Indicator for Top-Coding Effects on a Household Survey Income Elasticity of Demand Estimates

Daniel K. Yang
Office of Survey Methods Research
U.S. Bureau of Labor Statistics
FCSM Government Advances in Statistical Programming Workshop
October 24, 2018
Overview

- Introduction: Consumer Expenditure Surveys (CE) and Statistical disclosure limitation (SDL)
- Data utility and economics model
- Impact on Income Elasticity of Demand and Proportion Ratio Indicator:
  - Log linear regression model for expenditures
  - Propensity model for consumption
- Conclusion
Consumer Expenditure Survey

- Consumer Expenditure Survey (CE) Collects information on the buying habits of U.S. consumers.
- CE’s goal: Provides data on expenditures, income, and consumer unit (families and single consumers) characteristics.
- Balance: confidentiality vs. satisfactory data utility.
CE SDL Process

- CE microdata release requires statistical disclosure limitation (SDL).
- Objective: Conceal personally identifiable information (PII) to preserve the confidentiality and anonymity of survey participants.
- Production Process: “top-coding” and numerical impact.
Top-coding

Olivia Rivers (F, 25),
Salary: $542,508
Silver Spring, Montgomery County, MD (pop 71,452)

Note: This example is completely fabricated. For illustration purposes only.
Top-coding Illustration

Extreme values

Critical value

Top-coding Illustration (cont.)

Top-coded values

e.g. Replace Olivia’s Salary with the top-coded value: $246,921

Data Utility Measures

■ Analysis-specific:
  ➢ Compare regression coefficients from confidential data vs. top-coded data for the same analysis.
  ➢ Confidence Interval Overlap (IO) or Ellipsoid Overlap (EO).

■ Global:
  ➢ Compare propensity score percentiles.
  ➢ Compare clusters in cluster analysis.
  ➢ Compare Empirical Cumulative Density Functions (CDF’s).
Economics Model: Income Elasticity of Demand

- Income Elasticity:

\[
\frac{\partial E(y \mid x)}{\partial x_j} \cdot \frac{x_j}{E(y \mid x)}
\]

- Here, \( y \) – Expenditure, \( x \) – covariates, \( x_j \) - household income.

Cragg’s Double-Hurdle Model

The unconditional expectation is

\[ E(y|x) = P(y > 0|x)E(y|x, y > 0) + P(y = 0|x)E(y|x, y = 0) \]

\[ = P(y > 0|x)E(y|x, y > 0) + 0 \]

\[ = P(y > 0|x)E(y|x, y > 0). \]

Probability of spending conditional on income: logistic regression model

Expected amount of spending conditional on income: log linear regression model
\[ E(y|x) = P(y > 0|x) \cdot E(y|x, y > 0) \]

- Assume a Logistic propensity model of consumption:
  \[ P(y > 0 \mid x) = \Psi(x\gamma) = \frac{e^{x\gamma}}{1 + e^{x\gamma}} \]

- If assume the outcome follows:
  \[ \log(y_i) \mid y_i > 0 = x_i\beta + \varepsilon_i, \varepsilon_i \mid x_i \sim N(0, \sigma^2) \]
  where \( x_i = [1, \log(x_j), x_{i,k \neq j}] \)

- Unconditional expectation of \( E(y \mid x) \) is
  \[ E(y \mid x) = \Psi(x\gamma)\exp(x\beta + \sigma^2/2) \]
Income Elasticity of Demand

\[ \tau_{x_j} = \frac{\partial E(y \mid x)}{\partial x_j} \frac{x_j}{E(y \mid x)} = \gamma_i \left[ 1 - \Psi(x \gamma) \right] x_j + \beta_j \]

- Coefficient from logistic model
- Coefficient of income from log linear model
Expenditure Data and Demographics

- CE Data: 2008 public released micro data and confidential data.
- Expenditure outcomes: Utilities, Domestic Services, Transportation, Shelter, Medical Supplies, Major Appliances, Other Vehicle, and New Cars and Trucks
- Covariates (adopted from Omori 2010): household (HH) income, family type (ref.: married couple), geographical region (ref.: Northeast), numbers of children (age 0-5, 6-12 and 12-18), reference person's: education attainment (ref.: Less than HS), Occupation (ref.: Other), ethnicity (ref.: White), age.

ref.: reference level, HS: high school
Log Linear Part of the Model

$-\beta \ln(\text{Household Income})$ and 95% CI (1)
Log Linear Part of the Model

$-\beta \ln(\text{Household Income})$ and 95% CI (2)
Logistic P.S. Part of the Model

$\gamma$ Household Income and 95% CI (1)

P.S.: propensity scores

2008 Logistic Propensity Model of Non-0 Expenditures vs. HH Income
Logistic P.S. Part of the Model

$\gamma_{\text{Household Income}}$ and 95% CI (2)

2008 Logistic Propensity Model of Non-0 Expenditures vs. HH Income
Propensity Scores: Domestic Services
Propensity Scores: Medical Supplies
Propensity Scores Curve: New Cars and Trucks
Income Elasticity (1)
Income Elasticity (2)
## Income Elasticity of Demand

<table>
<thead>
<tr>
<th>Expenditures</th>
<th>Ratio of Top-coded over Confidential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilities</td>
<td>1.012</td>
</tr>
<tr>
<td>Domestic Services</td>
<td>1.034</td>
</tr>
<tr>
<td>Transportation</td>
<td>1.009</td>
</tr>
<tr>
<td>Shelter</td>
<td>1.006</td>
</tr>
<tr>
<td>Medical Supplies</td>
<td>1.176</td>
</tr>
<tr>
<td>Major Appliances</td>
<td>1.268</td>
</tr>
<tr>
<td>Other Vehicle (i.e. Motorcycle and Airplanes)</td>
<td>2.056</td>
</tr>
<tr>
<td>Cars and Trucks, New</td>
<td>1.922</td>
</tr>
</tbody>
</table>
Proportion Ratio Indicator:

- Proportion Ratio Indicator (PRI) of Top-coding:

\[ PRI = \frac{\text{proportion of top-coded purchaser}}{\text{proportion of nontop-coded purchaser}} - 1 \]
## Income Elasticity of Demand

<table>
<thead>
<tr>
<th>Expenditures</th>
<th>Ratio of Top-coded over Confidential</th>
<th>Proportion Ratio Indicator (PRI) of Top-coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilities</td>
<td>1.012</td>
<td>0.00026</td>
</tr>
<tr>
<td>Domestic Services</td>
<td>1.034</td>
<td>0.6840</td>
</tr>
<tr>
<td>Transportation</td>
<td>1.009</td>
<td>0.0021</td>
</tr>
<tr>
<td>Shelter</td>
<td>1.006</td>
<td>0.0037</td>
</tr>
<tr>
<td>Medical Supplies</td>
<td>1.176</td>
<td>0.2420</td>
</tr>
<tr>
<td>Major Appliances</td>
<td>1.268</td>
<td>0.2313</td>
</tr>
<tr>
<td>Other Vehicle (i.e. Motorcycle and Airplanes)</td>
<td>2.056</td>
<td>0.3882</td>
</tr>
<tr>
<td>Cars and Trucks, New</td>
<td>1.922</td>
<td>0.4313</td>
</tr>
</tbody>
</table>
Summary

- No difference in log linear model between confidential and top-coded data.
- Differences from some of the propensity models. This translates into some differences in income elasticity of demand for some expenditures.
- Proportion Ratio Indicator (PRI) appears to reflect those differences.
Acknowledgements and Disclaimer

We would like to acknowledge our fellow researchers and managers from the program office.

The views expressed in this paper are those of the author(s) and do not necessarily reflect the policies of the Bureau of Labor Statistics.
THANK YOU!
Contact Information

Daniel K. Yang
Research Mathematical Statistician
Office of Survey Methods Research (OSMR)

www.bls.gov/osmr/home.htm
yang.daniel@bls.gov

Disclaimer: Any opinions expressed in this paper are those of the author(s) and do not constitute policy of the Bureau of Labor Statistics.