

---

# **Numerical Impact of a Simple Random Subsample on Consumer Spending for Children**

**Daniel K. Yang**

**Office of Survey Methods Research (OSMR)**

**U.S. Bureau of Labor Statistics**

**Jeffrey M. Gonzalez**

**Agency for Healthcare Research and Quality (AHRQ)**

*Content does not represent BLS policy*

# Consumer Expenditure Survey

---

- Consumer Expenditure (CE) Survey  
Collects information on the buying habits of U.S. consumers.
- Provides data on expenditures, income, and consumer unit (families and single consumers) characteristics.
- Two components: Quarterly Interview Survey (focus of this study) and the Diary Survey

# Motivation

---

- The Division of Consumer Expenditure Survey (DCES) has been conducting CE redesign research.
- As a part of the redesign, CE had been exploring a variety of data collection methods to increase data quality and not add new burden on respondents.
- One of the considered method is the split questionnaire design.
- A simple random subsample is akin to implementing a split questionnaire design (see Raghunathan and Grizzle 1995)

# Motivation cont.

---

- The DCES and researchers apply different economic models to analyze consumer expenditures.
- OSMR has been conducting a numerical evaluation of the impact of a simple random subsample on various types of economic analyses:
  - ▶ Understand how data users and their respective economic analyses might be affected by a simple random subsample.
  - ▶ Learn about the process of assessing the impact of a simple random subsample on data users.
  - ▶ The outputs of this process will inform the work of other DCES teams and provide beneficial inputs to the CE redesign.

# Objective

---

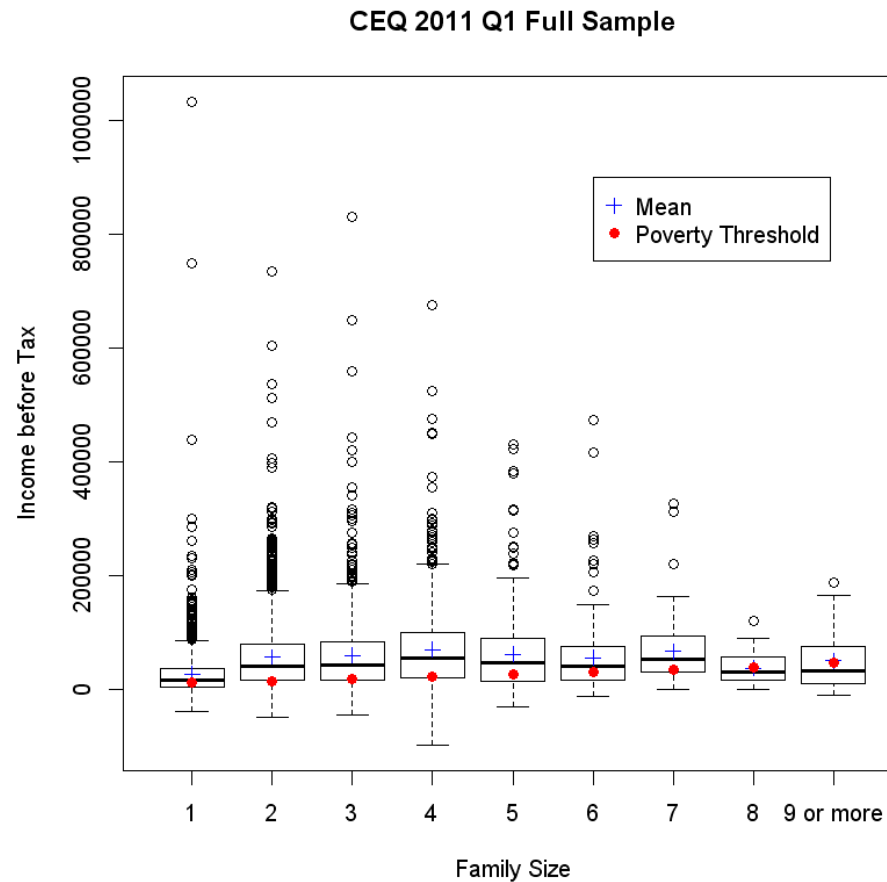
- What are the specific sensitivities of utilizing a simple random subsample to conduct the various economic analyses? e.g., are there particular parameters of the economic models that are compromised?
- Provide preliminary recommendations for general changes to the design and statistical methods used to collect and analyze the CE data.

# Data Sources

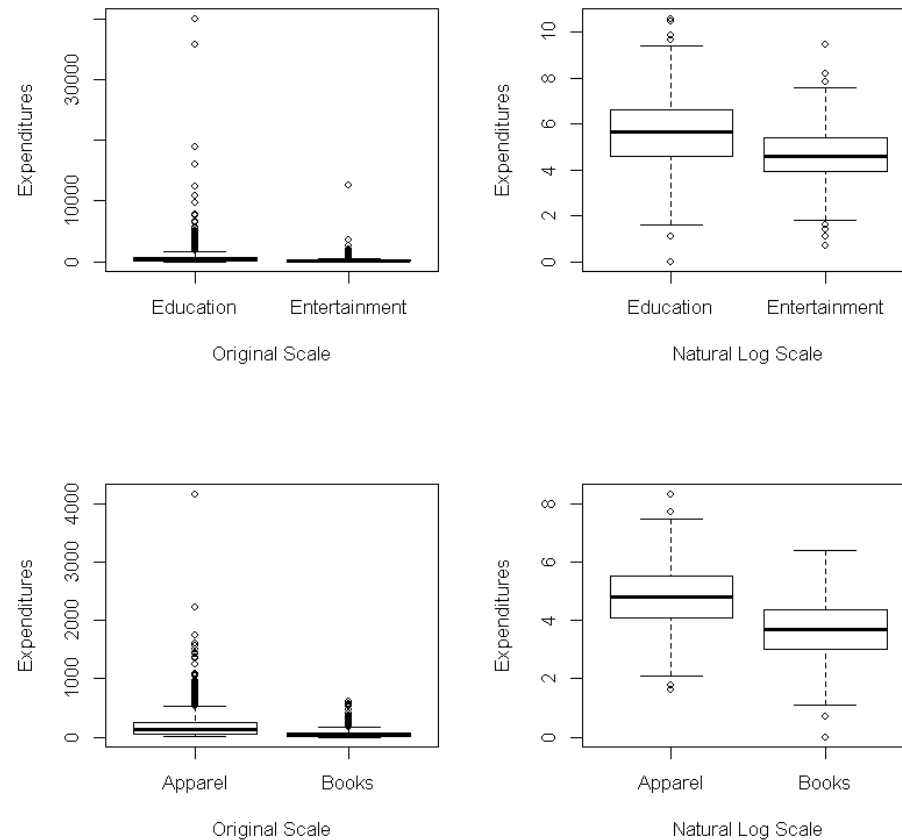
---

- Our analyses focus on household spending patterns for children.
- Selected the first quarter of 2011 CEQ data (CEQ 2011 Q1), excluded consumer units (CU) with:
  - ▶ missing values for demographic and expenditure variables,
  - ▶ households with adult children,
  - ▶ and households with expenditures on gifts for non-CU children.

# Figure 1: Household incomes before tax, poverty threshold among family size



# Figure 2: Box Plot of Children's Expenditures: Original Scale vs. Natural Log Scale





# Methodology

Model	Estimates
Logistic and linear regression model	coefficient and their associated standard errors (SE)
two-stage regression Cragg's model	1 <sup>st</sup> stage: probability of purchase and predicted expenditure (buyers only) 2 <sup>nd</sup> stage: marginal propensity to consume (MPC) and elasticity
two-stage hierarchical generalized linear mixed (HGLM) Cragg's model to account for the variation among geographical regions	1 <sup>st</sup> stage: probability of purchase and predicted expenditure (buyers only) 2 <sup>nd</sup> stage: MPC and elasticity

Note: May include variable transformations such as the natural logarithm (to insure model assumptions are met).

# Model Specifications for logistic and linear models

- Expenditure categories on children: education, entertainment, books, and apparel.
- Logistic regression:  $\text{logit}(\mathbf{P}) = \mathbf{Xb}$
- Linear regression:  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}$
- Notation:
  - ▶  $\mathbf{P}$  – whether a household reported a nonzero expenditure on children.
  - ▶  $\mathbf{Y}$  – nonzero expenditure on children
  - ▶  $\boldsymbol{\beta}, \mathbf{b}$  – corresponding coefficients
  - ▶  $\mathbf{X}$  – covariates matrix (income percentile, socioeconomic and demographic characteristic variables)

# Socioeconomic and Demographic Variables for logistic and linear models

---

- Household type (married couple, single mother, single father and Cohabiting)
- Geographical region (Northeast, Midwest, South, West)
- Number of children in the household: ages 0-5, ages 6-12, and ages 13-18.
- Reference person's:
  - ▶ Age
  - ▶ Ethnicity (White, African-American, Hispanic, Other)
  - ▶ Education attainment (Less than high school, High school, Some college, College and higher)
  - ▶ Occupation level (Managerial/professional, Administrative, Other)

**Table 1. Logistic regression results: likelihood of expenditures on items in selected categories (partial)**

Parameter	Analysis	Education		Entertainment	
		Estimate	95% CI	Estimate	95% CI
Household type: <sup>2</sup> Single mother	CEQ 2011 Q1 (wt)	1.023	(1.021, 1.026)	0.987	(0.985, 0.989)
	SRSWORSS (wt)	1.438	(1.433, 1.442)	1.081	(1.078, 1.085)
Single father	CEQ 2011 Q1 (wt)	0.756	(0.753, 0.760)	0.582	(0.580, 0.585)
	SRSWORSS (wt)	1.138	(1.130, 1.145)	1.040	(1.033, 1.047)
Cohabiting	CEQ 2011 Q1 (wt)	0.640	(0.638, 0.641)	0.697	(0.695, 0.699)
	SRSWORSS (wt)	0.607	(0.605, 0.609)	0.686	(0.683, 0.688)
Income percentile	CEQ 2011 Q1 (wt)	1.012	(1.012, 1.012)	1.020	(1.020, 1.020)
	SRSWORSS (wt)	1.011	(1.011, 1.011)	1.016	(1.016, 1.016)
Education: <sup>3</sup> High school	CEQ 2011 Q1 (wt)	1.391	(1.387, 1.395)	1.509	(1.505, 1.513)
	SRSWORSS (wt)	1.840	(1.832, 1.847)	1.306	(1.301, 1.310)
College and higher	CEQ 2011 Q1 (wt)	2.443	(2.437, 2.449)	2.412	(2.406, 2.418)
	SRSWORSS (wt)	3.574	(3.561, 3.588)	2.623	(2.613, 2.632)
Occupation: <sup>4</sup> Managerial/professional	CEQ 2011 Q1 (wt)	1.285	(1.283, 1.287)	1.546	(1.544, 1.549)
	SRSWORSS (wt)	1.466	(1.463, 1.470)	1.562	(1.559, 1.566)
Administrative	CEQ 2011 Q1 (wt)	0.916	(0.914, 0.919)	1.088	(1.085, 1.091)
	SRSWORSS (wt)	0.971	(0.967, 0.974)	1.259	(1.254, 1.264)

**Table 2. OLS regression results: estimates of (the natural logarithm of) quarterly expenditures on items in selected categories (partial)**

Parameter	Analysis	Education		Entertainment	
		Estimate	SE	Estimate	SE
Intercept	CEQ 2011 Q1 (wt)	4.456	<sup>7</sup> 0.378	3.145	<sup>7</sup> 0.268
	SRSWORSS (wt)	4.950	<sup>7</sup> 0.502	3.164	<sup>7</sup> 0.367
Household type: <sup>2</sup>					
Single mother	CEQ 2011 Q1 (wt)	-0.337	<sup>9</sup> 0.165	-0.228	0.119
	SRSWORSS (wt)	-0.634	<sup>8</sup> 0.210	-0.200	0.161
Single father	CEQ 2011 Q1 (wt)	-0.659	0.340	0.210	0.235
	SRSWORSS (wt)	-0.938	<sup>9</sup> 0.442	0.321	0.312
Cohabiting	CEQ 2011 Q1 (wt)	-0.229	0.181	-0.237	0.124
	SRSWORSS (wt)	-0.171	0.250	-0.175	0.176
Income percentile	CEQ 2011 Q1 (wt)	0.006	<sup>8</sup> 0.002	0.006	<sup>7</sup> 0.001
	SRSWORSS (wt)	0.007	<sup>8</sup> 0.002 <sup>†</sup>	0.006	<sup>7</sup> 0.002
Education: <sup>3</sup>					
High school	CEQ 2011 Q1 (wt)	0.494	<sup>9</sup> 0.218	0.118	0.156
	SRSWORSS (wt)	0.154	0.318	0.285	0.220
College and higher	CEQ 2011 Q1 (wt)	0.971	<sup>7</sup> 0.200	0.378	<sup>9</sup> 0.146
	SRSWORSS (wt)	0.736	<sup>8</sup> 0.293	0.443	<sup>9</sup> 0.207
Occupation: <sup>4</sup>					
Managerial/professional	CEQ 2011 Q1 (wt)	0.317	<sup>8</sup> 0.110	0.287	<sup>7</sup> 0.077
	SRSWORSS (wt)	0.183	0.146	0.401	<sup>7</sup> 0.110
Administrative	CEQ 2011 Q1 (wt)	-0.053	0.205	0.052	0.138
	SRSWORSS (wt)	-0.194	0.269	-0.099	0.187

## Table 3: Cragg's model estimates under "ceteris paribus" (partial)

Expenditure, Ceteris paribus criteria	Analysis	Married couple	Single mother
<b>Education</b>			
Probability of purchase	CEQ 2011 Q1	0.45	0.30
	SRSWORSS	0.42	0.37
Predicted expenditure (buyers only)	CEQ 2011 Q1	336.91	203.54
	SRSWORSS	394.8	190.57
Marginal propensity to consume	CEQ 2011 Q1	0.06	0.01
	SRSWORSS	0.01	0.01
Elasticity	CEQ 2011 Q1	0.15	0.11
	SRSWORSS	0.22	0.24
<b>Entertainment</b>			
Probability of purchase	CEQ 2011 Q1	0.51	0.33
	SRSWORSS	0.50	0.37
Predicted expenditure (buyers only)	CEQ 2011 Q1	120.81	73.9
	SRSWORSS	125.7	79.8
Marginal propensity to consume	CEQ 2011 Q1	0.05	0
	SRSWORSS	0.01	0
Elasticity	CEQ 2011 Q1	0.30	0.06
	SRSWORSS	0.30	0.14

## Table 4: Hierarchical Generalized Linear Mixed (HGLM) Cragg's model estimates under "ceteris paribus" (partial)

Expenditure, Ceteris paribus criteria	Analysis	Married couple	Single mother
<b>Education</b>			
Probability of purchase	CEQ 2011 Q1	0.61	0.57
	SRSWORSS	0.60	0.59
Predicted expenditure (buyers only)	CEQ 2011 Q1	335.6	198.31
	SRSWORSS	389.45	178.52
Marginal propensity to consume	CEQ 2011 Q1	0.05	0.01
	SRSWORSS	0.01	0.01
Elasticity	CEQ 2011 Q1	0.17	0.18
	SRSWORSS	0.27	0.33
<b>Entertainment</b>			
Probability of purchase	CEQ 2011 Q1	0.62	0.58
	SRSWORSS	0.62	0.59
Predicted expenditure (buyers only)	CEQ 2011 Q1	120.54	72.60
	SRSWORSS	124.95	76.93
Marginal propensity to consume	CEQ 2011 Q1	0.04	0*
	SRSWORSS	0	0
Elasticity	CEQ 2011 Q1	0.32	0.10
	SRSWORSS	0.33	0.21

# Evaluated Economic Analysis Models

---

- Linear regression and generalized logistic regression models: different coefficient estimates and generally higher SE
- Two-stage regressions Cragg's model: 1<sup>st</sup> stage estimates, 2<sup>nd</sup> stage estimates
- Two-stage Cragg's model using generalized linear mixed regressions to account for geographical region variations: 1<sup>st</sup> stage estimates, 2<sup>nd</sup> stage estimates, comparing to classic Cragg's estimates
- This is a work in progress and some statements may change.



# Preliminary Recommendations

---

- Oversampling: increase d.f. for var. est.
- Dynamic interview: identify groups, oversample and apply weights
- Pooling additional quarters/years of: e.g. 4 years data
- Implement Bootstrap: large number of subsamples
- Hierarchical modeling with random components: reflect the complexity of CE data, combine at least one year data

# Future research

---

- The DCES has developed a new CE redesign proposal. Further studies are needed to evaluate those economic models under this proposed redesign condition.
- Future analyses might also include drawing repeated simple random subsamples.
- Recent developments in economics research illustrate the interest of generating a composite statistic from multi-dimensional measurements.

# Acknowledgement

---

- Branch of Research and Program Development (BRPD), DCES.
  - ▶ Adam N. Safir, Chief
  - ▶ Laura A. Paszkiewicz, Economist
- Branch of Information and Analysis (BIA), DCES
  - ▶ Steve W. Henderson, Chief
  - ▶ Geoffrey D. Paulin, Senior Economist
- Office of Survey Methods Research (OSMR)
  - ▶ John L. Eltinge, Associate Commissioner
  - ▶ Wendy L. Martinez, Director, Mathematical Statistics Research Center (MSRC)

# THANK YOU!

---

## Federal Committee on Statistical Methodology

4-6 November, 2013  
Washington, D.C.



# Contact Information

---

**Daniel K. Yang**

Research Mathematical Statistician  
Office of Survey Methods Research (OSMR)

[www.bls.gov/osmr/home.htm](http://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.*

