Estimation of Discrete Choice Models with Aggregate Data: An Application to the Adoption of Conservation Tillage

Tara Wade*, Luba Kurkalova*, Silvia Secchi**

*North Carolina A & T State University
** Southern Illinois University Carbondale

USDA/ERS Workshop, August 8, 2012

Outline

- Introduction
- Economic model
- Econometric model
- Application to 2002 and 2004 Iowa data
- Conclusions

Conservation tillage

- Conservation tillage (CT) is any tillage system that leaves at least 30% of crop residue on the soil surface at the time of planting
- Umbrella term used for
  - Mulch till (disturbs the entire soil surface)
  - No-till (disturbs only the minimal amount of soil)
    - Strip-till
    - Vertical tillage
    - Fluffing harrows
- CT is known to reduce soil erosion, reduce nutrient runoff, increase soil carbon sequestration, and improve overall soil health

Costs of the adoption of conservation tillage (CT)

- Farmers adopt CT if
  - Profits from CT exceed those from conventional till plus a premium due to uncertainty (risk aversion, option value or incentive to learn more)

$$\pi_{CT} \geq \pi_0 + P$$

- Opportunity cost of adopting CT (subsidy) is the minimum compensation that is needed to entice a farmer to use CT instead of the conventional one
  - The cost is zero for those already using CT

$$S = \max\left\{\pi_0 - \pi_{CT} + P,\ 0\right\}$$

Econometrics: binary choice model

$$Y = \begin{cases} 1, & \text{if CT is adopted} \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} \Pr(Y=1) &= \Pr\left(\pi_{CT} - \pi_0 - P \geq \sigma_\eta \eta\right) \\ &= \Pr\left(\boldsymbol{\alpha}'\mathbf{w} - \pi_0 - \boldsymbol{\gamma}'\mathbf{z} \geq \sigma_\eta \eta\right) = \Pr\left(\boldsymbol{\beta}'\mathbf{x} \geq \eta\right) \end{aligned}$$

Here

$$\boldsymbol{\beta}' = \left(\frac{1}{\sigma_\eta}\boldsymbol{\alpha}', \; -\frac{1}{\sigma_\eta}\boldsymbol{\gamma}', -\frac{1}{\sigma_\eta} \right), \quad \mathbf{x} = \begin{pmatrix} \mathbf{w} \\ \mathbf{z} \\ \pi_0 \end{pmatrix}.$$

Standard Logit model

N farmers

$$Y_i = \begin{cases} 1, & \text{if CT is adopted} \\ 0, & \text{otherwise} \end{cases} \quad , i = 1, \ldots, N.$$

$$\Pr(Y_i = 1) = \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)} \qquad E[Y_i] = \frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)}$$

Likelihood of observing tillage choice $y_i\,(1 \text{ or } 0)$ by the i^{th} farmer:

$$L\Bigg(\underbrace{\beta}_{parameters} \Bigg| \underbrace{y_i, x_i}_{data}\Bigg) = \left(\frac{\exp(\beta' x_i)}{1 + \exp(\beta' x_i)}\right)^{y_i} \left(\frac{1}{1 + \exp(\beta' x_i)}\right)^{1 - y_i}$$

Data on CT use

- “Old” NRCS-NRI – up to year 1992
- NRI-CEAP – very limited data
- USDA-NASS ARMS – limited access, relatively small sample size
- CTIC (Purdue University) – county-average CT use rates, by crop, 1989-2004

Grouped dependent variable logistic model (GDVLM)

- N farmers, each farming a_i acres, are divided into groups G_j , $j = 1,\ldots, J$
- The observed group-average proportion of CT adoption represents an expert opinion which is thought to be the expected value of the group-average of the binary outcome variable, subject to a random noise

$$p_j = \frac{1}{\sum_{i \in G_j} a_i} \sum_{i \in G_j} \frac{a_i \exp\left(\boldsymbol{\beta}'\mathbf{x}_i\right)}{1+\exp\left(\boldsymbol{\beta}'\mathbf{x}_i\right)} + \varepsilon_j \ , \varepsilon_j \sim N\left(0, \sigma_\varepsilon\right)$$

Estimating GDVLM

- Likelihood for the j-th group of observations

$$L(\boldsymbol{\beta}, \sigma_\varepsilon \,|\, p_j, \mathbf{x}_i (i \in G_j)) = -\frac{1}{2}\ln(2\pi\sigma_\varepsilon^2)$$

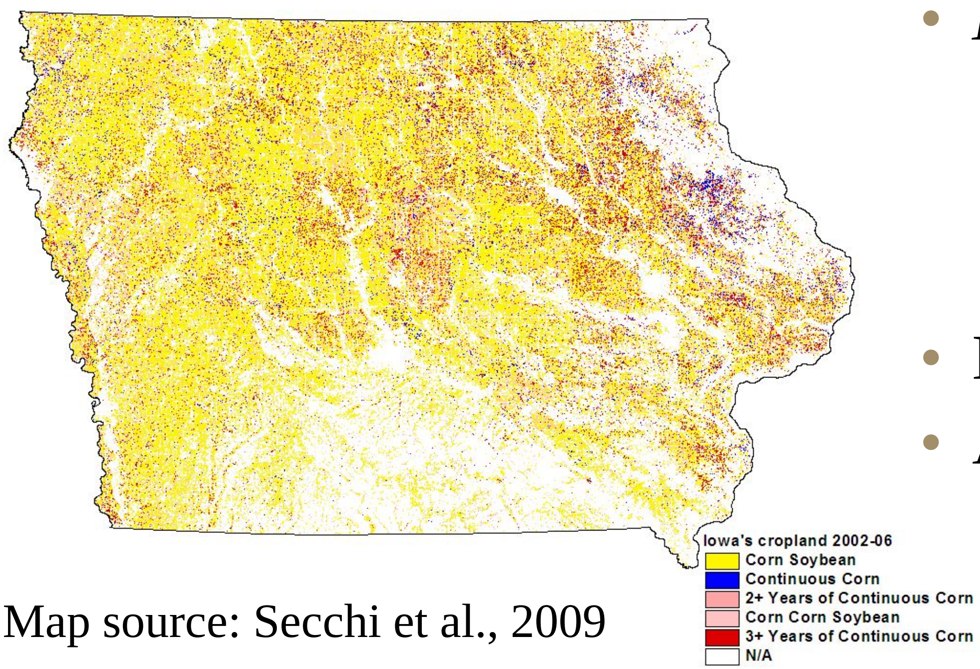
$$-\frac{1}{2\sigma_\varepsilon^2}\left\{ p_j - \frac{1}{\sum_{i \in G_j} a_i} \sum_{i \in G_j} \frac{a_i \exp\left(\boldsymbol{\beta}'\mathbf{x}_i\right)}{1+\exp\left(\boldsymbol{\beta}'\mathbf{x}_i\right)} \right\}^2$$

Application to Iowa: CT use data

- Collected by CTIC
- County-average proportion of CT use in 2002 and 2004
- Corn and soybeans
- J = 396

Explanatory variables: crop and previous crop

- USDA/NASS, GIS-based, CCM
- 2002 (2001) and 2004 (2003) crop and previous year crop data
- Overlaid with digital soils maps
- *N* = 123,157 fields
  - 15% in CC
  - 39% in CS
  - 46% in SC
- Median group size: 333 fields
- Average field size (a_j): 303 ac



Map source: Secchi et al., 2009

Estimation of the GDVLM on Iowa data

- Dependent variable: county-average proportion of CT use in 2002 and 2004, corn and soybeans
- Explanatory variables:
  - Crop and previous crop grown
  - Field and soil properties
    - Slope, permeability, flood frequency
  - π_0 varying with soil quality
  - Climatic factors
    - Average temperature, precipitation amount and variability
  - Farmer and farm characteristics from Census of Ag.
    - Experience, gender, farm ownership

Model estimated

$$\mathbf{\beta'x} = \beta_1 \cdot I_S + \beta_2 \cdot I_{CS} + \beta_3 \cdot SLOPE + \beta_4 \cdot PERM + \beta_5 \cdot FLOOD$$
$$+ \beta_6 \cdot PRCP + \beta_7 \cdot TMAX$$
$$+ \{(\beta_8 \cdot I_C + \beta_9 \cdot I_S) \cdot \pi_0 + (\beta_{10} \cdot I_C + \beta_{11} \cdot I_S) \cdot EXP$$
$$+ (\beta_{12} \cdot I_C + \beta_{13} \cdot I_S) \cdot MALE + (\beta_{14} \cdot I_C + \beta_{15} \cdot I_S) \cdot CORP\} \cdot \sigma_{prcp}^2$$
$$+ \beta_{16} \cdot \pi_0$$

- Relative error of the estimated proportion of acres in CT ranges from 2% (NW) to 10% (Central)
- Counterpart of $R^2 = 0.433$

Estimated costs of CT adoption

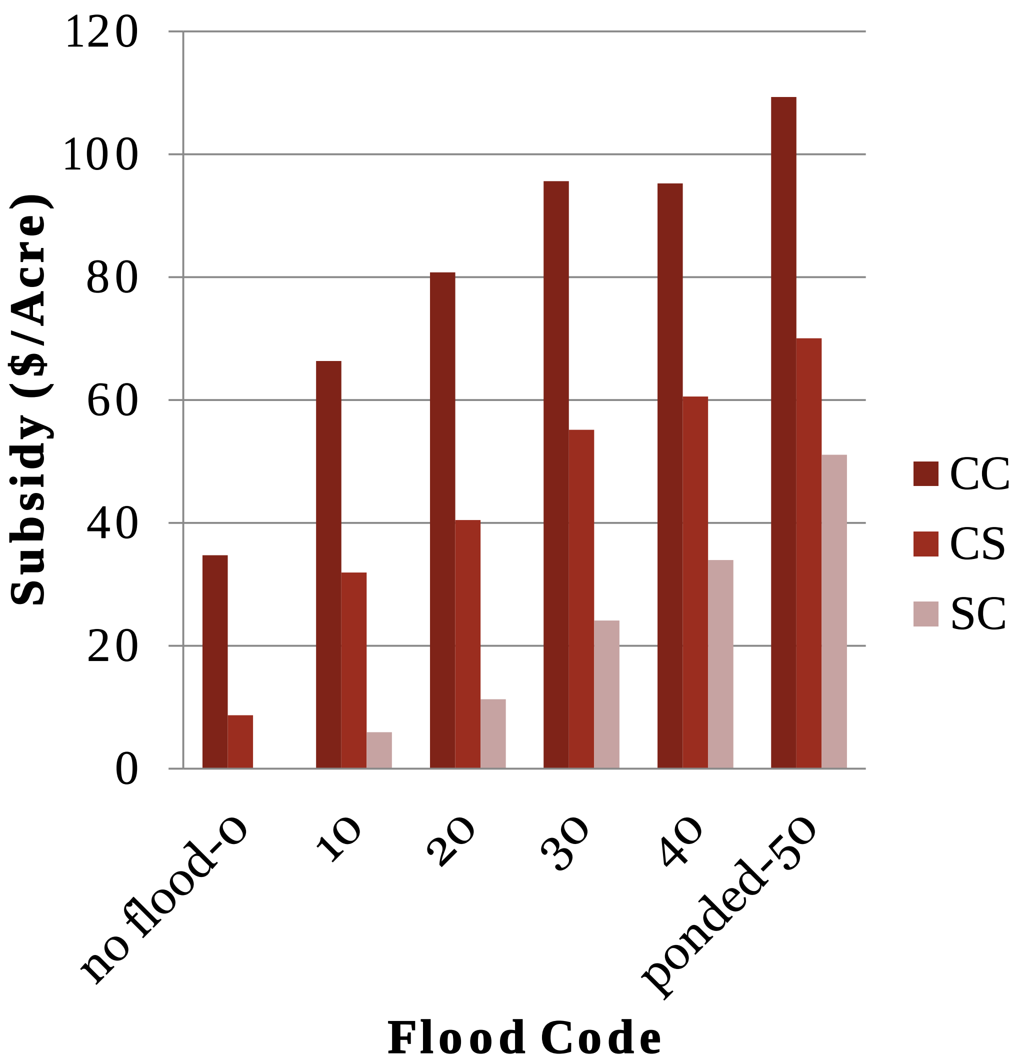
$$S = \max\left\{\left(\pi_0 - \widehat{\pi_{CT}}\right) + \hat{P},\ 0\right\}$$

Cropping sequence	Acres, %, observed	CT Acres, %, estimated	Cost, per acre (s.d.), estimated
CC	15	23	$43 ($31)
CS	39	49	$14 ($19)
SC	46	76	$2.6 ($9.5)

Willingness to adopt CT within the given subsidy ranges

Subsidy, $/ac	All	Corn	Soybeans
S < 10	67%	47%	92%
10 ≤ S < 20	9%	13%	3%
20 ≤ S < 30	6%	10%	2%
30 ≤ S < 40	5%	8%	1%
40 ≤ S < 50	4%	6%	1%
50 ≤ S	10%	16%	1%

Subsidies grouped by flood frequency



- Estimating individual probabilities allows flexibility in policy analysis
- Permits the consideration of other groupings
  - Watersheds

Conclusions

- Estimating individual probabilities expands the scope of the type of analysis that may be done with the aggregated choice data
- Average 2002, 2004 costs of CT adoption estimated at $13/acre
- The methodology of combining the grouped and individual- or field-level data could be applied to other BMPs

We are grateful to the USDA/ERS for providing partial funding for this project. The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. Department of Agriculture.

trwade@ncat.edu
lakurkal@ncat.edu
ssecchi@siu.edu