Characterization, Modeling and Management of Inferential Risk, Data Quality Risk and Operational Risk in Survey Procedures

John L. Eltinge
Bureau of Labor Statistics
Eltinge.John@bls.gov

NISS Affiliates' Meeting
April 27, 2007

Disclaimer and Acknowldegements:

The views expressed in this paper are those of the author and do not necessarily reflect the policies of the Bureau of Labor Statistics. The author thanks John Bosley, Alan Dorfman, Larry Ernst, Tom Guterbock, Ben Hansen, Stas Kolienkiov, Keith Muller, Joe Sedransk, Ron Wasserstein, Kirk Wolter and Stan Young for helpful comments on some of the issues considered here.

I. Overview

A. Usual Statistical Approach to Survey Design and Inference

1. Univariate estimand (often a mean, total or ratio)

2. Prior information on population structure

3. Loss function

4. Under complex stochastic structure induced by design or (hierarchical) model, optimize procedure (design and estimators)

B. Conjecture: Traditional approach (A) misses dominant features of the full scope of many multipurpose government surveys

   1. Multidimensional utility functions

   2. Constraints (often uncertain)

C. Several rich classes of research questions arise from (B)

   - Usual cautions about need for well-defined, tractable problems

II. Multidimensional Utility Functions and Priors for Large-Scale Surveys

A. Origins

1. Many estimands

2. Many stakeholders
   - Different utility functions
   - Different degrees of conditioning
   - Different priors

B. Consequences

1. Survey design and operations as optimization exercises subject to multiple constraints, multiple utility functions, uncertainty:

   Taken literally, an ill-defined problem

2. With multiple utility functions, may observe satisficing, minimax behavior or other forms of risk management, rather than optimization as such

3. Conjectures: (Under conditions) given a set of multiple stakeholders with divergent utility functions

a. Simple design-based procedures constitute (approximate) minimax solutions

b. Current design-based and model-based approaches both are problematic, due primarily to narrow scope of risk factors considered

C. Expectations and Behaviors of Stakeholders: Consistent with Literature on Adoption and Diffusion of Technology

1. Large and controversial literature
   (Rogers, 1995; Katz, 2004)

2. View literature with caution, but useful to consider distribution of stakeholders re:
   a. Pace and hurdle rates for adoption
   b. Degree of standardization
   c. Observable risk-reward trade-offs

3. In contrast with many other technologies, applied statistical methodology has

limited feedback loops

for risk and reward

4. Need to find concrete and transparent ways in which to link survey data quality and risk management with perceived value for stakeholders

    a. Limited: Regulatory approaches (OMB, IRBs)

    b. Preferable: More direct linkage with stakeholders' underlying utility functions

III. Suggestion: Expand Scope of Rigorous Statistical Work to Cover Four Dimensions of Survey Risk

A. Usual survey risk (“total survey error”)

B. Inferential risk

C. Data quality risk

D. Operational risk

IV. Usual Survey Risk (Total Survey Error)

A. Standard Decomposition:

Total Survey Error =

= (frame error)

+ (sampling error)

+ (nonresponse effects)

+ (measurement error)

+ (processing effects)

B. Idealized Goal: Under specified models *M*

for each component, allocate agency resources to minimize

$$E_M\{(\hat{\theta} - \theta_P)^2\}$$

for a given total cost

C. Caution on Local and Global Optimization

1. Scientific ideal

   a. Parsimonious model with small number of (controllable) predictors that explain a large proportion of the variability in one or more error components

   b. High degree of generalization across surveys

   c. Response surface analogue:
   Simple main-effects model

   d. Global optimization tractable

2. Perception by many survey program managers and field managers

   a. Optimization, if any, is highly conditional and constrained

   b. Many high-order interactions among predictors

   c. Response surface analogue: Highly irregular surface with many local optima

   d. Little or no prospect for generalization across surveys

3. Distribution of empirical results between (1) and (2)?

4. Diagnostics to evaluate (3)?

V. Inferential Risk

A. "Indications" and "Conclusions":

Tukey (1962, *Ann. Math. Stat.*) contrast between two uses of data analysis

1. "Conclusions" (Box, 1986: "Proof")

   Consistent with traditional approaches to formal inference

Ex: Classical significance testing

Ex: Interval est with prespecified min coverage rate

2. "Indications"

   More consistent with higher-risk exploratory analysis

B. Analogy: Traditional experimental design

1. Distinguish between exploratory and confirmatory experiments (and related analysis methods)

2. Less readily accomplished with large-scale surveys, high fixed costs, severe operational constraints

C. Concern: Do we distinguish clearly between “indication” and “conclusion” in our inferences from survey data?

D. Cf. related concerns about low rates of replicability in epidemiological findings

Young et al. (2006, NISS Workshop)

Ioaioannidis (2005, *J. Am. Med. Assoc.*)

Austin et al. (2006, *J. Clin. Epi.*)

E. Arguably of special concern for a policy-neutral, multiple-stakeholder government statistical agency

VI. Data Quality Risk

A. Government survey agencies generally acknowledge multiple components of data quality

Ex: Brackstone (1999) six components:

1. Accuracy (all components of total survey error, and related inferential issues)
2. Timeliness
3. Relevance
4. Interpretability
5. Accessibility
6. Coherence

B. Risk: Failure in one or more components of data quality

VII. Operational Risk

A. Literature on project management uses the equivalent term, “execution risk”

B. For survey organizations: Risk that the survey procedure will not be carried out as specified

1. Distribution of deviations from specified procedure

2. Impact of deviations on overall properties of procedure

C. Examples:

Frames and administrative records: Changes in laws, admin procedures or vendor degrade data quality

Sample design: Fail to select units with specified probabilities

Instrument design: Incorrectly programmed skip pattern for special subpopulation

Nonresponse follow-up: Fail to use specified protocol for callbacks

Measurement error: Fail to assign interviewers at random

Modeling of survey data: Exploratory work identifies all features of data important for final inference?

D. Distinguish between:

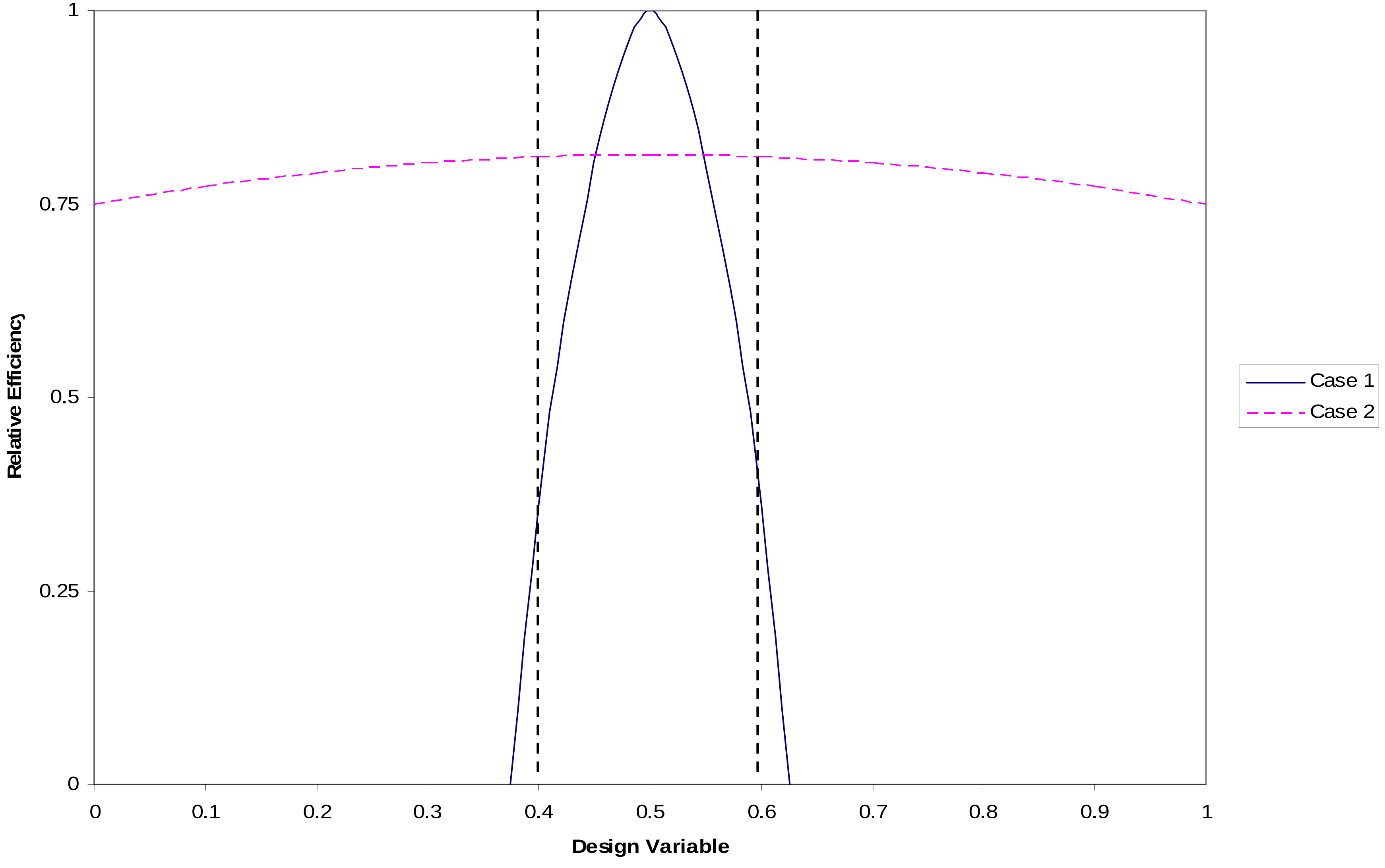
1. Incremental failure
   (14 instead of specified 15 callbacks)

2. Catastrophic failures
   (blunder in programming instrument)

E. Conjectures:

1. Program and field managers often avoid adoption of proposed new procedures due to concerns about operational risk, despite large nominal gains in efficiency under correct implementation

   - cf. Dillman (1996, *JOS)*



Relative Efficiency
1
0.75
0.5
0.25
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Design Variable
Case 1
Case 2

2. Expansion of research on survey methodology to incorporate more systematic assessment of operational risk:

a. May help to smooth and accelerate development and adoption of new procedures by field managers

b. Is potentially a rich area for methodological research

F. Formal Development

1. Current approach: Procedure $\hat{\theta}(X,Y)$

where X = Specification of procedure

Y = Reported data

Loss function: $L\{\hat{\theta}(X,Y),\theta_P\} = \{\hat{\theta}(X,Y) - \theta_P\}^2$

Evaluate

$E_D[L\{\hat{\theta}(X,Y),\theta_P\}]$ wrt design

$E_{\xi D}[L\{\hat{\theta}(X,Y),\theta_P\}]$ wrt design and model

2. Define “operational error” $x - X_0$
where X_0 = nominal specification

x = true implementation

Evaluate: $E_x(E_{\xi D}[L\{\hat{\theta}(X,Y),\theta_P\}])$

a. Distribution of operational error conditional on current agency practice and discretionary resource base

b. Sensitivity analysis

c. Elicit priors for distribution of $x - X_0$

d. Empirical approximation of dist?