

Paradata in the Consumer Expenditure Survey: Overview and recent research

Laura Paszkiewicz

Senior Economist
Division of Consumer Expenditure Survey
Bureau of Labor Statistics
FCSM Conference
January 10-12, 2012



www.bls.gov

Overview

- Overview of paradata files/CHI
- Newly available paradata files on the public use CD.
- Recent research projects in CE using paradata



CE Paradata

- Contact History Instrument (CHI)
- Automatically collected data in Computer Assisted Personal Interview (CAPI)
- Field representatives' (FR) observations in "the back" of CAPI.



3

Contact History Instrument (CHI)

- FR's asked to complete questions after each contact attempt
- Used by multiple surveys including
 - ▶ National Health Interview Survey (NHIS)
 - ▶ Current Population Survey (CPS)
 - ▶ Consumer Expenditure Survey (CE)

* VARIABLES NOT UNIQUE TO CE DATA *



4

Contact History Instrument (CHI)

- Examples of types of data collected with CHI:
 - ▶ strategies for attaining the interview
 - ▶ respondent concerns/reasons for non-response
 - ▶ mode of contact attempt (telephone, in person)
 - ▶ description of non-contact by mode
 - ▶ language of the contact attempt
 - ▶ date and time of contact attempt



5

Contact History Instrument (CHI)

Respondent Concerns:

| | |
|--|--|
| Not interested / Does not want to be bothered | Hostile or threatens FR |
| Too busy | Other household members tell respondent not to participate |
| Interview takes too much time | Talk only to specific household member |
| Breaks appointments (puts off FR indefinitely) | Family issues |
| Scheduling difficulties | Respondent requests same FR as last time |
| Survey is voluntary | Gave that information last time |
| Privacy concerns | Asked too many personal questions last time |
| Anti-government concerns | Last interview took too long |
| Does not understand survey | Too many interviews |
| Survey content does not apply | Intends to quit survey |
| Hang-up / slams door on FR | No concerns |



Other Paradata in the CE

- Data collected by CAPI
 - ▶ Time spent in each sections
 - ▶ Key Strokes (not included on public-use)
- Observed data by FR
 - ▶ Use of records by respondent
 - ▶ Use of Infobook by the respondent
- Other data about the survey
 - ▶ Response rates (in CE documentation)
 - ▶ Information about costs

PARADATA FILES ON THE PUBLIC USE MICRODATA

Paradata Files in the CE

- MCHI data set
 - ▶ Data collected in the CHI instrument
 - ▶ One record per contact attempt
 - ▶ Multiple records per CU per Wave
- FPAR data set
 - ▶ Data collected in CAPI
 - ▶ One record per CU per Wave (up to 5 total interview per CU)
 - ▶ Relates to the final contact attempt



9

Paradata Files in the CE

- Each file contains 9 quarters of data (i.e., 2009Q1 – 2011Q1 on 2010 CD)
 - ▶ Up to 5 interviews per Consumer Unit (CU)
 - ▶ Interviews (“waves”) 1 – 5
 - ▶ Completed interviews and Incomplete interviews (Type A)



10

Considerations

- Wave 1 interviews will ONLY have paradata included in the files (MCHI and FMIS).

- No demographic
- No income
- No expenditures
- No weights

If looking at the data independently, must match with other files to obtain the information (if available)



11

Considerations

- Non-Interviews will only have the MCHI data

- No demographic
- No income
- No expenditures
- No weights

If looking at the data independently, must match with other files to obtain the information, if available



12

RECENT RESEARCH USING THE PARADATA



Recent Research

- Effects of pre-paid incentives on first occurrence of non-response
- Optimal number of contact attempts to reduce cost/maintain quality
- Use of interviewer observed info to adjust weights for non-response bias
- Using CHI to identify reluctant respondents



Effect of incentives on 1st drop out

Background: Incentive study in 2005, 4 treatments, all prior to wave 1 interview

- ▶ \$40 prepaid card with letter priority mail
- ▶ \$20 prepaid card with letter priority mail
- ▶ Advance letter priority mail
- ▶ Advance letter regular mail (Control)

Research Objective: How incentives affect sample units participation across the 5 wave panel

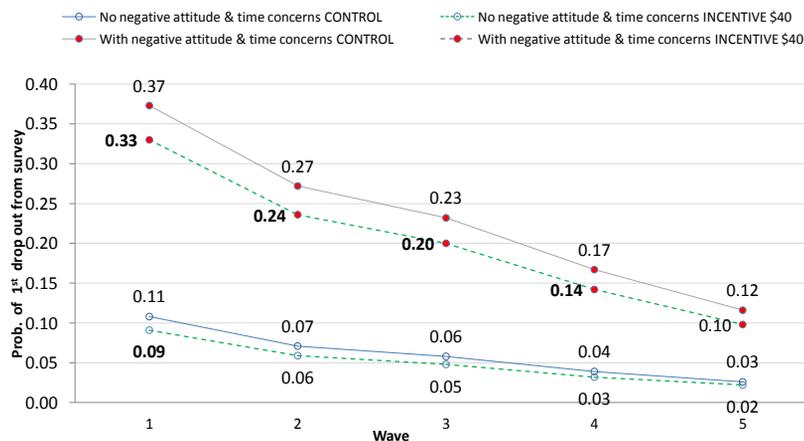


15

Effect of incentives on 1st drop out

Methodology: discrete time survival analysis for non-repeatable events (Singer & Willett 1993)

Hazard profiles for Control and \$40 incentive treatment groups, conditioned on selected concerns about survey participation observed in Wave 1



Adjusting weights for non-response bias

Background:

- CE currently adjusts weights to mitigate effects of nonresponse bias using a variety of information
- Little and Vartivarian (2005) suggest that most effective auxiliary information to use to adjust weights will be highly correlated with response propensity and the survey outcome (expenditures)

Research objectives:

- Investigate value in adding interviewer observed information (tenure and perceived concerns) in CE non-responses adjustment



17

Adjusting weights for non-response: Response propensities and correlations

| Tenure | | | |
|------------|----------------|-----------------|--------|
| | Respondents | Non-Respondents | Total |
| Owner (1) | 13,604 (73.8%) | 4,816 | 18,420 |
| Renter (0) | 6,705 (76.7%) | 2,034 | 8,739 |
| Total | 20,309 (74.8%) | 6,850 | 27,159 |

$$\rho(z, p) = -0.0309$$

| Perceived Concerns Index | | | | |
|--------------------------|--------|---------|---------|--------|
| | N | Mean | Std Dev | Median |
| Respondents (1) | 20,309 | -0.2681 | 2.451 | -1.105 |
| Nonrespondents (0) | 6,850 | 0.7949 | 3.061 | -0.279 |

$$\rho(z, p) = -0.1737$$



18

Adjusting weights for non-response: Survey outcome correlations

| Expenditure Category | Tenure | PCI |
|--|---------------|---------|
| Total (sum of all major categories) | 0.2529 | -0.0273 |
| Housing (including shelter, utilities, etc.) | 0.1657 | 0.0064 |
| Transportation | 0.1155 | -0.0239 |
| Food (at and away from home) | 0.2128 | 0.0268 |
| Personal insurance and pensions | 0.2025 | -0.0460 |
| Health care | 0.2109 | -0.0146 |
| Entertainment | 0.1119 | -0.0318 |
| Cash contributions | 0.0707 | -0.0010 |
| Apparel and services | 0.0832 | -0.0231 |
| Education | 0.0421 | -0.0058 |
| Miscellaneous | 0.0598 | -0.0172 |
| Tobacco and smoking | -0.0284 | -0.0287 |
| Alcoholic beverages | 0.0415 | -0.0268 |
| Personal care | 0.1530 | -0.0215 |
| Reading | 0.1144 | -0.0344 |



19

Adjusting weights for non-response: Nonresponse adjusted mean expenditures

| Expenditure Category | Traditional | Tenure | PCI | Both |
|----------------------|-------------|-----------|-----------|-----------|
| Total | 11,524.15 | 11,573.63 | 11,484.78 | 11,522.55 |
| Housing | 3,995.30 | 4,007.63 | 3,999.57 | 4,008.78 |
| Transportation | 1,882.61 | 1,892.17 | 1,873.79 | 1,881.36 |
| Food | 1,752.30 | 1,757.46 | 1,758.05 | 1,761.83 |
| Insurance | 1,174.84 | 1,182.58 | 1,156.75 | 1,162.63 |
| Health care | 771.62 | 776.84 | 769.82 | 773.73 |
| Entertainment | 581.61 | 585.03 | 574.12 | 576.69 |
| Cash contributions | 402.55 | 405.20 | 397.32 | 399.35 |
| Apparel | 257.38 | 257.98 | 255.80 | 256.37 |
| Education | 241.64 | 242.88 | 239.96 | 241.04 |
| Miscellaneous | 184.04 | 185.01 | 181.62 | 182.37 |
| Tobacco | 93.18 | 93.03 | 92.59 | 92.49 |
| Alcohol | 83.94 | 84.12 | 83.12 | 83.25 |
| Personal care | 76.04 | 76.42 | 75.57 | 75.85 |
| Reading | 27.11 | 27.29 | 26.89 | 26.82 |



20

Adjusting weights for non-response: Ratio of variance estimates (PCI)

| Expenditure Category | Variance (SE) Traditional | Variance (SE) Tenure | R | Variance (SE) PCI | R | Variance (SE) Both | R |
|----------------------|------------------------------|-------------------------|-------|----------------------|-------|-----------------------|-------|
| Total | 26,892.10 (163.99) | 28,424.06 (168.59) | 0.946 | 29,171.34 (179.80) | 0.922 | 26,924.15 (164.09) | 0.999 |
| Housing | 3,357.74 (57.95) | 3,325.00 (57.66) | 1.010 | 3,311.96 (57.55) | 1.014 | 3,337.44 (57.77) | 1.000 |
| Transportation | 2,109.53 (45.93) | 2,293.32 (47.89) | 0.920 | 2,102.10 (45.85) | 1.004 | 2,099.06 (45.82) | 1.005 |
| Food | 345.72 (18.59) | 359.56 (18.96) | 0.962 | 364.55 (19.09) | 0.948 | 413.00 (20.32) | 0.837 |
| Insurance | 968.39 (31.12) | 914.84 (30.25) | 1.059 | 1,560.95 (39.51) | 0.620 | 1,296.50 (36.01) | 0.747 |
| Health care | 296.41 (17.19) | 349.69 (18.70) | 0.845 | 289.35 (17.01) | 1.021 | 310.78 (17.63) | 0.951 |
| Entertainment | 237.95 (15.10) | 227.66 (15.09) | 1.001 | 310.43 (17.62) | 0.734 | 269.49 (16.42) | 0.846 |
| Cash contributions | 512.51 (22.64) | 530.76 (23.04) | 0.966 | 517.79 (22.76) | 0.990 | 509.26 (22.57) | 1.006 |
| Apparel | 37.95 (6.16) | 38.71 (6.22) | 0.981 | 39.43 (6.28) | 0.963 | 38.33 (6.19) | 0.990 |
| Education | 139.80 (11.82) | 137.53 (11.73) | 1.017 | 147.76 (12.16) | 0.946 | 142.02 (11.92) | 0.984 |
| Miscellaneous | 33.08 (5.75) | 34.69 (5.89) | 0.954 | 37.26 (6.10) | 0.888 | 34.71 (5.89) | 0.953 |
| Tobacco | 9.25 (3.04) | 9.01 (3.00) | 1.026 | 8.58 (2.93) | 1.078 | 8.53 (2.92) | 1.094 |
| Alcohol | 10.62 (3.26) | 10.45 (3.23) | 1.016 | 12.21 (3.49) | 0.870 | 11.88 (3.45) | 0.894 |
| Personal care | 3.80 (1.95) | 3.57 (1.89) | 1.063 | 4.49 (2.12) | 0.847 | 4.02 (2.01) | 0.944 |
| Reading | 0.92 (0.96) | 0.98 (0.99) | 0.941 | 1.03 (1.02) | 0.891 | 0.96 (0.98) | 0.959 |

'Optimal' number of contact attempts

Background: Disproportionate effort in collecting data from difficult-to-interview respondents – costly.

Research Objective: Determine an optimal contact attempt threshold by evaluating tradeoffs of survey performance measures along the continuum of contact attempts

- ▶ Response rate, sample representativeness, reporting quality, and cost

'Optimal' number of contact attempts

Methodology

| Survey performance measures | | Evaluation of recommendation: 7 |
|-----------------------------|--------------------------|---|
| Response rate (↑) | Reporting quality (↑) | Effect on relative NR bias |
| Total expenditures (↑) | Collection costs (↓) | Effect expenditure estimates • Reweighting macro |

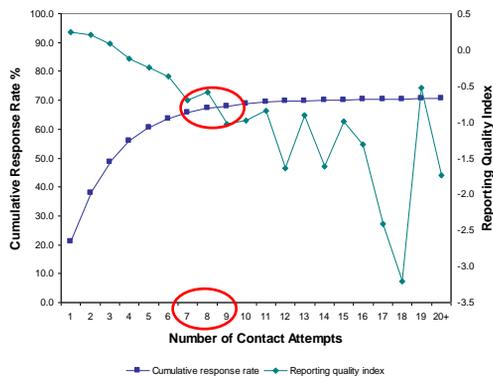
- Formed aggregated groups: 1-4, 5-7, 8+ attempts (cluster analysis)
- Data: Wave 1, April 2006 – March 2008



'Optimal' number of contact attempts

| Survey performance measures | Number of contact attempts | | |
|--|----------------------------|------------------|---------------|
| | 1-4 (n=7,698) | 5-7 (n=1,368) | 8+ (n=673) |
| Cumulative response rate (%) | 55.8 | 65.7 | 70.6 |
| Reporting quality index -- ordinal | 0.16 | -0.38 | -1.01 |
| Total expenditures reported (\$) -- not sig. diff. | 5,196 | 4,710 | 6,205 |
| % of Collection costs | 51.5 | 25.3 | 23.2 |

About 85% of eligible sample units resolved by 7th attempt



Indicator of reluctant respondents

Motivation

- 2009 CE Interviewer Survey (Mockovak, Edgar, & To 2010): “ *reluctant respondents hurt data quality and increase collection costs*”
- Existing **CV indicator**: CAPI backend question “*Was this a converted refusal?*”
 - ⊖ Strict criteria applied, asked only of completed interviews
- New from April 2010: processed demographic Wave 1 (Phase 2), new CHI variable, new CAPI paradata (refined levels: info book, record use)

Methodology

- Reluctant respondent: “any reluctance reported in contact attempt history”
- **CHI indicator** : compare alternative indicators constructed from CHI
 - ▶ Selection criteria
 1. identify a larger number of respondents as reluctant than the CV indicator
 2. differentiates between reluctant & cooperative respondents
- Compare characteristics of coop R. with reluctant R. using CHI indicator & CV indicator on: *demographics, data collection effort, reporting quality*
 - ▶ examined trends across the groups
- Data: CEQ April – Sept 2010



Indicator of reluctant respondents

| Candidate indicators: various forms of reluctance reported in the CHI | No. reporting (N=21,538) | No. resolved as Interviews | Row percent |
|---|--------------------------|----------------------------|-------------|
| 1. Soft refusal (CTTYPE, NONINTR3) | 2,921 | 628 | 21.5 |
| 2. Interim hard refusal (COUTCOME) | 2,737 | 10 | 0.004 |
| Hostile respondent | 1,112 | 3 | |
| Time related excuses | 601 | 3 | |
| Language problems | 1,024 | 4 | |
| Other | 0 | 0 | |
| 3. >=1 concern & 0 “no concern” (RSPDNTn) ★ | 8,570 | 4,546 | 53.0 |

| Composition of respondent concerns | All respondents (N=17,152) | Reluctant respondents | | Nonrespondents (N=4,431) |
|---|----------------------------|-------------------------|------------------------|--------------------------|
| | | CHI Indicator (N=4,546) | CV Indicator (N=1,895) | |
| Summary index of concerns (PCI) | -0.68 | 1.75 | 1.19 | 2.61 |
| PC1 – survey attitude/hostility/privacy | -0.85 | 1.40 | 1.07 | 3.29 |
| PC2 – prior wave interview | -0.37 | 2.37 | 1.53 | 1.43 |
| PC3 – time issues | -0.36 | 2.36 | 1.20 | 1.39 |



References

- Tsai, S. & Tan, L. (2010). An exploratory study on the effect of prepaid incentives on the first occurrence of nonresponse in the Consumer Expenditure Interview Survey. (Internal report)
- Gonzalez, J.M., and F. Kreuter. (2010). Evaluating the Impact of Interviewer Observed Auxiliary Information in Nonresponse Adjustments (JPSM presentation).
- Safir, A., and Tan, L. (2009). Using Contact Attempt History Data to Determine the Optimal Number of Contact Attempts: an Exploratory Study. (BLS Statistical Survey Paper)
- Tan, L. & Tsai, S. (2011). An exploration on the use of contact history information to identify reluctant respondents. (Internal report)



27

Contact Information

Laura Paszkiewicz
Senior Economist
Division of Consumer Expenditure
Survey

www.bls.gov/cex

202-691-5119

paszkiewicz.laura@bls.gov

