First in Class? The Performance of Latent Class Model[1]

Min Chen[2], Frank Lupi[3]

Department of Agricultural, Food, and Resource Economics
Michigan State University, East Lansing, MI 48824

Poster Prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

[1] Previously titled "First in Class? Relative Performance of Latent Class, Conditional Logit and Nested Logit RUM"
ID: 13825

[2] E-mail address: chenmin5@msu.edu
[3] E-mail address: lupi@msu.edu

First in Class? The Performance of Latent Class Model

Min Chen[1], Frank Lupi[1]

[1]Department of Agricultural, Food, and Resource Economics

Abstract

- Researchers have been using the latent class model (LCM) to value recreational activities for years. Several studies have compared LCM to other models using field data. We conduct Monte Carlo simulations to test if the latent class model is able to recover the truth. The simulation results show that although LCM reliably recovers population average values, it is less reliable at recovering the true underlying population segments.

Motivation

- LCM is widely used in empirical work to capture heterogeneity.
  - Boxall & Adamowicz (2002), Scarpa & Thiene (2005), Morey et al (2006), Owen & Videras (2007), Patunru et al (2007), Burton & Rigby (2009)
- Estimates from LCM have been compared with other models based on field data.
  - Greene & Hensher (2003), Provencher & Bishop (2004), Hynes et al (2008), Kosenius (2010)
- We want to know how well LCM will recover known population segments.
- Our goal: To investigate the performance of LCM through Monte Carlo simulations

Model Description

- Random Utility Model of recreation demand

$$U = X\beta + \varepsilon$$

  - X: covariate of travel cost, site characteristics, and possibly demographics and their interactions
  - β: vector of preferences over variables in X
  - ε: individual and site unobserved factors
- Latent Class Model
  - Multiple classes of people: different preferences
  - Probability of individual i belonging to class m: π
    - $\pi = \pi_m$: fixed share of each class
    - $\pi = \pi_{im}$: decided by individual-level variables
  - With fixed shares, M classes, and true class membership unknown
    - Probability of individual i choosing site j
      - Weighted average of probabilities from each class: $Pr_i(j) = \sum_{m=1}^{M} \pi_m \, Pr_i(j|m)$
    - Welfare measures
      - Weighted averages of welfare estimates from each class.

Methods---Monte Carlo Simulations

True World
- 1,000 people and 3 recreational sites
- Utility specification
- Two segments of different preferences
- Population shares of both segments
- Generate travel cost and site quality

$U = \beta_{TC} Travel\ Cost + \beta_Q Site\ Quality + \varepsilon$

700 People: $\beta_{TC}^1 = -0.06, \beta_Q^1 = 0.49$

300 People: $\beta_{TC}^2 = -0.10, \beta_Q^2 = 0.21$

3,000 Travel Cost and 3 Site Quality Data Randomly Drawn from Uniform Distribution

Generate Choices
- Draw i.i.d. errors from type I extreme value distribution
- Pick one's travel costs and site qualities
- Combine with one's known preference
- Choose site that maximize one's utility

One Person with travel cost parameter -0.06 and site quality parameter 0.49:

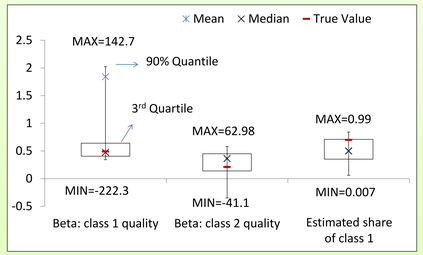
Pseudo Data

Site	1	2	3
Travel Cost	7.79	61.90	31.95
Site Quality	1.02	0.64	1.71
Error	-0.12	0.54	0.62
Utility	-0.09	-2.86	-0.46
Observation	1	0	0

Estimate with LCM
- Assume two latent classes
- Fixed shares across people
- Estimate parameters
- Compute welfare measures

Compare with Truth
- 1,000 iterations
- Descriptive statistics of parameter estimates and welfare estimates
- Compare mean, median, quartiles, etc. with true values

Parameter Estimates



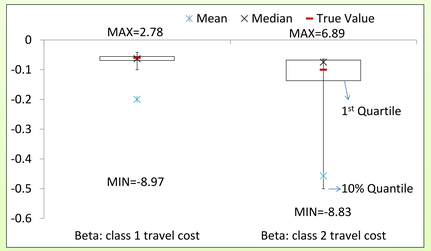


Figure 1a & 1b: About half of the time, estimated parameters of individual classes are fairly close to true values, as true values all lie inside the quartile boxes, mostly in the middle. Large deviations of means from true values are caused by extreme values, as indicated by MAX and MIN and the 10% and 90% quantiles.

Welfare Estimates

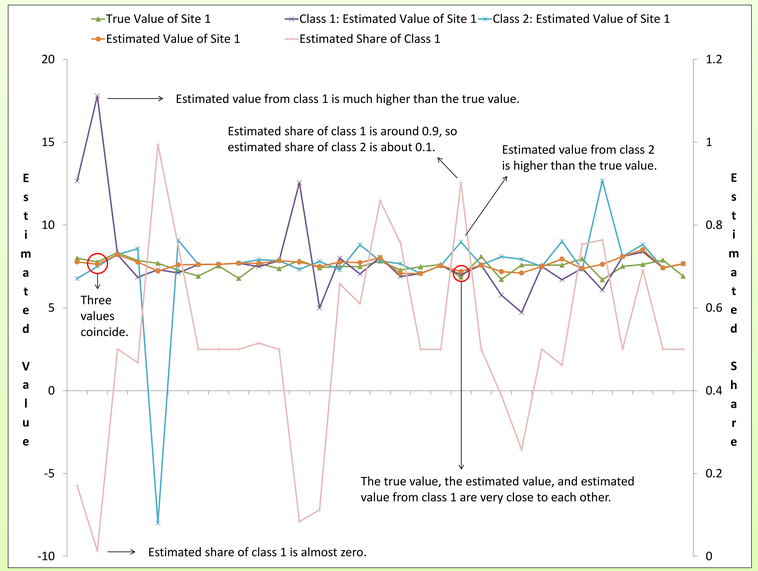


Figure 2: Despite the variation and extreme values for some of the estimated population segments, it is the countervailing effect of the estimated share (which can be very far from the true share 0.7) that makes the weighted averages have good performance.

Conclusions

- One of the strengths of latent class models is the potential for statistically identifying segments or classes of a population with distinct preferences. We find that although our estimated LCM models perform well in terms of recovering population average values, they did not do as well in terms of recovering the values and sizes of the population segments. The errors for the segments are at times quite large.
- Nonetheless, the latent class model worked well on average, because when a population segment was estimated with outlier values, the estimated share of individuals in the segment tended to be small.

References

[1] Boxall, P. C., W. L. Adamowicz (2002). "Understanding heterogeneous preferences in random utility models: A latent class approach." Environ. Resour. Econ. 23(4): 421-446.;
[2] Scarpa, R., M. Thiene (2005). "Destination Choice Models for Rock Climbing in the Northeastern Alps: A Latent-Class Approach Based on Intensity of Preferences." Land Econ. 81(3): 426-444.
[3] Morey, E. et al. (2006). "Using angler characteristics and attitudinal data to identify environmental preference classes: A latent-class model." Environ. Resour. Econ. 34(1): 91-115.
[4] Owen, A. L., J. R. Videras (2007). "Culture and public goods: The case of religion and the voluntary provision of environmental quality." J. Environ. Econ. Manag. 54(2): 162-180.
[5] Patunru, A. A. et al. (2007). "Who Cares about Environmental Stigmas and Does It Matter? A Latent Segmentation Analysis of Stated Preferences for Real Estate." Am. J. Agr. Econ. 89(3): 712-726.
[6] Burton, M., D. Rigby (2009). "Hurdle and Latent Class Approaches to Serial Non-Participation in Choice Models." Environ. Resour. Econ. 42(2): 211-226.
[7] Greene, W. H., D. A. Hensher (2003). "A latent class model for discrete choice analysis: contrasts with mixed logit." Transport. Res. B-Meth. 37(8): 681-698.
[8] Provencher, B., R. C. Bishop (2004). "Does accounting for preference heterogeneity improve the forecasting of a random utility model? A case study." J. Environ. Econ. Manag. 48(1): 793-810.
[9] Hynes, S. et al. (2008). "Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models." Am. J. Agr. Econ. 90(4): 1011-1027.
[10] Kosenius, A. K. (2010). "Heterogeneous preferences for water quality attributes: The Case of eutrophication in the Gulf of Finland, the Baltic Sea." Ecol. Econ. 69(3): 528-538.