Correcting Endogenous Stratification in Random Utility Models:

Application to the Marine Recreational Fishing Statistics Survey in the SE United States

Paul Hindsley
Coastal Resource Managment
East Carolina University

Brad Gentner
NMFS

Picture:http://www.shorebirdworld.org/fromthefield/Images/coastal%20development.JPG

Purpose of Study

Theoretical Background

1. Review the theory of sampling with endogenous stratification
   (Manski & Lerman, 1977; Manski and McFadden, 1981; Cosslett 1981a; Cosslett 1981b; Cosslett 1993)

Empirical Implications

2. Assess the impact of an intercept sampling strategy on estimation procedures for site/mode/species choice of Saltwater Anglers in the Southeast US.
   - RDD Telephone Survey
     - Simple Random Sample of Coastal Saltwater Recreational Anglers
   - Intercept Survey
     - Potential Bias: Endogenous Stratification
     - More complete information than Telephone Survey

3. If necessary, identify methods to correct sampling bias

Estimation in Preliminary Stages

Behavioral and Econometric Model

- Behavioral Model: Random Utility Model

$$
\begin{aligned}
P_{ni} &= \Pr(U_{ni} > U_{nj}, \forall j \neq i) \\
&= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall j \neq i) \\
&= \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}, \forall j \neq i)
\end{aligned}
$$

- Econometric Model: Conditional Logit

$$
P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} = \frac{e^{x_{ni}\beta_i}}{\sum_j e^{x_{nj}\beta_j}}
$$

- Welfare Measure

$$
WTP_i = -\frac{\beta_{Attribute}}{\beta_{TravelCost}}
$$

The Choice Model

- Species/Mode/Site Choice a function of
  - Travel Costs & Opportunity Costs
  - Site Characteristics
  - Site Quality (Depicted by Catch Rate)
  - Angler Characteristics

Sampling Recreational Fishermen

- Endogenous variables

Site Choice Among All Potential Sites

$i \in C \qquad C = \{1,..., M\}$

Angler Characteristics, Site Characteristics, Travel Cost

- Exogenous variables

$z \in Z \qquad Z = \{1,..., K\}$

- Joint Density

$$f(i,z) \equiv P(i \mid z, \beta_0)\, p(z) \equiv q(z \mid i) Q(i) \quad [(i,z) \in C \times Z]$$

The Contingen ng Examples

Endogenous Stratification: Sampling by Column or Columns

Exogenous Stratification: Sampling by Row or Rows

	i_1	i_2		i_j	
z_1	$P(i_1 \mid z_1, \beta_0)p(z_1)$	$P(i_2 \mid z_1, \beta_0)p(z_1)$	$\cdots$	$P(i_j \mid z_1, \beta_0)p(z_1)$	$p(z_1)$
z_2	$P(i_1 \mid z_2, \beta_0)p(z_2)$	$P(i_2 \mid z_2, \beta_0)p(z_2)$	$\cdots$	$P(i_j \mid z_2, \beta_0)p(z_2)$	$p(z_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
z_k	$P(i_1 \mid z_k, \beta_0)p(z_k)$	$P(i_2 \mid z_k, \beta_0)p(z_k)$	$\cdots$	$P(i_j \mid z_k, \beta_0)p(z_k)$	$p(z_k)$
	$Q(i_1)$	$Q(i_2)$	$\cdots$	$Q(i_j)$	1

Sampling Space

- Finite Space of Strata $\mathbf{b}\,2\,\mathbf{B}$

  $\mathbf{A_b}\mathrm{j}\quad \mathbf{C}\,\mathrm{B}\,\mathbf{Z}$

- Random Sample $\mathbf{B} = {}'1^{\curlyvee}\ and\ \mathbf{A}_1 = \mathbf{C}\,\mathrm{B}\,\mathbf{Z}$

- Exogenous Sample $\mathbf{A_b} = \mathbf{C}\,\mathrm{B}\,\mathbf{Z_b}$

- Endogenous Sample $\mathbf{A_b} = \mathbf{C_b}\,\mathrm{B}\,\mathbf{Z}$

Consequences of Choice Based Samples for Estimation of Saltwater Angler Preferences

- Sampling of Choices not Representative of Population Choices

$$f_c(i,z,b) = \frac{P(i \mid z, \beta_0)\, p(z) H(i)}{Q(i)}$$

$$where \quad H(i) \neq Q(i)$$

$$Q(i) = \sum_{\forall z} P(j \mid y, \beta_0)\, p(y)$$

- Sampling of Angler Characteristics not Representative of Population Characteristics

$$p(z_{sample}) \neq p(z_{pop})$$

Most Common Empirical Methods for Correcting for Endogenous Stratification

- CMLE (Manski and McFadden, 1981)
- WESMLE (Manski and Lerman, 1977)

$$\max_{\theta \in \Theta_0} \sum_{n=1}^{N} w(i_n) \ln P(i_n \mid z_n, \theta)$$

$$where \quad w(i) = \frac{Q(i)}{H(i)}$$

Data

- Marine Recreational Fishery Statistics Survey (MRFSS). Sampled September 2003 through the end of August 2004.
- Field intercept survey
  - Stratified by Site and Mode
  - Random sample of trips
- RDD telephone survey of coastal households
  - Random sample of coastal anglers
- Observations limited to those anglers on single day trips.

Preliminary Estimation

	RDD	Intercept
Travel Cost	**-.0266801 (.0022758)**	**-.0166358 (.0014447)**
Travel Time	**-.7952985 (.0410839)**	**-.8727553 (.0228217)**
Intercept Points in Aggregate Site(ln)	**.660892 (.0462976)**	**.990396 (.0214628)**
Big Game	**.6437769 (.0846256)**	**.656 (.0370**
Small Game	**.9696457 (.0235405)**	**1.191 (.0105**
Bottom Fish	**.4528924 (.0207658)**	**.042 (.0153**
Flat Fish	-.0955117 (.1133694)	**.3994 (.038**
Observations	2398	122

Why the negative value?

Misspecification from Species Aggregations?

Low Sample Size? Flat Fish less than 3% of sample

P-Value – over 0.4

Significant Results in Bold.

Are these parameter estimates actually different?

H_0 : The parameter estimates are the same.

$$-2(\mathrm{LL}(\beta_c) - \mathrm{LL}(\beta_u)) = -2(-44396.303 - (-44074.6685))$$
$$= 643.269$$

$$\text{Critical Value}: \chi^2_6 = 22.46$$

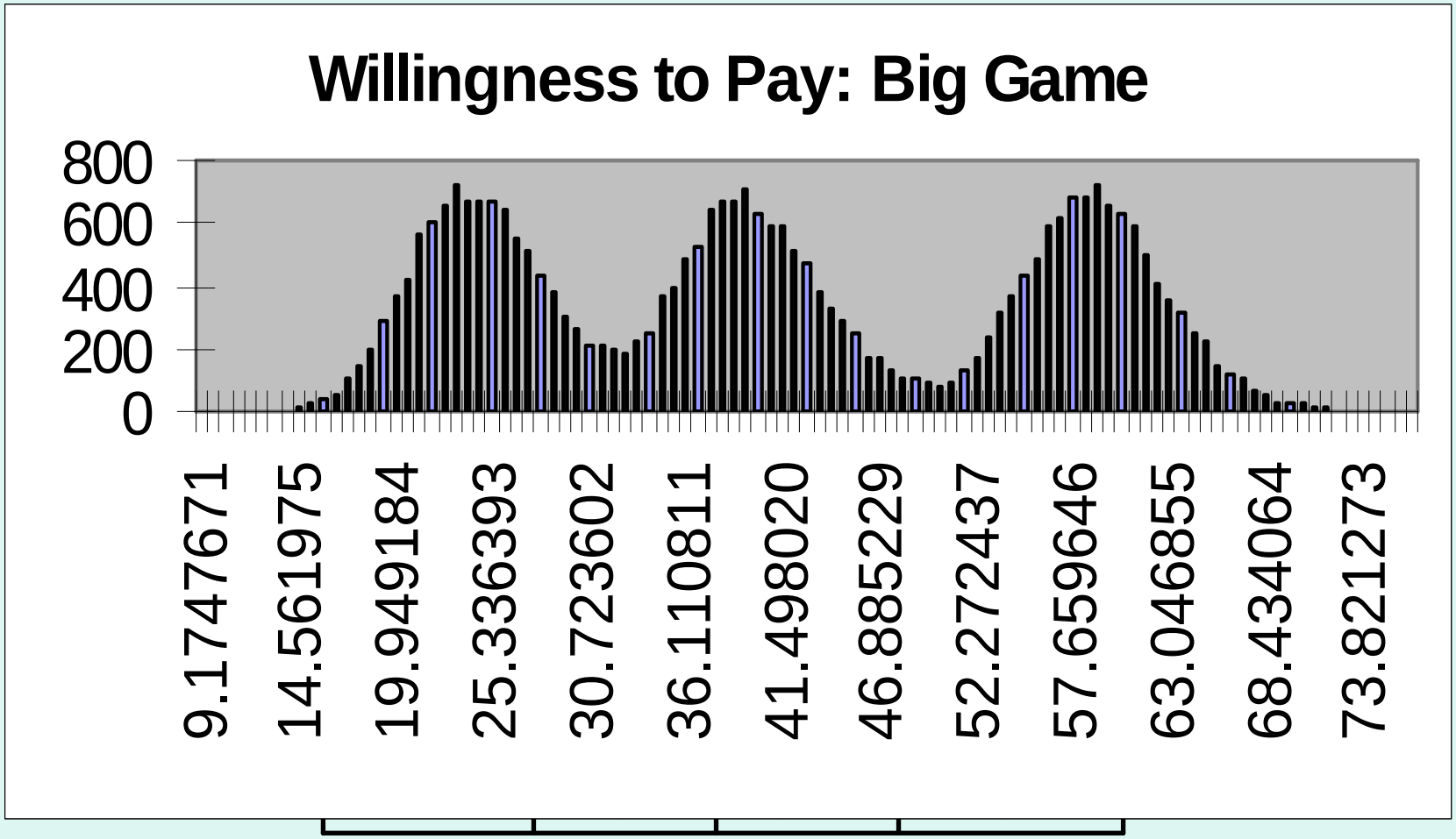
Reject H_0

Preliminary Estimation

	RDD	Intercept	WESMLE
Travel Cost	**-.0266801 (.0022758)**	**-.0166358 (.0014447)**	**-.0147732 (.0008789)**
Travel Time	**-.7952985 (.0410839)**	**-.8727553 (.0228217)**	**-.7849341 (.0138737)**
Intercept Points in Aggregate Site(ln)	**.660892 (.0462976)**	**.990396 (.0214628)**	**.8974661 (.0139749)**
Big Game	**.6437769 (.0846256)**	**.6569 (.0370265)**	**.8645981 (.0264782)**
Small Game	**.9696457 (.0235405)**	**1.191289 (.0105956)**	**1.344704 (.0076813)**
Bottom Fish	**.4528924 (.0207658)**	**.0422836 (.0153152)**	**.0972897 (.011185)**
Flat Fish	-.0955117 (.1133694)	**.3994468 (.038837)**	**.433936 (.0314082)**
Observations	2398	12260	12260

Significant Results in Bold.

Distribution of WTP for Big Game Catch



Conclusions

- Parameter estimates from RDD and Intercept are not the same.
- WESMLE does not correct Intercept WTP estimates in the direction of the RDD.
- Causes?
  - Differences in Representative Samples
  - Differences in exogenous characteristics

Extensions

Truncate Intercept Sample to Include Only Anglers from Coastal Counties.

Test for Differences Between Exogenous Characteristics Within Samples (Correct if necessary)

- Perform Non-Parametric Test between 2 Samples for Exogenous Variables
- If necessary, create a probability weight to correct for differences. Example: Inverse Probability Weighting (IPW) (Wooldridge, 2002)

Extensions

Apply estimation procedures when the quality variable is Expected Catch - a count model estimate of catch based on historic catch, site characteristics, and angler characteristics.

More detailed look at model specification

- Can the model specification be improved?
- Should different species aggregations be used?
- Will a Nested Logit improve estimation?
- Can we use models that incorporate heterogeneity?