Discussion of “Split Questionnaire Methods for the Consumer Expenditure Surveys Program”

John L. Eltinge
Office of Survey Methods Research, BLS

December 9, 2010
The views expressed here are those of the author and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics.
Overview

I. Partitioned Designs

II. Context

III. Methodological Issues

IV. Empirical Issues
I. Partitioned Designs

A. Primary Questions

1. For a specified resource base, can we improve the balance of quality/cost/risk in CE data products by assigning some sample units to limited data collection?
I. Partitioned Designs (continued)

B. Possible Examples

1. Split questionnaire – receive only some sections of current CE Quarterly instrument

2. Split questionnaire from (1), with some re-alignment of items from current diary/interview/TPOPS, or some globals

3. Collection of some data (with permission) through administrative records (e.g., grocery loyalty cards)
I. Partitioned Designs (continued)

C. Principal Evaluation Factors

Perceived burden/invasiveness
Cost (fixed and marginal components)
Impact on quality of specific CE products
Other risk factors?

D. How to modify estimation methods to account for partitioned design features?

- Weighting and imputation for CPI cost weights, commonly produced tables
- Construction of public-use datasets
II. Context

A. Survey Design (Defined Broadly)

1. Balances a wide range of factors
   - Cost
   - Respondent burden
   - Quality (coverage, bias, variance)
   - Confidentiality/privacy

2. Requires substantial investments
   (intangible capital)
II. Context (Continued)

B. Features of the Estimands, Measures and Design Factors:

Many practical measures are, at best, approximations to much more complex constructs.

Many design factors and quality characteristics are high-dimensional.

Many components of variability and cost.

Case-specific empirical results are crucial.
III. Methodological Issues

A. Assignment of Instrument Sections (Defined Broadly) to Sample Units

1. Randomized
   - Optimal (or improved) probabilities with multiple estimands

2. Contingent
   - Requires careful evaluation of conditioning events and conditional distributions, as in responsive design

3. Both (1) and (2) can be viewed as extensions of two-phase sample designs
III. Methodological Issues (continued)

B. Construction of Public-Use Datasets

1. Preserve some “full instrument” sample units?

2. Construction of “synthetic datasets”?  
   - cf. Reiter (2009), Raghunathan et al. (2003) and others
   a. Requires clarity on inferential goals, limits
      Ex: GLM fit for central 90% of pop?
   b. Related: Impact of dimensionality
C. Importance of Clarity on Sources of Variability Considered in Evaluation of Bias, Variance and Other Properties

1. Sources:
   - Superpopulation effects
   - Sample design (including subsampling)
   - Unit, wave and item nonresponse
   - Reporting error
   - Imputation effects
     - (including model lack of fit)

2. Conditioning and integration
IV. Empirical Issues

A. Impact on Perceived Burden and Data Quality from:

\[(\text{Aggregate length}) \times (\text{Number of waves}) \times (\text{Cognitive complexity}) \times (\text{Perceived invasiveness})\]
IV. Empirical Issues (continued)

A.1. Examples:

- Elapsed time
- Number of items
- Direct record search/use by respondent
- Informed consent to access records
- Level of detail
- Balance/reconciliation burden (June, 2010 CE Data Users’ Forum)
  - Aggregate level: Income/Expenditure/Savings
  - Section level: Global and detailed follow-up
IV. Empirical Issues (continued)

A.2. Interaction of factors from (A.1) with

a. Consumer Unit Characteristics, e.g.,
   CU structure
   Socioeconomic status

b. Groups of Items:
   Saliency
   Sensitivity
   Complexity of Expenditure
IV. Empirical Issues (continued)

B. Efficiency Improvements Obtained Through:

1. Imputation of non-observed items based on observed demographics, global responses, and individual items
   
   Issues: Predictive performance
   Stability over time

2. Modification of assignment probabilities
IV. Empirical Issues (continued)

C. Cost Structures

1. Examples: Fixed and marginal costs of:
   Unit initiation
   Wave contact
   Incremental sections/time

2. Variation of results from (1) over characteristics of consumer unit

3. Additional costs and risks of modified collection
V. Summary

Fascinating Topic, with Many Open Questions

A. Methodological Issues

B. Empirical Issues

C. Address (A) and (B) in Context Defined by:

1. Quality/cost/risk balance

2. Multidimensional needs of stakeholders, program operations
Contact Information

John L. Eltinge
Associate Commissioner
Office of Survey Methods Research
www.bls.gov
202-691-7404
Eltinge.John@bls.gov