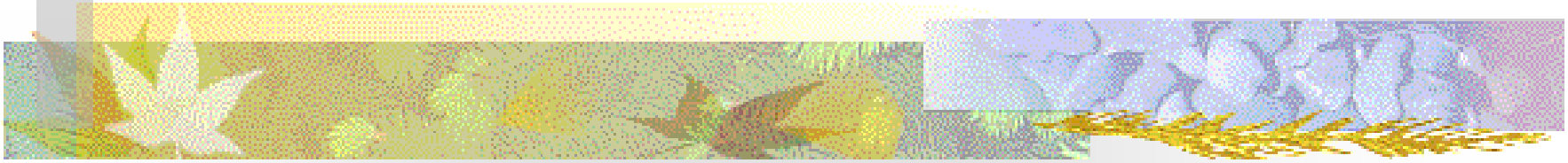
Estimation of a Discrete Choice Model when Individual Choices are Not Observable

Luba Kurkalova and Sergey Rabotyagov

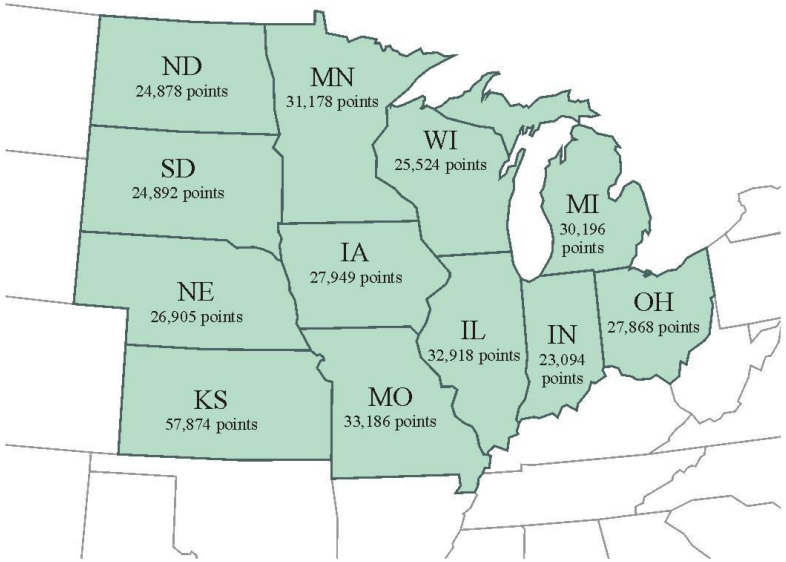


Midwest Economics Association Meeting

St. Louis, Missouri, March 29, 2003

Research Interest

- Economic analysis of policies that pay farmers for adoption of conservation tillage
- Basic data: National Resource Inventory (NRI), 1992 and 1997



Model of conservation tillage adoption

$$\Pr[adopt] = \Pr[\pi_1 \geq \pi_0 + P + \sigma_\varepsilon \varepsilon]$$

$$= \Pr[\boldsymbol{\beta}\mathbf{x} \geq \bar{\pi}_0 + \boldsymbol{\alpha}\mathbf{z}\sigma_{pr} + \sigma_\varepsilon \varepsilon]$$

$$= \Pr\left[\varepsilon \leq \frac{1}{\sigma_\varepsilon}\left\{\boldsymbol{\beta}\mathbf{x} - \bar{\pi}_0 - \boldsymbol{\alpha}\mathbf{z}\sigma_{pr}\right\}\right]$$

$$= \frac{\exp\left(\frac{1}{\sigma_\varepsilon}\left\{\boldsymbol{\beta}\mathbf{x} - \bar{\pi}_0 - \boldsymbol{\alpha}\mathbf{z}\sigma_{pr}\right\}\right)}{1 + \exp\left(\frac{1}{\sigma_\varepsilon}\left\{\boldsymbol{\beta}\mathbf{x} - \bar{\pi}_0 - \boldsymbol{\alpha}\mathbf{z}\sigma_{pr}\right\}\right)}$$

Probability of tillage choice y_i

$$L\left(\underbrace{y_i, \mathbf{x}_i, \mathbf{z}_i, \bar{\pi}_{0,i}, \sigma_{pr,i}}_{data}, \underbrace{\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_{\varepsilon}}_{parameters}\right) = \left(\frac{\exp w_i}{1+\exp w_i}\right)^{y_i} \left(\frac{1}{1+\exp w_i}\right)^{1-y_i},$$

where

$$w_i \equiv \frac{1}{\sigma_{\varepsilon}}\left\{\boldsymbol{\beta}\mathbf{x}_i - \bar{\pi}_{0,i} - \boldsymbol{\alpha}\mathbf{z}_i\sigma_{pr,i}\right\}$$

Problem:
in 1997, the y_i are not available,

But,
grouped data on y_i are available

Grouped data on conservation tillage choices

$$\bar{y}^{G_j} \equiv \frac{1}{N^{G_j}} \sum_{i \in G_j} y_i$$

- CTIC (Purdue) collects expert opinion surveys on adoption of conservation tillage
- Reports percentage of area in conservation tillage by state, county, and crop, 1989- current
- Conversion of area percentage data into count data

  NRI: 92 corn points in Boone county, IA

  CTIC: 55% corn in County#15 is in CT

  Count data (CTIC+NRI): 92*0.55=51 NRI corn points from Boone county must be in CT

Linear versus nonlinear model

	Linear model	Nonlinear model
Farm-level	$Y_i = \boldsymbol{\beta}'\mathbf{x}_i + \eta_i$	$Pr[Y_i = 1] = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_i)}$
Aggregated	$\bar{Y}^{G_j} = \boldsymbol{\beta}'\bar{\mathbf{x}}^{G_j} + \bar{\eta}^{G_j}$	$\bar{Y}^{G_j} = \frac{\exp(\boldsymbol{\alpha}'\bar{\mathbf{x}}^{G_j})}{1+\exp(\boldsymbol{\alpha}'\bar{\mathbf{x}}^{G_j})} + \xi_j$

Where

$$\bar{Y}^{G_j} \equiv \frac{1}{N^{G_j}} \sum_{i \in G_j} Y_i,$$

$$\bar{\mathbf{x}}^{G_j} \equiv \frac{1}{N^{G_j}} \sum_{i \in G_j} \mathbf{x}_i$$

Proposed approach

Express the likelihood of observing $\bar{y}^{G_j}$ as a function of

- original (farm-level) model parameters, and
- Farm-level data on explanatory variables

Probability of observing 1 point in conservation tillage out of 3 points in a group

$$\Pr\left[\bar{Y}^{G_j} = 1/3\right]$$

$$= \Pr\left[1 \textit{ out of } 3 \textit{ individuals in group } G_j \textit{ choose } A\right]$$

$$= \Pr\left[1st \textit{ chooses } A,\ 2nd \textit{ choses } B,\ \textit{and } 3nd \textit{ chooses } B\right]$$

$$+ \Pr\left[1st \textit{ chooses } B,\ 2nd \textit{ choses } A,\ \textit{and } 3nd \textit{ chooses } B\right]$$

$$+ \Pr\left[1st \textit{ chooses } B,\ 2nd \textit{ choses } B,\ \textit{and } 3nd \textit{ chooses } A\right]$$

$$= \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_1)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_1)} \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_2)} \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_3)}$$

$$+ \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_1)} \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_2)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_2)} \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_3)}$$

$$+ \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_1)} \frac{1}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_2)} \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_3)}{1+\exp(\boldsymbol{\beta}'\mathbf{x}_3)}.$$

Application to 1992 NRI data

- Random sub-sample (1,339 observations) of Iowa 1992 NRI data (soil and tillage) supplemented with Census of Agriculture (farmer characteristics) and climate data of NCDC
- 63% of farmers use conservation till
- Grouped the observed individual choice data into 240 groups by county and crop
- Pretended that we do not observe individual choices

Results

Variable	Aggregated data model	Discrete data model
Corn dummy	44.7 (6.9)	41 (11)
Slope	0.56 (0.18)	0.22 (0.12)
Permeability	0.85 (0.37)	0.63 (0.31)
Water capacity	0.87 (0.32)	0.73 (0.29)
σ_ε	5.69 (0.12)	6.0 (1.6)

Problems

- Programming ☺
- Area percentage versus percent of points
  - IA, County #5, corn, 1992 NRI

TILL	0	0	1	1	1	1	1	1
Acres	19	19	2	19	20	20	22	32

  - 75% area in CT could be represented by minimum 5 and maximum 7 points

Accuracy of aggregate data: IOWA, 1992

	Cons. till acres	Cons. till percent	Total acres
Corn			
NRI	8,668	65%	13,377
CTIC	4,474	35%	12,784
Soybeans			
NRI	5,542	66%	8,424
CTIC	4,687	57%	8,265

Conclusions

- Method allows recovery of individual-level model parameters with aggregated information on choices
- Worked very satisfactory on artificially aggregated data