

Potential Topics for Study in Evaluating Uses of External Administrative Data for the Consumer Expenditure Surveys

John L. Eltinge
Office of Survey Methods Research,
BLS

February 8, 2011



Disclaimer

- The views expressed here are those of the author and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics, nor the FCSM Subcommittee on Statistical Uses of Administrative Records

Overview

- I. Conceptual Background
- II. Methodological Issues
- III. Empirical Issues
- IV. Management Issues

I. Conceptual Background

A. Primary Question

For a specified resource base, can we improve the balance of

quality/cost/risk

in CE data products by obtaining some data through administrative records?

I. Conceptual Background (continued)

B. Possible Examples

1. Sales data from retailers, other sources
 - Aggregated across customers, by item
 - Possible basis for imputation of missing items or disaggregation of global reports
2. Collection of some data (with permission) through administrative records (e.g., grocery loyalty cards), linked with sample consumer units

I. Conceptual Background (continued)

- C. Framework: Population of Consumer Purchases Defined by Cross-Classification of:
- Classification of product/service, time, geography
 - Characteristics of purchaser (Consumer? Demographics?)
 - Admin: Outlet, intermediaries (financial, other)
- D. How to modify estimation methods to incorporate administrative data?
- Weighting and imputation for CPI cost weights, commonly produced tables
 - Construction of public-use datasets

II. Methodological Issues

- A. Methods for Evaluation of Properties of Prospective Administrative Record Sources
 - 1. Population aggregates (means, totals)
 - 2. Variable relationships (regression, GLM)
 - 3. Cross-sectional and temporal stability of (1) and (2)

- B. Methods for Integration for Sample and Administrative Record Data

II. Methodological Issues (Continued)

- C. Importance of Clarity on Sources of Variability Considered in Evaluation of Bias, Variance and Other Properties
 - 1. Sources:
 - Superpopulation effects
 - Sample design (e.g., subsampling, matching)
 - Unit, wave and item missingness or time lags
 - Aggregation effects (temporal, cross-sectional)
 - Reporting error (definitional, temporal, other)
 - Imputation effects (including model lack of fit)
 - 2. Conditioning and integration

III. Empirical Issues

A. Properties of Input Data and Final Estimators

B. Cost Structures

1. Obtaining Data:

Contractual costs with provider
Agency personnel (expertise)

2. Modification and maintenance
of production systems

C. Case studies are important, but may not allow
inference to broader populations, variables

IV. Managerial Issues

A. Central Issue: Management of Costs and Risks

- Methodological risks (commonly studied)
- **Operational risks ("execution risks")**

B. Contractual Structure:

1. Performance Criteria and Incentives for Data Provider (Timely Delivery, Quality, Notice on Changes)
2. Stability of Prospective Sources (AOL in 1999, 2011)
3. Changes in BLS Requests (New Products, New Channels)

C. Agency Personnel:

Skills, Incentives and Institutional Culture

IV. Managerial Issues (Continued)

D. Contrast Between

1. Incremental risks
(per standard statistical methodology)

2. Systemic risks

cf. literature from Perrow (1984, 1999)

and others on risks in

“complex and tightly coupled systems”

Contact Information

John L. Eltinge

Associate Commissioner
Office of Survey Methods Research

www.bls.gov

202-691-7404

Eltinge.John@bls.gov

