - \lambda \mathbf{W})]^{-1}
\]
Complications of the Generalized Regression Model

1. Potentially very large N – GPS data on agriculture plots
2. Estimation of \( \lambda \). There is no natural residual based estimator
3. Complicated covariance structure – no simple transformations
Panel Data Application

E.g., N countries, T periods (e.g., gasoline data)

\[ y_{it} = x'_{it} \beta + c_i + \epsilon_{it} \]

\[ \epsilon_t = \lambda W \epsilon_t + v_t = N \text{ observations at time } t. \]

Similar assumptions

Candidate for SUR or Spatial Autocorrelation model.
Spatial Autocorrelation in a Panel

Spatial Lags in Health Expenditures

Moscone, Knapp, and Tossetti (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:

\[ y_t = \gamma_i + \rho W y_t + X_t \beta + u + \varepsilon_t, \]

where \( u \) is a \( 148 \times 1 \) vector of random effects and \( i \) is a \( 148 \times 1 \) column of ones. For each local authority,

\[ y_{it} = \gamma_t + \rho (w_i' y_t) + x_{it} \beta + u_i + \varepsilon_{it}, \]

where \( w_i \) is the \( i \)th row of the contiguity matrix, \( W \). Contiguities were defined in \( W \) 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

\[ y_t = \gamma_i + X_t \beta + u + \varepsilon_t, \]

\[ \varepsilon_t = \lambda W \varepsilon_t + v_t. \]
Alternative Panel Formulations

**Pure space-recursive** - dependence pertains to neighbors in period t-1

\[ y_{i,t} = \gamma [W_y]_{t-1,i} + \text{regression} + \epsilon_{it} \]

**Time-space recursive** - dependence is pure autoregressive and on neighbors in period t-1

\[ y_{i,t} = \rho y_{i,t-1} + \gamma [W_y]_{t-1,i} + \text{regression} + \epsilon_{it} \]

**Time-space simultaneous** - dependence is autoregressive and on neighbors in the current period

\[ y_{i,t} = \rho y_{i,t-1} + \lambda [W_y]_{t,i} + \text{regression} + \epsilon_{it} \]

**Time-space dynamic** - dependence is autoregressive and on neighbors in both current and last period

\[ y_{i,t} = \rho y_{i,t-1} + \lambda [W_y]_{t,i} + \gamma [W_y]_{t-1,i} + \text{regression} + \epsilon_{it} \]
Analytical Environment

1. Generalized linear regression
2. Complicated disturbance covariance matrix
3. Estimation platform
   - Generalized least squares
   - Maximum likelihood estimation when normally distributed disturbances (still GLS)
Discrete Choices

1. Land use intensity in Austin, Texas – Intensity = 1,2,3,4
2. Land Usage Types in France, 1,2,3
3. Oak Tree Regeneration in Pennsylvania Number = 0,1,2,... (Many zeros)
4. Teenagers physically active = 1 or physically inactive = 0, in Bay Area, CA.
Discrete Choice Modeling

1. Discrete outcome reveals a specific choice

2. Underlying preferences are modeled

3. Models for observed data are usually not conditional means
   - Generally, probabilities of outcomes
   - Nonlinear models – cannot be estimated by any type of linear least squares
Discrete Outcomes

1. Discrete Revelation of Underlying Preferences
   - Binary choice between two alternatives
   - Unordered choice among multiple alternatives
   - Ordered choice revealing underlying strength of preferences

2. Counts of Events
Simple Binary Choice: Insurance

![Bar Chart: Choice of Public Insurance](image_url)

- Frequency of Public Insurance: 26624
- Frequency of Non-Public Insurance: 6656
- Total Observations: 33280
Redefined Multinomial Choice

Mode Choice for Sydney/Melbourne Travel

- **Fly**
  - Air: 72
  - Train: 54
  - Bus: 36
  - Car: 18

- **Ground**
  - Air: 36
  - Train: 72
  - Bus: 18
  - Car: 54
Multinomial Unordered Choice - Transport Mode

Mode Choice for Sydney/Melbourne Travel

<table>
<thead>
<tr>
<th>Choice</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIR</td>
<td>54</td>
</tr>
<tr>
<td>TRAIN</td>
<td>72</td>
</tr>
<tr>
<td>BUS</td>
<td>18</td>
</tr>
<tr>
<td>CAR</td>
<td>54</td>
</tr>
</tbody>
</table>
Health Satisfaction (HSAT)

Self administered survey: Health Care Satisfaction? (0 – 10)

Histogram for NewHSAT - Full Sample

Continuous Preference Scale
Ordered Preferences at IMDB.com

User ratings for
National Treasure: Book of Secrets

41,771 IMDb users have given a weighted average vote of 6.6

Demographic breakdowns are shown below.

Votes

Percentage

Rating

Votes

Average

Males

33,644

6.5

Females

5,464

7.2

Aged under 18

2,492

7.6

Males under 18

1,795

7.5

Females under 18

695

8.1

Aged 18-29

26,045

6.7

Males Aged 18-29

22,603

6.6

Females Aged 18-29

3,372

7.3

Aged 30-44

8,210

6.3

Males Aged 30-44

7,216

6.3

Females Aged 30-44

936

6.7

Aged 45+

2,258

6.6

Males Aged 45+

1,814

6.5

Females Aged 45+

420

7.0

IMDb staff

8

6.4

Top 1000 voters

309

6.0

US users

14,792

6.8

Non-US users

24,283

6.5

IMDb users 41,771

Arithmetic mean = 6.9. Median = 7
Counts of Events

Feeeway Accidents by Region and Season

Frequency

0 1 2 3 4 5 6 7 8 9 10

0 371 742 1113 1484

WINTER SUMMER
Modeling Discrete Outcomes

1. “Dependent Variable” typically labels an outcome
   - No quantitative meaning
   - Conditional relationship to covariates

2. No “regression” relationship in most cases

3. The “model” is usually a probability
Simple Binary Choice: Insurance

Decision: Yes or No = 1 or 0
Depends on Income, Health, Marital Status, Gender
Multinomial Unordered Choice - Transport Mode

Decision: Which Type, A, T, B, C.
Depends on Income, Price, Travel Time
Health Satisfaction (HSAT)

Self administered survey: Health Care Satisfaction? (0 – 10)

Outcome: Preference = 0,1,2,...,10
Depends on Income, Marital Status, Children, Age, Gender
Counts of Events

Outcome: How many events at each location = 0,1,...,10
Depends on Season, Population, Economic Activity
Nonlinear Spatial Modeling

1. Discrete outcome $y_{it} = 0, 1, ..., J$ for some finite or infinite (count case) $J$.
   - $i = 1,...,n$
   - $t = 1,...,T$

2. Covariates $x_{it}$.

3. Conditional Probability $(y_{it} = j)$
   $= a$ function of $x_{it}$. 
Two Platforms

1. Random Utility for Preference Models
   Outcome reveals underlying utility
   - Binary: $u^* = \theta'x \quad y = 1$ if $u^* > 0$
   - Ordered: $u^* = \theta'x \quad y = j$ if $\mu_{j-1} < u^* < \mu_j$
   - Unordered: $u^*(j) = \theta'x_j, y = j$ if $u^*(j) > u^*(k)$

2. Nonlinear Regression for Count Models
   Outcome is governed by a nonlinear regression
   - $E[y|x] = g(\theta, x)$
Probit and Logit Models

\[ \text{Prob}(y_i = 1 \text{ or } 0|x_i) = F(\theta'x_i) \text{ or } [1- F(\theta'x_i)] \]
Implied Regression Function
Estimated Binary Choice Models: The Results Depend on F(ε)

<table>
<thead>
<tr>
<th>Variable</th>
<th>LOGIT</th>
<th>PROBIT</th>
<th>EXTREME VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t-ratio</td>
<td>Estimate</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.42085</td>
<td>-2.662</td>
<td>-0.25179</td>
</tr>
<tr>
<td>X1</td>
<td>0.02365</td>
<td>7.205</td>
<td>0.01445</td>
</tr>
<tr>
<td>X2</td>
<td>-0.44198</td>
<td>-2.610</td>
<td>-0.27128</td>
</tr>
<tr>
<td>X3</td>
<td>0.63825</td>
<td>8.453</td>
<td>0.38685</td>
</tr>
<tr>
<td>Log-L</td>
<td>-2097.48</td>
<td></td>
<td>-2097.35</td>
</tr>
<tr>
<td>Log-L(0)</td>
<td>-2169.27</td>
<td></td>
<td>-2169.27</td>
</tr>
</tbody>
</table>
Effect on Predicted Probability of an Increase in X1

\[ \alpha + \beta_1 (X1 + 1) + \beta_2 (X2) + \beta_3 X3 \ (\beta_1 \text{ is positive}) \]
# Estimated Partial Effects vs. Coefficients

<table>
<thead>
<tr>
<th>Variable</th>
<th>LOGIT</th>
<th>PROBIT</th>
<th>EXTREME VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t-ratio</td>
<td>Estimate</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.42085</td>
<td>-2.662</td>
<td>-0.25179</td>
</tr>
<tr>
<td>X1</td>
<td>0.02365</td>
<td>7.205</td>
<td>0.01445</td>
</tr>
<tr>
<td>X2</td>
<td>-0.44198</td>
<td>-2.610</td>
<td>-0.27128</td>
</tr>
<tr>
<td>X3</td>
<td>0.63825</td>
<td>8.453</td>
<td>0.38685</td>
</tr>
<tr>
<td>Log-L</td>
<td>-2097.48</td>
<td></td>
<td>-2097.35</td>
</tr>
<tr>
<td>Log-L(0)</td>
<td>-2169.27</td>
<td></td>
<td>-2169.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOGIT</th>
<th>PROBIT</th>
<th>EXTREME VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>t ratio</td>
<td>Estimate</td>
</tr>
<tr>
<td>.00527</td>
<td>7.235</td>
<td>.00527</td>
</tr>
<tr>
<td>-.09844</td>
<td>-2.611</td>
<td>-.09897</td>
</tr>
<tr>
<td>.14026</td>
<td>8.663</td>
<td>.13958</td>
</tr>
</tbody>
</table>
Applications: Health Care Usage

German Health Care Usage Data, 7,293 Individuals, Varying Numbers of Periods

Variables in the file are

Data downloaded from Journal of Applied Econometrics Archive. This is an unbalanced panel with 7,293 individuals. They can be used for regression, count models, binary choice, ordered choice, and bivariate binary choice. This is a large data set. There are altogether 27,326 observations. The number of observations ranges from 1 to 7. (Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987). (Downloaded from the JAE Archive)

- **DOCTOR** = 1(Number of doctor visits > 0)
- **HOSPITAL** = 1(Number of hospital visits > 0)
- **HSAT** = health satisfaction, coded 0 (low) - 10 (high)
- **DOCVIS** = number of doctor visits in last three months
- **HOSPVIS** = number of hospital visits in last calendar year
- **PUBLIC** = insured in public health insurance = 1; otherwise = 0
- **ADDON** = insured by add-on insurance = 1; otherwise = 0
- **HHNINC** = household nominal monthly net income in German marks / 10000.
  (4 observations with income=0 were dropped)
- **HHKIDS** = children under age 16 in the household = 1; otherwise = 0
- **EDUC** = years of schooling
- **AGE** = age in years
- **FEMALE** = 1 for female headed household, 0 for male
- **EDUC** = years of education
An Estimated Binary Choice Model

Binomial Probit Model
Dependent variable: DOCTOR

<table>
<thead>
<tr>
<th>Index function for probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant: 1.04721***</td>
</tr>
<tr>
<td>AGE: .00708***</td>
</tr>
<tr>
<td>EDUC: .56427D-04</td>
</tr>
<tr>
<td>HSAT: -1.4285***</td>
</tr>
<tr>
<td>MARRIED: .11391**</td>
</tr>
<tr>
<td>HHNINC: -.32547**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DOCTOR</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>z</th>
<th>Prob.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.04721***</td>
<td>.15754</td>
<td>6.65</td>
<td>.0000</td>
<td>.73844 - 1.35597</td>
</tr>
<tr>
<td>AGE</td>
<td>.00708***</td>
<td>.00189</td>
<td>3.76</td>
<td>.0002</td>
<td>.00339 - .01078</td>
</tr>
<tr>
<td>EDUC</td>
<td>.56427D-04</td>
<td>.08686</td>
<td>0.01</td>
<td>.9949</td>
<td>-1.7312D-01 - .17425D-01</td>
</tr>
<tr>
<td>HSAT</td>
<td>-1.4285***</td>
<td>.01018</td>
<td>-14.03</td>
<td>.0000</td>
<td>-.16282 - -.12289</td>
</tr>
<tr>
<td>MARRIED</td>
<td>.11391**</td>
<td>.04830</td>
<td>2.36</td>
<td>.0184</td>
<td>.01924 - .20858</td>
</tr>
<tr>
<td>HHNINC</td>
<td>-.32547**</td>
<td>.12788</td>
<td>-2.55</td>
<td>.0109</td>
<td>-.57610 - -.07484</td>
</tr>
</tbody>
</table>

Note: nnnnn.D-xx or D+xx => multiply by 10 to -xx or +xx.
Note: ***, **, * => Significance at 1%, 5%, 10% level.

Partial Effects for Probit Probability Function
Partial Effects Averaged Over Observations
* => Partial Effect for a Binary Variable

<table>
<thead>
<tr>
<th>(Delta method)</th>
<th>Partial Effect</th>
<th>Standard Error</th>
<th></th>
<th></th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>.00254</td>
<td>.00067</td>
<td>3.78</td>
<td>.00122</td>
<td>.00386</td>
</tr>
<tr>
<td>EDUC</td>
<td>.00002</td>
<td>.00318</td>
<td>.01</td>
<td>-.00621</td>
<td>.00625</td>
</tr>
<tr>
<td>HSAT</td>
<td>-.05122</td>
<td>.00340</td>
<td>15.04</td>
<td>-.05789</td>
<td>-.04454</td>
</tr>
<tr>
<td>MARRIED</td>
<td>.04124</td>
<td>.01762</td>
<td>2.34</td>
<td>.00669</td>
<td>.07578</td>
</tr>
<tr>
<td>HHNINC</td>
<td>-.11669</td>
<td>.04574</td>
<td>2.55</td>
<td>-.20634</td>
<td>-.02703</td>
</tr>
</tbody>
</table>
An Estimated Ordered Choice Model

Ordered Probability Model
Dependent variable: HLTHSAT
Log likelihood function: -5403.29308
Restricted log likelihood: -5527.56097
Chi squared [5 d.f.]: 248.53578
Significance level: 0.0000
McFadden Pseudo R-squared: 0.224815
Estimation based on N = 4243, K = 8
Inf Cr. AIC = 10822.586 AIC/N = 2.551
Model estimated: May 10, 2011, 20:56:05

Underlying probabilities based on Normal

<table>
<thead>
<tr>
<th>HLTHSAT</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>z</th>
<th>Prob.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
</table>
| Index function for probability
| Constant | 1.59363***  | .14115         | 11.31 | 0.0000 | 1.31919 | 1.87248 |
| AGE     | -.01982***  | .00153         | -13.00 | 0.0000 | -0.2281 | -0.1683 |
| EDUC    | .04018***   | .00781         | 5.14  | 0.0000 | .02487  | .05549  |
| HHNINC  | .10593      | .10743         | .99   | .3241 | -1.0683 | .31649  |
| PUBLIC  | -.04299     | .05317         | -.81  | .4188 | -1.4720 | .06123  |
| MARRIED | -.02737     | .04005         | -.68  | .4944 | -1.0589 | .05115  |

| Threshold parameters for index
| Mu(1) | .84699*** | .01816 | 46.65 | .0000 | .81140 | .88257 |
| Mu(2) | 2.00185*** | .02361 | 84.78 | .0000 | 1.95558 | 2.04813 |

Note: ***, **, * --> Significance at 1%, 5%, 10% level.

Partial Effects Analysis for Ordered Probit Probability Y = 3

| df/dAGE (Delta method) | Partial Effect | Standard Error | |t| | 95% Confidence Interval |
|------------------------|----------------|----------------|-----|-----|------------------------|
| AFE Prob(y= 0)         | .00400         | .00032         | 12.53 | .00337 | .00462     |
| AFE Prob(y= 1)         | .00327         | .00023         | 13.92 | .00281 | .00372     |
| AFE Prob(y= 2)         | -.00166       | .00016         | 10.30 | -.00201 | -.00137 |
| AFE Prob(y= 3)         | -.00557       | .00043         | 13.01 | -.00641 | -.00473 |
# An Estimated Count Data Model

## Poisson Regression

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>HOSPVIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood function</td>
<td>-1909.85190</td>
</tr>
<tr>
<td>Restricted log likelihood</td>
<td>-1920.40214</td>
</tr>
<tr>
<td>Chi squared [ 3 d.f.]</td>
<td>21.10048</td>
</tr>
<tr>
<td>Significance level</td>
<td>.00010</td>
</tr>
<tr>
<td>McFadden Pseudo R-squared</td>
<td>.0054938</td>
</tr>
</tbody>
</table>

Estimation based on N = 4243, K = 4

Inf.Cr.AIC = 3827.704 AIC/N = .902

Model estimated: May 10, 2011, 20:59:10

Chi-squared = 14906.14910 RsqF= .0818

G-squared = 3039.60991 RsqD= .0069

Overdispersion tests: g=μ(i): 3.484

Overdispersion tests: g=μ(i)^2: 3.043

<table>
<thead>
<tr>
<th>HOSPVIS</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>z</th>
<th>Prob.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-53014</td>
<td>.37318</td>
<td>-1.42</td>
<td>.1554</td>
<td>-1.26156 - .20128</td>
</tr>
<tr>
<td>AGE</td>
<td>-.00611</td>
<td>.00309</td>
<td>-1.57</td>
<td>.1169</td>
<td>-.01374 - .00151</td>
</tr>
<tr>
<td>EDUC</td>
<td>-.09728***</td>
<td>.02623</td>
<td>-4.30</td>
<td>.0000</td>
<td>-.14164 - .05292</td>
</tr>
<tr>
<td>PUBLIC</td>
<td>-.20033</td>
<td>.13553</td>
<td>-1.48</td>
<td>.1394</td>
<td>-.46597 - .06531</td>
</tr>
</tbody>
</table>

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

---

Partial derivatives of expected val. with respect to the vector of characteristics.
Effects are averaged over individuals.
Observations used for means are All Obs.
Conditional Mean at Sample Point .1275
Scale Factor for Marginal Effects .1275

<table>
<thead>
<tr>
<th>HOSPVIS</th>
<th>Partial Effect</th>
<th>Standard Error</th>
<th>z</th>
<th>Prob.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>-.00078</td>
<td>.00051</td>
<td>-1.52</td>
<td>.1274</td>
<td>-.00178 - .00022</td>
</tr>
<tr>
<td>EDUC</td>
<td>-.01240***</td>
<td>.00321</td>
<td>-4.37</td>
<td>.0001</td>
<td>-.01869 - .00612</td>
</tr>
<tr>
<td>PUBLIC</td>
<td>-.02762</td>
<td>.02069</td>
<td>-1.33</td>
<td>.1820</td>
<td>-.06818 - .01294</td>
</tr>
</tbody>
</table>

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.
# 210 Observations on Travel Mode Choice

<table>
<thead>
<tr>
<th>CHOICE</th>
<th>ATTRIBUTES</th>
<th>CHARACTERISTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODE</td>
<td>TRAVEL</td>
<td>HINC</td>
</tr>
<tr>
<td>AIR</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>TRAIN</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>BUS</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>CAR</td>
<td>1.0000</td>
<td>35.000</td>
</tr>
<tr>
<td>AIR</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>TRAIN</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>BUS</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>CAR</td>
<td>1.0000</td>
<td>35.000</td>
</tr>
<tr>
<td>AIR</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>TRAIN</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>BUS</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>CAR</td>
<td>1.0000</td>
<td>35.000</td>
</tr>
<tr>
<td>AIR</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>TRAIN</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>BUS</td>
<td>.00000</td>
<td>35.000</td>
</tr>
<tr>
<td>CAR</td>
<td>1.0000</td>
<td>35.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INVC</th>
<th>INVT</th>
<th>TTME</th>
<th>GC</th>
</tr>
</thead>
<tbody>
<tr>
<td>59.000</td>
<td>100.00</td>
<td>69.000</td>
<td>70.000</td>
</tr>
<tr>
<td>31.000</td>
<td>372.00</td>
<td>34.000</td>
<td>71.000</td>
</tr>
<tr>
<td>25.000</td>
<td>417.00</td>
<td>35.000</td>
<td>70.000</td>
</tr>
<tr>
<td>10.000</td>
<td>180.00</td>
<td>.00000</td>
<td>30.000</td>
</tr>
<tr>
<td>58.000</td>
<td>68.000</td>
<td>64.000</td>
<td>68.000</td>
</tr>
<tr>
<td>31.000</td>
<td>354.00</td>
<td>44.000</td>
<td>84.000</td>
</tr>
<tr>
<td>25.000</td>
<td>399.00</td>
<td>53.000</td>
<td>85.000</td>
</tr>
<tr>
<td>11.000</td>
<td>255.00</td>
<td>.00000</td>
<td>50.000</td>
</tr>
<tr>
<td>127.00</td>
<td>193.00</td>
<td>69.000</td>
<td>148.00</td>
</tr>
<tr>
<td>109.00</td>
<td>888.00</td>
<td>34.000</td>
<td>205.00</td>
</tr>
<tr>
<td>52.000</td>
<td>1025.0</td>
<td>60.000</td>
<td>163.00</td>
</tr>
<tr>
<td>50.000</td>
<td>892.00</td>
<td>.00000</td>
<td>147.00</td>
</tr>
<tr>
<td>44.000</td>
<td>100.00</td>
<td>64.000</td>
<td>59.000</td>
</tr>
<tr>
<td>25.000</td>
<td>351.00</td>
<td>44.000</td>
<td>78.000</td>
</tr>
<tr>
<td>20.000</td>
<td>361.00</td>
<td>53.000</td>
<td>75.000</td>
</tr>
<tr>
<td>5.0000</td>
<td>180.00</td>
<td>.00000</td>
<td>32.000</td>
</tr>
</tbody>
</table>

An Estimated Unordered Choice Model

Discrete choice (multinomial logit) model
Dependent variable: Choice
Log likelihood function: -184.50669
Estimation based on N = 210, K = 7
Inf.Cr.AIC = 383.013, AIC/N = 1.824
Model estimated: May 10, 2011, 21:00:53
R2=1-LogL/LogL* Log-L fcn R-sqrd R2Adj
Constants only -283.7588 .3498 .3425
Chi-squared[ 4] = 198.50415
Prob [chi squared > value] = 0.0000
Response data are given as ind. choices
Number of obs. = 210, skipped 0 obs

<table>
<thead>
<tr>
<th>MODE</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>z</th>
<th>Prob.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td>.06930***</td>
<td>.01743</td>
<td>3.97</td>
<td>.0001</td>
<td>.03513 - .10346</td>
</tr>
<tr>
<td>TME</td>
<td>-.10365***</td>
<td>.01094</td>
<td>-9.48</td>
<td>.0000</td>
<td>-.12509 - -.08221</td>
</tr>
<tr>
<td>INVC</td>
<td>-.08493***</td>
<td>.01938</td>
<td>-4.38</td>
<td>.0000</td>
<td>-.12292 - -.04694</td>
</tr>
<tr>
<td>INVT</td>
<td>-.01333***</td>
<td>.00252</td>
<td>-5.30</td>
<td>.0000</td>
<td>-.01827 - -.00840</td>
</tr>
<tr>
<td>A_AIR</td>
<td>5.20474***</td>
<td>.90521</td>
<td>5.75</td>
<td>.0000</td>
<td>3.43056 - 6.97893</td>
</tr>
<tr>
<td>A_TRAIN</td>
<td>4.36060***</td>
<td>.51067</td>
<td>8.54</td>
<td>.0000</td>
<td>3.35972 - 5.36149</td>
</tr>
<tr>
<td>A_BUS</td>
<td>3.76323***</td>
<td>.50626</td>
<td>7.43</td>
<td>.0000</td>
<td>2.77098 - 4.75548</td>
</tr>
</tbody>
</table>

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Elasticity wrt change of X in row choice on Prob[column choice]

<table>
<thead>
<tr>
<th>GC</th>
<th>AIR</th>
<th>TRAIN</th>
<th>BUS</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIR</td>
<td>4.9664</td>
<td>-2.1466</td>
<td>-2.1466</td>
<td>-2.1466</td>
</tr>
<tr>
<td>TRAIN</td>
<td>-2.1912</td>
<td>6.8310</td>
<td>-2.1912</td>
<td>-2.1912</td>
</tr>
<tr>
<td>BUS</td>
<td>-1.0547</td>
<td>-1.0547</td>
<td>6.9321</td>
<td>-1.0547</td>
</tr>
<tr>
<td>CAR</td>
<td>-1.8020</td>
<td>-1.8020</td>
<td>-1.8020</td>
<td>4.8097</td>
</tr>
</tbody>
</table>
Maximum Likelihood Estimation
Cross Section Case
Binary Outcome

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

\[
\begin{align*}
\text{Prob} \left( y_1 = j \mid x_1 \right) &= \text{Prob} \left( \varepsilon_1 \leq \text{or} > \theta'x_1 \right) \\
\text{Prob} \left( y_2 = j \mid x_2 \right) &= \text{Prob} \left( \varepsilon_2 \leq \text{or} > \theta'x_2 \right) \\
&\ldots \\
\text{Prob} \left( y_n = j \mid x_n \right) &= \text{Prob} \left( \varepsilon_n \leq \text{or} > \theta'x_n \right)
\end{align*}
\]

We operate on the marginal probabilities of \( n \) observations.
Log Likelihoods for Binary Choice Models

\[
\log l(\theta | X, y) = \sum_{i=1}^{n} \log F[(2y_i - 1) \theta' x_i]
\]

- Probit

\[
F(t) = \Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp(-t^2 / 2)\,dt = \int_{-\infty}^{t} \phi(t)\,dt
\]

- Logit

\[
F(t) = \Lambda(t) = \frac{\exp(t)}{1 + \exp(t)}
\]
Spatially Correlated Observations
Correlation Based on Unobservables

\[ y_1 = \theta' x_1 + u_1 \]
\[ y_2 = \theta' x_2 + u_2 \]
\[ \vdots \]
\[ y_n = \theta' x_n + u_n \]

\[ \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \rho W \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \sim f \left( \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right), (\rho^2 W W') \]

\[ W = \text{the usual spatial weight matrix.} \]
In the cross section case, \( W = I \). Now, it is a full matrix. The joint probably is a single n fold integral.
**Spatially Correlated Observations**

**Correlated Utilities**

\[
\begin{pmatrix}
  y_1^* \\
  y_2^* \\
  \vdots \\
  y_n^*
\end{pmatrix} = \rho W 
\begin{pmatrix}
  y_1^* \\
  y_2^* \\
  \vdots \\
  y_n^*
\end{pmatrix} + \begin{pmatrix}
  \theta' x_1 + \varepsilon_1 \\
  \theta' x_2 + \varepsilon_2 \\
  \vdots \\
  \theta' x_n + \varepsilon_n
\end{pmatrix} = (I - \rho W)^{-1} \begin{pmatrix}
  \theta' x_1 + \varepsilon_1 \\
  \theta' x_2 + \varepsilon_2 \\
  \vdots \\
  \theta' x_n + \varepsilon_n
\end{pmatrix}
\]

\( W = \) the usual spatial weight matrix.

In the cross section case, \( W = I \). Now, it is a full matrix. The joint probably is a single \( n \) fold integral.
Log Likelihood

1. In the unrestricted spatial case, the log likelihood is one term,

\[ \log L = \log \text{Prob}(y_1|\mathbf{x}_1, y_2|\mathbf{x}_2, \ldots, y_n|\mathbf{x}_n) \]

2. In the discrete choice case, the probability will be an \( n \) fold integral, usually for a normal distribution.
LogL for an Unrestricted BC Model

\[ \text{LogL}(\theta|X, y) = \log \int_{-\infty}^{\theta'x_n} \cdots \int_{-\infty}^{\theta'x_1} \phi_n \left( \begin{array}{c} q_1 \epsilon_1 \\ q_2 \epsilon_2 \\ \vdots \\ q_n \epsilon_n \end{array} \right) \left( \begin{array}{cccc} 1 & q_1 q_2 w_{12} & \cdots & q_1 q_n w_{1n} \\ q_1 q_2 w_{21} & 1 & \cdots & q_2 q_n w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_1 q_n w_{n1} & q_2 q_n w_{n2} & \cdots & 1 \end{array} \right) \left( \begin{array}{c} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{array} \right) \right) \]

\( q_i = -1 \) if \( y_i = 0 \) and \( +1 \) if \( y_i = 1 \).

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

1. Distinguish between private and social shocks and use pseudo-ML
2. Approximate the joint density and use GMM with the EM algorithm
3. Parameterize the spatial correlation and use copula methods
4. Define neighborhoods – make $W$ a sparse matrix and use pseudo-ML
5. Others ...
Pseudo Maximum Likelihood


Spatial Autoregression in Utilities

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

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

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

\[ = A \]

\[ = D + A - D \quad \text{where } D = \text{diagonal elements} \]

\[ y^* = AX\theta + \begin{array}{c} D\varepsilon \text{ Private} \\ (A - D)\varepsilon \text{ Social} \end{array} \]

Suppose individuals ignore the social "shocks." Then

\[
\text{Prob}[y_i = 1 \text{ or } 0 \mid X] = F\left[ (2y_i - 1) \frac{\sum_{j=1}^{n} a_{ij} (\rho) \theta' x_j}{d_i} \right], \text{ probit or logit.}
\]
Pseudo Maximum Likelihood

1. Assumes away the correlation in the reduced form
2. Makes a behavioral assumption
3. Requires inversion of \((I-\rho W)\)
4. Computation of \((I-\rho W)\) is part of the optimization process - \(\rho\) is estimated with \(\theta\).
5. Does not require multidimensional integration (for a logit model, requires no integration)
GMM


\[ y^* = X\theta + \varepsilon, \quad \varepsilon = \rho W\varepsilon + u \]
\[ = \begin{bmatrix} I - \rho W \end{bmatrix}^{-1} u \]
\[ = Au \]

Cross section case: \( \rho = 0 \)

Probit Model: FOC for estimation of \( \theta \) is based on the generalized residuals \( \hat{u}_i = y_i - E[\varepsilon | y_i] \)

\[
\sum_{i=1}^{n} x_i \left( \frac{(y_i - \Phi(\theta'x_i))\phi(\theta'x_i)}{\Phi(\theta'x_i)[1 - \Phi(\theta'x_i)]} \right) = 0
\]

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

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

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

$$= Au$$

Autocorrelated Case: \( \rho \neq 0 \)

Probit Model: FOC for estimation of \( \theta \) is based on the generalized residuals \( \hat{u}_i = y_i - E[\varepsilon | y_i] \)

$$\sum_{i=1}^{n} z_i \left( y_i - \Phi \left[ \frac{\theta'x_i}{a_{ii}(\rho)} \right] \Phi \left[ \frac{\theta'x_i}{a_{ii}(\rho)} \right] \Phi \left[ \frac{\theta'x_i}{a_{ii}(\rho)} \right] (1 - \Phi \left[ \frac{\theta'x_i}{a_{ii}(\rho)} \right]) \right) = 0$$

Requires at least \( K + 1 \) instrumental variables.
GMM Approach

1. Spatial autocorrelation induces heteroscedasticity that is a function of $\rho$.

2. Moment equations include the heteroscedasticity and an additional instrumental variable for identifying $\rho$.

3. LM test of $\rho = 0$ is carried out under the null hypothesis that $\rho = 0$.

Basic Logit Model

\[ y_i^* = \beta'x_i + \varepsilon_i, \quad y_i = 1[y_i^* > 0] \] (as usual)

Rather than specify a spatial weight matrix, we assume \([\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]\) 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 \(C[u_1 = \Lambda(\varepsilon_1), u_2 = \Lambda(\varepsilon_2), \ldots, u_n = \Lambda(\varepsilon_n) | \theta]\)
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):

$$\Lambda(V_1 < v_1, V_2 < v_2, \ldots, V_q < v_q, \ldots, V_Q < v_Q)$$

$$= \left[ \prod_{q=1}^{Q} \Lambda_q(v_q) \right] \times \left[ 1 + \sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} \theta_{qk} \cdot (1 - \Lambda_q(v_q)) (1 - \Lambda_k(v_k)) \right], \quad (7)$$

where $\theta_{qk}$ 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:

\[
z_q^* = \beta' x_q + \varepsilon_q
\]

\[
\begin{cases}
0 & \text{if } z_q^* < 0 \\
1 & \text{if } z_q^* \geq 0
\end{cases}
\]

where \(z_q^*\) is an unobserved propensity variable, \(\beta\) is a vector of coefficients to be estimated, and \(\varepsilon_q\) is a logistically distributed idiosyncratic error term with a scale parameter of \(\sigma_q\) (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

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

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

(11)
Parameterization

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

The parameter \( \sigma_q \) in Eq. 11 is next parameterized as:

\[ \sigma_q = g(\lambda' \varpi_q) = \exp(\lambda' \varpi_q), \]

where \( \varpi_q \) 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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Binary (aspatial) logit model</th>
<th>Copula-based spatially correlated and heteroscedastic model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>t statistic</td>
</tr>
<tr>
<td>Constant</td>
<td>-5.534</td>
<td>-7.45</td>
</tr>
<tr>
<td>Individual demographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.238</td>
<td>1.18</td>
</tr>
<tr>
<td>Caucasian</td>
<td>0.722</td>
<td>2.42</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.457</td>
<td>0.95</td>
</tr>
<tr>
<td>Driver's license</td>
<td>0.661</td>
<td>3.02</td>
</tr>
<tr>
<td>Household demographics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household size</td>
<td>0.562</td>
<td>5.40</td>
</tr>
<tr>
<td>Single parent family</td>
<td>1.264</td>
<td>2.95</td>
</tr>
<tr>
<td>Presence of bicycle</td>
<td>-0.266</td>
<td>-0.93</td>
</tr>
<tr>
<td>Household location and season variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>San Francisco County</td>
<td>1.309</td>
<td>1.84</td>
</tr>
<tr>
<td>Summer</td>
<td>0.816</td>
<td>3.94</td>
</tr>
<tr>
<td>Fall</td>
<td>4.265</td>
<td>8.47</td>
</tr>
<tr>
<td>Physical environment measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zonal structure, density, and race composition variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fraction of multi-family dwelling units</td>
<td>1.100</td>
<td>1.57</td>
</tr>
<tr>
<td>Household density</td>
<td>-0.308</td>
<td>-3.54</td>
</tr>
<tr>
<td>Fraction of African–American population</td>
<td>-1.299</td>
<td>-0.59</td>
</tr>
<tr>
<td>Fraction of Asian population</td>
<td>0.152</td>
<td>0.17</td>
</tr>
<tr>
<td>Zonal activity opportunity, housing cost, and transportation network variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of physically active recreation centers such as fitness centers, sports centers, dance and yoga studios</td>
<td>0.031</td>
<td>1.07</td>
</tr>
<tr>
<td>Average of median housing value</td>
<td>0.132</td>
<td>2.09</td>
</tr>
<tr>
<td>Bicycling facility density (miles of bike lanes per square mile)</td>
<td>0.033</td>
<td>0.66</td>
</tr>
<tr>
<td>Number of zones within 4 non-motorized mode miles</td>
<td>0.032</td>
<td>2.16</td>
</tr>
</tbody>
</table>
Table 1  Estimation results for teenagers’ weekday physical activity participation choice

<table>
<thead>
<tr>
<th>Variables</th>
<th>Binary (aspatial) logit model</th>
<th>Copula-based spatially correlated and heteroscedastic model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>t statistic</td>
</tr>
<tr>
<td>(Spatial) heteroscedasticity variables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single parent family</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Presence of bicycle</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Fraction of multi-family dwelling units</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Spatial correlation variables (δ) in the θ parameter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverse of distance between zonal centroids</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of observations</td>
<td>722</td>
<td></td>
</tr>
<tr>
<td>Log-likelihood at convergence</td>
<td>−318.323</td>
<td></td>
</tr>
</tbody>
</table>
Other Approaches


Ordered Probability Model

\[ y^* = \beta' x + \varepsilon, \text{ we assume } x \text{ contains a constant term} \]

\[ y = 0 \text{ if } y^* \leq 0 \]

\[ y = 1 \text{ if } 0 < y^* \leq \mu_1 \]

\[ y = 2 \text{ if } \mu_1 < y^* \leq \mu_2 \]

\[ y = 3 \text{ if } \mu_2 < y^* \leq \mu_3 \]

\[ ... \]

\[ y = J \text{ if } \mu_{j-1} < y^* \leq \mu_j \]

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

\[ \mu_{-1} = -\infty, \quad \mu_0 = 0, \quad \mu_J = +\infty, \quad \mu_{j-1} < \mu_j, \quad j = 1,...,J \]
Outcomes for Health Satisfaction

Histogram for Health - Full Sample

Frequency

0 1 2 3 4
0 1911 3822 5733 7644

MALE  FEMALE
A Spatial Ordered Choice Model


Core Model: Cross Section
\[ y_i^* = \beta'x_i + \varepsilon_i, \quad y_i = j \text{ if } \mu_{j-1} < y_i^* \leq \mu_j, \quad \text{Var}[\varepsilon_i] = 1 \]

Spatial Formulation: There are R regions. Within a region
\[ y_{ir}^* = \beta'x_{ir} + u_i + \varepsilon_{ir}, \quad y_{ir} = j \text{ if } \mu_{j-1} < y_{ir}^* \leq \mu_j \]

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

Spatial Autocorrelation Across Regions
\[ u = \rho Wu + v, \quad v \sim N[0, \sigma_v^2 I] \]
\[ u = (I-\rho W)^{-1}v \sim N[0, \sigma_v^2 \{(I-\rho W)'(I-\rho W)\}^{-1}] \]

The error distribution depends on 2 parameters, \( \sigma_v^2 \) and \( \rho \)

Estimation Approach: Gibbs Sampling; Markov Chain Monte Carlo

Dynamics in latent utilities added as a final step: \( y^*(t) = f[y^*(t-1)] \).
# OCM for Land Use Intensity

Table 1 Data Description for Land Development Intensity Level Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTLV</td>
<td>Development intensity level</td>
</tr>
<tr>
<td>ELEVTN</td>
<td>Average elevation of the 300m grid cell (km)</td>
</tr>
<tr>
<td>SLOPE</td>
<td>Average slope of the 300m grid cell (%)</td>
</tr>
<tr>
<td>NSCHOOL</td>
<td>Number of K-12 schools in the neighborhood</td>
</tr>
<tr>
<td>POP</td>
<td>Population (thousand) in the neighborhood</td>
</tr>
<tr>
<td>WORKER</td>
<td>Number of workers (thousand) living in the neighborhood</td>
</tr>
<tr>
<td>INC</td>
<td>Average household income (thousand dollars) in the neighborhood</td>
</tr>
<tr>
<td>EMPTT</td>
<td>Travel time to nearest major (top 15) employer (hours)</td>
</tr>
<tr>
<td>CBDTT</td>
<td>Travel time to CBD (hours)</td>
</tr>
<tr>
<td>AIRT</td>
<td>Travel time to nearest airfield (hours)</td>
</tr>
<tr>
<td>RDTT</td>
<td>Travel time to nearest highway (hours)</td>
</tr>
</tbody>
</table>
# OCM for Land Use Intensity

## Table 2 Summary Statistics for Land Development Intensity Analysis

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

---
## Estimated Dynamic OCM

### Table 3 Estimation Results for Model of Land Development Intensity Levels

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>POP</td>
<td>-0.024</td>
<td>0.036</td>
<td>-0.668</td>
</tr>
<tr>
<td>WORKER</td>
<td>0.089</td>
<td>0.067</td>
<td>1.327</td>
</tr>
<tr>
<td>INC</td>
<td>0.019</td>
<td>0.002</td>
<td>9.143</td>
</tr>
<tr>
<td>EMPTT</td>
<td>-0.232</td>
<td>0.130</td>
<td>-1.778</td>
</tr>
<tr>
<td>CBDTT</td>
<td>-4.365</td>
<td>0.851</td>
<td>-5.126</td>
</tr>
<tr>
<td>AIRT</td>
<td>-2.867</td>
<td>0.248</td>
<td>-11.550</td>
</tr>
<tr>
<td>RDTT</td>
<td>2.309</td>
<td>0.385</td>
<td>6.001</td>
</tr>
<tr>
<td>NSCHOOL</td>
<td>0.039</td>
<td>0.017</td>
<td>2.305</td>
</tr>
<tr>
<td>ELEV</td>
<td>-0.239</td>
<td>0.696</td>
<td>-0.343</td>
</tr>
<tr>
<td>SLOPE</td>
<td>-0.034</td>
<td>0.010</td>
<td>-3.394</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.561</td>
<td>0.019</td>
<td>30.005</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.857</td>
<td>0.074</td>
<td>11.612</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.871</td>
<td>0.222</td>
<td>3.931</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.834</td>
<td>0.011</td>
<td>-77.231</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>2.235</td>
<td>0.031</td>
<td>71.393</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>4.361</td>
<td>0.034</td>
<td>130.167</td>
</tr>
</tbody>
</table>
Figure 2 Distribution of Region-Specific Error Term Estimates ($\theta$) for Land Development Intensity Levels

Statistically Significant (at a 0.05 significance level)

Mean of $\theta$

-0.584 - 0.020
0.021 - 0.851
0.852 - 1.666
1.667 - 2.688
2.689 - 4.996
Unordered Multinomial Choice

Core Random Utility Model

- Underlying Random Utility for Each Alternative
  \[ U(i,j) = \beta_j' x_{ij} + \varepsilon_{ij}, \quad i = \text{individual}, \; j = \text{alternative} \]
- Preference Revelation
  \[ Y(i) = j \text{ if and only if } U(i,j) > U(i,k) \text{ for all } k \neq j \]
- Model Frameworks
  Multinomial Probit: \[ [\varepsilon_1, \ldots, \varepsilon_J] \sim N[0, \Sigma] \]
  Multinomial Logit: \[ [\varepsilon_1, \ldots, \varepsilon_J] \sim \text{iid type I extreme value} \]
Multinomial Unordered Choice - Transport Mode

Decision: Which Type, A, T, B, C. Depends on Income, Price, Travel Time

Utility Functions, land parcel i, usage type j, date t
\[ U(i,j,t) = \beta' x_{ijt} + \theta_k + \epsilon_{ijt} \]

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>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Scale</th>
<th>Source of the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>land use</td>
<td>land use (= 1 if agriculture, 2 if urban, 3 if forest and 0 if no-use)</td>
<td>Parcel</td>
<td>TERUTI survey</td>
</tr>
<tr>
<td>NRSEC90</td>
<td>number of second homes</td>
<td>County</td>
<td>INSEE population census</td>
</tr>
<tr>
<td>aver</td>
<td>average rain</td>
<td>County</td>
<td>The Climate Database of Europe at the resolution of 50 km</td>
</tr>
<tr>
<td>avesl</td>
<td>average slope</td>
<td>County</td>
<td>The Digital Elevation Model of Europe at the resolution of 1 km</td>
</tr>
<tr>
<td>REV</td>
<td>average household income</td>
<td>County</td>
<td>Income tax survey <em>Impôt sur le revenu des communes</em></td>
</tr>
<tr>
<td>whyd</td>
<td>wheat yield</td>
<td>Region</td>
<td>AGRESTE</td>
</tr>
<tr>
<td>grpop</td>
<td>population growth between 1990 and 1999</td>
<td>County</td>
<td>INSEE population census</td>
</tr>
<tr>
<td>network</td>
<td>travel time to the nearest highway</td>
<td>County</td>
<td>Microsoft Autoroute 2007</td>
</tr>
<tr>
<td>TEXT1</td>
<td>Soil quality 0, if coarse texture (clay &lt;18% and sand &gt; 65%); 1, otherwise</td>
<td>County</td>
<td>The French Soil map at the scale of 1/1,000,000</td>
</tr>
<tr>
<td>Variable</td>
<td>mean</td>
<td>std</td>
<td>min</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>grpop</td>
<td>12.83</td>
<td>17.23</td>
<td>-11.40</td>
</tr>
<tr>
<td>NRSEC90</td>
<td>211.84</td>
<td>1,103.54</td>
<td>6.00</td>
</tr>
<tr>
<td>aver</td>
<td>2.38</td>
<td>0.53</td>
<td>1.01</td>
</tr>
<tr>
<td>avesl</td>
<td>2.49</td>
<td>1.31</td>
<td>0.24</td>
</tr>
<tr>
<td>REV</td>
<td>14,320.08</td>
<td>4,513.06</td>
<td>5,102.07</td>
</tr>
<tr>
<td>whyd</td>
<td>106.17</td>
<td>14.74</td>
<td>80.00</td>
</tr>
<tr>
<td>network</td>
<td>21.88</td>
<td>15.55</td>
<td>1</td>
</tr>
<tr>
<td>text1</td>
<td>0.48</td>
<td>0.50</td>
<td>0</td>
</tr>
</tbody>
</table>
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.
Modeling Counts

Feeway Accidents by Region and Season

Frequency

0 1 2 3 4 5 6 7 8 9 10

WINTER  SUMMER
Canonical Model


Poisson Regression
\[ y = 0, 1, \ldots \]
\[ \text{Prob}[y = j|x] = \frac{\exp(-\lambda)\lambda^j}{j!} \]
Conditional Mean \( \lambda = \exp(\beta'x) \)
Signature Feature: Equidispersion
Usual Alternative: Various forms of Negative Binomial
Spatial Effect: Filtered through the mean
\[ \lambda_i = \exp(\beta'x_i + \theta_i) \]
\[ \theta_i = \rho \sum_{m=1}^{n} w_{im} \theta_m + \varepsilon_i \]
Fig. 1  Locations of the 38 mixed-oak stands in central Pennsylvania
Grazie!